Essays on the Impact of Universal Credit

Sam Grant

Presented in fulfillment of the requirements
for the degree of Doctor of Philosophy

Department of Economics
University of Strathclyde
April 2023

Declaration

Acknowledgements

First of all, I am deeply indebted to my supervisors Markus Gehrsitz and Stuart McIntyre for supporting me through the PhD process. This thesis would not have been possible without the time and energy you both invested. Working with you has been a thoroughly rewarding and enriching experience, both personally and professionally. It has also been a lot of fun, and I look forward to future projects together. Stuart, I am grateful for you taking me under your wing as an undergraduate, it has had a major impact on my life.

I would also like to thank all other academic staff within the Economics Department. Virtually all of you contributed to the production of this thesis, either directly through your roles as discussants or indirectly through your teachings when I was an undergraduate or master's student. I have enjoyed the camaraderie that exists within the department. It made coming to the office a delight. A special thanks to Alex Dickson and Julia Darby, for your efforts building and supporting the PhD community.

I am grateful to those who helped support me financially during my studies. First, the Joseph Rowntree Foundation and the Economic and Social Research Council for joint funding me over the last four and a half years. Both organisations kindly extended my stipend for four months due to delays accessing secure data. Second, I thank Graham Galloway, Pippa Stone, Dette Cowden, Georgia-Lee Smith and the rest of the Business Investment and Trade Economics team at the Scottish Government for giving me the opportunity to intern with them. You were very supportive and welcoming.

I also thank Donal Cairns for collating data from Public Health England to use in my first empirical chapter, and for answering my countless data queries in subsequent emails. Further, I would like to thank Data First and the ONS for providing me with the criminal justice data I use in my third empirical chapter. Thank you Claire Melson and your colleagues within ONS statistical support for checking and clearing my outputs.

I would also like to thank past and present PhD colleagues and friends, Annie, Arnold, Ben, Beth, Césarine, Geoff, Grant, Iswat, Jon, Lateef, Luigi, Orion, Rory, Ross, Sharada, Slawek, Yuri and Zhan for all the laughs and great conversations. Additionally, I'd like to express my gratitude to Calum and Ciara from the Fraser of Allander Institute for the support and for keeping me sane in the office in the final months leading up to my submission. I may have put you off returning to academia.

Finally, a special thank you to my mother Fiona, my sister Rosie, my mother's fiancé Richard and to the wider Grant and Laverick family for the continuous love and support. I could not have done it without you.

Abstract

This thesis evaluates the impacts of Universal Credit, the UK Government's flagship welfare reform. After providing an overview of the policy, I make use of applied econometric methods to study its potential unintended consequences on mental health, local crime rates and prisoner recidivism. My findings are intended to contribute to the ongoing policy debate around Universal Credit, as well as the wider economics literature on welfare reform.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The Welfare Reform Act (2012) legislated the UK Government's flagship Universal Credit (UC) programme, and with it initiated the most radical change to the benefit system in a generation. By 2024, UC is expected to directly impact the lives of 8 million low-income families - that is, around 1 in 3 working-age UK households (Brewer et al., 2019). One of its twin objectives is to simplify the benefit system, which it addresses by streamlining six-out-of-work benefits into a single benefit. Its second objective is to recast the role of social security (Department for Work and Pensions, 2010). Indeed, it reprioritises claimant responsibility and the transition into employment as the central tenets of the UK benefit system. To achieve this, the Department for Work and Pension have adopted a suite of reforms so that the programme "mirrors the world of work" (Department for Work and Pensions, 2010). Payments of the benefit, for example, will no longer be staggered throughout the month but will instead be made monthly in arrears. Other changes include increasing the conditionality attached to receiving the benefit; levying stricter sanctions; moving to a fully digitalised system; and paying the housing component of the benefit to tenants, as opposed to landlords directly.

A reform of this magnitude has significant economic and social implications, making it crucial for UK policymakers to better understand its impacts.[1] In recent years the urgency for evidence has become even more pressing. Indeed, the impact of the Covid-19 pandemic on employment led to UC

[1] This need for evidence was highlighted, for example, by the inquiry into the economics of UC conducted by the House of Lords Economic Affairs Committee in February 2020 (House of Lords, 2020).

playing a major role in the UK Government's strategy to mitigate the economic consequences of an economy in 'lockdown'. As a result, more than six million people have recent experience of being in receipt of UC and subject to its rules and regulations. This number is destined to rise as the next phase of the programme's implementation - involving transferring over claimants from the "legacy" system it replaces - is set to resume in 2023.

Beyond relevance to the UK, UC represents a unique opportunity to study the socio-economic effects of welfare reforms. Benefit systems are rarely overhauled in the way UC has to the legacy system. Less common still are such reforms introduced in a manner that enables causal inference. UC, with its size and phased delivery, offers both of these advantages for scholars to exploit. As such, the programme can help further our understanding of the impacts of welfare reform and update a literature largely focused on a U.S. policy from thirty years ago. Providing evidence on these matters is also particularly timely. Across the OECD there has been a renewed interest in active labour market policies (ALMP) (OECD, 2021), many of which share overlapping elements with UC. Policymakers internationally have therefore taken great interest in watching the UC's development (Wickham et al., 2020). By evaluating the programme's wider socio-economic impacts, this thesis can help policymakers better understand the true costs and benefits of welfare reform.

1.1 Objectives

The aim of this thesis is to provide quantitative evidence on the impact of the rollout of UC on mental health and crime. It seeks to answer the following questions:

1) What is UC's impact on objective measures of mental health?

2) What is UC's impact on local crime rates?

3) What is UC's impact on prison leavers' propensity to reoffend?

To address each question, I will seek to provide causal estimates in my analysis. Doing so for question 1 and 3 will represent the first of its kind in the literature. For question 2, I will focus on addressing limitations arising within the existing UC literature in order to provide a new contribution to the field. Overarching the objectives is an aim of improving our understanding of UC and contributing to its on-going policy discussion.

1.2 Thesis Outline and Contributions

This thesis will provide novel quasi-experimental evidence on the impacts of UC on mental health and crime. It unfolds as follows.

Chapter 2: Universal Credit

This chapter overviews UC. I begin with a short description of the Welfare Reform Act (2012) and UC's headline reform. I then outline UC's objectives and its reforms to achieve them. Finally, I describe how UC was rolled out across the UK; this is key to the methodologies I employ in each empirical chapter.

Chapter 3: The Wrong Prescription? Universal Credit and Mental Health

The first empirical chapter will provide the most comprehensive evaluation of the effects of UC on mental health to date. A small literature supports the position that UC has led to a deterioration in self-reported mental health among the unemployed. I greatly expand on this work by exploring

a) a far wider array of outcomes, and

b) whether increased psychological distress translates to a deterioration in objectively measurable mental health.

In doing so, this chapter aims to shed light on whether UC has caused an increase in mild, moderate or severe levels of clinical depression. This has clear policy relevance. UC has a key aim of improving the labour market attachment amongst those receiving welfare. At the same time, there are well-established links between mental health and employment. Adverse effects on wellbeing from UC may therefore undermine one of its core objectives.

Exploiting the programme's staggered roll-out across local authorities within a difference-in-difference (DD) framework, I find no robust evidence indicating UC has led to an increase in mild, moderate or severe levels of clinical depression among claimants in communities following its introduction. That said, one possible exception pertains to individuals with pre-existing and complex mental health problems. I find suggestive evidence of the programme leading to a 2.4% increase in the number of open referrals to the National Health Service (NHS) secondary care mental health services.

Chapter 4: Punitive Welfare Reform and Crime

The second empirical chapter provides new evidence on the impact of UC on local crime rates. Previous studies within this literature have exploited the programme's staggered rollout using two-way fixed-effect estimators that rely on strict assumptions of treatment effect homogeneity. In this chapter, I employ new difference-in-differences methods that carry a number of advantages over these estimators, the most significant being they are unbiased when treatment effects are dynamic. In addition to this contribution, I provide the first criminological estimates of the Full Service, the version of UC that opened up the benefit to several million households and is in place today.

I find suggestive evidence of the programme leading to a rise in acquisitive types of crime during the period at which it was only available to single jobseekers. However, this finding does not hold up to robustness checks. I also find no link between the Full Service and any offence type.

Chapter 5: Sittin' in the Dock Without Pay? Welfare Reform and Reoffending

The final empirical chapter examines the relationship between UC and prisoners' propensity to reoffend. In doing so it provides the first quasi-experimental evidence on links between welfare reform and prisoner recidivism, outside of a U.S. context. There are strong theoretical reasons to assume that a relationship between UC and reoffending may exist. For prison leavers this is particularly true, since many face a range of barriers to employment and thus disproportionately rely on welfare; 54% claim out of work benefits within their first month of release, compared to the 13% of the population (Ministry of Justice, 2014). Facing a cut in entitlement under UC, it seems plausible that there may unintended consequences for recidivism.

To test this hypothesis, I leverage rich offender-case level data pertaining to the universe of prison spells, court cases and probationary periods within England and Wales during the period of UC's rollout. These data enable me to build a timeline of offenders' criminal incidences following release from prison. To establish causality with UC, I employ a regression discontinuity design and exploit the fact that marginal differences in release dates sees prison leavers exposed to either UC or the legacy system, depending on which local authority they undertake their probation in.

Overall, I find no evidence to suggest the introduction of UC led to an increase in reoffending rates among single, working-age prisoners - the demographic most likely eligible for the benefit during the period of the rollout I study. My main analysis produces no estimates that are statistically different from zero for total, violent, and acquisitive forms of crime. However, my estimates are imprecise, and thus only able to rule out large criminological effects from the programme.

Chapter 6: Conclusion

Chapter 6 concludes this thesis by summarising the aims, methods and results of each empirical chapter.

Chapter 2

Universal Credit

The Cameron-led UK Coalition Government introduced the Welfare Reform Act (2012) as part of its broader programme of austerity. Among the provisions of the Act were controversial changes to the housing benefit – introducing an under-occupancy penalty – and the replacement of the Personal Independence Allowance with the Disability Living Allowance. Other changes included the scrapping of a Council Tax benefit and the introduction of a hard "benefit cap" at £350 per week for single claimants and £500 per week for those with children or in a couple (Department for Work and Pensions, 2021).

However, by far the most radical and high-profile component of the legislation was UC. Still ongoing to date, the programme has an ambition of combining 6 means-tested working-age benefits and tax credits into a single payment. This includes the primary means of support for those in or out of work: income-based Jobseekers Allowance (JSA), income-related Employment Support Allowance (ESA), Income Support, Housing Benefit (HB), Child Tax Credits (CTC) and Working Tax Credits (WTC). In doing so, UC is replacing the legacy benefit system. The effect of this is that welfare recipients no longer have to make a separate claim to each benefit, rather they now apply for and receive a single UC transfer. While simple in theory, this has fundamentally changed the way claimants interact with the benefit system.

UC is available to jobseekers and those in work but on low earnings or part time hours. It is also available to people unable to work, for example due to health reasons. Claimants must have less than £16,000 in assets to be eligible to apply. The standard allowances, currently in 2023, is £265.31 (£334.90) for single claimants under (over) 25 years old and £416.45 (£525.72) for couples under (or at least one is over) 25 years old. While comparable to the various legacy benefits, final entitlement can markedly differ due to differences in the way each regime treats particular circumstances, such as: child and caring responsibilities, housing situation, disability and health, employment, and earnings. In the following section I will highlight some of the notable differences.

2.1 Objectives and Reforms

2.1.1 Simplification

By streamlining six benefits into one, a core objective of UC is to simplify social security. In doing so it addresses a longstanding issue of complexity within the legacy system. Dating back to the 1960s, governments had made incremental changes to the legacy benefits due to a myriad of political, economic and administrative reasons (Timmins, 2016). As a result, by the mid-2000s there was a period in which the welfare state comprised of 51 separate benefits, compared to 27 in 1979 and 7 in 1948 (Centre for Social Justice, 2009). Within each of these were varying eligibility criteria, timetables, withdrawal rates and tests - many of which also interacted with each other. It thus became widely acknowledged both within and out of government that the benefit system had become incredibly difficult to navigate (Timmins, 2016). This was true not only for its users but for the Department for Work and Pension (DWP), its administrator. In February 2005, the National Audit Office (NAO) published a report with its head, Sir John Bourne, warning that complexity "is one of the most important issues impacting on the performance of the Department [DWP]" (NAO, 2005).

This objective of UC has generally been seen as a technocratic change rather than a political one, as evidenced by its cross-party support prior to legislation (Timmins, 2016). While the programme would not dissolve most of the benefits in the legacy system, it was to amalgamate the six major working-age ones that, together, accounted for more than £60 billion in expenditure per year Timmins (2016).

2.1.2 Transition to Work

UC's other main objective is to "transform" the benefit system into a service that prioritises claimants moving out of welfare. This ties in with the simplification aim in that it was argued the legacy system's complexity had itself become a barrier to work, with claimants unaware of the in-work support available or how working would impact their finances (Department for Work and Pensions, 2010). Moreover, there was said to be a "familiar security" of claiming out-of-work benefits (Department for Work and Pensions, 2010). UC aims to address this by combining out of work support with working tax credits.

A second means to achieve this goal is to ensure work, no matter how little, always pays more than inactivity. This is in part to be achieved by reducing the overall benefit entitlement by £2 billion per year (Brewer et al., 2019). However, it has also meant introducing a 'work allowance' and a more generous benefit withdrawal rate, which allows claimants to keep 63p of their entitlement for every additional £1 in (post-tax) earnings. By doing so, UC is scrapping legacy's so-called 'hours-rules' that saw virtually no support available for lone parents working less than 16 hours a week and couples 30 hours a week. These changes have substantially reduced the marginal participation tax that claimants faced under legacy, which had been as high as 96% (Timmins, 2016).[1]

The more generous support for those in work has been traded off with a reduction in support for other, mainly out-of-work groups. Analysis in Brewer et al. (2019) shows that, as a whole, nearly 25% of workless households lose more than £1,000 p.a under UC versus the 3% that stand to gain as much. Among those most affected are disabled claimants deemed able to work: where entitlement can fall by £2,230 p.a (Brewer et al., 2019). One notable exception to UC's focus on inactivity has been its treatment of the self-employed. The programme has introduced a 'Minimum Income Floor' which sees some low earners within this group have their benefit entitlement calculated as if they were earning the minimum wage for 35 hours week. As a result, a single, self-employed individual can lose up to £8,200 p.a on UC compared to legacy (Brewer et al., 2019).

[1]Once in work, out-of-work benefits would be quickly withdrawn, off-setting earnings from work. This was compounded by the fact that as earnings progressed, claimants would lose other benefits such as free school meals and free prescriptions, which could cost more than the gain from WTC (Timmins, 2016).

Beyond simplifying and offering financial incentives to work, UC's entire design has been tailored to achieve its employment goals. This has meant introducing several sub-policies to create a benefit system that mirrors the world of the labour market, in theory making the transition from welfare easier (Department for Work and Pensions, 2010). Key to this is giving claimants greater personal responsibility for handling their finances and changing their relationship with the state to mirror that of an employee-employer. These sub-policies have represented the most significant cultural shift from the legacy system. They have also been a main source of UC's controversy. Five of the major changes are as follows.

First, UC claimants receive their payment monthly and in arrears, as many salaries would be. This contrasts with the legacy system whereby benefits are paid on a weekly or fortnightly basis - in practice this will be even more frequent if individuals made multiple legacy claims.

Second, the housing element of UC is paid to the tenant, thereby making claimants in social housing, who previous had their rent paid directly to the landlord, take responsibility for meeting their rent obligations.

Third, UC services are 'digital by default', meaning the initial application and regular communications with the Job Centre are web-based rather than in person, by post or telephone. As well as lower administrative costs, this is meant to help ready claimants for the modern world of work.

Fourth, job-search requirements are more stringent for the unemployed and the disabled relative to the legacy system. In addition, UC extends these requirements to claimants already in part-time or low-paid work, affecting nearly 900,000 people who would have previously received payments unconditionally.[2] The rationale behind this is that the state now has a direct financial interest in moving people into full-time work and off welfare entirely. In practice, most UC claimants not in full-time work must now spend up to 35 hours a week searching for a job or finding additional work, and document their efforts in their "Claimant Commitment" either in person or online. The effect of these changes is that the UK now has the second strictest job-search requirements in the OECD (Immervoll and Knotz, 2018).

[2]StatXplore, the DWP's public database on benefits, provides information on the number of claimants in each conditionality regime.

Finally, benefit sanctions enforcing job-search are more punitive under UC. This is mainly for two reasons. First, sanctions under UC are consecutive rather than concurrent. Second, hardship payments - payments claimants can apply for whilst sanctioned – are deducted at a 40% rate from future benefit instalments, where previously they were treated as a grant. As a result, an equivalent sanction under UC lasts 2.5 times longer in real terms for people claiming hardship payments than under the legacy system Webster (2018). Other key changes to the system include lone parents with pre-school age children now being exposed to a risk of losing their entire allowance, where previously it would have been capped at 20% under Income Support Webster (2018).

2.2 UC Rollout

Overhauling the benefit system with a new main in and out-of-work benefit has been an enormous administrative task for the DWP. In addition to transferring several million claimants, it has meant building a benefit that can handle as many as 1.6 million changes in claimant circumstances every month, each triggering different payment amounts and (potentially) job-search requirements (Public Accounts Committee, 2013). The size and complexity of this change has necessitated a gradual phasing in of UC. In practice, this has meant slowly rolling out the benefit across Job Centres and withholding eligibility to certain types of claims. (Department for Work and Pensions, 2011).

To date, UC is still restricted to *new* claims only. This has meant that when the programme is introduced to a given Job Centre, only local residents who were eligible for UC – e.g. met the earnings and savings threshold, along with other criteria stated at the beginning of Chapter 2 – and also experienced a change in circumstances could claim the new benefit. These changes in circumstances could mean anything from losing a job to entering a relationship, having a child, being evicted, moving house or having a change in health or disability status. The effect of this decision was that claimants of the legacy system would not immediately transfer over to UC once it was launched in their community. Rather, they would be transferred over only if their circumstances changed in a manner such that they were required to make a new claim to the benefit system. This provided the DWP with some insurance that the UC system would not be immediately overwhelmed once implemented, since the first month often involved less than a dozen claimants being enrolled. Over time however, the caseload increased gradually as new claims were directed to UC.

This process of new claims only was known as ‘natural migration’ and was conducted via a “twin-track” rollout across local authorities. It involved two versions of the programme, the Live Service (LS) and the Full Service (FS), with the former being rolled out first and the latter replacing it. A single Job Centre in Ashton-under-Lyne was chosen to pilot the LS in April 2013. It restricted the benefit to individuals who met initial LS “gateway” conditions: single, unemployed, no children, did not own a home or require any housing benefit. These were claims that formerly would have been made to JSA and were deemed the simplest for the UC system to manage. Over the following year, 13 other Job Centres, known as “pathfinders”, piloted the LS in the North West of England.

From April 2014 the LS was further expanded across the North West and rolled out nationally, covering all UK local authorities by May 2016. This phase of the rollout also saw the introduction of the Claimant Commitment, a core part of UC missing during the piloting. Thus, for the first time the programme began enforcing job search requirements and benefit sanctions. By May 2016, around 270,000 individuals were claiming the benefit. The effect of the LS gateway conditions meant that, in practice, two-thirds of these claimants were men, most of whom (100,000) were under the age of 25.[3]

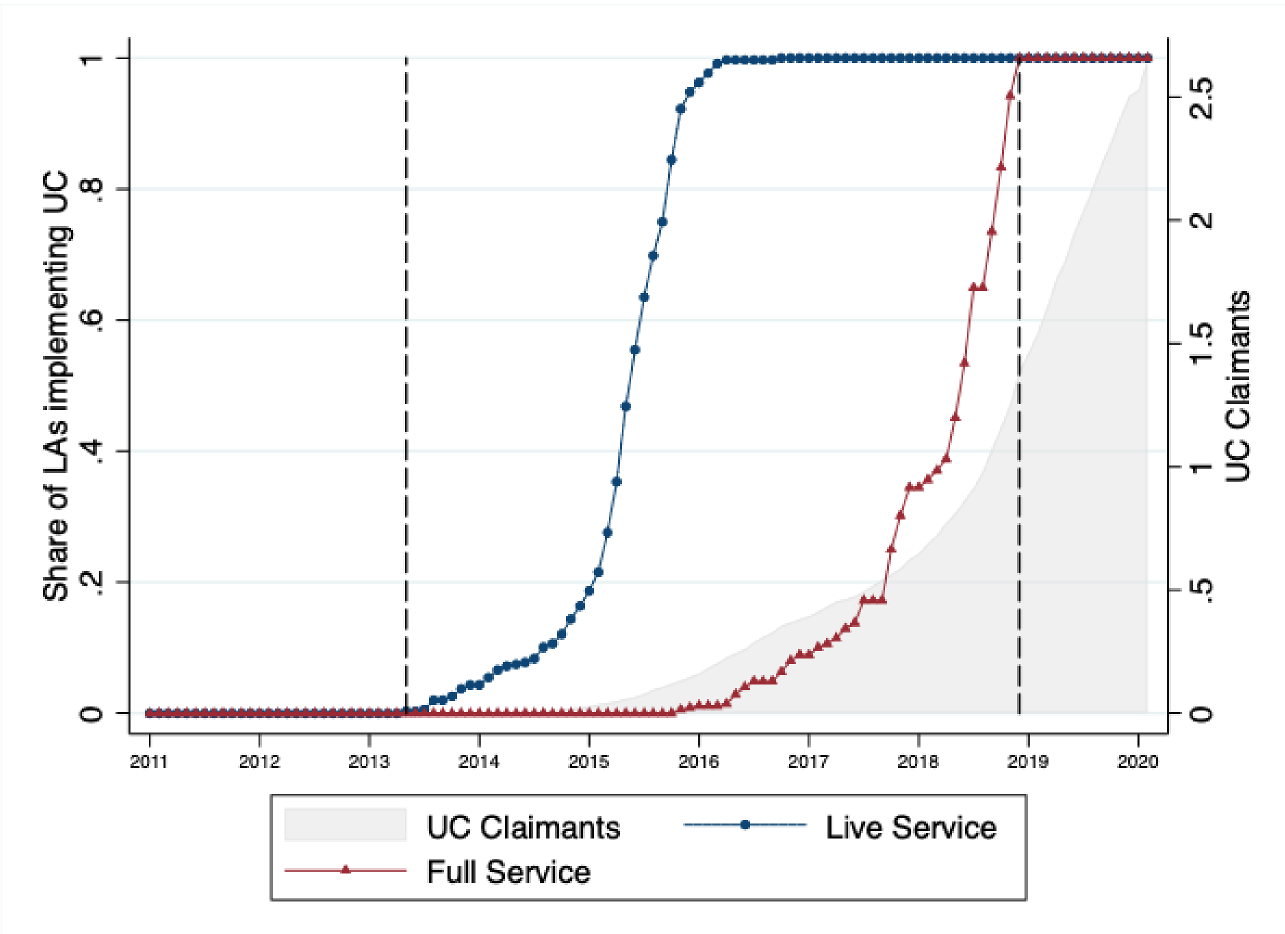
While the LS was being rolled out, the FS was launched in June 2015 in Croydon. Unlike its predecessor, the FS enabled anyone who met the UC eligibility criteria to make a new claim, not just those who met the gateway conditions. Figure 2.1 shows that once the FS began its national rollout - opening up UC to couples, those in work and those with more complex cases - the number of people claiming the benefit increased dramatically. By December 2018, all local authorities had converted from the LS to the FS, and 1.5 million people were in receipt of UC payments.

The conclusion of the FS rollout meant that UC had completed its first stage of ‘natural migration’. The second stage of ‘managed migration’ - involving transferring over recipients from the out-going legacy system - was initially piloted in July 2019 in Harrogate, but was postponed due to the Covid-19

[3]StatXplore provides a breakdown by age and gender of the UC caseload.

pandemic. It has since resumed in Bolton and Medway and is expected to be fully completed across the UK in 2024.

Figure 2.1: UC Rollout Timeline



Notes: Dashed lines represent the start and end of the UC rollout in Great Britain (April 2013 to December 2018).

Chapter 3

The Wrong Prescription? Universal Credit and Mental Health

3.1 Introduction

Poor mental health is a widespread and growing problem. The World Health Organisation (WHO) project that by 2030, mental health problems will pose the largest health burden of any disease in affluent countries (Mathers and Loncar, 2006).[1] This increase in prevalence is particularly acute in the UK, where 1 in 6 people now report experiencing mental health problems in any given week, a rise of 20% since 1993 (McManus et al., 2016).[2] For UK policymakers this is of growing concern. In addition to adverse human and social impacts, mental health problems impose a substantial economic burden on the economy, estimated to be greater than £101 billion per year, that is, 5% of the UK's GDP (McDaid et al., 2022). These costs pertain to, but are not limited to, days off work (Alonso et al., 2011; Lim et al., 2000; Bubonya et al., 2017), health care (Lim et al., 2008) and education (Cornaglia et al., 2015). Among those most often affected are the disabled, poor and unemployed (Murphy and Athanasou, 1999) - populations typically more dependant on welfare.

[1]WHO define a "Global Burden of Disease" metric that quantifies health loses in terms of mortality and disability across countries and time (World Health Organisation, 2023).

[2]The UK ranked 20th out of 27 EU countries in the WHO-5 mental health wellbeing index (Department of Health, 2014)

This raises the question of whether recent UK welfare reforms have played a role in exacerbating this trend. UC, in particular, has been subject to fervent debate in this context. Concerns have regularly been raised in Parliament and by third-sector organisations elsewhere about the distress moving to the new benefit is causing claimants, particularly those already most vulnerable. Doctors have attributed an increase in their workload and a rise in the number of patients struggling with new or pre-existing mental health problems to the programme (Arie, 2018). Qualitative studies, interviewing claimants, likewise consistently warn of a negative impact on wellbeing (Rabindrakumar and Dewar, 2018; Griffiths et al., 2020), and further report that many risk falling into poverty, destitution (Andersen, 2020; Johnsen and Blenkinsopp, 2018) and even suicidality (Dwyer, 2018; Cheetham et al., 2019; Wright et al., 2022). These findings have implications beyond claimant welfare. Prior research has shown poor mental health to represent a key barrier to employment (Björklund, 1985; Danziger et al., 2000; Danziger and Heflin, 2000; Meara and Frank, 2006; Ettner et al., 1997; Jayakody et al., 2000). As such, if UC has an adverse effect on mental health, it may seriously undermine its own welfare-to-work goals.

In this chapter I explore the effects of UC on mental health in England using administrative data from NHS Digital and Public Health England. To do so, I exploit the plausibly exogenous variation arising from the programme's Full Service rollout across local authorities between June 2015 and December 2018. I employ a difference-in-difference (DiD) research design that enables identification of credibly causal mental health estimates.

Wickham et al. (2020) and Brewer et al. (2022) provide the only other quasi-experimental evidence on UC's mental health impact. These studies, focusing on self-reported depression among the unemployed, produce findings that appear to substantiate the anecdotal concerns regarding the rollout. Wickham et al. (2020) find that the programme increased the levels of psychological distress of 63,674 people, with 21,760 potentially reaching the diagnostic threshold for depression. Brewer et al. (2022) find that single adults and lone parents experience a decline in mental health quality by 8.4%-13.9% of a standard deviation. This study builds on this work in three respects.

First, I examine a broader array of mental-health related outcomes. As such, my analysis represents the most comprehensive evidence available on UC's impact on wellbeing to date. My results distinguish between its impact on mild, moderate and severe depression, based on the National Institute for Health and Care Excellence classifications. This provides novel insight into whether UC leads to or exacerbates *clinical* depression. It also, in some cases, sheds new light on the types of claimants affected – specifically, those with or without pre-existing health conditions.

Second, my analysis employs weaker identifying assumptions on two fronts. One is that my results are not contingent on treatment effect homogeneity, as I employ updated and robust econometric methods. The second is that I do not assume UC has no labour supply effect. This is reassuring for the internal validity of my results given that UC changes labour supply incentives. Wickham et al. (2020) implicitly make this assumption by allowing within their research design people to move between control and treatment groups as their employment status changes in subsequent survey waves. Brewer et al. (2022) do so as well, albeit to a lesser degree, by comparing employment exits across UC and legacy operating areas. In this study, I sidestep this issue by focusing on the causal effect of UC being launched in a local authority.

Third, I investigate whether the adverse impact on self-reported depression, as reported in Wickham et al. (2020) and Brewer et al. (2022), leads to a deterioration in *objectively* measurable health outcomes. The distinction between subjective and objective measures of wellbeing is crucial in this context. A claimant's personal assessment of the programme, for instance, may influence the former without affecting the latter (Caliendo, 2019) Conversely, an individual could under report subjective feelings when being surveyed – this is a known problem with these types of data, which could be exacerbated by the stigma associated with mental health (Caliendo, 2019). My findings will not be affected by these issues, nor others concerning the interpretation of subjective measures more generally. Moreover, they should also help provide an improved identification of UC's impact, due to the fact that the sourced objective outcomes have a monthly frequency, enabling a one-for-one synchronisation with the monthly roll-out schedule. Finally, when evaluating UC, policy-induced changes in objective health outcomes can often be tied to tangible monetary costs (savings). Policies should, when possible, be evaluated in such a manner to account for true societal costs (benefits). Given that a justification for UC has been rooted in its potential cost-savings for the tax payer, objective health outcomes are particularly interesting to explore because of their links to a publicly financed health service.

Turning to my main results, I find no robust evidence of UC leading to an increase in mild, moderate or severe levels of clinical depression in communities following its Full Service rollout. Specifically, my most conservative estimates rule out increases in the prescribing of antidepressants by 1.2% and referrals to the main NHS talking therapy programme by 5.1% (with 95% confidence). Looking more broadly at mental health, my results suggest UC has not influenced mental health related hospitalisations, suicide rates, nor the prescribing of drugs associated with anxiety, pain or substance addiction. This suggests the adverse effects found Wickham et al. (2020) and Brewer et al. (2022) have not extended to reach clinical levels of depression.

One possible exception pertains to individuals with pre-existing and complex mental health problems. I find suggestive evidence of the programme leading to a 2.4% increase in the number of open referrals to the National Health Service (NHS) secondary care mental health services. These are services that are only available to individuals with a doctor's referral, and typically treat complex or enduring mental disorders. My analysis suggests that only patients with pre-existing problems were affected, as no link with UC was found for new referrals to these services. I argue that this particular finding must be interpreted with caution, however.

These findings contribute to a strand of the economics literature concerning the health effects of welfare reform (for a review see Blank (2002)). With few exceptions, these studies have focused on the U.S. Temporary Assistance for Needy Families (TANF) reform implemented in 1996 (e.g. Currie and Grogger (2002); Bitler et al. (2005)). Establishing causality has proven challenging in this setting (Blank, 2002). TANF was implemented across states at almost the same time (within 9 months), created considerable heterogeneity in welfare policies within these treated states, and coincided with an economic boom and multiple policy changes (e.g. minimum wage and tax credits). This chapter studies welfare reform in a setting more favourable for identification: UC is invariant across treated districts, was rolled out over a 6 year period of relatively stagnant growth, and is over-hauling an existing benefit system. Separately, UC is also being implemented in a country where health care is free of charge (in almost all cases), which provides an ideal setting to study the effects of welfare on objectively health outcomes pertaining to treatment demand. Taken together, I argue my estimates provide some of the most credibly causal in the literature.

The remainder of this chapter is organised as follows. In the next section, I describe why the UC reforms might be expected to detrimentally impact claimants' mental health. I also outline in this section how mental health care is provided in the UK, as the different stages of treatment enable my outcomes to capture distinct variation in mild, moderate and severe levels of depression. Section 3.3 describes the existing economics literature on welfare reform and health. Section 3.4 describes my data. Section 3.5 outlines and tests my empirical strategy. Section 3.6 presents my results and Section 3.7 subjects them to robust checks. Section 3.8 discusses my results and Section 3.9 concludes.

3.2 Background

In this section I outline the rationale for why UC may adversely affect claimants' mental health. I then provide a background as to how health care for mental illness is provided in the UK. This highlights how the objective health measures I source, pertaining to treatments and interactions with the NHS, capture the different stages of clinical mental health deterioration as classified by the National Institutional for Health and Care Excellence (2017).

3.2.1 Potential Links between UC and Mental Health

As discussed in Chapter 2, UC is consolidating six separate means-tested working-age 'legacy' benefits into a single payment and attempting to provide a stepping stone for claimants to transition into work. To achieve this latter goal, elements of the programme have been designed with the purpose of fostering financial autonomy amongst claimants. These features have resulted in UC being considered controversial in so far as they may unintentionally cause claimants hardship. It is conceivable that some features will only cause distress to particularly vulnerable claimants, such as those with disabilities or pre-existing mental health problems. Others however may undermine the mental health of the broader UC population by inducing financial strain. I will discuss each of these features in turn.

First, as a part consequence of the monthly payment schedule, UC has an inbuilt waiting period of 5 weeks before the initial payment is received. During this period claimants are often left with few or no sources of income, as payments from any claimed legacy benefits are stopped once the UC application has been sent; analysis by the DWP found that between 2014 to 2018, 49% of UC households had no earnings in the three months prior to claiming (National Audit Office, 2020). Compounding matters further, there are often additional delays. For example, in 2017, 40% of claimants waited at least 11 weeks to receive their first payment in full, and 10% waiting at least 28 weeks (National Audit Office, 2018). Third-sector organisations maintain this has led to debt, foodbank usage, and demand for both payday loans and crisis grants (Citizens Advice, 2017, 2018; Trussell Trust, 2019).[3]

Second, the monthly payment schedule itself carries a risk of undermining the budgeting strategies of many low-income families who are accustomed to budgeting on a weekly or fortnightly basis (Hartfree, 2014). What is more, if claimants struggle to adapt and exhaust their benefit too quickly, UC inadvertently raises the consequences of doing so. This is because the benefit incorporates any claimed housing benefit, meaning it is easier to fall into rent arrears. Empirical evidence has previously shown that households on welfare have quasi-hyperbolic preferences and thus consume benefits too quickly (Stephens Jr, 2003; Shapiro, 2005; Stephens Jr and Unayama, 2008), in turn leading to poor health outcomes such as a reduction in caloric intake (Shapiro, 2005; Wilde and Ranney, 2000) and hospitalizations from drug abuse and fatality (Riddell and Riddell, 2006; Dobkin and Puller, 2007). Recent evidence on UC suggests the programme has indeed led to an increase in rent arrears (Hardie, 2021).

Third, UC's intensive work-related conditionality may directly affect recipients' mental health, either through the threat of sanction (Dwyer, 2018) or it reducing available leisure time. DWP survey evidence indicates that the stringency of these conditions is enforced in practice: less than two-thirds of responding claimants thought their job-search conditions were achievable (Department for Work and Pensions, 2018*b*). One group that can be assumed to be at particular risk are lone parents with

[3]The Joseph Rowntree Foundation estimated that 1 in 3 households in the bottom quintile of the income distribution had no savings at all between 2014-16 (Joseph Rowntree Foundation, 2018).

pre-school age children; this demographic are newly exposed to conditionality and have been linked to developing depressive symptoms from job-search requirements in an RCT study (Morris, 2008).

Fourth, the 'digital by default' ethos underlying UC has raised barriers to welfare for claimants without a smartphone, computer or lacking IT skills. According to the DWP's own survey in June 2018, only 54% of recipients in UC Full Service areas were able to apply online without assistance (Foster et al., 2018). Citizens Advice, the largest UK independent advice provider on benefits, similarly reported that 52% of its clients found the online application process challenging (Foley, 2017). Claimants have described the online process and lack of face-to-face support as anxiety inducing in interviews (Wright et al., 2022; Cheetham et al., 2019).

Finally, the net £2 billion reduction in benefit generosity under UC leaves the average claimant with less of a financial buffer to deal with unexpected costs and adverse shocks to income (such as being sanctioned). What compounds matters further is the fact that these changes are regressive (Brewer et al., 2019). Lone parents, the unemployed and the disabled are three groups in particular that typically see a reduction in entitlement (Brewer et al., 2019). Notably for this study, these groups also have a higher baseline risk of mental illness.

3.2.2 Mental Health Care in the UK

In the UK, health care is provided by the National Health Service (NHS) and is free at the point of use to all residents. Provision of mental health services within this is split into 3 tiers: primary, secondary and tertiary care. Each tier differs by the stage of mental health deterioration they treat and the direct accessibility of its services to patients. I focus on the provision of treatment for depression and anxiety in particular within primary and secondary care.

Primary Care

Primary care provides entry level care for mental health problems. It includes doctors based in the community, known as General Practitioners (GPs), and, in England, the Improving Access to Psychological Therapies (IAPT) service. GPs usually represent the first line of contact for people seeking treatment. To be seen by a GP, an individual must register as a patient; there are low barriers to doing so: it is free, and proof of address or having ID is not required (National Health Service, 2023). Being registered with multiple GPs simultaneously is not possible (National Health Service, 2023). Virtually all UK residents are registered to a GP: official patient counts in fact exceed local mid-year ONS population estimates each year (roughly by 3 million).[4]

Depression ranks as the third in most common reasons to visit a GP (Whitty and Gilbody, 2005). According to a survey in Mind (2018), depression and other mental health disorders accounts for 40% of all GP appointments. In assessing whether a patient is to be offered treatment, GPs are advised to adhere to guidance by the National Institute for Health and Clinical Excellence (NICE). NICE recommends GPs to first diagnose whether a patient meets the clinical criteria for depression (National Institutional for Health and Care Excellence, 2017). This threshold is met if 5 out of the 9 symptoms listed in Table 3.1 have been present for at least two weeks, including at least one of the following.

- Persistent sadness or low mood.
- Marked loss of interests or pleasure

If a patient is deemed clinically depressed, the type of treatment offered by GPs will depend on the severity of the depression. NICE classifies depression into three levels: mild, moderate, and severe. Judgement for diagnosis should be based on the number and severity of symptoms and the degree to which they interfere with daily functioning (National Institutional for Health and Care Excellence, 2017). However, the criteria for assessing the latter two elements is somewhat ambiguous. National Institutional for Health and Care Excellence (2017) provide the following classifications:

- Mild depression is characterized by the presence of 5 to 7 symptoms listed in Table 3.1 and only "minor" functional impairment.

[4]Official guidance on this states that this is likely due to administrative error from GPs, as well as there being different definitions (between ONS and GPs) about who counts as a 'resident' in the UK (House of Commons Library, 2016).

- Moderate depression is when functional impairment or symptoms in Table 3.1 are between "mild" and "severe".

- Severe depression is identified by the presence of most symptoms in Table 3.1, with "marked interference" in functioning.

For patients with mild to moderate depression or an anxiety disorder, National Institutional for Health and Care Excellence (2017) recommends offering psychological treatments such as cognitive behavioural therapy, self-help support, and counselling. The IAPT program has been the main NHS provider of these services for working-age people in England since 2008, with approximately 1.2 million people participating in a given year - equivalent to 3.1% of the working-age population (House of Commons Library, 2018). Referrals to IAPT are recommended over antidepressants or anxiety medication for patients with mild to moderate depression (National Institutional for Health and Care Excellence, 2017). Additionally, referrals may be made to patients who have not been diagnosed with clinical depression but have persistently experienced 4 symptoms shown in Table 3.1 National Institutional for Health and Care Excellence (2017).

For patients with moderate to severe depression or those who have not responded adequately to talking therapy, GPs are recommended to prescribe selective serotonin reuptake inhibitors (SSRIs), a type of antidepressant, in combination with the talking therapy (National Institutional for Health and Care Excellence, 2017). SSRIs are the first-line medication offered by GPs for depression (National Institutional for Health and Care Excellence, 2017). Survey evidence suggests GPs adhere to SSRI guidelines in practice, with around 90% of patients with moderate or severe depression being given SSRIs compared to 14.5% for those with mild depression (Gyani et al. (2012)). Other types of antidepressants are generally not recommended due to potential side effects and increased risk of overdose (National Institutional for Health and Care Excellence, 2017).

Importantly for this study, most legacy and UC claimants are exempt from prescription charges. As such, claimants' entire interaction with health care, from initial registration to receiving treatment, sees them subject to no financial barrier that otherwise may have deterred access. Under the legacy system, recipients of any of three main out-of-work benefits - Income Support, income-based JSA and income-related ESA - are granted immediate entitlement to free prescriptions. For in-work benefits, WTC and CTC, free prescriptions are available if claimants receive a child or disability top-up to their tax credit and have an income of less than £15,276.

UC broadly continues this policy. For a start, it mirrors legacy in the sense that unemployed recipients are exempt from fees - specifically, anyone with no earnings or earnings net UC less than £435 is exempt. Second, UC gives exemptions to in-work recipients claiming a disability or child element, again consistent with legacy. The main difference is that UC withdraws the exemption for people whose net earnings in the previous month was over £936. This has two effects. First, claimants who experience large upswings in income within this group are more likely to have their exemption withdrawn, simply due to the more frequent assessment period (monthly versus annual). The second is that claimants further up the income distribution become liable for prescription charges. Assuming a constant monthly salary of £936, UC withdraws this right for people earning £11,220 or over; under legacy the equivalent was £15,276.[5]

Secondary Care

Secondary care is the next stage of care for people whose depression is treatment-resistant, atypical, reoccurring or represents a significant risk (Mind, 2023). In some instances, GPs may decide to refer patients with severe depression straight to secondary care, especially if they are suffering from comorbidities or have complicating problems, such as being unemployed or having financial difficulties. In general, access is restricted to patients with a referral from a health professional; this includes GPs, IAPT staff and NHS local urgent helplines. These services include hospital care and community-based mental health teams, and are usually staffed by psychiatrists, psychologists, social workers occupational therapists. The mental health teams include the following support services.

- "Crisis Resolution" teams for people who are at a significant risk of self-harm, self-neglect, harm to others or suicide.

[5]StatXplore provides a count of the number of UC claimants by monthly £100 award amounts. Unfortunately, it is not possible to use these data to estimate the number of claimants who would have their free prescription entitlement removed under UC. The reason is that the £15,276 threshold relates to net earnings, not just welfare payments.

- "Community Mental Health Teams" that provide support for both short and long term mental health disorders, not just depression.
- "Early Intervention Teams" for people who experience psychosis for the first time.

3.3 Literature Review

This section reviews the existing evidence base on the impact of welfare reform on health. The vast bulk of this literature focuses on TANF – a U.S. welfare-to-work reform that markedly increased levels of work-related conditionality for single women. One strand of this research explores whether TANF led to changes in women's health care insurance coverage or health care utilisation[6]. Another strand investigates whether there were implications for women's health directly[7]. I restrict my attention to the quasi-experimental studies among the latter. These typically report triple DiD estimates of health effects for single women by exploiting variation in state-level welfare policies and employing married women as a control cohort. An important caveat to these studies, as discussed in Blank (2002), is that it is challenging to establish causality in this setting.

Upon review, it is clear that there is little consensus on the direction of TANF's impacts on health. A body of evidence suggests that the reform had no significant impact on both physical and mental health (Bitler et al., 2005; Currie and Grogger, 2002; Kaestner and Kaushal, 2003; Kaestner and Tarlov, 2006), while a number of studies report negative effects on various physical outcomes. I present the main findings from studies pertaining to the latter. Reichman et al. (2005) report sanctioning under TANF increased the propensity for mothers to report having poor physical health (but not mental health). Looking at work-related conditionality more broadly, Haider et al. (2003) concludes that breast-feeding rates six months after birth would have been 5.5% higher in absence of the reform. Knab et al. (2006) find an increase in welfare generosity had in fact perverse effects on mother's self-reported maternal health. Evidence in Kaplan et al. (2005) suggests the reforms were associated

[6]Studies generally find the reform leading to a decline in coverage (Bitler et al., 2005; Borjas, 2003; DeLeire et al., 2006; Kaestner and Kaushal, 2003; DeLeire et al., 2003)

[7]The literature concerning the effects of welfare reform on children's health is not discussed here. See Grogger et al. (2005) for a review.

with an increase in cholesterol, blood pressure, body mass index and several other physical measures predictive of adverse health outcomes. Overall, these studies appears to rule out TANF having positive effects on physical health - only a zero or negative effect is found.

To the best of my knowledge, only three TANF studies report a statistically significant effect for outcomes pertaining to direct measures of mental health. One of these is Herbst (2013), who reports that TANF led to improvements in single mothers' subjective wellbeing by 5-11% of a standard deviation. The remaining two papers suggest the intensification of work-related conditionality led to negative impacts, however: Knab et al. (2008) report an increase in depressive symptoms and Davis (2019) finds an increase in the number of days recipients report having mental ill-health. Several other studies attempt to measure TANF's impact on mental health indirectly by looking at changes in rates of harmful health-related forms of consumption. Basu et al. (2016) report the reform being associated with a 4-percentage point increase in binge drinking, contradicting findings in Kaestner and Tarlov (2006).[8] In addition to the previously noted physical health findings, Kaplan et al. (2005) find a positive association with smoking rates. Evidence in Knab et al. (2008) suggests this relationship may have been part driven by changes in welfare generosity. Corman et al. (2013) report that the welfare reform was associated with a reduction in illicit drug use of 10-21%. Taken together, estimates on the effect of TANF on mental health are generally negative, where statistically significant.

Turning to the UK, a handful of studies outside of the economics literature find a negative relationship between pre-UC welfare reforms and mental health (Barr et al., 2015; Williams, 2021). It is difficult to attribute causality to their findings, however, since observational data is used in the absence of a natural experiment. One exception is Katikireddi et al. (2018). This study uses DiD to analyse the effect of the UK Government lowering the child age threshold at which a lone parent must seek work to claim welfare. Parents subsequently required to seek work scored lower on a mental health index than parents unaffected by the reforms. This is a noteworthy finding in the context of UC, as the programme further increases conditionality for this demographic.

In recent years, there has been a growing body of research using quasi-experimental methods to

[8]Estimates in Kaestner and Tarlov (2006) suggest the reform had little effect on health or health behaviour. The only exception was that scholars found TANF to be associated with less binge drinking.

document the socio-economic effects of UC. Exploiting the program's staggered rollout across communities in the UK, these studies have linked UC to various negative outcomes, such as increased reliance on food banks (Reeves and Loopstra, 2021), higher rates of criminal activity (d'Este and Harvey, 2022; Tiratelli et al., 2022; Lim et al., 2020), more frequent landlord evictions (Hardie, 2021), and poorer (self-reported) mental health (Wickham et al., 2020; Brewer et al., 2022)

This study builds in particular on the work of Wickham et al. (2020) and Brewer et al. (2022), who use the UK Household Longitudinal Study to track individuals' mental health over time. Wickham et al. (2020) provide the first quantitative evidence of UC adversely affecting on mental health, while Brewer et al. (2022) extend their analysis to explore how treatment effects differ by claimant type (e.g. lone parents, couples) and the potential underlying mechanisms. In contrast to this study, their focus concerns subjective indicators of mental health. Moreover, their research designs differ from mine. Most notably, the approach in Wickham et al. (2020) compares the unemployed (treated) to the employed (control) within UC areas, before and after the Live Service launch. My design instead exploits the staggered rollout, defining treated and control groups based on when districts implements the UC Full Service. Further, in Wickham et al. (2020), unemployed respondents switch from the treated to control group if they report finding employment in a subsequent survey wave. This risks inducing upward bias if UC helps people find work, and work in turn improves mental health.

My research design bears closer resemblance to Brewer et al. (2022) in so far as they also exploit the staggered rollout. Their design conditions on survey participants exiting employment to be included in their analytic sample. They argue that although this is likely endogenous, their method compares individuals exiting employment in UC areas (treated) to those doing the same in legacy areas (control). In turn, they maintain that they estimate the relative effect of being exposed to the UC regime, under assumptions of parallel trends, treatment effect homogeneity and rollout exogeneity.

This approach may be flawed however if the implementation of UC leads to an increase in labour market attachment. Specifically, if UC causes individuals to have a lower propensity to exit employment, variation in untreated potential outcomes across the two welfare regimes may not be quasi-random. Empirical evidence on UC's labour effects is scarce, meaning it is difficult to substantiate or reject this hypothesis.[9] However, economic theory suggests that UC, by subsidising employment and reducing the generosity of benefits to those out of work, would increase the relative cost of unemployment, thereby increasing the demand for longer employment spells and (potentially) affecting the comparability of those exiting employment across the regimes. In such a scenario, rollout exogeneity is not strong enough to recover quasi-randomisation, since the treatment itself affects the probability of being in the sample (i.e. exiting employment). Unbiased estimates can still be recovered in the absence of quasi-randomisation if parallel trends holds across individuals treated and untreated. Yet, this assumption would also become more tenuous given the design of their study.

My study's estimates of UC are robust to labor supply effects because I focus on intent-to-treat effects at the local authority level, in contrast to Wickham et al. (2020) and Brewer et al. (2022) who do so at the individual level. At the local authority level, UC's treatment effects can't affect the probability of units (local authorities) being included in a control group, treatment group, or sample overall. This sidesteps the potential problem of a labour supply response. At the same time, a natural concern of such approach is whether "reduced form" estimates can detect treatment effects. That is, treatment estimates may be too diluted within my analysis, as a function of not focusing on the population directly affected (claimants themselves). For studying UC, this is less problematic. Figure 2.1 showed that there was substantial take-up of the benefit following its Full Service rollout, the version of the programme that opened up claims to the greater welfare-seeking population, not just single jobseekers. By December 2018, the end of the rollout, 1.5 million people were claiming the benefit. As such, the "first stage" in this setting is significant.

[9]The only quantitative evidence on this matter that I am aware of is a DWP report stating that UC claimants are 4 percentage points more likely to have found a job within 3 months than their matched counterparts claiming legacy. While this does not shed light on its impact on job retention or individuals' propensity to exit employment, it does suggest UC is having an impact on labour market outcomes (Department for Work and Pensions, 2018*a*).

3.4 Data

This section describes the data used for my analysis. In sourcing an array of mental health outcomes, data were often available at different levels of geography and windows of time. Following data cleaning, the final sample consists of two levels of geography: local authority and clinical commissioning group. This section will discuss this further.

3.4.1 UC

I track the expansion of the UC Full Service across England by using a monthly rollout timetable published by the DWP.[10] I also source monthly data from StatXplore on the number of people claiming UC in each local authority. These are combined with mid-year ONS population estimates to construct a measure of the percentage of people claiming within the local population.

3.4.2 GP Prescribing Data

A key set of outcomes for this study are GP prescribing rates of mental health related medications. These data are sourced from NHS digital, which publishes monthly prescribing counts for every unique drug, dose-strength and state (i.e. solid or liquid) product combination within each GP practice in England. The unit of measurement for these prescriptions are single supplies of medication that typically last patients 1 month. As such, these data not only capture the universe of prescribed medication by GPs, but also infer the length of patients' course of treatment.

To capture UC's effect on moderate to severe depression, I filter out items classed as SSRIs according to the British National Formulary code; as covered in Section 3.2.2, SSRIs represent the first-line of medication that GPs will offer to treat this stage of depression. I then repeat this procedure for NICE recommended anxiety treatments (hypnotics and anxiolytics); nicotine, alcohol and opioid addiction treatments; and painkillers (analgesics). For additional insight I distinguish between non-opioid and opioid painkillers, as the former is used to treat mild or musculoskeletal pain while the latter is used

[10] This can be found here: Department for Work and Pensions (2018*c*)

for moderate to severe pain (National Institutional for Health and Care Excellence, 2021).

I aggregate these prescription items to the local authority level and generate monthly rates of prescribing per 1,000 GP patients using month-year registration data from NHS digital. To assign GP practices to local authorities, I make use of catchment area shapefiles provided by NHS digital. These catchment areas, illustrated in Figure 3.1, are more informative than GP practice locations for this study, as they help identify the predominant local authority where registered patients reside.[11] This is particularly useful since GPs have a degree of autonomy in determining their catchment area, which in turn leads to overlap and non-conformance with local authority boundaries. Accounting for these potential treatment effect spillovers, I exclude from my sample all GP practices that don't have at least 50% of their catchment area nested within a single local authority.[12]

My final analytic sample consists of the prescribing rates in 326 local authorities between April 2013 and July 2019. Table 3.2 shows there is significant variation in average rates across medications. It shows that 51.06 antidepressant (SSRI) items are prescribed per month and 1,000 GP patients on average during the course of my sample. Figure 3.2 shows this rate steadily increased between 2013-2018, the period UC was rolled out.

3.4.3 IAPT Data

I next source monthly data on the number of referrals to IAPT between April 2013 and July 2019 from NHS Digital. These data are available at the clinical commissioning group (CCG) level, a higher level of geography than local authority and of which 203 were active during my sample. Fortunately, since both these geographies are constructed from ONS Lower Layer Super Output areas, it is possible to aggregate the UC claimant count from the local authority level to the CCG level. Monthly referral rates per 1,000 working-age population are then constructed using CCG population estimates sourced

[11] In 2015, the UK Government initiated a "Choice of GP practice" policy in 2015 which allowed patients to register to any practice of their choice. Evidence from GP surveys suggest patient uptake has been incredibly limited. In 2018, for example, NHS England reported the majority of practices having "no out of area patients at all" (National Health Service, 2015).

[12] All results are robust to increasing this threshold in 10 percentage point increments to 100%. Note, Figure 3.3 provides a visual for the distribution of these catchment areas by percentage covering one local authority.

from NOMIS, an ONS open data source.[13] Table 3.2 shows that there is an average monthly referral rate of 2.66 in England.[14]

3.4.4 Drug and Alcohol Rehabilitation Programmes

Providers of NHS drug and alcohol rehabilitation programmes in England submit data on a monthly basis to the National Drug and Alcohol Treatment Monitoring System (NDTMS). Following a request, Public Health England have kindly provided me with a local authority level dataset containing the number of new clients aged 18+ presenting to these providers in a given month. These data span April 2014 to December 2019. Table 3.2 shows that on average 0.36 new clients attend these rehabilitation programmes per month and 1,000 aged 18+ population.

3.4.5 NHS Secondary Care

I next utilise NHS Digital data on the number of new referrals and contacts to NHS secondary care mental health services. Contacts to NHS secondary care refers to the number of people with an open referral to these services (henceforth I refer to contacts as open referrals). As with new referrals, it is not possible to distinguish which kind of NHS secondary service these referrals are for (e.g. Early Intervention Access). Only an aggregate of these new referrals are available.

I also source data on the number of people with an open hospital spell due to a mental health disorder. This outcome represents the most progressed stage of treatment for mental health deterioration in my dataset. For example, patients may be admitted to hospital if they are at risk of self-harming or acting on suicidal thoughts (National Health Service, 2022).

Data for new referrals, open referrals and hospitalisations are available monthly and at the CCG level from January 2016. They are also available separately for children and adults aged 18-65, so I construct rates of the latter per 1,000 people aged 18-65. These data in my panel span between January-2016 and July-2019. Table 3.2 shows that there are on average 18.25 open referrals, 4.93 new

[13] Working age being 18-65 years old.

[14] Figure 3.4 shows the distribution of referral rates for IAPT and claimant rates for UC in England in the year 2018.

referrals, and 0.48 open hospital spells per working-age population in England in a given month.

3.4.6 Suicides

ONS publishes annual suicide statistics at the local authority level, which I use in conjunction with ONS mid-year local authority population estimates to generate annual suicide rates per 100,000 population, over the period 2013-2018. Table 3.2 shows that the mean suicide rate is 8.83.

3.4.7 Controls

I source local authority level demographic and economic controls from NOMIS, an open data source provided by ONS. The economic controls include quarterly unemployment rates and the number of people claiming any of the legacy benefits that UC is replacing. Demographic controls include the percentage of a local authority aged between 16-24, 25-49, 50-64, over 65, as well as the percentage of females aged over 16. For the GP prescribing analysis, I follow Spence et al. (2014) and Williams (2019) in including the rate of antibiotic prescriptions per 1,000 GP patients. This provides a proxy for GP's propensity to prescribe medication.

3.5 Empirical Method

In this section I describe my DiD methodology and identifying assumptions, test the exogeneity of the UC Full Service rollout, and describe my estimator of choice.

3.5.1 DiD

My empirical strategy to identify UC's effect on mental health is to adopt a DiD framework. This method exploits the fact that the rollout of UC has been gradual, with some local authorities implementing the programme several years before others. It does so by employing the change in outcomes among districts that adopt the Full Service later in the rollout (control) as the counterfactual for the change experienced by districts adopting it earlier (treated). The comparison over time enables DiD to eliminate district-specific, time-invariant confounders. The comparison across districts then eliminates bias from nation wide shocks. Together, the effect of UC can be isolated.

Key to DiD establishing causality in this setting are two assumptions. First, there can be no anticipatory mental health response to UC being launched. For example, patients demanding additional or new medications just before the programme's launch would violate this assumption (if the two are connected). This is needed so that periods observed before UC represent untreated potential outcomes. Note, however, that what can be accommodated are claimants having a mental health response in anticipation of claiming UC. This is because the treatment here is defined at the district level, not individual. Since UC's introduction only *enables* new claims to be made, the vast majority of its take-up occurs in subsequent months and years ahead. This distinction is likely important. It seems plausible that the anticipation of making a new claim to UC could induce anxiety; it is arguably less so for the launch date. My analysis will test the robustness of this assumption regardless.

The second identifying assumption is parallel trends. That is, early UC adopters would have experienced the same average trend in mental health outcomes as the late adopters, had they not in fact adopted the policy. To increase the liklihood of this holding, the *timing* of when districts implement UC needs to be exogenous to mental health trends. The case for this would be undermined, for example, if the rollout was designed to minimise UC's caseload while the programme was in its infancy, perhaps first launching in areas where unemployment was trending downwards.

While the rationale for the order of the rollout is not in the public domain, I argue that the timing of when districts were treated can be considered as quasi-random. This stems from the fact that the rollout was highly chaotic, with variation in treatment timing largely a function of delays by the DWP. These delays were in fact substantial and frequent. In 5 successive years between 2014-2018, the DWP pushed back the completion of the Full Service rollout (twice it did so by two years) (Timmins, 2016). Importantly, these setbacks were not due to changing socio-economic conditions. Rather, they have been acknowledged as being a result of two factors internal to the DWP.

First, the DWP had been struggling with a 'learning-by-doing' approach to delivering UC, having little to no experience with this style of project implementation. This led to a lack of a coherent blueprint for the programme which in turn resulted in it having to be rebuilt from scratch in 2014 (and having £34 million in IT assets written off) National Audit Office (2013). Second, stable governance was a persistent issue for UC. Usually UK civil service projects are assigned a senior responsible owner for the duration of the project (Timmins, 2016). By contrast, UC had 6, with at one point 5 within a single year (Timmins, 2016). Taken together, these administrative and governance related factors suggest that the delays, and thus rollout timing, represent an exogenous intervention, at least with respect to mental health.

3.5.2 Testing Rollout Exogeneity

To empirically test the rollout exogeneity, I check whether local authority characteristics help explain the timing of when districts implement the UC Full Service. To do so, I estimate using OLS the bivariate regression:

$$RolloutDate_i = \beta_1 \Delta LocalChar_i + \epsilon_i \tag{3.1}$$

where $Rolldate_i$ is the month-year local authority i implemented the UC Full Service; $LocalChar_i$ is one of the following local authority characteristics: unemployment rate, number of legacy benefit claimants, ethnicity, gender, age demographics; and ν_i is the unexplained error. The characteristics in $LocalChar_i$ represent potential determinants of wellbeing. Finding an association with the rollout timing would therefore undermine the plausibility of exogeneity holding. For a more direct measure of wellbeing, I also run the following bivariate regression

$$RolloutDate_i = \beta_2 \Delta MentalHealthPresc_i + \nu_i \tag{3.2}$$

where $MentalHealthPresc_i$ represents the GP prescribing per 1,000 patients of one of the following mental health medications: antidepressants, anxiety medication, opioid painkillers, non-opioid pain killers, treatments for alcohol/nicotine addiction; and ν_i is the unexplained error. Both sets of regressions employ changes in the independent variable's value between 2013 and 2014, the closest years prior to the rollout. This focus on the change, rather than levels, enables a more direct test of the exogeneity assumption in question, i.e. no correlation with *trends* in mental health.

Table 3.3, Panel A shows each local authority characteristic has no statistically significant effect on the month-year a district adopts the UC Full Service. This suggests the rollout timing is quasi random with respect to local economic and demographic factors. Interestingly, Column (2) shows that even changes in the legacy benefit caseload are unrelated. This provides assurances that areas experiencing negative economic shocks, or areas with Job Centres working close to capacity, were not withheld from the early stages of the rollout. Panel B similarly suggests there is no relationship with GP prescribing rates. This supports the main hypothesis that the rollout was uncorrelated with mental heath trends

3.5.3 Estimating the Mental Health Effects of UC

To implement DiD and thus estimate UC's effect on mental health, I employ the estimator proposed in Callaway and Sant'Anna (2021). This estimator extends the canonical DiD model with 2 groups (a treated and control) and 2 time periods (pre and post) to a setting with multiple time periods and staggered treatment adoption - as is this setting. It does so by estimating an average treatment effect on the treated (ATT) for every group g and time t period in the panel, where a group is defined as a cohort of units i (local authorities) implementing the treatment in the same period g (month-year). The advantage of this approach is that it allows treatment effects to be heterogeneous - that is, the group-time ATT can very by g and time t (henceforth $ATT(g,t)$). This contrasts the traditional means of implementing DiD, two-way fixed effects, which is biased in such settings (Goodman-Bacon, 2021).

In practice, the estimator computes each $ATT(g,t)$ by comparing the outcome evolution of group g from period $g-1$ to t to groups not-yet treated by period t. It then aggregates these 'mini' treatment effects in various ways to summarise the impact of the policy. In my main analysis, I aggregate to an overall ATT with weights commensurate to the sample share of each group in a given period. This provides a natural summary of the effect of implementing UC. In Section 3.7, I aggregate the $ATT(g,t)$ by relative time in order to explore how treatment effects evolve dynamically. This is particularly appropriate because of how the UC Full Service increases the claimant count rapidly once launched.

For $ATT(g,t)$ estimates to be unbiased, the Callaway and Sant'Anna (2021) estimator requires the no anticipation and parallel trends assumptions to hold. Although fundamentally untestable, it is possible to probe their validity using the estimator. In particular, $ATT(g,t)$ estimates in pre-treatment periods can be viewed as a test of whether outcomes moved in parallel from period t to period g-1. Finding a zero and statistically insignificant effect here suggests that parallel trends holds. For periods closer to the treatment, a differential trend indicates a violation of the no anticipation assumption - for example, a "jump" in the estimate for period g-1 would suggest there was an adverse mental health in advance of UC. I examine these pre-treatment estimates via an event study plot in Section 3.7.

To increase the likelihood of parallel trends holding, my analysis conditions on covariates for some specifications. Specifically, I use the "doubly robust" procedure recommended in Callaway and Sant'Anna (2021). This involves using covariates in the following two ways.

First, it employs a regression adjustment procedure. Intuitively, this method isolates $ATT(g,t)$ by subtracting an estimate of treated groups' counterfactual change in outcomes from that actually observed by these groups. The counterfactual estimate is obtained by estimating the conditional expectation of the outcome change among control groups (using covariates X_i), and then applying the estimated parameters to the empirical distribution of X_i for treated groups (Roth et al., 2022). When the conditional expectation function is correctly modelled, the estimator identifies $ATT(g,t)$.

Second, the estimator uses covariates to conduct a two-step inverse probability weighting procedure. The first stage involves running a probit model to estimate propensity scores - that is, the probability of being treated in period t given X_i. In the second stage, units with similar propensity scores are given a greater weight when $ATT(g,t)$ is estimated. In essence, this method sidesteps the issue of correctly specifying a conditional expectation function and instead assumes parallel trends holds unconditionally among units deemed particularly similar.

If either the regression adjustment or propensity score method is specified correctly, the doubly robust method consistently estimates $ATT(g,t)$. This effectively gives two shots for unbiased estimation. In cases where both are correctly specified, the Callaway and Sant'Anna (2021) estimator achieves the semi-parametric efficiency lower bound. Thus, taken together, the Callaway and Sant'Anna (2021) estimator provides a strong balance between less stringent identifying assumptions and efficient estimation.

3.6 Results

In this section I present my estimates for the effect of the UC rollout on objective measures of mental health. I focus first on the impact on GP prescriptions, before turning to the effect on therapy usage and referrals, NHS secondary care mental health service, and finally suicides.

3.6.1 GP Prescribing

Table 3.4 displays UC estimates for all GP prescriptions under study. For each prescription there are three models presented, the contents of which are shown at the foot of the table. The first model is the baseline model with no controls. The second model includes the following demographic controls: % aged 16-24, % aged 25-49, % aged 50-64, % aged 65+, % White and UK national and % female over the aged of 16. It also includes the rate of antibiotic prescriptions as a means of controlling for variation in GP's propensity to prescribe both over time and across England. The third model further adds the unemployment rate and number of legacy benefit claimants. While estimates from Table 3.3 suggest these factors do not have predictive power for when local authorities adopt the UC Full Service, I include them in case parallel trends only holds conditional on their presence once the programme has been implemented. Standard errors are clustered at the local authority level throughout.

Panel A shows estimates for antidepressant and anxiety medication prescribing. Column (2), the model with demographics, suggests the implementation of the UC Full Service led to an additional 0.408 antidepressants prescribed per 1,000 GP patients. The estimate is statistically significant at the 5% significance level. However, the size of this effect is small: it represents a 0.7% increase in prescribing relative to the mean antidepressant prescribing rate (51.06). Column (3) shows it is also not robust to the inclusion of economic controls - the estimate reduces considerably to 0.02 and is insignificant at any reasonable statistical level. I therefore conclude there is no robust evidence of UC causing an increase in these prescriptions. Turning to anxiety medication in Columns (4)-(6), all UC estimates are close to zero and statistically insignificant, suggesting there was no change in their prescribing attributable to the programme. As with the antidepressant analysis, these zero effects are precisely estimated and not a consequence of a lack of statistical power. The largest pairing of standard error and UC estimates, found in Column (5), can reject at the 5% significance level an increase in anxiety medication prescribing of 0.282 percentage points or greater. Thus, this model, the most conservative of the three, can this rule out a 1.2% increase (or above) in the prescribing rate relative to its mean (22.79).

Panel B shows results for the prescribing of painkillers (analgesics). Column (1)-(3) presents estimates for nonopioids, the weaker type of painkillers generally used for musculoskeletal conditions (e.g. paracetamol and aspirin). Across the three specifications the models produce precisely estimated zero and insignificant effects. Consistent with the results for antidepressant prescribing, controlling for the unemployment rate and number of legacy claimants greatly improves estimation precision yet also reduces the size of the UC estimate by a greater degree. Estimates for opioid forms of painkillers are also approximately zero and insignificant.

Lastly, Panel C documents estimates for substance abuse medication. Columns (1) and (2) show statistically significant and relatively large UC effects for prescriptions related to alcohol addiction. Surprisingly, the estimates are negatively signed and in the opposite direction to those for antidepressant prescribing. The raw DiD estimate of -0.017 in Column (1) is statistically significant at the 1% level and represents a reduction in prescribing by 8.1% relative to the mean prescribing rate (0.21). The estimate in Table 3 however shows this outcome is correlated with the rollout timing, thus the estimate in Column (1) may be biased. Regardless, Column (3) shows that UC's effect on alcohol-related prescriptions is not robust to the inclusion of economic controls; the estimate is reduced to a negligible magnitude, -0.001, and is no longer statistically significant at even the 10% level. Estimates for treatments for nicotine addiction are also approximately zero effects and insignificant across the three specifications. Column (3), the model with the full set of controls, can rule out a 2.4% increase in these prescriptions.

3.6.2 Therapy for Mental Health and Substance Abuse

I next explore whether UC has led to an increase in the number of people being referred to IAPT. As discussed in Section 3.2.2, IAPT is a talking therapy programme that typically represents the first line of treatment for patients exhibiting depressive symptoms or suffering from anxiety. I also explore in this section UC's effect on the number of people newly presenting themselves to providers of substance abuse rehabilitation programmes. These programmes usually include therapy sessions but may also include medication or NHS secondary care.

Column (1)-(3) in Table 3.5 show UC estimates for IAPT referrals. Column (1), the baseline model with no controls, estimates a statistically insignificant impact of -0.023. Column (2) adds the age, gender and ethnicity controls and produces a positive estimate of 0.012, again statistically insignificant. The inclusion of the unemployment rate and the number of legacy claimants in Column (3) does not change the size of the estimate and only very marginally changes the standard error. These insignificant estimates are less precisely estimated than that of the GP prescribing in Table 3.4. For this analysis, the most conservative model, Column (3), can rule out a 5.1% increase in referrals to the therapy programme.

Equivalent specifications for estimates of new clients to drug abuse treatments, Columns (4)-(6), are also statistically insignificant at any reasonable significance level. They are also negatively signed, consistent with the GP prescription estimates for substance abuse in Table 3.4. Estimates from Column (6), the model with the full set of controls, indicate that an increase in new clients above 4.8% of the mean (0.36) can be ruled out at the 5% significance level.

3.6.3 NHS Secondary Care Mental Health Services

So far, I have found no compelling evidence to suggest that UC results in changes in mental health-related GP prescriptions or referrals/uptake in therapy or rehabilitation programmes. I now turn to outcomes within the NHS secondary care for mental health, which are more likely to capture UC's impact on people with more serious psychological conditions. Consistent with the previous analysis, I present three models for each outcome: a baseline model, a model with age, ethnicity, and gender demographics, and a model that incorporates economic controls. The dependent variables are expressed as rates per 1,000 population, and the analysis is conducted at the CCG level. Since all CCGs have a nonzero UC claimant count by the start of sample, I define the treatment as turning "on" when least 1% of the population are claiming the benefit and remaining "off" otherwise. Standard errors are clustered at the CCG level.

Columns (1)-(3) in Table 3.6 suggest UC has led to an increase in open referrals to NHS secondary care mental health services among the working age population. Column (1) estimates an increase of 1.037 percentage points and is statistically significant at the 1% significance level. Adding demographic controls in Column (2) reduces the estimate size to 0.688 and is significant at the 5% significance level. Column (3) adjusts for differences in the unemployment rate and legacy claimant count across control and treatment groups. It produces a point estimate of 0.619 and a standard error of 0.346, thus statistically significant at the 10% level. On average, there are 18.25 open referrals to these services per 1,000 population, so this estimate translates to an average increase of 3.4%.

Columns (4)-(6) suggest this finding is not attributable to new referrals to mental health services. The estimate for the baseline model in Column (4) is 0.137, which is reduced in magnitude to 0.060 and then -0.002 once demographic and economic controls are included, respectively. All models are statistically insignificant. Column (6), the model with the full set of controls, rules out increases in new referrals above 3.7%. This upper threshold is low enough to be informative given the UC estimates for the "stock" of referrals shown in Columns (1)-(3). In particular, I would expect the effect on new referrals (i.e. the "inflow") to exceed this figure if the estimates in Column (1)-(3) were being driven by new people being referred. In light of this, the estimates in Columns (1)-(3) may only be interpreted as UC prolonging the treatment period for patients with *pre-existing* referrals.

Columns (7)-(9) show no evidence that UC had led to an increase in the number of people being hospitalised due to mental health problems. These estimates are imprecisely estimated however.

Column (9), for example, can only rule out increases in the hospitalisation rate by 6.9%.

3.6.4 Suicides

Lastly, I explore UC's impact on suicide rates per 100,000 population. This analysis is at the local authority level and thus can utilise the UC dummy variable that equals 1 when the Full Service has been implemented and zero otherwise. Consistent with most other outcomes in the analysis, Table 3.7 shows no evidence of UC leading to an increase in suicide rates. The estimates are in fact negatively signed and, for Column (1)-(2), even statistically significant (or borderline) at the 10% level. Adding economic controls produces an estimates of 0.299. Taken with the standard error, this model rules out an increase in suicides of 3.5% relative to the mean.

3.7 Robustness

3.7.1 Event study

In this section, I subject my analysis to several robustness checks. I start by assessing the plausibility of parallel trends and no anticipation, the two identifying assumptions underlying the Callaway and Sant'Anna estimator (2012). I focus mainly on testing their validity for open referrals to NHS secondary care mental health services (shown in Table 3.6), as this was the only outcome I found to be statistically significant across all three specifications in my results section (albeit at the 10% level in the final model). The Callaway and Sant'Anna (2021) event-study plot in Figure 3.5 provides a visual test for these assumptions. As discussed in Section 3.5, each pre-treatment estimate represents the residual of a long difference between the outcome in that period and the period prior to treatment, across both treated and not-yet treated areas. In doing so, they test for differential (non-parallel) trends prior to the treatment.

The pre-treatment estimates in Figure 3.5 appear to lend support to parallel trends holding. They show that, in the 13 months before CCGs received the treatment, the number of open referrals to secondary care mental health services evolved, on average, in parallel between areas receiving the treatment and areas receiving it later. In addition to being statistically insignificant at the 5% level, the estimates do not exhibit a trend consistently above or below the zero effect mark. Figure 3.5 also shows no sharp break in the pre-treatment estimates in the months just before the treatment. As such, there is no evidence of an increase in the number of open referrals in anticipation of UC being adopted locally. This suggest the second identifying assumption has also not been violated.

Beyond testing robustness, Figure 3.5 provides evidence of dynamic treatment effects. The post-treatment estimates suggest UC has had a growing effect on the rate of open referrals over time. This would be consistent with the idea of the caseload building over time once the Full Service has been launched. Estimates from the 6th post period month to the 27th (the last) are statistically significant. The estimate for the 27th relative time period suggests the cumulative effect of the policy over this period increased the number of open referrals by 2.7 per 1,000 population. This translates to a 14.8% increase relative to the mean (18.25).

In Appendix A I show the same plots for all other outcomes under study. For all of them, parallel trends and no anticipation appear to hold. For GP prescriptions in particular, my data enable me to show this for up to 3 years prior to UC. In regard to dynamic effects, the plots further suggest that the zero effects shown in Section 3.6 are not masking a treatment effect that takes time to accumulate. Indeed, the plots show small and insignificant treatment effects for all relative periods. The only exception to this is that for IAPT referrals. This plot shows a growing effect from UC from months 13 to 27 of an order of magnitude around 2.5% of the mean. This is surprising given the negative estimate for the overall average treatment effect shown in Table 3.5, Column (1). The discrepancy is likely due to the overall ATT in Column (1) putting relatively more weight on treatment effects over shorter horizons - by definition of a staggered rollout, there will be a greater share of treated groups observed in over short horizons than long. Regardless, since the IAPT estimates did not hold across the three specifications, this plot should be interpreted with caution.

3.7.2 Placebo

It is possible that the effect found in Table 3.6 for open referrals to secondary care could still be driven by omitted confounding variables. For example, it could be that the set of controls used for propensity score matching, as part of the Callaway and Sant'Anna (2021) estimator, were not extensive enough for parallel trends to hold conditionally in the post-period. The same reasoning holds for why a zero effect was estimated for GP prescriptions. To test this hypothesis, I conduct a placebo test for both sets of outcomes. The key idea here is to find an outcome plausibly subject to the same confounding factors yet also unlikely to be affected by UC.

Following Williams (2021), I use the prescribing of asthma inhalers as a placebo for the GP prescriptions I have studied. These products are highly unlikely to be affected by UC since asthma often starts in childhood and symptoms are generally triggered by allergies or changes in weather. Their prescribing, however, may be correlated with the prescribing for mental health treatments. For example, both are subject to NHS resource constraints. Further, asthma inhalers may capture latent patient demand for health care, particularly in regard to GP visits.

For an open referral placebo, I exploit the fact that UC is a working-age benefit and use data on the number of open referrals to the same mental health services for people aged 18 or under. Variation in the supply of these services is likely to highly correlated with the equivalent provision for those aged 18-65. While children could still conceivably be affected by UC, possibly via its impact on the finances or health of parents, I argue these indirect effects would be small given the lack of consistent evidence I have found for mental health effects on the population directly affected.

Table 3.8 displays UC's estimated effect on the placebos using the Callaway and Sant'Anna (2021) estimator and equivalent controls to my main analysis. As expected, Columns (1)-(3) show zero and statistically insignificant effects for asthma inhaler prescriptions. More importantly, however, the precision of the estimates rule out the possibility of confounders biasing my estimates to zero in Table 3.4 (that is, at least confounders shared between asthma inhalers and mental health medications). Column (3), for example, the most conservative estimate of the three specifications, places a lower bound of treatment effects to be -0.78 percentage points, which translates to -1.26% of the mean prescribing rate (61.95). This further supports the suggestive patterns shown in Appendix A that parallel trends holds in the post period.

Turning to open referrals for under 18s, estimates in Column (4)-(6) are also statistically insignificant at any reasonable significance level. However, they do not provide the same level of assurance as the placebo estimates for asthma inhalers. For one, the estimates are of a similar magnitude to that in Table 3.6 in the main analysis. For example, Column (4) of Table 3.8 shows an estimate of 1.32 additional open referrals per 1,000 people under 18, which is in fact larger than the 1.04 estimate it acts as a placebo for, shown in Table 3.6. What is more, a further comparison shows the standard errors to br around a factor of 10 larger in the placebo models. Consequently, the final model in Column (6), with a point estimate of 0.078, can only rule out increases in open referrals for children above 9.4% of the mean referral rate. Table 3.6 therefore does not provide strong evidence to rule out unaccounted for bias affecting open referral rates.

3.8 Discussion

My analysis has studied the effects of UC on a variety of objectively measurable mental health outcomes. In summary, I find no robust evidence of policy increasing demand for mental health care once its Full Service version has been implemented in communities. This lack of evidence was most strongly inferred within the analysis for NHS primary care, the first line of services for patients struggling with their wellbeing or exhibiting mild to moderate depressive symptoms. Specifically, no link was detected between UC and treatments for depression or anxiety, whether that be therapy or medication, demand for painkillers, substance abuse medication or drug rehabilitation programmes. Estimates for medication prescribing in particular were precise enough to rule out even tiny changes in their provision due to UC (ranging from 1% to 2% on average). As such, they do not support the anecdotal evidence or fears of UC increasing demand for GPs due to poor mental wellbeing (e.g. Arie, 2018; Walton, 2018).

At the severest end of the mental health spectrum, my analysis also found no evidence of UC leading to increased hospitalisations or suicidality. Estimates for the latter were able to rule out increases larger than 3.4%, thus not supporting concerns in a number of qualitative papers (Dwyer, 2018; Cheetham et al., 2019; Wright et al., 2022). I caution however that these estimates were derived from data with an annual frequency, which makes it difficult to exploit variation in suicides arising from the monthly UC rollout. Further, estimates for hospitalisations could only rule out changes in these rates by 6.9% or greater (relative to their average).

It is important to qualify that these results do not undermine the case others have made for UC negatively impacting wellbeing (see, Johnsen and Blenkinsopp (2018); Andersen (2020); Rabindrakumar and Dewar (2018)). Rather, it suggests UC has not led to claimants demanding additional treatment for these issues. This is an important distinction. It is clearly possible for UC to exacerbate mental health without leading claimants to require health care. Further, receiving treatment is a function of many factors, not just health - for example, time constraints, patient awareness, and doctors' propensity to misdiagnose. The most relevant factor in this context is likely time constraints. It is possible, for instance, that UC's stringent work-related conditionality may have induced a lower propensity to visit GPs or other providers. My data unfortunately does not enable me to test this. With that said, I argue such an effect would be unlikely to fully crowd out an increase in demand due to ill health. The precision of my estimates would also enable me to detect a true health effect even if the demand was partially offset. Thus, I believe it is reasonable to extrapolate my results to state that UC is unlikely to have led claimants to becoming *clinically* depressed, relative to what they would have experienced under the legacy system (i.e. the counterfactual).

My analysis produced one possible exception to this conclusion. Table 3.6 provided suggestive evidence of the programme leading to a 0.619 percentage point increase in the number of open referrals to NHS secondary care mental health services per 1,000 working-age people. This translates to a 3.4% relative to the mean open referral rate, a non-negligible increase. Noteably, these services are restricted to people who have received a GP referral and have typically not recovered following initial antidepressant or IAPT treatments. They therefore suggest UC exacerbated the mental health problems of those with more serious psychological impairments. My analysis did not support the hypothesis of the policy causing *new* referrals to these services. As such, UC could only reasonably be attributed to prolonging the need for secondary care treatment among people with *pre-existing* mental health conditions, based on this evidence.

It's worth noting that this finding does not necessarily contradict the lack of found evidence for an increase in demand for antidepressants (or primary care services more generally). The two treatments may capture distinct mental health problems, one acute and the other longer term. This is because a major component of secondary care includes care for people suffering with an immediate and severe mental health episode - this is provided by mental health "crises teams" and community mental health teams. Antidepressants, by contrast, are not intended for instantaneous relief, rather, they raise baseline levels of serotonin and are only effective following a sustained period of dosage National Institutional for Health and Care Excellence (2017). In this regard they provide a different type of treatment. In the context of UC, some of its most disruptive aspects are short term in nature - e.g. the five week waiting period or initial online claiming process - and thus the programme could conceivably lead to demand for acute mental health treatments without spilling into those for more longer term issues.

Nonetheless, these estimates for open referrals to secondary care must be interpreted with caution. This is for three reasons. First, the inclusion of the unemployment rate and legacy benefit claimants reduced the statistical significant to this outcome to the 10 % significance level. Second, estimates from the placebo test were too imprecise to provide a meaningful robustness check. Lastly, due to data being available from 2016, the UC dummy variable used for this analysis was defined as being 1 if more than 1% of a CCG's population were claiming the benefit. Ideally, this would have been any nonzero number of claimants.

3.9 Concluding Remarks

This chapter investigated the mental health impact of UC, a major and controversial UK welfare reform linked to causing hardship. In contrast to most studies to date, I explored this relationship within a quasi-experimental framework to provide credibly causal policy estimates. My analysis builds on the work Wickham et al. (2020) and Brewer et al. (2022), who also adopt a DiD framework to exploit UC's staggered rollout. I sought to advance this work in three respects. First, I expanded on the number of outcomes investigated to provide a fuller picture of UC's impact on mental health. The high frequency of these outcomes meant that their variation would be fully synchronised with UC's monthly roll-out. Second, I relaxed potentially important identifying assumptions - namely, treatment effect homogeneity and an absent labour supply response from UC being launched. Third, I tested whether the reported deterioration of subjective measures of wellbeing, maintained in Wickham et al. (2020) and Brewer et al. (2022), translated to objective indicators. These measures, relating to primary care demand, secondary care demand, and suicides, in principal correspond to a spectrum of depressive severity, which in turn provides greater insight into the extent UC has exacerbated wellbeing.

In summary, my analysis found no robust evidence of UC increasing demand for treatments for mild, moderate or severe forms of depression. Specifically, my estimates found no UC-related increase in the prescribing of antidepressants, talking therapy sessions or, at the extreme end, suicide rates. In addition, no evidence was found of UC influencing the prescribing of treatments for poor wellbeing more broadly - namely, anxiety, pain or substance addiction. My results are robust across specifications and tests of underlying model assumptions, which provides assurances that the research design is credible.

My analysis found one potential outlier to this conclusion. Estimates in Table 3.6 found suggestive evidence of the policy leading to 3.4% increase in the number of open referrals to NHS secondary care mental health services. These services require a referral from a GP and are typically for people who have been treatment-resistant or represent severe cases. My analysis did not find an increase in new referrals to these services, which suggests UC, based on these estimates, could only reasonably be attributed to pro-longing the need for care for those with *pre-existing* mental health issues. I maintain that this finding should be interpreted with caution: this outcome was statistically significant at the 10% level in the final specification, and its placebo estimate was of a similar magnitude.

My results provide, in general, a more positive outlook of UC than that of previous evidence. I highlight however that this does not undermine the findings in qualitative studies (see, (Dwyer, 2018; Cheetham et al., 2019; Wright et al., 2022)). UC can be a source of emotional strain without leading to clinical levels of depression or the need for mental health treatment more generally. One possibility is that UC, through intensifying job-search requirements, imposes time constraints on claimants such that it reduces attendance at IAPT sessions or visits to the pharmacy or GP practice. In this scenario my estimates would understate the policy's true impact. I argue, however, that this is unlikely to be the case, as even marginal changes in these outcomes were not found.

Table 3.1: Symptoms of Clinical Depression

	DSM-IV major/minor depressive disorder
1	Depressed mood by self-report or observation made by others*
2	Loss of interest or pleasure*
3	Fatigue/loss of energy
4	Worthlessness/excessive or inappropriate guilt
5	Recurrent thoughts of death, suicidal thoughts
6	Diminished ability to think/concentrate or indecisiveness
7	Psychomotor agitation or retardation
8	Insomnia/hypersomnia
9	Significant appetite and/or weight loss

Notes: This table lists the 9 depressive symptoms of clinical depression listed in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV), reported in National Institutional for Health and Care Excellence (2017). A patient is diagnosed with a depressive disorder if they self report at least 5 or more of these symptoms, including at least one * symptom.

Table 3.2: Summary Statistics

	Mean	Std Dev	Min	Max	N
GP Prescriptions per 1,000 Patients					
Antidepressants	51.06	14.57	16.04	111.46	24,776
Anxiety medication	22.79	7.38	5.82	56.53	24,776
Opioid painkillers	34.16	14.09	8.14	97.78	24,776
Non opioid painkillers	56.04	19.03	12.93	136.49	24,776
Nicotine addiction medication	1.60	0.97	0.01	10.25	24,776
Alcohol addiction medication	0.21	0.19	0.01	2.03	24,776
Antibiotics	47.96	9.40	16.93	92.52	24,776
Therapy Referrals and Usage per 1,000 Population					
IAPT referrals	2.66	1.00	0.00	9.31	12,887
New clients for substance abuse treatments	0.36	0.21	0.00	2.76	8,288
NHS Secondary Care Usage per w/age 1,000 Population					
Open referrals with mental health services	18.25	8.37	0.35	185.69	7,973
New referrals to mental health services	4.93	1.98	0.05	54.95	7,973
Open hospital spells (mental health related)	0.48	0.23	0.00	1.52	7,973
Suidices per 100,000 Population					
Suicides	8.83	3.36	0.00	43.63	23,256
Demographics					
GP patients in local authority	165,567	121,664	2,119	1,268,933	24,776
Local authority population	173,165	118,189	2,224	1,141,816	24,776
% Female & aged 16+	51.15	1.41	38.50	100.00	24,776
% Aged 16-25	13.24	2.98	3.30	26.90	24,776
% Aged 25-49	40.82	7.18	14.20	78.30	24,776
% Aged 50-64	23.45	3.68	8.80	56.30	24,776
% Aged 65+	22.53	5.52	5.18	42.30	24,776
Economic					
Unemployment rate	5.48	2.42	1.10	17.70	24,776
Economically inactive	20.92	4.67	7.00	76.00	24,776

Notes: This tables shows summary statistics for all variables used in the analysis. IAPT is an acronym for Improving Access to Psychological Therapy. In some areas, the registered GP population exceeds the estimated overall populations. Official guidance on this states that this is likely due to administrative error from GPs, as well as there being different definitions (between ONS and GPs) about who counts as a 'resident' in the UK (House of Commons Library, 2016).

Rates for IAPT are constructed using per 1,000 population aged 18-65. Rates of new clients for substance abuse treatments are constructed using per 1,000 population aged 18+.

Table 3.3: Rollout Endogeneity Check

Panel A: Economic and Demographic Trends	(1)	(2)	(3)	(4)	(5)	(6)
Δ Unemployment rate	0.110					
	(0.410)					
Δ Legacy claimants		0.00003				
		(0.930)				
Δ % Aged 16-24			-0.039			
			(-0.160)			
Δ % Aged 25-49				-0.055		
				(-0.320)		
Δ % Aged 50-65					0.206	
					(0.690)	
Δ % Female & aged 16+						-0.505
						(-2.640)
Observations	326	326	326	326	326	326

Panel B: GP Prescribing Trends	(1)	(2)	(3)	(4)	(5)	(6)
Δ Antidepressants	0.006					
	(0.012)					
Δ Anxiety medication		0.001				
		(0.020)				
Δ Opioid painkillers			-0.004			
			(0.012)			
Δ Nonopioid painkillers				-0.001		
				(0.007)		
Δ Alcohol medication					0.773	
					(0.376)	
Δ Nicotine medication						0.087
						(0.096)
Observations	326	326	326	326	326	326

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.10$. Heteroscedasticity-robust standard errors adjusted for clustering at the local authority level are reported in parentheses.

This table shows results from estimating bivariate, cross-sectional regressions that use the month-year of the Full Service's implementation as the dependant variable. The independent variables are first-differenced in order to test whether changes in their values have predictive power for when the Full Service was implemented. I take the first difference between the year 2013 and 2014 as this represents the earliest pair of consecutive years during which no local authority had adopted the Full Service.

Table 3.4: GP Prescribing

Panel A: Depression and Anxiety						
	Antidepressants			Anxiety medication		
	(1)	(2)	(3)	(4)	(5)	(6)
UC	0.296	0.408**	0.020	−0.049	0.086	0.062
	(0.219)	(0.207)	(0.084)	(0.149)	(0.100)	(0.106)
Mean	51.06	51.06	51.06	22.79	22.79	22.79
Observations	24,776	24,776	24,776	24,776	24,776	24,776
Unit of analysis	LA	LA	LA	LA	LA	LA
Panel B: Painkillers						
	Non-opioids			Opioids		
	(1)	(2)	(3)	(4)	(5)	(6)
UC	−0.044	0.031	0.003	−0.021	0.028	0.114
	(0.250)	(0.240)	(0.097)	(0.165)	(0.160)	(0.097)
Mean	56.04	56.04	56.04	34.16	34.16	34.16
Observations	24,776	24,776	24,776	24,776	24,776	24,776
Unit of analysis	LA	LA	LA	LA	LA	LA
Panel C: Substance Abuse						
	Alcohol addiction			Nicotine addiction		
	(1)	(2)	(3)	(4)	(5)	(6)
UC	−0.017***	−0.016***	−0.001	−0.002	−0.001	−0.003
	(0.006)	(0.006)	(0.004)	(0.042)	(0.042)	(0.019)
Mean	0.21	0.21	0.21	1.60	1.60	1.60
Observations	24,776	24,776	24,776	24,776	24,776	24,776
Unit of analysis	LA	LA	LA	LA	LA	LA
Demographics		Yes	Yes		Yes	Yes
Antibiotic prescribing		Yes	Yes		Yes	Yes
Economic controls			Yes			Yes

Notes: *** p<0.01, ** p<0.05, * p<0.10. Heteroscedasticity-robust standard errors adjusted for clustering at the local authority level are reported in parentheses.

This table shows the results from using the Callaway and Sant'Anna (2021) estimator. The binary UC variable that takes on the value of 1 when UC has been implemented in a local authority and 0 otherwise. Each model is estimated from a panel dataset at the local authority (LA) level and uses dependent variables that are expressed as rates per 1,000 GP patients.

For each outcome there are three models. The first is the baseline model that includes no controls and assumes parallel trends holds unconditionally. The second includes demographics (% aged 16-24, % aged 25-49, % aged 50-64, % aged over 65, % female and aged over 16) and antibiotic prescribing per 1,000 GP patients. The third adds economic controls (% the unemployment rate and number of legacy claimants). The Callaway and Sant'Anna (2021) estimator uses the period 1 values of these controls to estimates a group's propensity score. It then places greater weight on comparisons with similar propensity score when estimating ATT(g,t).

Table 3.5: Therapy Referrals and Usage

	IAPT referrals			New clients for drug abuse treatments		
	(1)	(2)	(3)	(4)	(5)	(6)
UC	−0.023	0.012	0.012	−0.005	−0.011	−0.018
	(0.074)	(0.062)	(0.064)	(0.015)	(0.018)	(0.018)
Mean	2.66	2.66	2.66	0.36	0.36	0.36
Observations	12,887	12,887	12,887	7,669	7,669	7,669
Unit of analysis	CCG	CCG	CCG	LA	LA	LA
Demographics		Yes	Yes		Yes	Yes
Economic controls			Yes			Yes

Notes: *** p<0.01, ** p<0.05, * p<0.10. Heteroscedasticity-robust standard errors adjusted for clustering at the local authority level are reported in parentheses.

This table shows the results for referrals to the NHS Instance Access to Pychological Therapy programme (IAPT) and the number of people newly presenting to rehabilitation programmes for substance abuse. The estimates are produced using the Callaway and Sant'Anna (2021) estimator. The binary UC variable that takes on the value of 1 when UC has been implemented in a district and 0 otherwise. For the IAPT analysis, the relevent district is the clinical commissioning group (CCG) level. For the new clients for drug abuse treatments analysis, the relevent district is the local authority (LA) level. The dependent variable for each model is expressed as a rate per 1,000 population.

For each outcome there are three models. The first is the baseline model that includes no controls and assumes parallel trends holds unconditionally. The second includes demographics (% aged 16-24, % aged 25-49, % aged 50-64, % aged over 65, % female and aged over 16) and the third add economic controls (% the unemployment rate and number of legacy claimants). The Callaway and Sant'Anna (2021) estimator uses the period 1 values of these controls to estimates a group's propensity score. It then places greater weight on comparisons with similar propensity scores when estimating ATT(g,t).

Table 3.6: NHS Secondary Care for Mental Health

	Open referrals			New referrals			Open hospital spells		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
UC	1.037***	0.688**	0.619*	0.137	0.060	−0.016	0.004	0.0003	−0.00007
	(0.367)	(0.331)	(0.346)	(0.094)	(0.094)	(0.100)	(0.013)	(0.012)	(0.017)
Mean	18.25	18.25	18.25	4.93	4.93	4.93	0.48	0.48	0.48
Observations	7,973	7,973	7,973	7,973	7,973	7,973	7,973	7,973	7,973
Demographics		Yes	Yes		Yes	Yes		Yes	Yes
Economic controls			Yes			Yes			Yes

Notes: *** p<0.01, ** p<0.05, * p<0.10. Heteroscedasticity-robust standard errors adjusted for clustering at the local authority level are reported in parentheses.

This table shows the results from using theCallaway and Sant'Anna (2021) estimator. The binary UC variable that takes on the value of 1 when UC has been implemented in a local authority and 0 otherwise. Each model is estimated from a panel dataset at the local authority (LA) level and uses dependent variables that are expressed as rates per 1,000 population.

The outcome "Open referrals" represents a measure of the number of people with an open referral to use these services. The outcome "New referrals" represents the number of new referrals to NHS secondary care mental health services within each month. These new referrals may be for people interacting with NHS secondary care for the first time or for existing users (i.e. contacts) who are referred to additional or alternative NHS secondary care services. The outcome "Open hospital spells" represents the number of people either occupying a bed in a mental health hospital or has a bed open for them as part of ongoing period of leave.

For each outcome there are three models. The first is the baseline model that includes no controls and assumes parallel trends holds unconditionally. The second includes demographics (% aged 16-24, % aged 25-49, % aged 50-64, % aged over 65, % female and aged over 16) and the third add economic controls (% the unemployment rate and number of legacy claimants). The Callaway and Sant'Anna (2021) estimator uses the period 1 values of these controls to estimates a group's propensity score. It then places greater weight on comparisons with similar propensity scores when estimating ATT(g,t).

Table 3.7: Suicides

	Suicides		
	(1)	(2)	(3)
UC	-0.448^{*}	-0.422	-0.299
	(0.242)	(0.257)	(0.309)
Mean	8.83	8.83	8.83
Observation	23,466	23,466	23,466
Unit of analysis	LA	LA	LA
Demographics		Yes	Yes
Economic controls			Yes

Notes: *** p<0.01, ** p<0.05, * p<0.10. Heteroscedasticity-robust standard errors adjusted for clustering at the local authority level are reported in parentheses.

This table shows the results from using the Callaway and Sant'Anna (2021) estimator. The binary UC variable that takes on the value of 1 when UC has been implemented in a local authority and 0 otherwise. Each model uses suicides per 100,000 population as the dependent variables and is estimated from a panel dataset at the local authority (LA) level.

The first model is the baseline model that includes no controls and assumes parallel trends holds unconditionally. The second model includes demographics (% aged 16-24, % aged 25-49, % aged 50-64, % aged over 65, % female and aged over 16) and the third model add economic controls (% the unemployment rate and number of legacy claimants). The Callaway and Sant'Anna (2021) estimator uses the period 1 values of these controls to estimates a group's propensity score. It then places greater weight on comparisons with similar propensity scores when estimating ATT(g,t).

Table 3.8: Placebos

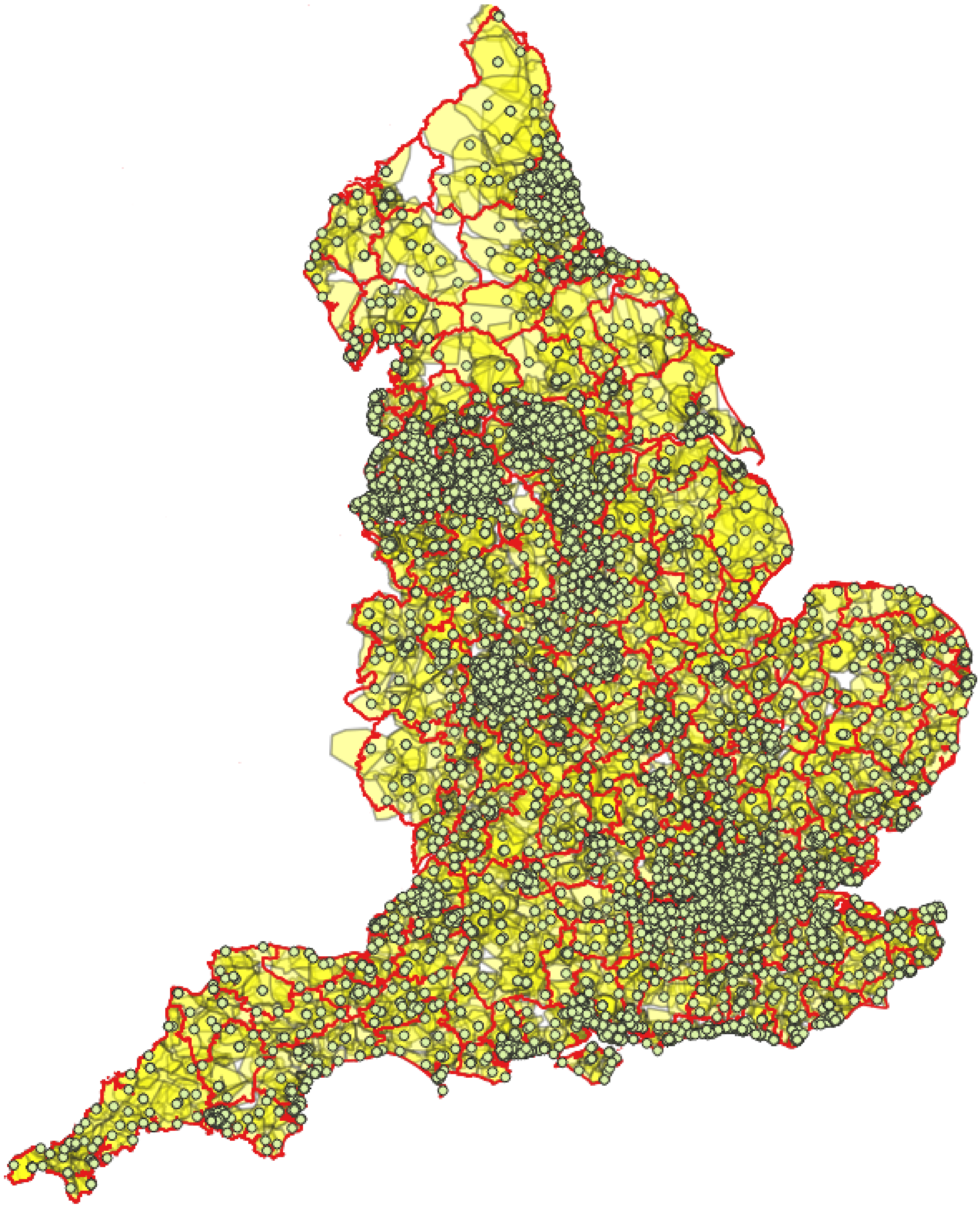
	Asthma inhalers			Open referrals (aged under 18)		
	(1)	(2)	(3)	(4)	(5)	(6)
UC	−0.148	−0.092	−0.316	1.320	1.160	0.078
	(0.187)	(0.196)	(0.237)	(0.879)	(0.973)	(1.02)
Mean	61.95	61.95	61.95	22.06	22.06	22.06
Observations	24,776	24,776	24,776	7,669	7,669	7,669
Unit of analysis	LA	LA	LA	CCG	CCG	CCG
Demographics		Yes	Yes		Yes	Yes
Economic controls			Yes			Yes

Notes: *** p<0.01, ** p<0.05, * p<0.10. Heteroscedasticity-robust standard errors adjusted for clustering at the local authority level are reported in parentheses.

This table shows the placebo results from using the Callaway and Sant'Anna (2021) estimator. The binary UC variable that takes on the value of 1 when UC has been implemented in a local authority and 0 otherwise. Asthma inhaler prescriptions per 1,000 GP patients are used as a placebo test for the GP prescribing outcomes shown in Table 3.4. Open referrals to NHS secondary care mental health services by people aged under 18 is used as a placebo test for equivalent open referrals by people of working age (shown in Table 3.6). This placebo is expressed as a rate per 1,000 people aged under 18.

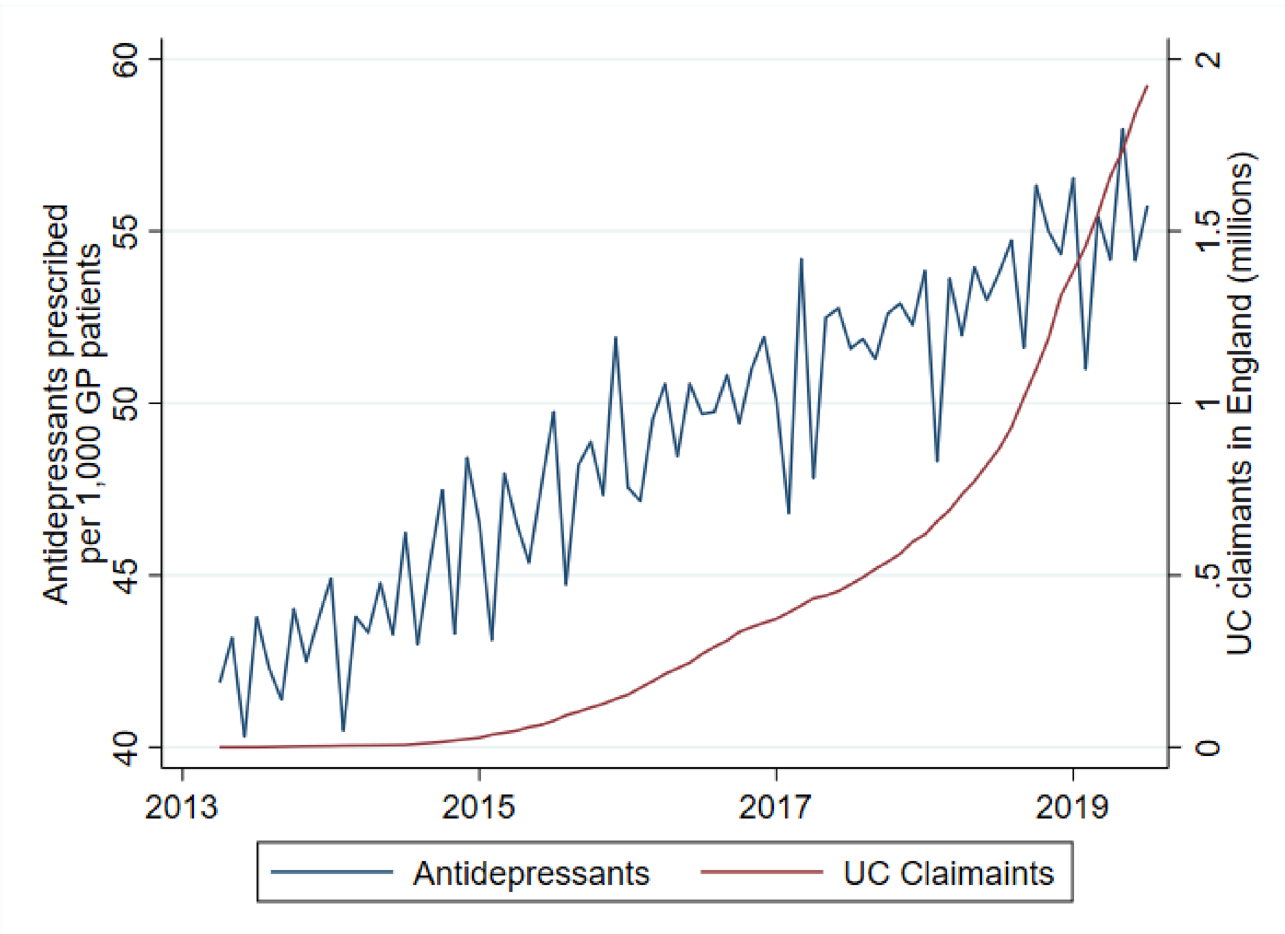
The first model is the baseline model that includes no controls and assumes parallel trends holds unconditionally. The second model includes demographics (% aged 16-24, % aged 25-49, % aged 50-64, % aged over 65, % female and aged over 16) and the third model add economic controls (% the unemployment rate and number of legacy claimants). The Callaway and Sant'Anna (2021) estimator uses the period 1 values of these controls to estimates a group's propensity score. It then places greater weight on comparisons with similar propensity scores when estimating ATT(g,t).

Figure 3.1: Map of GP Practices and Catchment Areas



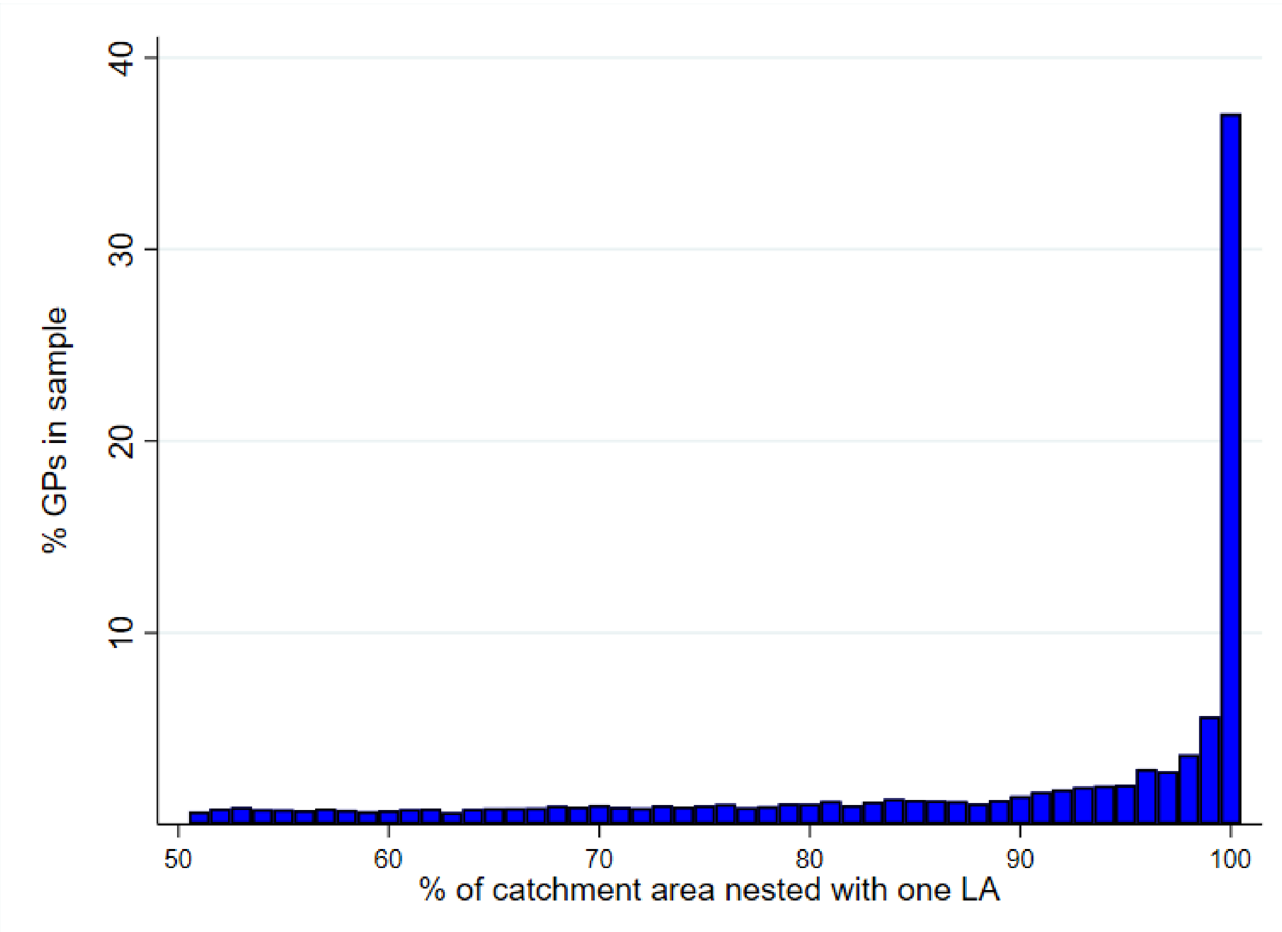
Notes: the light green dots show the location of GP practices, the yellow areas show GP catchment areas and the red boundaries are local authority districts. It can be seen from this figure that GP catchment areas do not conform to local authority boundaries

Figure 3.2: Antidepressant Prescribing and UC Claimants



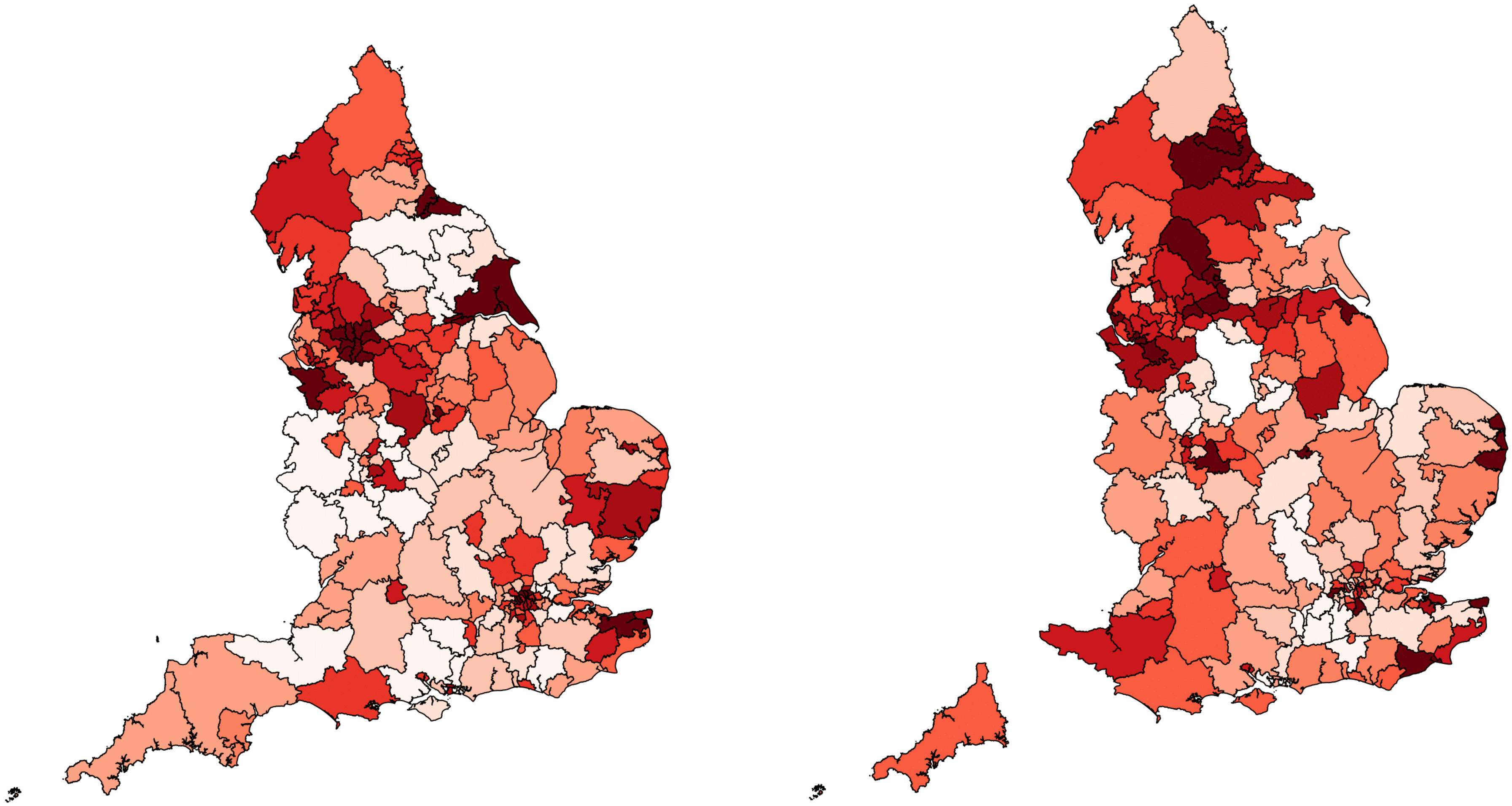
Notes: The count of antidepressants is given by the number of prescribed selective serotonin reuptake inhibitor (SSRIs) items. An item refers to a single supply of medicine that typically takes 1 month to exhaust.

Figure 3.3: Distribution of GP Catchment Area Nesting



Notes: This figure shows the distribution of GP catchment area nesting among the GP practices included in my analytic sample. Only GP practices with more than 50% of their catchment area nested within a single local authority are included in the analytic sample. The figure shows, for example, that 37% of GP practices in England have their boundary completely nested (i.e. 100%) within a unique local authority.

Figure 3.4: IAPT Referrals and UC claimants

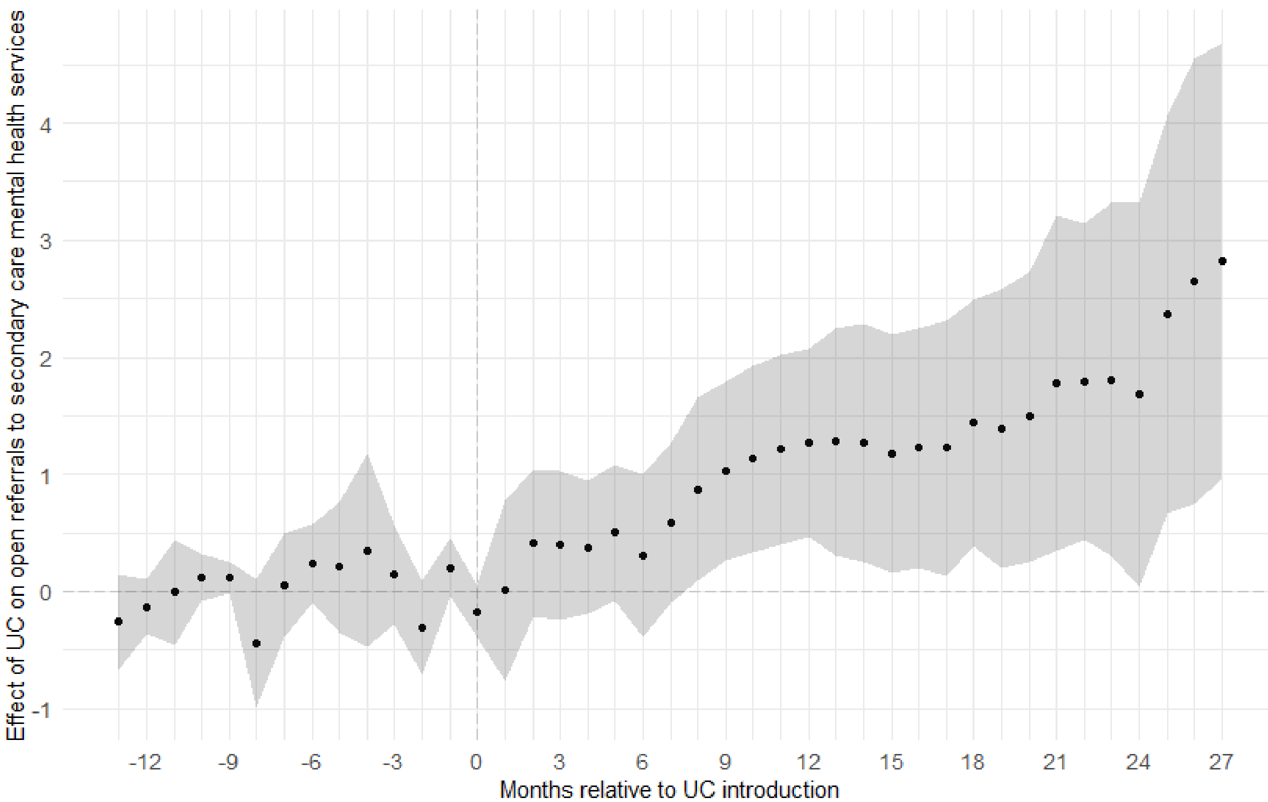


(a) IAPT referral rate

(b) UC claimant rate

Notes: This figure shows a map of 203 NHS clinical commissioning groups (CCG) in England. The left figure shows the distribution of referrals per 1,000 working-age population to the Instant Access to Psychological Therapy (IAPT) programme in England. The right figure shows the same for UC claimants. Colours represent deciles in the corresponding distribution in the year 2018. Claimant data were missing for Devon CCG

Figure 3.5: Event Study Plot - Open Referrals to NHS Secondary Care Mental Health Services



Notes: This figure shows the estimated impact of UC on open referrals to secondary care mental health services, using the Callaway and Sant'Anna (2021) estimator. Each dot represents a point estimate. Shaded area represents 95% confidence intervals.

Chapter 4

Punitive Welfare Reform and Crime

4.1 Introduction

Does punitive welfare reform cause crime? Such reforms are often legislated on the basis of increasing labour supply among the unemployed; a belief that has been substantiated in the empirical literature (Moffitt, 1985; Katz and Meyer, 1990; Van den Berg et al., 2004; Card et al., 2015). Nevertheless, Becker (1968) posited that crime occurs when its expected net benefits exceed those available from legal sources of income. From this perspective, welfare reforms may have important secondary effects on crime, in changing its returns relative to legal sources. An empirical evidence base shows that cuts in benefit entitlement do induce a crime response, and a smaller literature suggests so do more stringent job search requirements.[1] However, causal estimates on these matters almost exclusively pertain to subgroups, such as ex-offenders (Yang, 2017*a*; Tuttle, 2019; Munyo and Rossi, 2015), juveniles (Deshpande and Mueller-Smith, 2022; Andersen et al., 2019), immigrants (Andersen et al., 2019), and fired employees (Bennett and Ouazad, 2020).

A remaining question is therefore whether these findings translate to the working-age welfare population at large. Part of the empirical challenge in answering this question has been that large-scale

[1] For benefit entitlement, see: Yang (2017*a*); Tuttle (2019); Munyo and Rossi (2015); Deshpande and Mueller-Smith (2022); Andersen et al. (2019); Bennett and Ouazad (2020); Watson et al. (2020); Fishback et al. (2010). For job search requirements, see Bennett and Ouazad (2020) and Fallesen et al. (2018).

welfare reforms are rare, and when they do occur, it is less common still that they provide the quasi-random geographic variation needed to identify causal effects for broader populations. Nonetheless, identifying these effects is of clear importance given large overlaps between the offender and welfare population, and the strong negative externalities that crime and its consequences impose on society (see, for example, Aizer and Doyle Jr (2015); Czabanski (2008); McCollister et al. (2010)).[2]

Universal Credit (UC) provides an ideal and interesting opportunity to answer this question. The programme's twin-track rollout of the Live Service (LS) and Full Service (FS) across the UK meant that between April 2013 and December 2018, working-age welfare populations in some local authorities were subject to a relatively more punitive welfare regime, while others in different local authorities were not. Figure 4.1 shows that its introduction also coincided with the reversal of a downward trend in UK crime rates, with violent crime in fact doubling during the course of both rollouts.

In recent years, a handful of papers using two-way fixed effects (TWFE) and event study models have documented that the programme indeed played a role in exacerbating crime rates in communities once implemented. d'Este and Harvey (2022) estimates that UC caused an additional 35,000 burglaries and 25,000 vehicle crimes. Lim et al. (2020) document an increase in robberies (albeit they describe the evidence as "weak"). Within the sociology literature, Tiratelli et al. (2022) find that with every 10 additional UC claimants per 1,000 population, total crime increased by 0.26 crimes per 1,000 population. Their estimates also suggest UC contributed to the rise in violent crime illustrated in Figure 4.1.

In this chapter, I revisit this question using monthly police-recorded crime data between December 2010 and February 2020 in England and Wales. My novel contribution is that I employ new difference-in-difference (DiD) estimators that solve key problems recently uncovered with TWFE. It is worth highlighting three advantages of these estimators, and why their use would seem particularly important in this setting.

First, TWFE has been shown to be inconsistent for the average treatment effect on the treated (ATT) when there is variation in treatment timing and treatment effects are dynamic (Goodman-Bacon, 2021;

[2]In the UK, data from 2014 show that 22% of out-of-work benefit claims were made by individuals who had previously been cautioned or convicted of an offence at some point since 1st January 2000 (Ministry of Justice, 2014).

Borusyak et al., 2022; De Chaisemartin and d'Haultfoeuille, 2020). Here, UC's introduction not only varies by time, but its treatment effects on population level outcomes, such as local crime rates, are almost guaranteed to be dynamic: UC is only available to new claims, meaning its caseload builds from zero once launched. New DiD estimators, in contrast, do not rely on homogeneous treatment effects for identification.

Second, in the presence of multiple treatments, a TWFE estimate for one treatment is 'contaminated' by the effects of the others (de Chaisemartin and d'Haultfœuille, 2020). In this context, TWFE may therefore be biased by UC being implemented via two distinct rollouts, the LS and FS. My main analysis employs the de Chaisemartin and d'Haultfœuille (2020) estimator which is robust to the presence of multiple treatments, and thus both rollouts.

Third, new DiD estimators can provide formal tests of core identification assumptions - no anticipation and parallel trends - in contrast to event studies which have biased "pre-trends" when treatments are staggered (Sun and Abraham, 2021).

By employing the de Chaisemartin and d'Haultfœuille (2020) estimator, my analysis provides arguably the most credible UC criminological estimates to date. By disentangling the effects of the two UC rollouts, I am moreover able to contribute the first, separate criminological estimates of the FS - the version of UC in operation at present and that expanded the program's reach to millions of low-income UK residents. As such, this study broadens our understanding of UC's impact.

I find that using these estimators produces results that contradict the existing UC-crime literature. Specifically, I find no robust evidence of the LS rollout leading to an increase in offending in the period up to 18 months after its implementation. Conservative estimates from my main analysis are able to rule out (with 95% confidence) that the LS increased violent crime by more than 3.63%. I find some suggestive evidence of the LS increasing acquisitive types of crime, however these estimates are not robust to the inclusion of local authority specific time trends.

I similarly find that areas adopting the FS experienced no significant change in any form of crime in the 24 months following its adoption. Conservative estimates from this analysis are able to rule out (with 95% confidence) increases in violent crime by more than 9.4%, and acquisitive crime by more than 6.3%.

The remainder of this chapter is structured as follows. In the next section, I describe the conceptual framework and recap institutional details of the LS and FS, previously outlined in Chapter 2. I then review the relevant empirical literature in Section 4.3, introduce my data in Section 4.4 and my methodology in Section 4.5. In Section 4.6 I present my results on the effects of the LS (in Section 4.6.1) and the FS (in Section 4.6.2). In Section 4.7 I test the robustness of these findings. Section 4.8 concludes.

4.2 Background

4.2.1 Conceptual Framework

Becker (1968) and Ehrlich (1973) provide the building blocks for analysing crime from the perspective of a rational offender. According to Becker, individuals choose to engage in criminal activity when the expected utility of doing so outweighs that from legal sources of income. For example, an increase in expected illicit earnings would increase one's criminal propensity, ceteris paribus. Conversely, if the probability of being apprehended or the severity of punishment increases, individuals will be more likely to engage in the legal sector. Ehrlich (1973) extends this dichotomous choice between crime and work to consider a model where individuals optimally allocate time between the two, analogous to the work and leisure tradeoff in neoclassical labour supply theory.

These models suggest UC could affect crime through two mechanisms. First, the programme's harsher sanction regime and five-week waiting period reduces *legal income* among claimants, which Becker's model would predict increases crime. The reduction in legal income may be amplified in two further ways. The first is that claimants may struggle to adapt to UC's monthly payment schedule and expend their benefit too quickly. The second is that UC represents a regressive change in entitlement relative to legacy, meaning that claimants more likely to be on the margin of crime will see a reduction in their monthly legal income. Note, those on the margin may also face significant barriers to employment, making them less likely to benefit from UC financially rewarding work (through its reduced withdrawal rate), or, by extension, any direct employment effects.

Second, UC may affect criminal behavior by altering the amount of *time available* that claimants have for illicit activities, ala Ehrlich (1973). The direction of this effect is ambiguous, however. On the one hand, increased job-search may lock claimants into more time spent on legal activities, leaving less time for crime. On the other hand, UC's digital application process, sanction regime, work-related conditionality or less generous entitlement may deter individuals from claiming welfare, potentially increasing their time available for crime.[3] While withdrawals or exits from welfare due to these reasons will likely be rare compared to UC take-up, it remains theoretically unclear which effect will dominate since both groups likely have different baseline risks of offending. For instance, some individuals may exit welfare because increased job-search incurs a particularly high criminal opportunity cost.

The models of Becker (1968) and Ehrlich (1973) are less informative when it comes to non-acquisitive crime, which do not involve financial gain. Theory from criminology and sociology however suggests that these types of crimes may still be influenced by UC. The concept of *time available* as a factor that contributes to criminal behaviour, for example, has been studied by criminologists, who argue that increased leisure and a lack of daily structure can lead to more offending regardless of incentives, due to increased opportunity to do so (Felson, 1998). This mechanism may therefore affect both non-acquisitive and acquisitive crime, though again the direction of which is unclear due to UC reducing time available for some individuals (recipients) and potentially increasing it for others (welfare exits).

Merton's (1938) strain theory provides another useful framework for analysing how UC may impact crime. While Becker's model focuses on rational choice and acquisitive crime, Merton's theory emphasises the role of a particular type of psychological strain, namely, that which can arise when individuals' feel unable to achieve their goals due to social, economic, or political systems. According to Merton, crime is more likely to occur when individuals experience such strain. Agnew (1992) extends this theory to acknowledge how this effect may spill to non-acquisitive forms of crime as well. Applied in this context, UC's negative impact on individuals' finances, or constraints on individual behaviour, may create psychological strain and in turn increase the likelihood of both violent and

[3] A meta analysis by Card et al. (2007) shows that reductions in benefit payments leads to spikes in welfare exits.

financially motivated crime occurring.

In summary, predictions from these models suggest that UC may increase crime by reducing claimants' legal income and/or inducing psychological strain. This could be amplified or offset by the time availability mechanism given its ambiguous (theoretical) direction of effect.

4.2.2 Live Service and Full Service

In this study I evaluate the criminological effects of the LS and FS. As discussed in Section 2.2, the key difference between these two versions of UC was that the LS restricted UC's introduction to a small number of new claims, namely: single tenants with no dependent children or housing benefit requirements. The effect of these so-called "gateway" conditions meant that by the end of the LS roll-out, two-thirds of UC recipients were men, many of whom were under the age of 25 (100,000 out of 270,000).[4] The FS subsequently opened up UC to all new claims from work-less or low-income households who otherwise would have been eligible for the 6 legacy benefits the programme is replacing.

Figure 4.2 provides a visual of how the LS and FS were rolled out across England and Wales. It shows that the LS (blue) was first implemented in the North West of England, before rapidly being rolled out to virtually all local authorities by the end of 2015 (the final launch date was in April 2016). It further shows that the FS (red) was first piloted in Croydon in 2015, before being rolled out nationally and concluding in December 2018. Figure 4.3 sheds more light on the amount of time that the LS was in operation in local authorities before being replaced by the FS. It shows that the modal number of months the LS was in place was 30-34 months (2.5 to 2.8 years).

4.3 Literature Review

The determinants of crime have been extensively explored in empirical literature across various academic disciplines. Within economics, there has been a narrower focus on the role of incentives, often to test Becker's theoretical premise (for a review, see Draca and Machin (2015)). For this chapter, I

[4]The DWP's open database on benefit statistics, StatXplore, provides a breakdown by age and gender of the UC caseload.

provide a review of findings within the economics literature on the role of reduced earnings and time availability as being criminogenic - two potential mechanisms in the UC-crime relationship described in Section 4.2.1. The most credibly causal estimates on their effects come from the literature on unemployment and welfare reform, so I review these in turn. Finally, in Section 4.3.3, I summarise three papers investigating UC's impact on crime and outline my contribution relative to these.

4.3.1 Unemployment

Early research into the causal effect of unemployment on crime struggled to provide insight into the factors linking this relationship due to using area level (usually U.S. state) crime and unemployment data (Raphael and Winter-Ebmer, 2001; Gould et al., 2002; Öster and Agell, 2007; Fougère et al., 2009; Dix-Carneiro et al., 2018; Dell et al., 2019).[5]

In recent years, however, a new wave of studies have had greater success in this regard, by combining employee-employer administrative data with criminal records. Their starting point has been to exploit mass layoffs as an exogenous unemployment shock to individuals, with the basic intuition being that, by being involuntary, unexpected and (broadly) firm-wide, these shocks should be uncorrelated with individuals' propensity to offend (Rose, 2018; Britto et al., 2022; Bennett and Ouazad, 2020; Rege et al., 2019; Khanna et al., 2021).

Regarding mechanisms, three of these studies isolate the effect of reduced earnings on crime (Rose, 2018; Britto et al., 2022; Bennett and Ouazad, 2020). According to their estimates, a reduction in earnings leads to an increase in crime. For example, Rose (2018) exploits a kink in the unemployment insurance (UI) system in Washingston State, where benefit entitlement is calculated as a fraction of one's pre-job loss earnings up until a maximum earnings threshold. Comparing fired workers on either side of the earnings kink, Rose (2018) estimates a recidivism elasticity with respect to UI as -0.5 for ex-offenders. Britto et al. (2022) explore the effects of earnings in a broader context, focusing on the effects of UI for all displaced workers, not just ex-offenders. Consistent with Rose (2018), Britto

[5]These studies attempted to overcome issues of endogeneity by using Bartick type instruments, whereby local fixed characteristics are interacted with national level employment shocks. They typically find positive (albeit small) statically significant links between unemployment and crime.

et al. (2022) exploit a discontinuity in UI policy, however they do so with respect to eligibility criteria rather than entitlement - that is, they focus on the extensive margin of insurance. Estimates from their study suggest that, in Brazil, eligibility for UI offsets entirely the estimated 23% increase in offending likelihood that occurs from job lay off (within one year of dismissal). Notably, they find that this effect vanishes once benefits expire, which is consistent (albeit stronger) with the effect found in Bennett and Ouazad (2020) in a Danish context. Further evidence to support the earnings channel is found in Khanna et al. (2021). This study exploits differences in credit access that resulted from a Columbian law that enabled some retail and finance outlets to operate as commercial credit institutions, while others could not. Using distance to these outlets as an instrument, they find that access to credit completely mitigates the crime response to job displacement.

Bennett and Ouazad (2020) and Rege et al. (2019) investigate the hypothesis that increased time available may lead to more crime. Bennett and Ouazad (2020) study the effect of Denmark's Act on an Active Labor Market reform, which introduced benefit conditionality for a period of 3 years following job loss. They find that displaced workers committed significantly less crime during the period of conditional benefits than unconditional benefits, indicating that increased leisure does increase crime. This finding is supported in Rege et al. (2019) who examine the relationship by looking at temporal patterns of crime across days of the week. This study finds that the violent crime response from job loss is only statistically different from zero on weekdays, which they interpret as suggestive evidence of daily structure and routine (i.e. less time available) playing a role in mitigating crime. Both of these studies are consistent with findings from Jacob and Lefgren (2003) who, in a separate literature, concludes that the incapacitation effect of school plays a significant role in reducing crime. Taken together, evidence from these studies would suggest that UC's increased work-related conditionality may decrease crime among recipients.

4.3.2 Welfare Reform

A related strand of literature examines the effects of welfare reform on crime.[6] These studies often focus specifically on the role of financial assistance, complementing the aforementioned job-loss literature. They report surges in acquisitive crime for a number of groups when assistance is reduced, including ex-offenders (Yang, 2017*a*; Tuttle, 2019; Munyo and Rossi, 2015), immigrants (Andersen et al., 2019), juveniles (Deshpande and Mueller-Smith, 2022; Andersen et al., 2019; Chioda et al., 2016), as well as the broader population (Watson et al., 2020; Fishback et al., 2010). A common approach has been to exploit laws that withdraw or dramatically reduce entitlement for these populations. For example, Yang (2017*a*) and Tuttle (2019) explore the effect of the 1996 U.S. ban on felony drug offenders accessing Supplemental Nutrition Assistance Program (SNAP) benefits. Tuttle (2019) employs a regression discontinuity design using Florida's policy enactment date as the discontinuity. He finds that drug traffickers subject to the ban are 60% more likely to return to prison after release than drug traffickers who have access to SNAP. This effect size is much larger than that found in Yang (2017*a*), who employs a triple DiD design to exploit both state-level variation in policy opt outs and offenders not affected by the reform. Her estimates suggest that the ban increased recidivism by 13% relative to her control group.[7][8] Among all of these studies, Watson et al. (2020) employs the most similar identification strategy to this chapter by exploiting variation in the timing of a policy implementation. Their paper documents the criminological effects of universal basic income, finding that property crime decreased by 12% in the fortnight after the cash transfer (no change is found for violent crime).

A smaller number of studies focus on the wider features of benefit systems. One includes payment schedules. Foley (2011) shows that 12 U.S. cities operating a monthly SNAP payment schedule observe increases in acquisitive crime by 6% at the end of the month. This temporal pattern was not observed in other jurisdictions with staggered payments, suggesting that welfare recipients consume

[6]The previously mentioned job loss studies focus primarily on the causal effect of unemployment, and exploit natural experiments within existing welfare states rather those arising through reforms (with the exception of Bennett and Ouazad (2020).

[7]Due to statistical power issues in Tuttle (2019), an effect size of 10% is in fact within this study's 95% confidence intervals, which suggests his results may not be at such odds with Yang (2017*a*).

[8]Another study by Luallen et al. (2018) combines the regression discontinuity and DiD methods but finds that the SNAP ban had zero effect on reoffending.

their benefits too quickly and turn to illicit income sources as compensation. Similarly, Carr and Packham (2019) find that the transition to a staggered SNAP timetable in Illinois reduced crime and theft in grocery stores by 17.5% and 20.9%, respectively. A second benefit feature studied in Machin and Marie (2006) is the role of benefit sanctions. This study provides descriptive, quasi-experimental and qualitative evidence that the introduction of the UK Jobseeker's Allowance (JSA) benefit, which appreciably increased sanction severity, led to increases in crime in the police force areas most affected by the reform. Together, these studies add to the evidence that personal finances influence individuals' propensity to offend.

4.3.3 UC

Recently, three papers have contributed further to the welfare reform evidence base by evaluating the impact of UC on local crime rates in England and Wales (d'Este and Harvey, 2022; Tiratelli et al., 2022; Lim et al., 2020). These authors use variants of TWFE to exploit the phased rolled out of the programme within a DiD framework. Tiratelli et al. (2022) provide the first contribution to the sociology literature. They use police recorded crime data at the Community Safety Partnership (CSP) area to estimate a TWFE model where the treatment variable is the UC claimant rate per 1,000 population. Their results suggest that for every 10 additional claimants per 1,000 population, there is an increase of 0.26 crimes per 1,000 population. This would correspond to a 4.6% increase in total crime, based on their descriptive statistics presented in their Table 1. Their estimates show that property crime is mainly driving this finding, though they do also find increases in violent crime. They do not present an event study to test for parallel trends or no anticipation, two key identifying assumptions underpinning their analysis. by 0.26 crimes per 1,000 population In concurrent work d'Este and Harvey (2022), provide the first evidence on this matter to the economics literature. The authors use police-recorded crime data at the Westminster constituency level, a more granular level of geography than CSP. Rather than estimating the effect of the marginal claimant, they adopt a traditional DiD setup whereby areas are deemed treated once they adopt the policy. Since neither the LS nor FS rollout were conducted at the Westminster Constituency (WC) level, they determine treatment status based on the first non-zero number of UC claimants in each WC, in effect estimating the impact of the LS (since the first claimants were always on the LS). The authors detect no effect on total or violent crime, in contrast to Tiratelli et al. (2022). However, they find increases in burglary ranging from 2.3% to 3%, and increases in vehicle crime ranging from 1.9% to 4.3%. Their event study plots suggest these effects are persistent for around 3 years post policy adoption. "Back-of-the-envelope" extrapolations from the authors suggest UC caused an additional 35,000 burglaries and 25,000 vehicle crimes.

Finally, Lim et al. (2020) use UC as a case study to test the propositions of a theoretical crime model they develop. The authors take a bottom-up approach to constructing local crime rates, sourcing crime incidence data at the street-level and reverse geocoding their longitude and latitude coordinates to the county level (a higher level of geography than local authority, CSP or WC). Their analysis combines TWFE with an instrument for the UC claimant rate, namely, non-UC benefit expenditure. The authors find that UC may have decreased rates of weapons possession and public disorder, though also increased robbery.

My analysis departs from previous studies by using new DiD estimators. This enables me to provide two new contributions to the literature. First, I provide an arguably more robust evaluation of UC, since my estimators are unbiased under treatment effect heterogeneity and the presence of the programme's twin-track rollout. Note, in my methodology sections, I explain the issues with TWFE estimators that have been identified in the recent DiD literature. Second, I am able to provide the first, separate estimates of the impact of the FS, the version of UC in place to date and that expanded the programme's reach to millions of individuals. Analyses in Lim et al. (2020), Tiratelli et al. (2022) and d'Este and Harvey (2022) do not provide separate estimates for the FS.

4.4 Data

4.4.1 Crime

The main source of data for this chapter are month-year police-recorded crime rates at the local authority level from December 2010 to February 2020 in England and Wales. These data are imported from UKCrimeStats, an open data platform of the Economic Policy Centre think tank. Each month UKCrimeStats collects street-level crime data from the UK Home Office and geocodes their coordinates to various geographical hierarchies. Since a local authority level aggregation is not available, I import more granular data reported at the Lower Super Output Area (LSOA) and aggregate accordingly. LSOAs are small statistical areas that nest perfectly within local authority boundaries, meaning they can be aggregated in this respect without risk of resulting bias.

The UK Home Office - and thus UKCrime Stats - groups crimes into 14 categories. Of these, I focus in my main analysis on the category violent and sexual offences, as well as an aggregation of all groups to measure total crime. To test the hypothesis that UC has led to an increase in financially motivated crime, I generate a single acquisitive crime variable that encompasses the following 7 categories: burglary, robbery, shoplifting, vehicle crime, bike theft, theft from the person and other theft. During the period of my sample the Home Office made two changes to the reporting of some of these categories. In September 2011, shoplifting, bike theft and several non-acquisitive crime types were separated from the category "other crimes".[9] As a result, the acquisitive crime measure in my panel can only reasonably be recorded from this month onwards (including "other crimes" prior to this date would over estimate the number of acquisitive crimes). Fortunately, this change occurred at least 3 years before the vast majority of local authorities implemented the LS, meaning there is still sufficient duration to test for pre-trends for this crime type.[10] The second change, initiated in May 2013, involved separating bike thefts and theft from the person from the category "other theft". Since this recategorisation only involved acquisitive types of crime, it is not problematic for my constructed measure of acquisitive crimes (it is possible to use the category "other theft" prior to this date without obtaining an incorrect count).

As with any study on crime, there is always the concern of substantial measurement error existing between true levels of crime and that which is detected and recorded by authorities. This is a somewhat intractable but long understood problem that this study is unfortunately no different from being limited by. Using data from the British Crime Survey was considered for the analysis, however the

[9] Criminal damage and arson; drugs; weapons possession and public order.
[10] 94% of local authorities implemented the LS in 2015.

police-recorded data from UKCrime stats was ultimately preferable due to the former's high level of aggregation making it near impossible to exploit UC's geographic rollout.[11] As noted in Tiratelli et al. (2022), the British Crime Survey is also a victimisation survey and so excludes many crimes relevant in the current case (e.g. shoplifting). Choosing police-recorded crime here is consistent with the existing UC-crime literature.

4.4.2 UC

Rollout dates for the FS are published at the Job Centre and local authority level on the Department for Work and Pensions (DWP) website.[12] The timetable enables me to identify that 42 out of 348 local authorities contain Job Centres implementing the FS across different months, with this being due to some Job Centres serving more than one local authority. For these 42 areas, the FS treatment will not turn "on" in a unique month, so I create a dummy variable equalling 1 for these areas to exclude them later from the analysis as part of a robustness check.

LS rollout dates have not been published at the local authority level by the DWP, to the best of my knowledge. I therefore use UC caseload data from the DWP's open database "StatXplore" and define each local authority's LS treatment date as the first month-year in which a non-zero number of UC claimants were in that area.

4.4.3 Controls

My analysis uses a number of economic, demographic and welfare caseload controls at the local authority level. The economic controls include quarterly unemployment rates, median house prices and median weekly pay. Demographic controls include the percentage of individuals in a local authority aged between 16-24, 25-49, 50-64, over 65, as well as the percentage of females aged over 16. Both sets of these controls are sourced from NOMIS, an open data source provided by ONS. Counts of the number of people claiming Jobseekers Allowance (JSA), Employment Support Allowance (ESA),

[11] Surveys interviewing victims of crime typically report higher levels of crime rates than police records.

[12] Department for Work and Pensions (2018*c*)

Income Support, Housing Benefit are sourced from StatXplore. Counts of the number of Child and Working Tax Credit claimants are sourced from HMRC.[13] These counts together give an estimate of the number of legacy claimants in each area in a given month.

4.4.4 Analytic Sample

My main sample consists of a balanced month-year panel dataset of 348 local authorities in England & Wales between December 2010 and February 2020. I pick the latter cut off to avoid distortions in criminal behaviour (and likelihood of police detection) due to the Covid-19 lockdown starting on 23rd March 2020. The former cut-off represents the first month UKCrimeStats data are available. As such, my data covers 26 months in which all local authorities had not adopted the LS, the 61 months in which the Live and FS were rolled out across the UK, and 12 months in which all local authorities had implemented the FS. This length of sample enables testing for pre-trends well before UC's adoption, as well as identifying any long-term effects of either version of the programme. Even minor changes in trends, pre or post UC, should be detected given the monthly frequency of the data being fully synchronised with the UC's monthly LS and FS rollout.

My panel dataset consists of 38,280 observations. All crime measures are transformed into per 100,000 population rates using the Office for National Statistics' mid-year population estimates. For the reporting reasons stated in Section 4.4.1, there are fewer observations for the acquisitive crime variable and some other crime types. Table 4.1 presents summary statistics. It shows that there were on average 813 crimes committed per month and local authority per 100,000 population. Antisocial behaviour (263) and acquisitive types (253) crime make up the vast majority of these, followed by violent crime (196).

Table 4.2 provides a pre-UC comparison of crime rates, demographics, economic outcomes, and legacy benefit caseloads in local authorities that were scheduled to adopt the LS relatively early (before the median adoption date, June 2015) to local authorities that were scheduled to adopt the LS relatively late (after the median adoption date). It shows that early adopters had higher crime rates,

[13] See HMRC (2022)

slightly poorer economic outcomes and higher welfare caseloads. For example, early adopters had a total crime rate of 863.83, compared to 763.14 for late adopters (a 13.2% difference). Other notable differences include median house prices (£184,025 versus £207,499) and the JSA claimant count (4,475 versus 2,427). Given that the LS was first rolled out in the North West of England, these differences may reflect the well-known economic disparities between the North and South of the country. Note, differences in levels between early and late adopters are not a threat in of themselves to my identification - I will discuss my empirical strategy Section 4.5. Table 4.2 shows that differences between early and late FS adopting local authorities were generally smaller. For example, the total crime rate in early FS adopting areas was 855.65, compared to 767.41 for areas adopting it later. The smaller differences here are not surprising given the more balanced FS rollout timing across the North and South - this can be seen in Figure 4.2.

4.5 Empirical Method

In this section I begin by outlining my DiD empirical approach and testing whether the rollout of the LS and FS were plausibly exogenous with respect to crime. Following this, in Section 4.5.2 I describe the problems of TWFE in settings where policy adoption is staggered. Lastly, in Section 4.5.3, I describe the de Chaisemartin and d'Haultfœuille (2020) estimator I adopt to estimate (and disentangle) the effects of the LS and FS.

4.5.1 DiD

To briefly recap Section 3.5.1, DiD isolates the effect of a policy by comparing the change in outcomes of treated units to the change in outcomes of untreated units. In this setting, this amounts to comparing local authorities treated for UC relatively early to local authorities treated for UC relatively late - I will discuss how this extends to the two rollouts in Section 4.5.3. Intuitively, by focusing on the changes in outcomes, this approach eliminates selection issues where heterogeneity within local authorities is fixed over time. By comparing the change across units, any country-wide developments that may affect crime are similarly eliminated. Thus, the effect of the policy is disentangled from potential confounders.

To establish causality, DiD requires two assumptions to hold. First, there can be no crime response in anticipation of UC being implemented. In practice this seems unlikely to be of concern. It would mean that offenders would have to anticipate not only the need to make a future benefit claim, but the timing around it in the context of the UC rollout. As discussed in Section 3.5.1, UC's delivery was set back on numerous occasions, leading to timetable rescheduling in each year from 2013 to 2018. This should, in theory, make anticipation even more difficult. Yet, even if it does occur, it is worth highlighting that this behavioural response would have to occur on a large enough scale to influence local authority level crime rates.

The second assumption is that parallel trends holds - that is, the average change in crime rates experienced by not-yet treated local authorities reflects that which treated local authorities would have experienced had they not in fact been treated. Parallel trends is more likely to hold if the timing of both rollouts are exogenous to changes in crime and its potential determinants. To test exogeneity empirically, I estimate the following bivariate regression using OLS:

$$LiveServiceRolloutDate_i = \beta_1 \Delta LocalChar_i + \epsilon_i \tag{4.1}$$

where $LiveServiceRolloutDate_i$ is the month-year local authority i implemented the LS; $LocalChar_i$ is one of the following local authority characteristics: total crime, acquisitive crime, violent crime, median house price, unemployment rate, median weekly pay, or the percentage of the population aged between 16-24%; and ϵ_i represents the unexplained error. Characteristics in $LocalChar_i$ are either direct measures of crime or factors which are likely criminogenic. Since parallel trends is a restriction on the *changes* in crime, testing for whether the change in these variables (not levels) is exogenous to the LS rollout is the most relevant test of this assumption. Here, the change under consideration is the difference in crime rates between 2011 and 2012 - the first pair of consecutive years in which no local authority had implemented the LS or FS. Following the same logic, I also estimate:

$$FullServiceRolloutDate_i = \beta_1 \Delta LocalChar_i + \epsilon_i \tag{4.2}$$

which tests for exogeneity of the FS month-year implementation date. Table 4.3 displays the estimates. Panel A shows that all LS estimates are small and statistically insignificant at any conventional level. As an example of the tiny effects, the estimate of -0.001 for total crime suggests that for every additional month a local authority is pushed further back the LS rollout schedule, there is an associated reduction in the change in crime of 0.00095% (relative to the mean change of -104.25). Put differently, a 2 year wait to adopt the LS is associated with a decline in the change of crime of 0.230%. Panel B similarly shows that all estimates pertaining to the FS are statistically insignificant at any conventional level. Thus, Table 4.3 supports the claim that the LS and FS can be treated as an exogenous intervention, at least with respect to crime. This in turn suggests that the DiD research design is credible.

4.5.2 TWFE Issues

Traditionally, both in the UC literature and applied work more broadly, DiD has been operationalised in a panel regression setting via two-way fixed effects models, the simplest of which is shown in Equation 4.3:

$$y_{it} = \alpha_i + \lambda_t + \beta D_{it} + \epsilon_{it} \tag{4.3}$$

where y_{it} is an outcome of interest for unit 'i' (e.g district) during period 't'; α_i is a unit fixed effect; λ_t is a period fixed effect; D_{it} is the treatment variable of interest, either binary or continuous, and ϵ_{it} is the unexplained error. The model maps to DiD intuition as follows. First, the fixed effect α_i allows for units to have their own unit-specific yet time-invariant relationship with y_{it}. In essence, where the first difference in DiD eliminates this heterogeneity, TWFE controls for it. Second, the period fixed-effect captures the idea of units' having outcomes that trend in parallel - since λ_t affects all units equally, it shifts their outcomes by an equal quantity. This clear link between TWFE and DiD led to these models being widely adopted by scholars.

In recent years, however, multiple authors have noted that the estimate β may not represent a straight-forward weighted average of unit level treatment effects, as was widely assumed (Goodman-Bacon, 2021; Borusyak et al., 2022; De Chaisemartin and d'Haultfoeuille, 2020). In short, these authors have found that in settings where there is variation in treatment timing, the assumption of parallel trends is in fact insufficient for identifying causal estimates. Goodman-Bacon (2021) demonstrates why this is the case by decomposing β to show exactly which comparisons between control and treatment groups drive the overall estimate. Confirming prior intuition, he shows that units treated later in the sample are indeed used as controls for units treated earlier. However, he also shows that units treated earlier are used as controls for those treated later, due to their treatment status being constant once treated (treatment is assumed to be an absorbing state). This is an important result. It means that if treatment effects vary over time, by definition there will be a violation of parallel trends for these comparisons, since already-treated control groups will be on a new outcome trajectory. Consequently, TWFE will often aggregate comparisons with "negative weights", even if all unit-level treatment effects are positive.[14] In turn, TWFE will only estimate the causal impact of a policy under an additional, previously unrecognised assumption of treatment effect homogeneity.

Since in this context there is variation in treatment timing for both the LS and FS, TWFE estimates are at risk of being biased if treatment effects are indeed dynamic. For UC, treatment effects on population level outcomes are guaranteed to be dynamic for mechanical reasons. The reason is that the UC caseload grows rapidly over time as a result of the benefit only being available to new claims. For example, my data show that in the first month of LS adoption, 16,038 claimants were enrolled in UC across England and Wales, yet 12 months on this figure was 196,070. For the FS, the equivalent was 471,848 and 1,816,520.

Furthermore, recent analysis in de Chaisemartin and d'Haultfœuille (2020) found that in settings with multiple treatments - in this case, the LS and FS – TWFE estimates of each treatment can be biased

[14]For example, if two units receive a treatment effect of 1 and 4 respectively, they may be weighted to produce a negative overall estimate: (2x1) - (4x1) <0.

by the effects of other treatments, even when all treatments are incorporated within specifications. Applied in this context, a TWFE estimate for the LS impact is, under parallel trends, the sum of two terms: i) a weighted sum of LS's effects in each month and timing group; and ii) a weighted sum of the effects of the FS. If treatment effects from the FS are dynamic, the weights in ii) do not sum to 0. Thus, TWFE estimates for the LS will be contaminated by the FS effect even under parallel trends.

4.5.3 de Chaisemartin and d'Haultfœuille (2020) Estimator

To consistently estimate the criminological effects of UC, I use new DiD estimators which are robust to heterogeneous treatment effects. My main analysis uses the de Chaisemartin and d'Haultfœuille (2020) "DiD_l" estimator, which has the additional advantage of being able to separately estimate the effects of the LS (treatment 1) and the FS (treatment 2). It can do so because both treatments satisfy the following criteria: i) they are binary, ii) units always experience treatment 1 before treatment 2, and iii) units cannot "switch off" either treatment (de Chaisemartin and d'Haultfœuille, 2020). By eliminating the contamination problem and being robust to heterogeneous treatment effects, DiD_l is therefore the most appropriate estimator to use among the new class of DiD estimators.

To estimate the average cumulative effect of having implemented the LS for l months, DiD_l, compares the outcome evolution between districts that started receiving the LS l months ago to districts not yet converted from the legacy system. This comparison is over $l + 1$ months, from the month before treatment to l months after. To avoid contamination from the FS, the estimator restricts its comparisons to only involve observations where the legacy system or LS is active. It then produces an event study graph in which the instantaneous treatment effect of the LS is shown at t=0 and dynamic effects l are shown to the right. In effect, this method is analogous to the estimator in Callaway and Sant'Anna (2021) (the estimator used in Chapter 3), with the exception that all periods where units have adopted the FS are excluded.

Estimating the equivalent dynamic effect l for the FS is trickier since districts have already received a prior treatment. DiD_l does so by first restricting its comparisons to groups that start receiving the LS in the same month, before comparing the outcome evolution of groups that started receiving the FS l

months ago to those that have not yet received it. Take, for example, Croydon and Bristol City, which both implement the LS in February 2015. Croydon subsequently adopts the FS in November 2015, whereas Bristol City does so 31 months later in June 2018. DiD_{30} would thus include a comparison of the outcome evolution in Croydon and Bristol City from October 2015 to May 2018 (31 months). In practice, I restrict comparisons to local authorities implementing the LS in the same quarter, rather than month, to reduce demands on my data.

Following the estimation of dynamics, DiD_l aggregates treatment effect estimates to provide a policy relevant parameter. In this setting with a staggered design and no always-treated units, this parameter is an estimate of the average of all instantaneous and dynamic treatment effects across local authorities. This is the standard average treatment effect on the treated across l periods.

Interpreting the aggregate and dynamic treatment effects as causal for both the LS and FS requires parallel trends and no anticipation to hold in each case. By conditioning on districts implementing the LS in the same month, DiD_l greatly improves on the likelihood of this occurring for the FS over traditional TWFE or alternative DiD estimators. The restriction means that only local authorities that have operated the LS for the same number of relative time periods are used within the estimation. Thus, control and treatment groups are more likely to be at the same 'stage' in their outcome evolution in the pre-FS period. Nonetheless, it is worth recognising that parallel trends holding for the FS analysis means placing a restriction on treated potential outcomes, since the treatment group has received the LS. In effect, it means that LS treatment effects, while allowed to be dynamic, must evolve in the same way (on average) in every timing group.

Formal testing for parallel trends and no anticipation for both the LS and FS is possible through the DiD_l event study graphs. To the left of t=0 in the graph, the estimates are long difference placebos that test parallel trends by comparing the outcome trends of those treated and not treated before the treatment group was treated. More precisely, they compare outcomes over the length of time that their counterpart's post treatment estimate would require in order to not violate parallel trends. As such, placebos are symmetric to the post treatment estimates by construction. Take for example the 5th placebo estimate, DiD_5^{pl}, which compares the outcome evolution across groups between relative period -7 and -1. If found to be statistically insignificant, it indicates that parallel trends held in the 6 months before treatment, a window long enough for DiD_5 to be unbiased if this also occurred post treatment. As some placebos may be statistically significant by chance, I conduct a joint significance test for all placebos in which the null hypothesis is parallel trends holds for l periods (i.e. no difference in pre-trends).

I include covariates within some model specifications to mitigate potential concerns of omitted variable bias. DiD_l leverages covariates in a different way from that of Callaway and Sant'Anna (2021). Indeed, where Callaway and Sant'Anna (2021) use period 1 values to estimate a propensity score for each local authority, DiD_l allows parallel trends to fail in ways that can be explained by linear changes in groups' covariate values. One benefit of this approach is that it enables my analysis to use time-varying covariates, such as local authority specific linear time trends. Note, however, that since UC may affect economic outcomes, I still use period 1 values for these covariates to avoid the issue of "bad controls" described in (Angrist and Pischke, 2009, p.64). I also do the same for legacy caseload covariates, since UC's rollout will affect these by design.

Finally, it is worth mentioning that the observation count DiD_l reports is higher than that contained within analytic samples (in this case, 38,280 observations). The reasons for this, in the context of this study, is that each unique month/local-authority cell will be used multiple times within the estimation process. The simplest example of this is how the last local authority treated for the LS, Hart, will be used as a control group for every other district for LS estimation. Observation counts within my results should be interpreted as the number of long differences DiD_l uses to estimate treatment effects.

4.6 Results

In this section I analyse the effects of the LS (Section 4.6.1) and FS (Section 4.6.2) on acquisitive, violent and total crime rates. I first explore descriptively how crime rates have evolved since the the programme was introduced across local authorities. I then present estimates of the overall average treatment effect using the DiD_l estimator, followed by an event study plot to show treatment effect dynamics.

4.6.1 Live Service Effects

Descriptive

Figure 4.4 plots rates of total, acquisitive, and violent crime with a linear trend for the 3 years before and after the Live Service (LS) was introduced. It also shows how the UC claimant rate - the number of claimants in England and Wales per 100,000 population - builds from zero once the LS is implemented in each local authority. Figure 4.4a suggests a relationship between the LS and acquisitive crime. It shows that in the 3 pre-period years, local authorities experienced a downward trend in acquisitive crimes before witnessing this trend reverse following LS adoption. The magnitude of this reversal is non-negligible: -11% during the pre-period and 9.9% during the post-period. It is noteworthy that the minimum rate of acquisitive crime occurs eight months after LS adoption, rather than the exact month itself. If there is in fact a causal relationship, this delayed minimum could be consistent with the idea of the UC caseload starting from zero and then accumulating over time, as shown in the figure. Figure 4.4b shows that violent crime was already on an upward trajectory before the LS, and its trend did not appear to be affected in the 3 years following the policy implementation. It therefore suggests that the LS rollout is unlikely to be a reason behind the rise in violent crime shown in Figure 4.1. Figure 4.4c shows that the quadratic trend observed for acquisitive crime is also present in the total crime rate (unsurprisingly given its large share). However, the relationship with total crime is less striking, as the downward trend stabilizes in the 6-8 months before the LS. This suggests that other forms of crime, outside of those financially motivated, may not have been influenced by the LS.

Estimation

Table 4.4 shows the main DiD_l results: the estimated average effect of the LS on acquisitive and violent crime rates within the programme's first 18 months of adoption. I exclude total crime from the estimation process and instead provide a breakdown of the LS effect on all crime categories in

Appendix B. The rationale for this is that acquisitive crime makes up a large share of total crime, and thus displaying results separately for each category is arguably more informative.

Panel A displays results for acquisitive crime. Checking first the validity of the DiD design, the panel shows that across all model specifications, the joint tests of significance for the placebos are statistically insignificant at any reasonable level, meaning that parallel trends was not violated in the pre-period. The raw difference-in-difference model in Column (1) produces a large but statistically insignificant estimate of -17.8. This is surprising given the upward trend observed in Figure 4.4. Adding demographic controls (% aged 16-24, % aged 25-49, % aged 50-64, % aged over 16 and female) in Column (2) produces a positive estimate of 5.2 that is statistically significant at the 10% significance level. Column (3) controls for the number of legacy recipients in each district; Column (4) adds economic controls (median house prices, unemployment rate and weekly pay); and Column (5) excludes the 48 local authorities containing job centres that implemented the LS across different months (leading to potential geographical treatment effect spillovers). These models estimate the LS to have caused an additional 6-7 acquisitive crimes per month and 100,000 population. The lower bound of this represents a 2.7% increase relative to the pre-LS crime mean of 253.36. Estimates in Column (3) and (5) are statistically significant at 0.05 level, whilst the estimate in Column (4) is borderline significant at this level. With the exception of Column (1), it is reassuring that these coefficients are of a similar magnitude as this supports the validity of the research design. The large discrepancy with the estimate in Column (1) suggests that the timing of the rollout is exogenous conditional on population demographics (parallel trends is a relatively weaker assumption, thus this is not inconsistent with "passing" the placebo joint test). Finally, Column (6) shows that this finding is not robust to the inclusion of local authority specific time trends. The estimate here is reduced to 4.33 and is statistically insignificant at even the 10% level. This suggests that the introduction of the LS correlates with other trends in local authority level crime rates, and it is not possible to disentangle the effect of the programme from these underlying trends. As a result, I interpret estimates from this panel as providing only suggestive evidence of the LS causing additional acquisitive crime.

Panel B shows the results for violent crime. In each specification, the placebo joint test again provides evidence of parallel trends holding. The baseline model in Column (1) produces an estimate of 15.9 that is statistically insignificant at any conventional level. Column (2) adds the demographic controls which, consistent with Panel A, leads to a substantial improvement in precision. Nonetheless, the estimate in this model, 1.7, is again statistically insignificant. After controlling for the number of legacy claimants and including economic controls, the estimate size in Column (3) and (4) remains stable with that reported in Column (2), at 1.8 and 1.7 respectively. Column (5) drops the 48 local authorities where there is a greater risk of treatment effect spillovers and Column (6) includes the local authorities specific linear time trends. These models estimate an effect of 2.3 and 0.9, respectively, which are again statistically insignificant. Taken together, while all point estimate are positive, there is no robust evidence suggesting the LS caused an increase in violent crime. A conservative estimate, taken from Column (5), can rule out increases in violent crime greater than 3.6% relative to the mean.

I now turn to the DiD_l dynamic treatment effect estimates. Figure 4.5 presents an event study plot for acquisitive crime using specification 5 from Table 4.4. At first glance, it is apparent that the overall ATT of 6.38 in specification 5 is not driven by a roughly uniform distribution of treatment effects over time. Rather, as expected, estimates are close to zero and statistically insignificant for the first several months since the LS is implemented (as the caseload builds), followed by a growing treatment effect between months 8-14 post adoption. At its peak, DiD_l estimates the LS to have caused 22 additional acquisitive crimes per month and 100,000 population. Estimates between months 8-14 post adoption are significant or borderline significant at the 5% level, with the exception of the month 11. Interesting, I find these effects are not persistent after 14 months. This is unlikely to be driven by a changing sample composition, since only 3.7% of local authorities in the sample are not observed operating the LS for at least 18 months.[15]

Importantly for identification, placebo estimates in the run-up to the LS are precisely estimated null effects, suggesting individuals' criminal propensity did not change in anticipation of the policy. There is also strong evidence of parallel trends holding for at least 12 months before the policy launch. Placebos from earlier periods are also statistically insignificant but less precisely estimated.

[15] Figure 4.3 shows the distribution of months between LS and FS across local authorities

Figure 4.6 shows the event study plot for violent crime using specification 5. Broadly consistent with acquisitive crime, estimates here are approximately zero in the first 8 post period months, before increasing in magnitude for 13 months post LS adoption. In contrast to Figure 4.5 however, all estimates in the post period are statistically insignificant at the 5% level (with exception of the 13th post period month). Thus, there is no strong evidence of the LS affecting violent crime within its first 18 months of implementation. Looking at the placebo estimates, parallel trends appears to hold in every relative time period, and there is also no evidence of anticipatory behaviour.

Appendix B presents event study plots for several other crimes types: antisocial behaviour; bike theft; burglary; criminal damage and arson; drug offences; other crimes; other theft; robbery; shoplifting; theft from the person and vehicle crime. For all of these plots, there is no crime type that exhibits statistically different from zero LS treatment effects for a sustained period.

4.6.2 Full Service Effects

Descriptive

Figure 4.7 shows descriptively how crime rates evolved relative to the FS introduction. To avoid issues of sample attrition, the post-period window is limited to 12 months to ensure all local authorities are observed. Additionally, I exclude local authorities implementing the LS at some point within the 24 months prior to the FS in order to provide a fair comparison between the two UC systems (this leaves 310 out of 348 districts, 89%). Figure 4.7a shows the trend for acquisitive crime. In contrast to the trend observed for the LS, the FS does not appear to affect acquisitive crime - if anything the upward trend observed in the pre-period begins to reverse once the FS has been implemented. This is surprisingly given my positive estimates for the LS impact, and the fact that the figure shows the UC claimant count rate trebling within the first year of the FS implementation. Similarly, Figure 4.7b and Figure 4.7c suggest the FS did not affect either violent or total crime.

Estimation

Table 4.5 shows my main ATT results for the FS over an 18 month window. Turning first to acquisitive crime in Panel A, the placebo joint test of significant strongly suggests that parallel trends holds in all specifications. The baseline difference-in-difference model in Column (1) produces a negative coefficient for the FS, -3.40, that is statistically insignificant. Column (2) adds demographic controls which changes the sign of the estimate to be positive and shrinks its magnitude to 0.96. Column (3) and (4) add legacy caseload and economic controls which produce FS coefficients of similar magnitude at 2.64 and 2.39, respectively. However, when local authorities at risk of treatment effect spillovers are excluded, the estimate drops to -3.60. In sum, all of the FS estimates are of a smaller magnitude than that for the LS and are statistically insignificant at any conventional level. The standard errors are around twice the size than that for the LS in Table 4.4 - this is likely due to DiD_l only comparing local authorities that implement the LS within the same quarter. The most conservative estimate from this analysis, shown in Column (3), rules out increases in acquisitive crime greater than 6.2% relative to the mean.[16]

Turning to violent crime in Panel B, there is no strong evidence of on an effect from the FS. Estimates in each model are positive but statistically insignificant. What supports the internal validity of these results is that the estimates are of similar magnitude in each column; they range in size from 1.2 to 2.9). These estimates are less precise than that for acquisitive crime, however - the most conservative estimate can rule out increases greater than 9.4%

Figure 4.8 and Figure 4.9 present further evidence of a zero effect from the FS on acquisitive and violent crime. These figures also further support the notion that both parallel trends and no anticipation hold. The number of placebo estimates displayed have been restricted to 12 months, due to low statistical power. Nonetheless, they show that these estimates are tightly clustered around the zero effect mark, supporting the validity of the design. In Appendix B I show that this result extends to all crime types under study.

[16] I do not controls for local authority linear time trends in the FS analysis due to demands on my data, having already made the restriction of comparing only local authorities that implement the LS within the same quarter.

4.7 Robustness

The analysis in Section 4.6.1 detected some evidence of the LS leading to additional acquisitive crime. However, these estimates were not robust to local authority specific linear time trends, nor were they statistically significant in relative time after 14 post period months (even without time trends). Thus, my results contradict the evidence in d'Este and Harvey (2022) and Tiratelli et al. (2022), who find robust and long lasting effects for these offences.

Regarding my FS results, it is possible that a causal effect on violent and acquisitive crime could have gone undetected due to a lack of statistical power. Indeed, my results could only rule out increases in violent crime greater than 9.38% and acquisitive forms of crime by more than 6.3%. This imprecision will be a function of the DiD_l estimator comparing only local authorities that implement the LS within the same quarter of each other. While theoretically this should help ensure that parallel trends holds for the FS analysis, it comes at a cost of limiting the number of observations used for the analysis. Therefore as a robustness check, I relax this condition and rerun my results using the full sample of observations. Doing so also enables the use of a wider variety of robust DiD estimators, not just de Chaisemartin and d'Haultfœuille (2020). I therefore check my results against 4 new DiD estimators in the following section.

4.7.1 Alternative DiD estimators

The estimators I choose to employ are those proposed in Callaway and Sant'Anna (2021), Sun and Abraham (2021), Roth and Sant'Anna (2021), De Chaisemartin and d'Haultfoeuille (2022) and Borusyak et al. (2022), all of which are robust to heterogeneous treatment effects. The key difference with de Chaisemartin and d'Haultfœuille (2020) is that when computing an overall ATT for the FS, these estimators will not solely rely on comparing local authorities that implement the LS within the same quarter of each other. This opens up the sample available for estimation, which should provide greater power to detect treatment effects.

Figure 4.10 visualises the FS estimates (and 95% confidence intervals) for each crime type and estimator. It shows that virtually all estimates are statistically insignificant and tightly clustered around the zero effect mark. For example, theft from the person; weapons possessions; vehicle crime; robbery; and burglary consistently have an estimate magnitude of less than 1 additional crime per month and 100,000 population. Estimates for criminal damage, bike theft, and shoplifting are marginally larger (1-2 additional crimes), yet not consistently signed across each estimator. Antisocial behavior (ASB) is one outlier in this regard: the estimators proposed in Callaway and Sant'Anna (2021) and Roth and Sant'Anna (2021) suggest the FS is linked to causing an additional 10 ASB crimes per month and 100,000 population. However, these estimates are statistically insignificant, and are not supported by the other DiD estimators - for example, the Borusyak et al. (2022) estimator computes a negative point estimate for ASB. In general, Figure 4.10 shows no robust evidence of the FS leading to additional crime.

4.8 Conclusion

This chapter estimated the criminological effects of UC using police-recorded crime data in England and Wales between December 2010 and February 2020. To establish causality, I exploited the programme's phased introduction across local authorities between April 2013 and December 2018 within a DiD framework. In contrast to the existing UC-crime literature, I employed novel DiD estimators that are robust to heterogeneous treatment effects and the presence of the programme's twin-track LS and FS rollout. Given how UC's caseload is designed to accumulate over time, thus meaning treatment effects on crime will do the same (if present to begin with), I argue that I provide a more robust evaluation of UC than previously covered in the crime literature. Further, I contribute the first separate, criminological estimates of the FS - the version of UC in place today and that directly affects several million low-income UK households.

My analysis found no robust evidence of the programme's LS increasing local crime for any offence type, at least in the period up to 18 months after its implementation. Some estimates from my analysis do suggest the LS increased acquisitive types of crime, however they were not robust to the inclusion of local authority specific linear time trends. Regarding violent crime, conservative estimates from my analysis were able to rule out increases in violent crime greater than 3.63% (with 95% confidence). I also estimate that the FS had no statistically significant effect on any form of crime in the 24 months following its adoption. Conservative estimates for this analysis were able to rule out (with 95% confidence) increases in violent crime by more than 9.38% and increases acquisitive crime by more than 6.3%. These results are at odds with the existing literature, in particular d'Este and Harvey (2022) and Tiratelli et al. (2022) who report robust and long lasting effects from UC on crime.

Table 4.1: Summary Statistics

	Mean	Std Dev	Min	Max	N
Crime Rates per 100,000 Population					
Total	813.16	513.87	214.49	13,333.33	38,280
Acquisitive	253.36	272.42	64.70	7,030.30	35,496
Antisocial Behaviour	263.16	158.32	2.53	2,893.94	38,280
Violent	149.61	97.58	1.45	1,597.22	38,280
Burglary	56.04	29.81	7.20	833.33	38,280
Robbery	6.79	11.29	0.00	237.11	38,280
Shoplifting	46.68	49.44	1.74	1,366.67	35,496
Other Theft Rate	76.29	143.05	2.16	5,030.30	35,496
Drug	21.58	28.44	1.07	1,015.15	35,496
Bike Theft Rate	12.32	28.44	0.00	950.82	28,536
Theft from the Person	10.71	40.62	0.00	1,134.02	28,536
Criminal Damage & Arson	73.47	30.63	18.50	462.69	35,496
Possession of Weapons	3.75	12.36	1.09	590.91	35,496
Vehicle	50.30	30.27	0.00	483.33	38,280
Demographic					
Population	166,338	113,122.49	2,200	1141800	38,280
% aged 16-24	13.22	3.14	3.30	26.90	38,280
% aged 25-49	40.63	7.16	14.20	78.30	38,280
% aged 50-64	23.64	3.89	8.40	56.30	38,280
% aged 65	22.57	5.57	5.20	42.30	38,280
% Female & aged 16+	51.23	1.52	38.50	100.00	38,280
Economic					
Median House Prices	226,400.33	124,165.72	70,000.00	1,450,000.00	38,280
Unemployment Rate	5.84	2.70	1.00	21.60	38,280
Median Weekly Pay	543.69	83.50	338.60	1,034.10	38,280
Legacy Claimants					
UC Claimants	1,217	3,105.99	0	84,200	38,280
JSA	2,213	3,274.98	5	51,522	38,280
ESA	4,913	4,904.23	5	52,980	38,280
Housing Benefit	12,050	11,551.26	268	117,367	38,280
Income Support	2,326	2,796.14	5	45,875	38,280
Child & Working Tax Credits	11,172	10,465.37	100	132,600	37,836

Notes: This table presents local authority level summary statistics. Crime rates are transformed to be per 100,000 population.

Table 4.2: Balance Checks (pre UC)

	Live Service		Full Service	
	Early	Late	Early	Late
Crime Rates per 100,000 Population				
Total	863.83	763.14	855.65	767.41
	(331.30)	(661.13)	(740.62)	(325.02)
Violent	101.55	95.28	101.20	95.42
	(42.99)	(72.04)	(78.68)	(43.64)
Acquisitive	265.42	242.77	277.58	232.81
	(148.37)	(365.35)	(418.27)	(138.71)
Antisocial Behaviour	346.57	289.99	327.76	303.37
	(158.22)	(176.37)	(194.46)	(149.93)
Demographic				
Population	188,066	136,501	161,417	156,051
	(131,211)	(73,067)	(112,833)	(98,504)
% Aged 16-24	14.03	13.29	13.56	13.65
	(3.13)	(3.02)	(3.02)	(3.15)
% Aged 25-49	42.16	40.61	41.55	41.07
	(6.80)	(6.76)	(6.98)	(6.69)
% Aged 50-64	22.80	23.65	23.14	23.40
	(3.72)	(3.86)	(3.85)	(3.80)
% Aged 65	21.03	22.53	21.87	21.88
	(5.13)	(5.33)	(5.34)	(5.26)
% Female & aged 16+	51.16	51.48	51.39	51.30
	(1.21)	(1.90)	(1.99)	(1.31)
Economic				
Median House Prices	184,025	207,499	191,385	201,707
	(101,094)	(79,247)	(77,166)	(98,14)
Median Weekly Pay	507.43	533.23	517.29	525.85
	(76.58)	(77.46)	(70.35)	(83.61)
Unemployment Rate	7.40	6.57	7.12	6.78
	(2.84)	(2.95)	(2.96)	(2.90)
Legacy Claimants				
JSA	4,475	2,427	3,597	3,111
	(5,178)	(2,500)	(4,471)	(3,658)
ESA	4,538	3,024	3,735	3,622
	(4,444)	(2,777)	(3,766)	(3,580)
Income Support	4,095	2,369	3,358	2,939
	(4,302)	(2,280)	(3,619)	(3,253)
Child & Working Tax Credits	15,589	10,248	13,153	12,076
	(13,562)	(7,624)	(11,876)	(10,093)

Notes: This table presents local authority level summary statistics disaggregated by "early" and "late" UC adopters, pertaining to the LS (left hand side) and FS (right hand side). A local authority is defined as an early (late) adopter if they adopt that particular version of UC before (after) the median local authority does.
Crime rates are transformed to be per 100,000 population.

Table 4.3: Rollout Endogeneity Check

Panel A: Live Service	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Δ Crime Rate	-0.001						
	(0.002)						
Δ Acquisitive Crime Rate		-0.00001					
		(-0.01)					
Δ Violent Crime Rate			-0.010				
			(0.015)				
Δ Median House Prices				-0.0001			
				(0.00008)			
Δ Unemployment Rate					-0.257		
					(0.225)		
Δ Median Weekly Pay						-0.018	
						(0.021)	
Δ % 16-24 year olds							0.089
							(0.150)
Observations	348	348	348	348	348	348	348

Panel B: Full Service	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Δ Crime Rate	-0.013						
	(0.008)						
Δ Acquisitive Crime Rate		0.017					
		(0.016)					
Δ Violent Crime Rate			-0.033				
			(0.0269)				
Δ Median House Prices				-0.000001			
				(0.00003)			
Δ Unemployment Rate					-0.098		
					(0.204)		
Δ Median Weekly Pay						−0.004	
						(0.023)	
Δ % 16-24 year olds							-0.269
							(0.220)
Observations	348	348	348	348	348	348	348

Notes: *** p<0.01, ** p<0.05, * p<0.10. Heteroscedasticity-robust standard errors adjusted for clustering at the local authority level are reported in parentheses.

This table shows results from estimating bivariate, cross-sectional regressions that use the month-year of the UC's implementation as the dependant variable. Panel A shows the results for the LS, Panel B shows the results for the FS. Independent variables are first-differenced in order to test whether changes in their values have predictive power for when UC was implemented. I take the first difference between the year 2011 and 2012; the earliest pair of consecutive years during which no local authority had adopted the LS or FS.

Table 4.4: Live Service - DiD_l Estimates

	Panel A: Acquisitive Crime					
	(1)	(2)	(3)	(4)	(5)	(6)
UC	-17.800	5.181*	6.794**	6.109*	6.380**	4.328
	(17.235)	(3.056)	(3.303)	(3.180)	(3.082)	(3.993)
Mean	253.36	253.36	253.36	253.36	253.36	253.36
Observations	88,851	88,851	88,851	88,851	64,939	88,851
P value placebo joint test	0.970	0.382	0.384	0.423	0.519	0.643
	Panel B: Violent Crime					
	(1)	(2)	(3)	(4)	(5)	(6)
UC	15.874	1.736	1.785	1.650	2.326	0.876
	(10.033)	(1.525)	(1.570)	(1.535)	(1.587)	(2.605)
Mean	149.61	149.61	149.61	149.61	149.614	149.614
Observations	88,851	88,851	88,851	88,851	64,939	88,851
P value placebo joint test	0.990	0.401	0.412	0.419	0.597	0.542
Demographics		Yes	Yes	Yes	Yes	Yes
Legacy Claimants			Yes	Yes	Yes	Yes
Economic Controls				Yes	Yes	Yes
Exclude UC Spillover LAs					Yes	
LA Linear Time Trends						Yes

Notes: *** p<0.01, ** p<0.05, * p<0.10. Heteroscedasticity-robust standard errors adjusted for clustered at the local authority level are reported in parentheses.
This table shows the results from DiD_l estimation of the LS impact on violent and acquisitive crime within its first 24 months of adoption. To estimate the LS effect, DiD_l compares the change in outcomes of local authorities treated for the LS to the outcomes of local authorities not-yet treated for the LS. Column (1) excludes all controls and assumes parallel trends hold unconditionally. Column (2) controls for the following demographics: % aged 16-24, % aged 25-49, % aged 50-64, % aged over 65, and % female and aged over 16. Column (3) adds period 1 values of the number of legacy benefit claimants (Jobseekers Allowance, Employment Support Allowance, Child and Working Tax Credits, Income Support and Housing Benefit). Column (4) further controls for the period 1 value of the unemployment rate, median house prices and weekly pay. Column (5) excludes the 48 local authorities that have job centres serving multiple districts. Column (6) includes local authority specific linear time trends. The joint placebo test is a multiple hypothesis test of whether the placebos are jointly equal to 0.

Table 4.5: Full Service - DiD_l Estimates

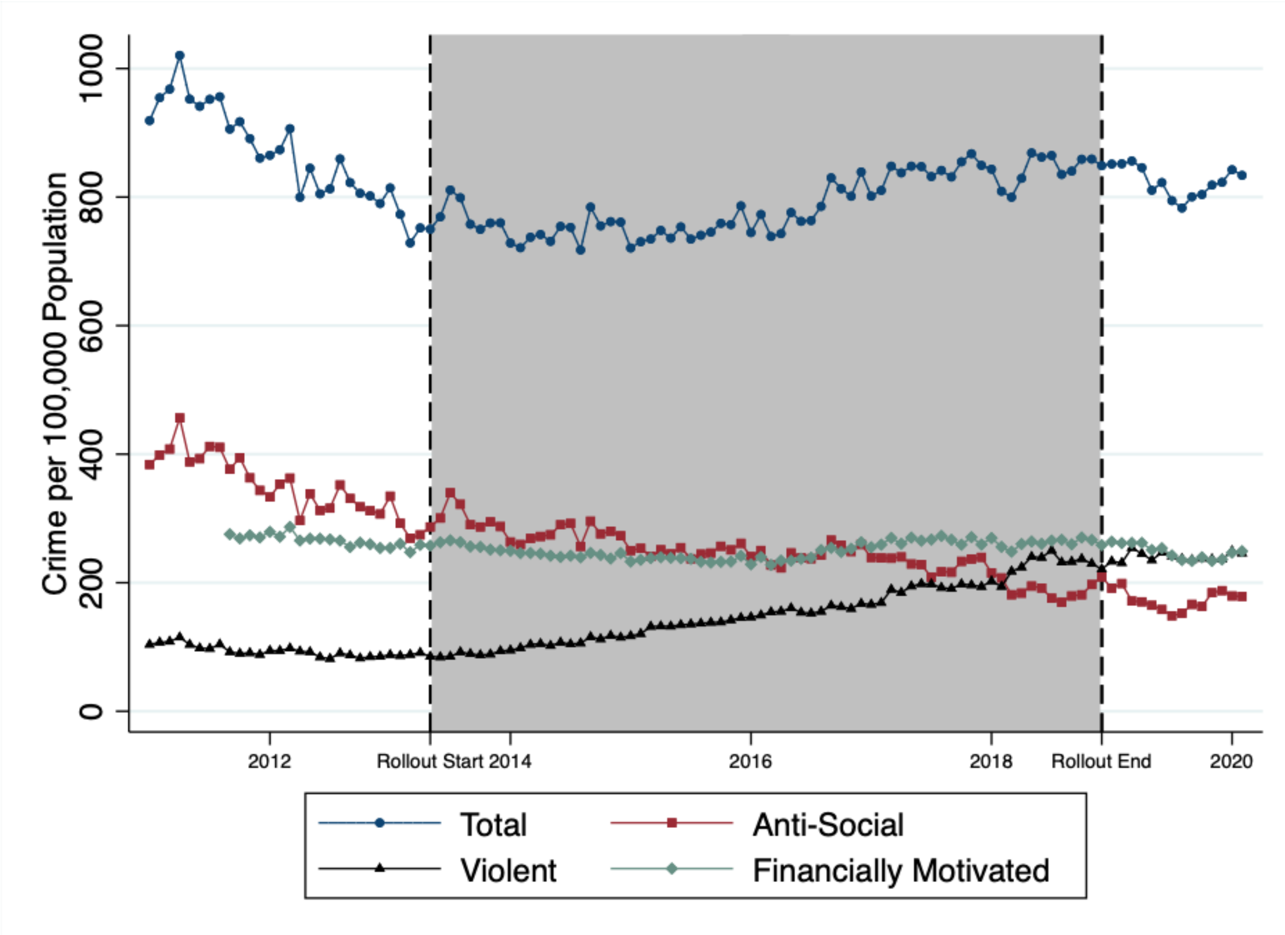
	Panel A: Acquisitive Crime				
	(1)	(2)	(3)	(4)	(5)
UC	-3.402	0.963	2.640	2.393	-3.596
	(4.561)	(5.988)	(6.751)	(6.631)	(5.168)
Mean	253.357	253.357	253.357	253.357	253.357
Observations	107,632	99,823	99,823	99,823	81,930
P value placebo joint test	0.437	0.959	0.959	0.970	0.943
	Panel B: Violent Crime				
	(1)	(2)	(3)	(4)	(5)
UC	2.738	1.243	1.595	1.685	2.940
	(4.141)	(5.520)	(5.384)	(5.349)	(5.662)
Mean	149.614	149.614	149.614	149.614	149.614
Observations	107,632	99,823	99,823	99,823	81,930
P value placebo joint test	0.714	0.131	0.164	0.160	0.251
Demographics		Yes	Yes	Yes	Yes
Legacy Claimants			Yes	Yes	Yes
Economic Controls				Yes	Yes
Exclude UC Spillover LAs					Yes

Notes: *** p<0.01, ** p<0.05, * p<0.10. Heteroscedasticity-robust standard errors adjusted for clustered at the local authority level are reported in parentheses.

This table shows the results from DiD_l estimation of the FS impact on violent and acquisitive crime within its first 24 months of adoption. To estimate the FS effect, DiD_l compares the outcomes of local authorities treated for the LS in the same quarter, yet treated for FS in different months.

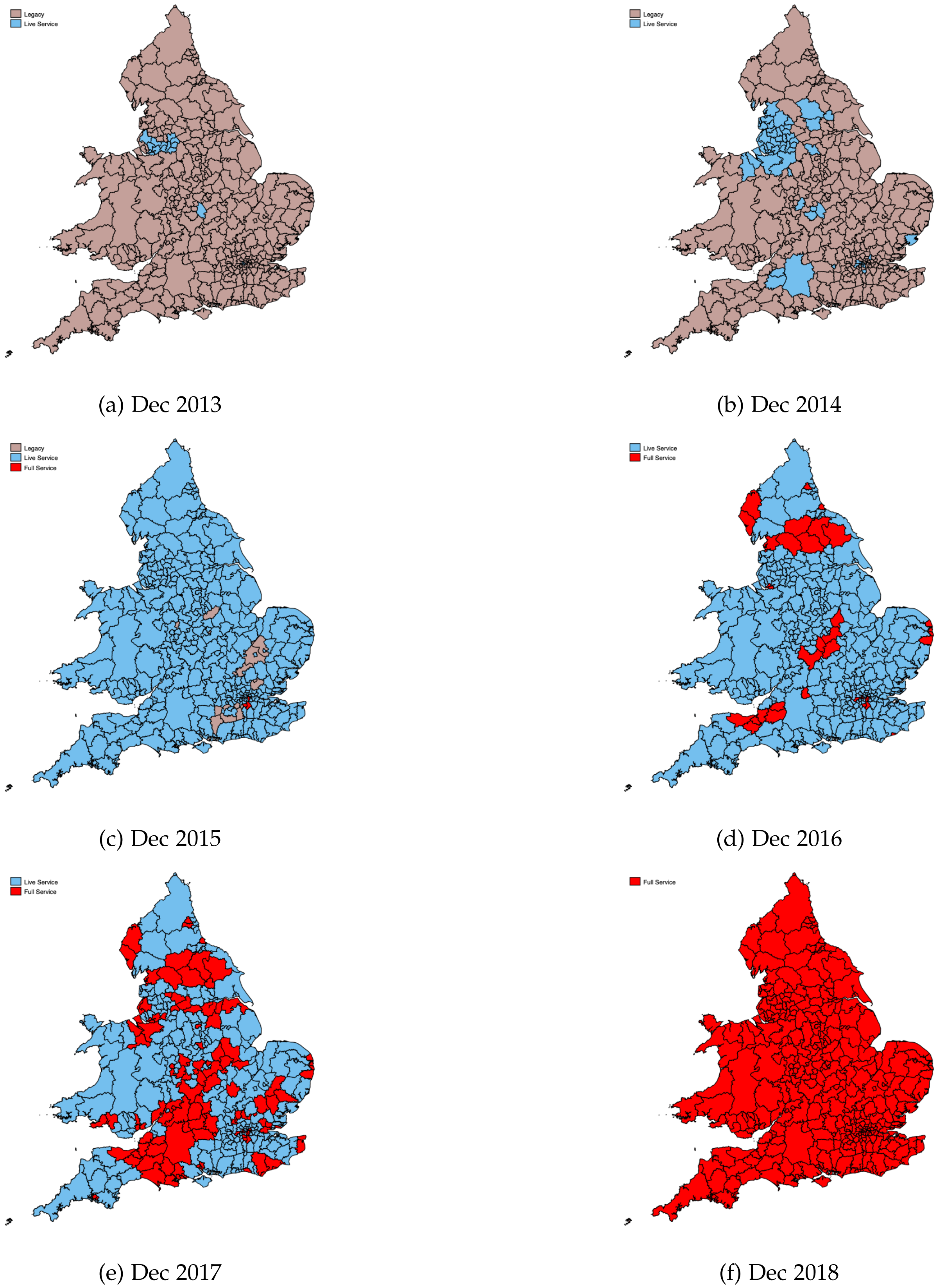
Column (1) excludes all controls and assumes parallel trends hold unconditionally. Column (2) controls for the following demographics: % aged 16-24, % aged 25-49, % aged 50-64, % aged over 65, and % female and aged over 16. Column (3) adds period 1 values of the number of legacy benefit claimants (Jobseekers Allowance, Employment Support Allowance, Child and Working Tax Credits, Income Support and Housing Benefit). Column (4) further controls for the period 1 value of the unemployment rate, median house prices and weekly pay. Column (5) excludes the 48 local authorities that have job centres serving multiple districts. The joint placebo test is a multiple hypothesis test of whether the placebos are jointly equal to 0. Standard errors clustered at the local authority level.

Figure 4.1: Crime Rates in England & Wales



Notes: Dashed vertical lines refer to the start and end of the UC twin-track rollout. Crime rates are seasonally adjusted.

Figure 4.2: UC Rollout



(a) Dec 2013

(b) Dec 2014

(c) Dec 2015

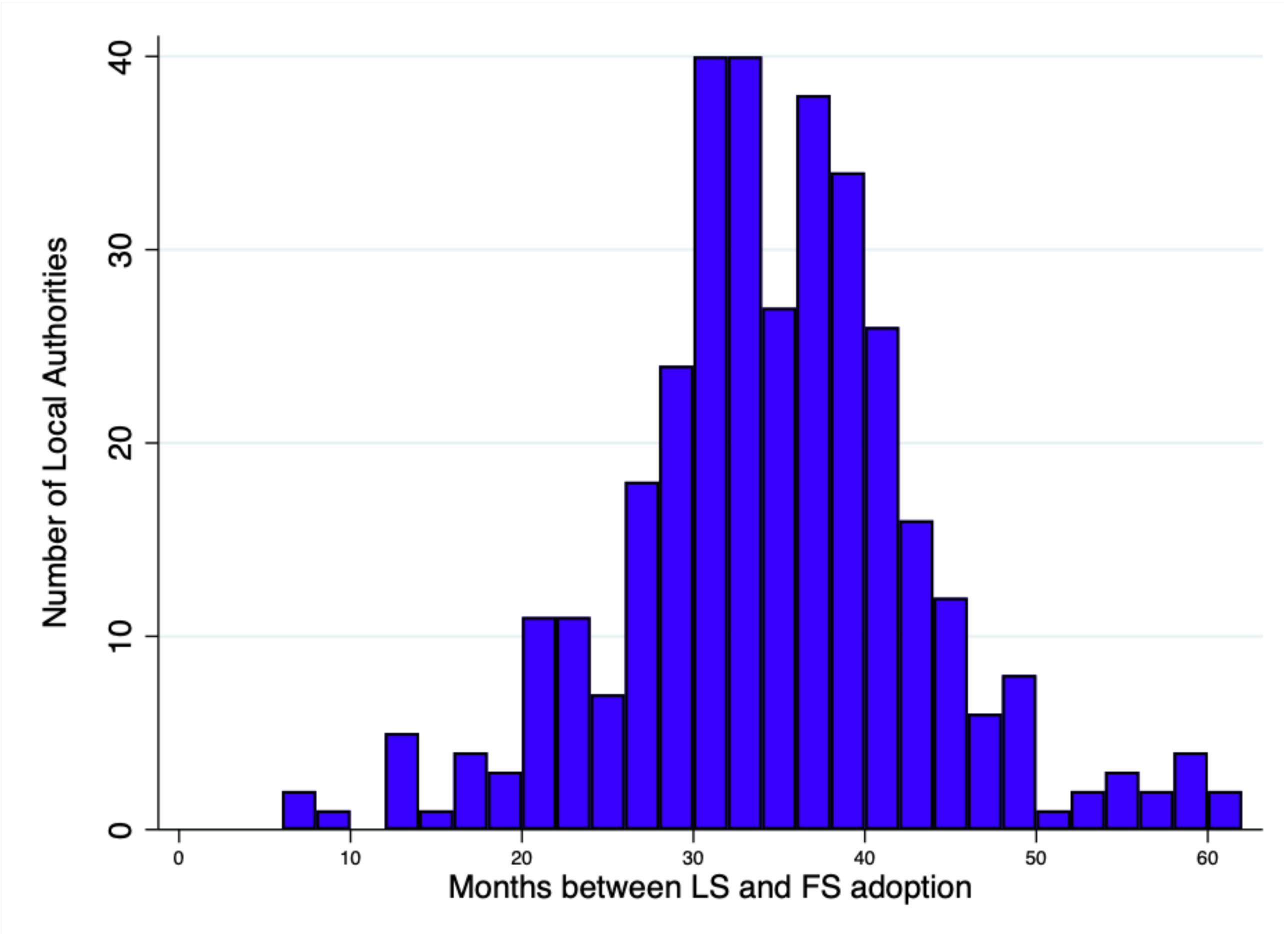
(d) Dec 2016

(e) Dec 2017

(f) Dec 2018

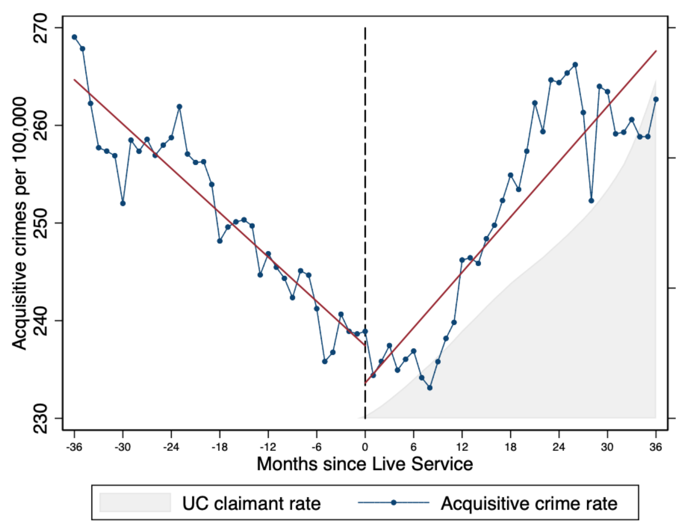
Notes: These maps show how the rollouts of the LS (blue) and FS (red) replaced the legacy system (rose).

Figure 4.3: Live Service and Full Service Rollout Timing

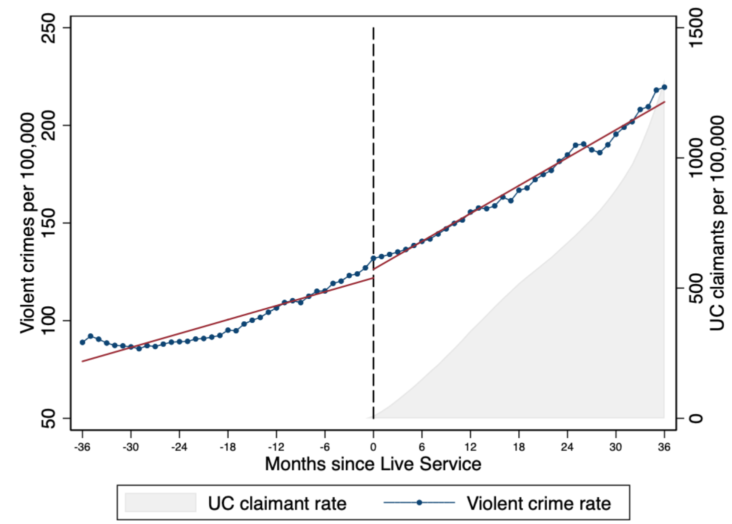


Notes: This figure shows the distribution of the number of months between the LS and FS adoption dates.

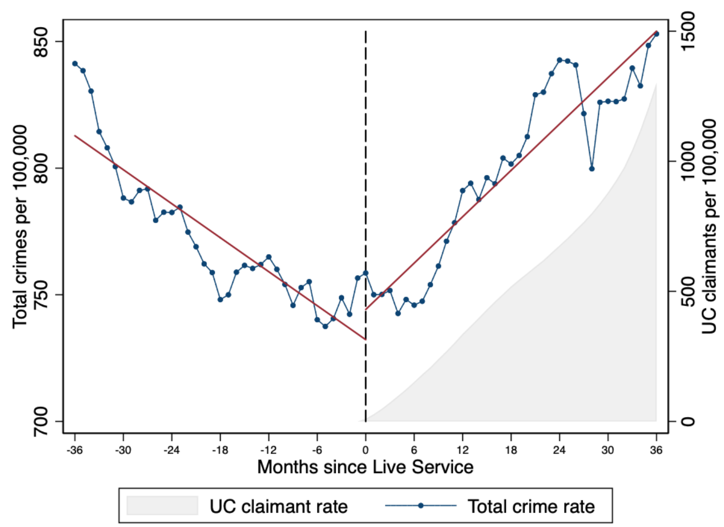
Figure 4.4: Crime Rates Relative to Live Service Introduction



(a) Acquisitive crime



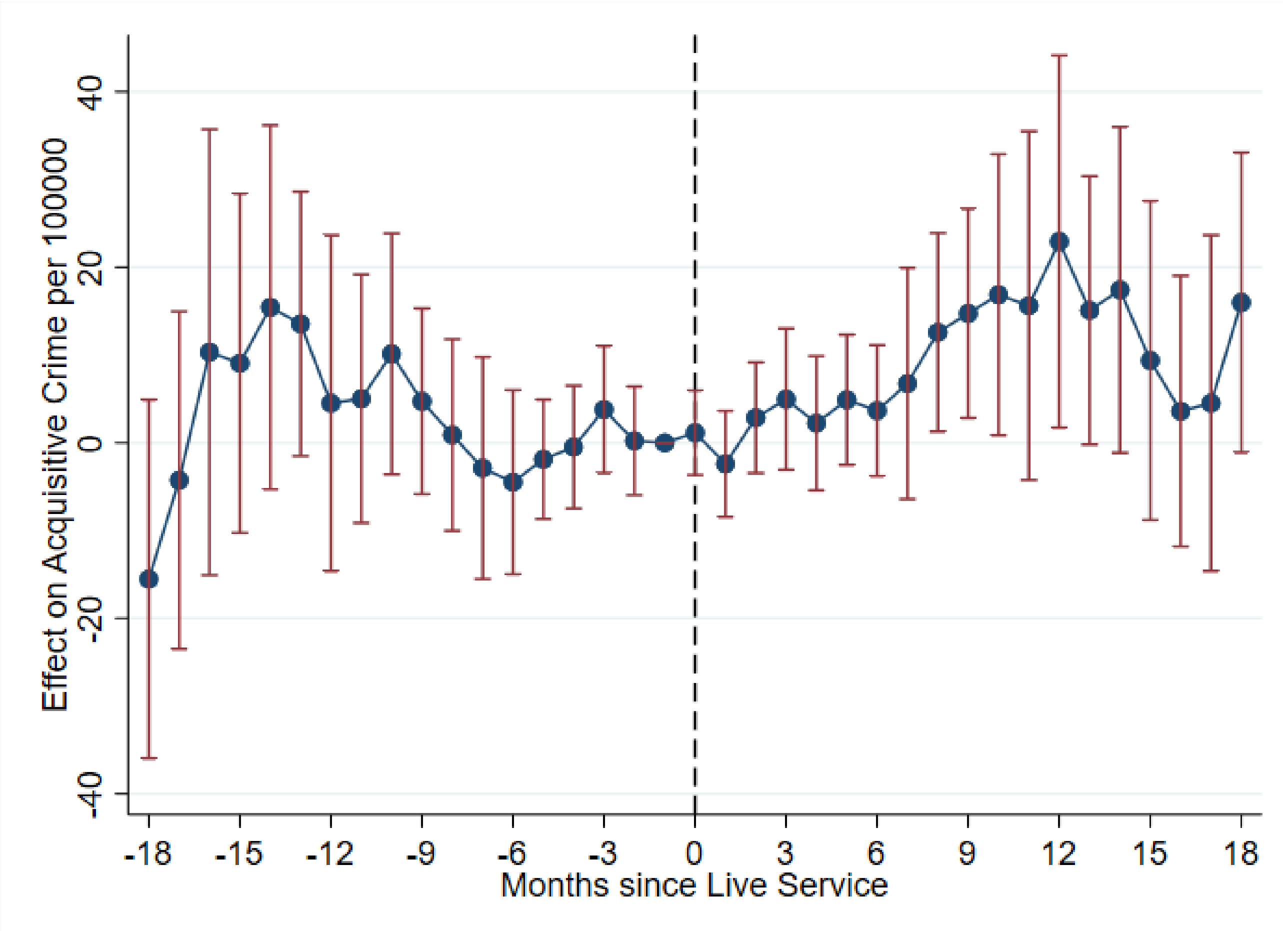
(b) Violent crime



(c) Total crime

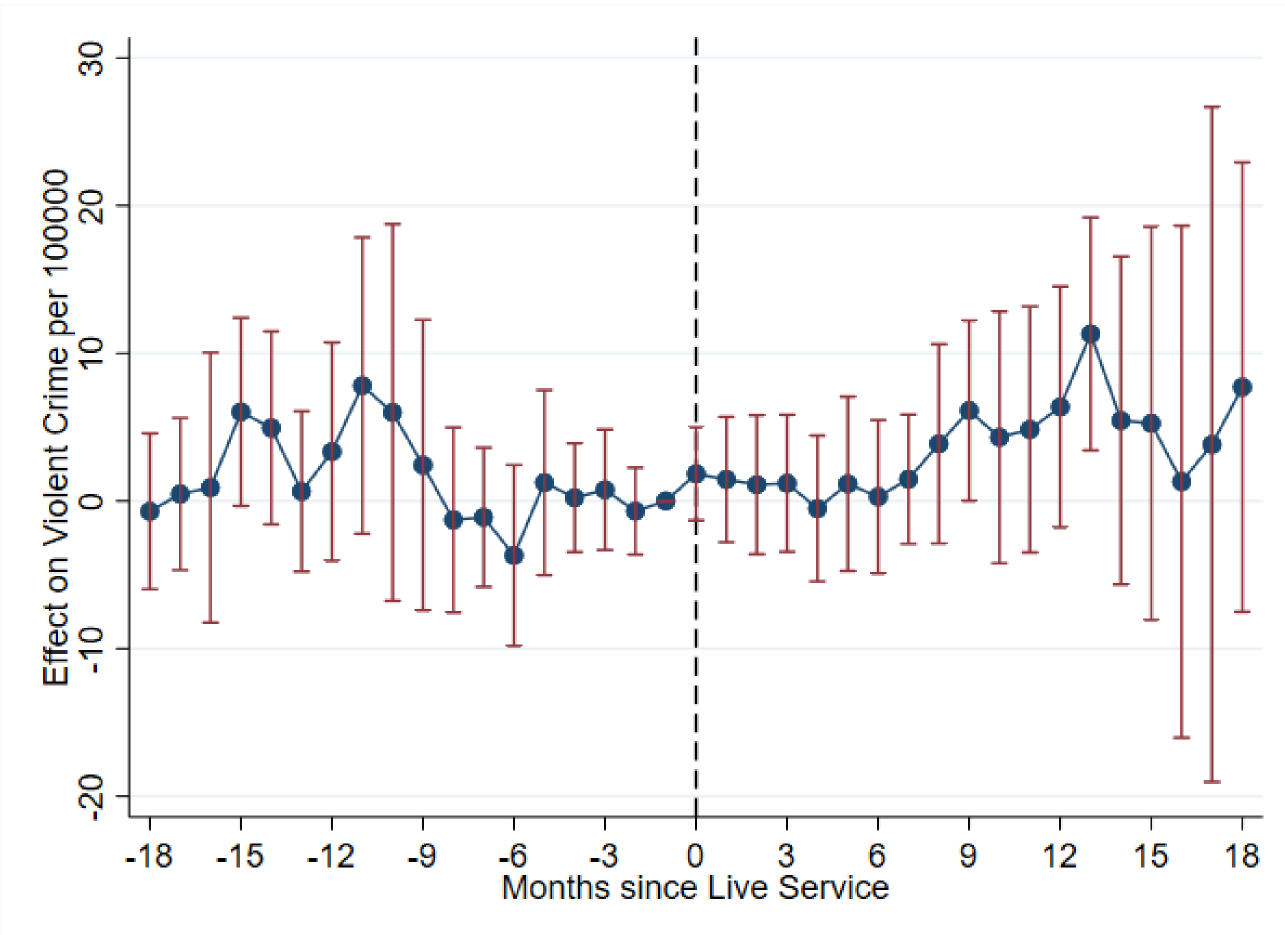
Notes: These figures plot the mean crime rate in the 36 months before and after the LS was introduced in each local authority. The red lines are predicted values from two linear regressions of crime - either acquisitive, violent or total - on months relative to LS, either side of relative period 0. The gray shaded area shows the build up of UC claimants per 100,000 population following the LS introduction.

Figure 4.5: Live Service Event Study - Acquisitive Crime



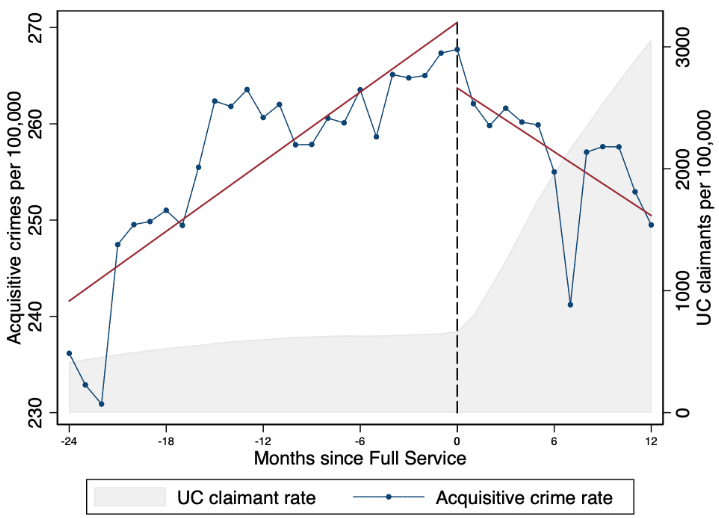
Notes: This figure shows how the average effect of the LS on acquisitive crime rates evolved over time across local authorities. 95% confidence intervals are shown as red lines. DiD_l estimates the dynamic effect l by comparing the change in acquisitive crime rates between local authorities that started receiving the LS l months ago to districts not yet converted from the legacy system. This comparison is over $l+1$ months, from the month before treatment to l month after. Placebo estimates on the left of t=0 test the assumptions of parallel trends and no anticipation. Standard errors are clustered at the local authority level.

Figure 4.6: Live Service Event study - Violent Crime

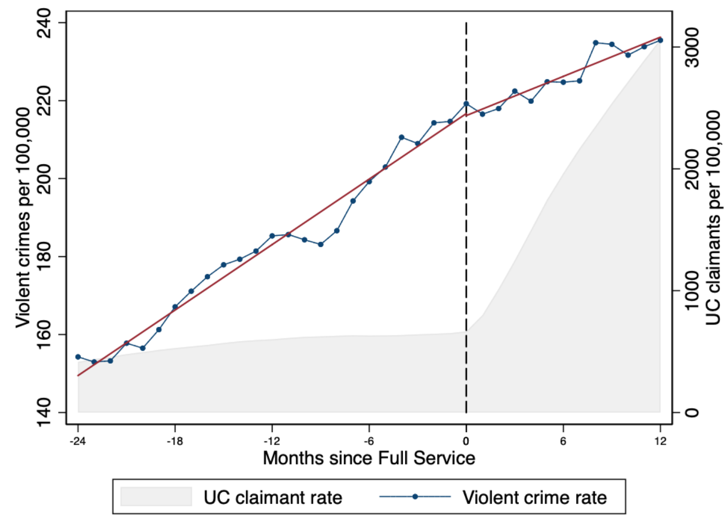


Notes: This figure shows how the average effect of the LS on violent crime rates evolved over time across local authorities. 95% confidence intervals are shown as red lines. DiD_l estimates the dynamic effect l by comparing the change in violent crime rates between local authorities that started receiving the LS l months ago to districts not yet converted from the legacy system. This comparison is over $l+1$ months, from the month before treatment to l month after. Placebo estimates on the left of t=0 test the assumptions of parallel trends and no anticipation. Standard errors are clustered at the local authority level.

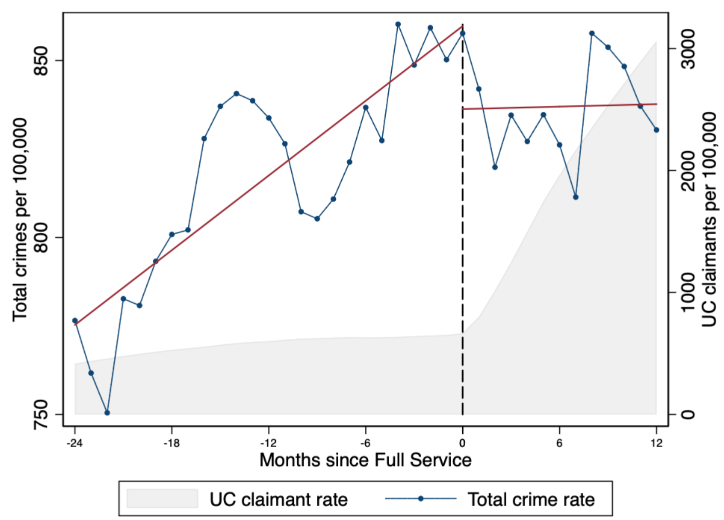
Figure 4.7: Crime rates Relative to Full Service Introduction



(a) Acquisitive crime



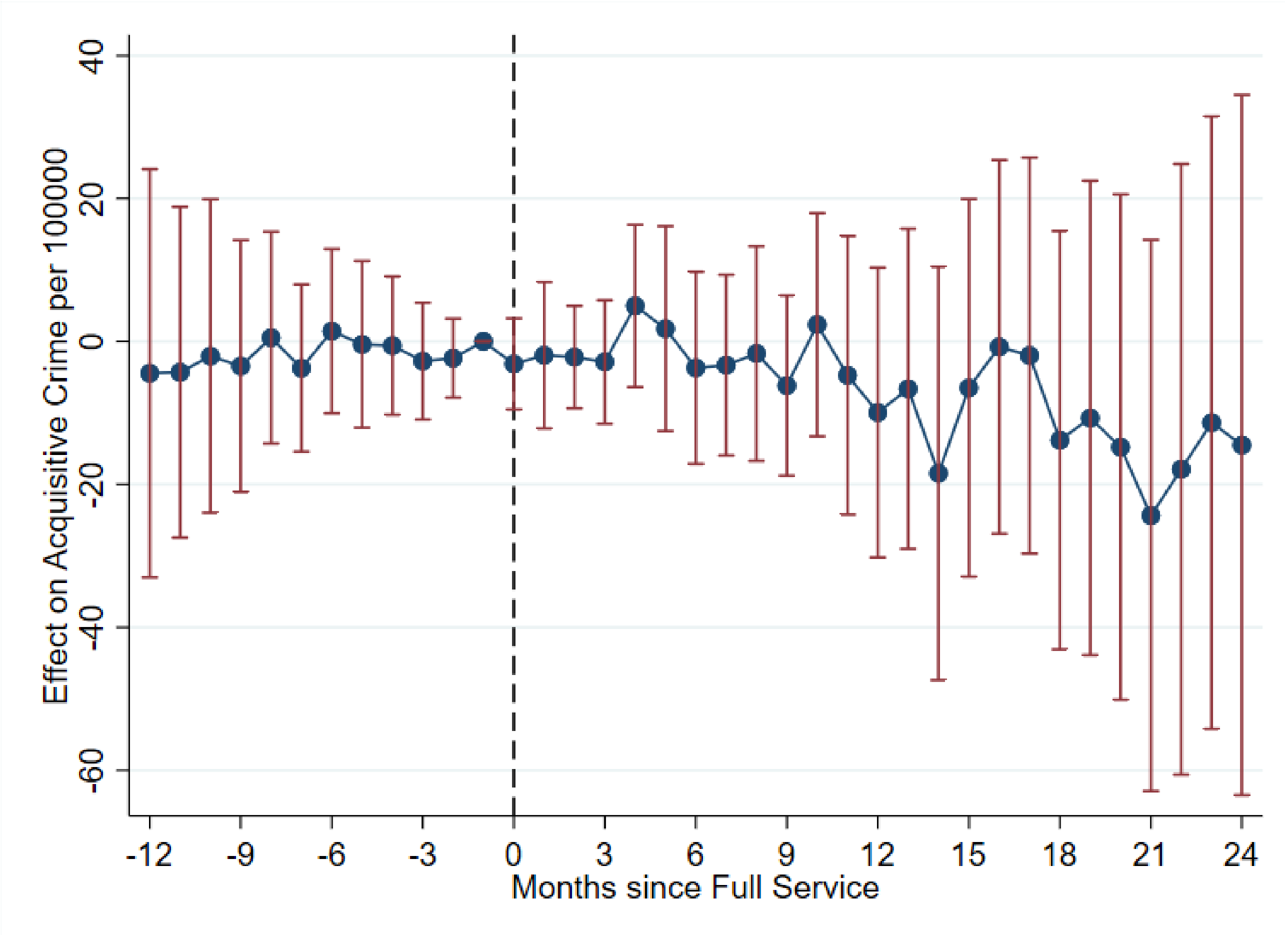
(b) Violent crime



(c) Total crime

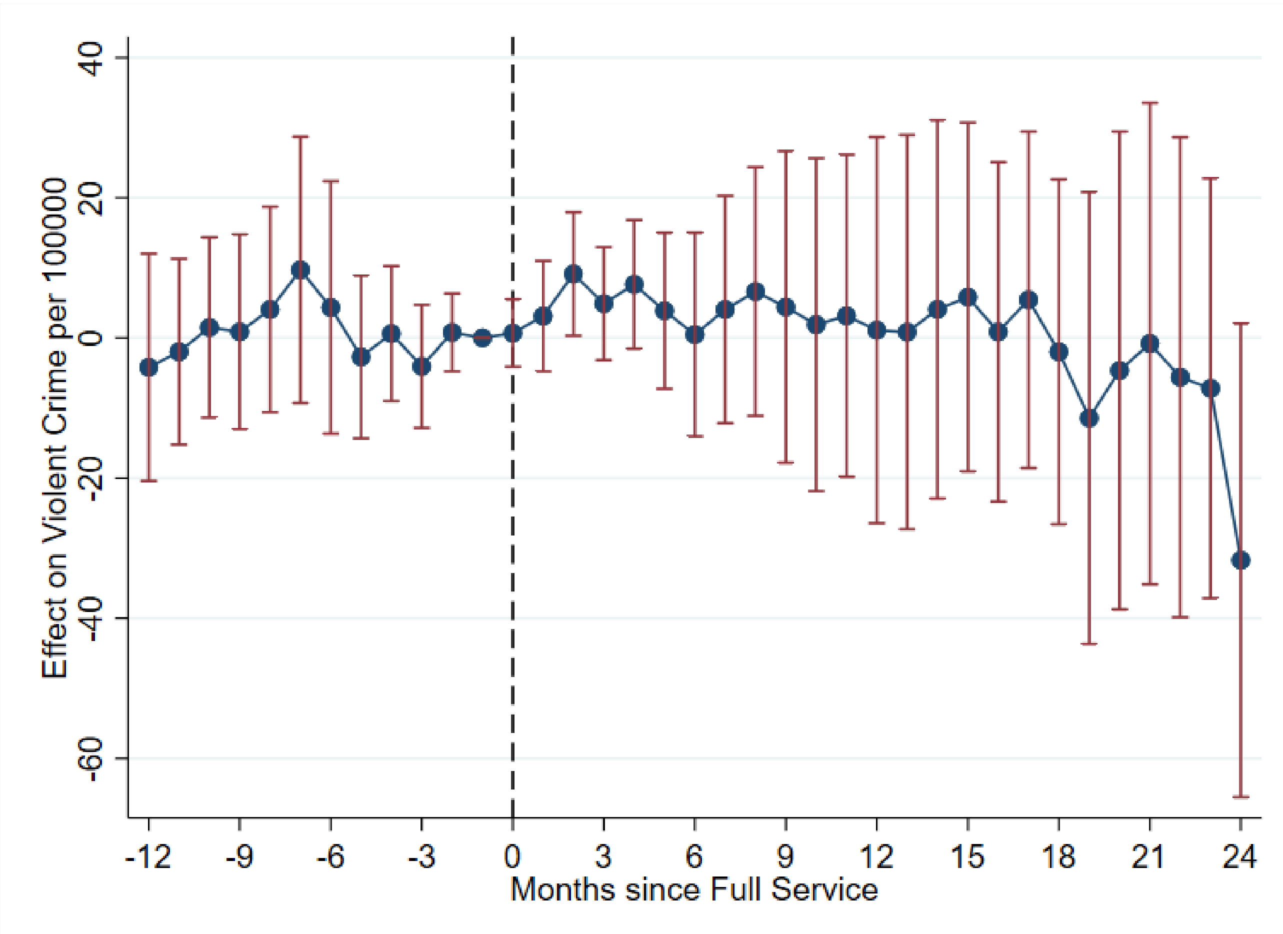
Notes: These figures plot the mean crime rates in both the 24 months before and 12 months after the FS was introduced in each local authority. The red lines are predicted values from two linear regressions of crime - either acquisitive, violent or total - on months relative to FS, either side of relative period 0. The gray shaded area shows the build up of UC claimants per 100,000 population following the FS introduction.

Figure 4.8: Full Service Event Study - Acquisitive Crime



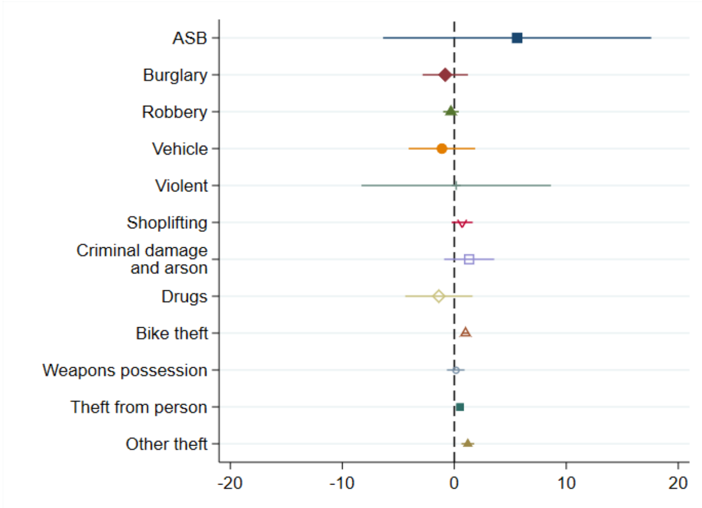
Notes: This figure shows how the average effect of the FS on acquisitive crime rates evolved over time across local authorities. 95% confidence intervals are shown as red lines. DiD_l estimates the dynamic effect l by comparing the change in acquisitive crime rates between local authorities that started receiving the FS l months ago to districts not yet converted from the legacy system. This comparison is over $l+1$ months, from the month before treatment to l month after. DiD_l restricts its comparisons to only be between local authorities that implemented the LS within the same quarter. Placebo estimates on the left of t=0 test the assumptions of parallel trends and no anticipation. Standard errors are clustered at the local authority level.

Figure 4.9: Full Service Event Study - Violent Crime

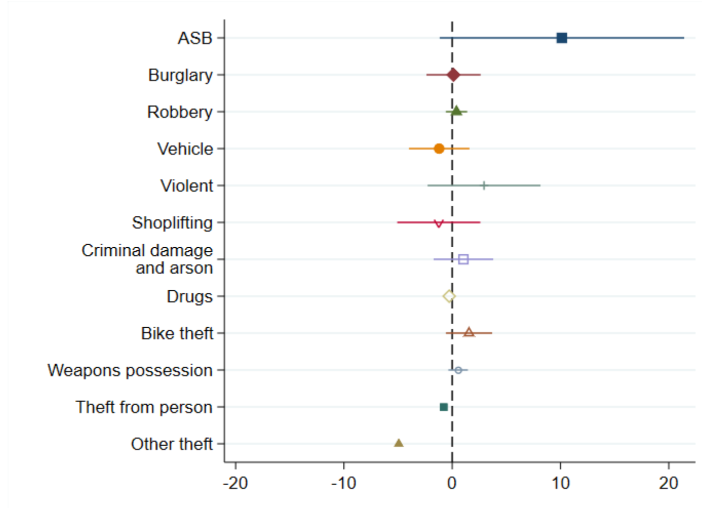


Notes: This figure shows how the average effect of the FS on violent crime rates evolved over time across local authorities. 95% confidence intervals are shown as red lines. DiD_l estimates the dynamic effect l by comparing the change in violent crime rates between local authorities that started receiving the FS l months ago to districts not yet converted from the legacy system. This comparison is over $l + 1$ months, from the month before treatment to l month after. DiD_l restricts its comparisons to only be between local authorities that implemented the LS within the same quarter. Placebo estimates on the left of t=0 test the assumptions of parallel trends and no anticipation. Standard errors are clustered at the local authority level.

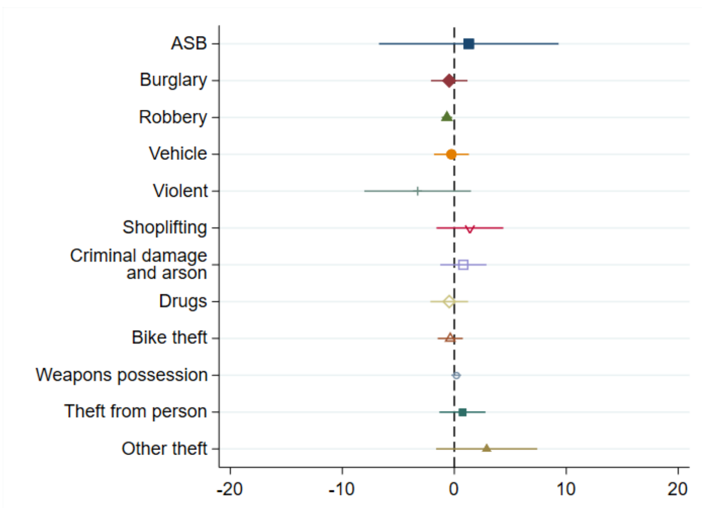
Figure 4.10: Robust Difference-in-Difference Estimators



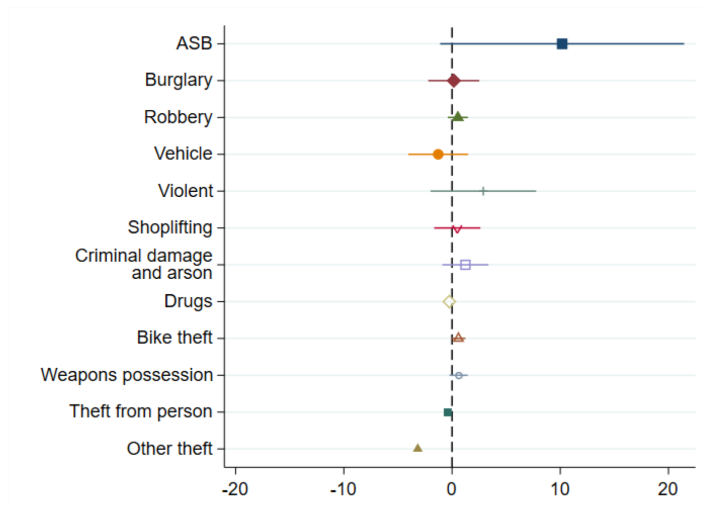
(a) de Chaisemartin & D'Haultfoeuille (2022)



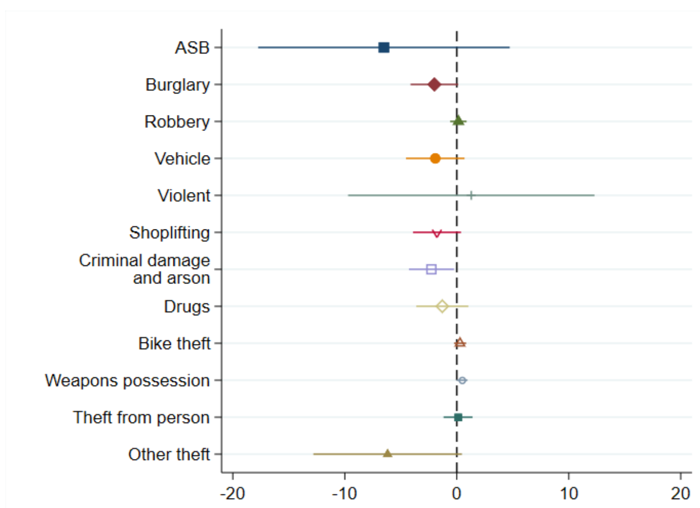
(b) Callaway & Sant'Anna (2021)



(c) Sun & Abraham (2021)



(d) Roth & Sant'Anna (2021)



(e) Borusyak (2022)

Notes: These figures show unconditional, robust DiD estimates for the FS for each crime type in the data. Point estimates are presented as symbols and 95% confidence intervals are shown as lines. Standard errors are clustered at the local authority level.

Chapter 5

Sittin' in the Dock Without Pay? Welfare Reform and Reoffending

5.1 Introduction

Prison leavers face well-documented challenges to obtaining employment.[1] They often experience issues synonymous with other disadvantaged job seekers, such as poor self-esteem, lack of basic skills, health or behavioural problems (Holzer et al., 2003; Travis et al., 2001; Hirsch et al., 2002). They are also less likely to be hired than otherwise similar candidates without a criminal record (Pager, 2003; Holzer et al., 2006, 2007; Agan and Starr, 2017). Several factors work against them. For example, they are required by law in many countries to disclose any 'unspent' criminal record if asked at any point during a hiring process, which may deter a potential employer (Pager, 2003). In addition, long periods of incarceration deprive prison leavers of recent work experience, prevent human capital accumulation and deteriorate bonds with legal job-finding networks (Schmitt and Warner, 2011; Schmitt et al., 2010). Finally, ex-offenders are barred from entering some forms of employment entirely (Bushway and Sweeten, 2007). Together, these factors contribute to as many as 60% to 75% of former offenders being out of legitimate work one year after release (Petersilia, 2003; Visher et al., 2008).

[1] Throughout this chapter I use the terms "prison leavers", "ex-prisoners", "ex-offenders" and "former offenders" interchangeably.

Consequently, welfare benefits and relapsing into illicit forms of activities represent two realistic sources of income for prison leavers. In the U.S., 70% of ex-prisoners have been shown to claim benefits within two months of release (Western et al., 2015). In England and Wales, the subject of this study, data from 2014 show 54% do so within a month (Ministry of Justice, 2014). Given this level of dependency, changes in benefit entitlement or benefit conditionality may influence ex-prisoners' propensity to re-engage in crime. Yet, despite this, the effect of welfare reform on recidivism has not been widely studied. Indeed, only three studies to my knowledge have explored this question within a quasi-experimental framework, in any setting (Yang, 2017*a*; Tuttle, 2019; Luallen et al., 2018). Moreover, all do so with respect to one reform: the Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (PRWORA). This lack of attention has not been due to a perceived lack of importance (Luallen et al., 2018). Rather, there has been a recognised challenge in obtaining data that links prisoners' offending outcomes longitudinally (Luallen et al., 2018; Sheely and Kneipp, 2015; Butcher and LaLonde, 2006; Holtfreter et al., 2004; Yang, 2017*b*).

This chapter overcomes this obstacle, as I will discuss below, and estimates the effects of Universal Credit (UC) on prisoner recidivism. This provides not only new evidence on the matter but arguably that which is of greater policy relevance, at least outside of the U.S. There are three reasons for this. First, in being an all encompassing reform to the welfare state, UC affects virtually *all* job-seeking and low income ex-prisoners. By contrast, the most notable reform affecting ex-prisoners under PRWORA was the banning of drug felons from accessing Supplemental Nutrition Assistance Program (SNAP) benefits.[2] In the U.S., drug felons constitute only around 16% of state prisoners (Yang, 2017*a*); in the UK, my data show that this fraction is 14%. As such, evidence from PRWORA concerns a minority - and specific type - of offender. Second, UC will constitute the majority of ex-prisoners' entire benefit income, rather than 20% as was estimated for SNAP (Western et al., 2015). Finally, UC is characterised by a range of policy tools - e.g. job search requirements and work incentives - that bear greater similarities to other OECD active labour market programme's than SNAP, which prohibited benefit

[2]PRWORA also legislated the Temporary Assistance for Needy Families (TANF) programme, which primarily affected low-income, single women. Studies exploring PRWORA in the context of prisoner recidivism have not focused on this element, due to the relatively small female prison population.

access. These features may offset or amplify any effect from a reduction in entitlement.

To explore UC's impacts, I leverage rich offender-case level administrative data covering the universe of prison sentences, court cases, and probation periods in England and Wales between 1st January 2011 and 31st December 2020. These data include a unique identifier that enables me to track the journey of prisoners as they move through the criminal justice system, from court to prison to probation, for both their initial and any subsequent offense within the sample period. This type of tracking has not previously been available to researchers; studies have at best relied on datasets linking only prison sentences, and even that is relatively rare (Yang, 2017*b*). One advantage of having data that links together the components of the prisoner journey through the criminal justice system is that it enables me to identify which community the prisoner returns to following release - this overcomes a key empirical obstacle highlighted within the wider literature on prisoner recidivism (Yang, 2017*b*).

I combine these novel data with a natural experiment to identify UC's intent-to-treat effects on re-offending. The experiment stems from the fact that small differences in release dates see prison leavers exposed to the two different welfare regimes: UC and the legacy system. This gives rise to a regression discontinuity (RD) design, wherein I compare the recidivism outcomes of prisoners released marginally before and after the date that their local authority of residence implemented the UC Live Service. The phased rollout of the Live Service between 29th April 2013 and 27th April 2016 works to strengthen this design. Namely, it should mitigate the impact of any shocks occurring simultaneously to the policy - a concern often raised with RD designs - since effects will be averaged out over multiple UC launch dates. Separately, incarceration greatly inhibits prisoners' ability to "sort" into their preferred welfare regime, as I will show. Taken together, this produces plausibly exogenous variation, enabling my analysis to estimate causal UC effects.

Overall, I find no evidence to suggest the introduction of the UC Live Service led to an increase in re-offending rates among single, working-age prisoners - the demographic most likely eligible for the benefit. My main analysis produced no estimates that were statistically different from zero for total, violent, and acquisitive forms of crime. In Appendix C I show this result extends to all crime categories under study (e.g. drug offences, criminal damage). My results are robust to alternative model specifications and all windows of recidivism up to 2 years following release from prison. Notably, my findings also indicate that UC's 5-week wait before initial payment did not contribute to an increase in criminal activity during this period.

However, my estimates are imprecise, and thus only able to rule out large criminological effects from the programme. For example, estimates from my main analysis can only reject (with 95% confidence) increases in total crime greater than 9.5% relative to the control group average. Estimates for subcategories of crime are more imprecise still. Exceptions to this are estimates pertaining to UC's five-week waiting period: my analysis can rule out increases in crime greater than 5.34% (with 95% confidence). In general, this imprecision is at least in part due to the small RD bandwidth chosen by the Calonico et al. (2014) selection method. This reflects the fact that prisoners released on either side of UC's implementation dates are not similar enough for comparison at distances greater than approximately 3 months. Specifically, a decline in national prosecution rates during the course of my sample meant that prison leavers released towards the end of my sample, who are also more likely to be exposed to UC, were less likely to reappear in court than prisoners released relatively earlier. While the staggered rollout of the Live Service works to mitigate this trend, it only does so to some extent.

Setting aside statistical power, this chapter builds on and tests the external validity of the nascent literature on welfare reform and recidivism among prisoners (Luallen et al., 2018; Tuttle, 2019; Yang, 2017*a*). As previously mentioned, these studies explore the recidivism effects of the SNAP ban on drug felons released from prison. Tuttle (2019) finds that drug traffickers banned from SNAP are 60% more likely to return to prison than drug traffickers with SNAP entitlement. Yang (2017*a*) similarly finds that eligibility for SNAP reduces one-year reoffending rates by up to 10% among convicted drug felons. While my results contrast these studies, they are consistent with the null effects reported in (Luallen et al., 2018).

Lastly, while UC is not targeted specifically at prisoners, my results are also relevant to re-entry policy. An experimental-based literature has shown these interventions have limited impact on reducing recidivism (Cook et al., 2015; Uggen, 2000; Mallar and Thornton, 1978; Redcross et al., 2011). What connects these studies is that they evaluate programmes that are supportive in nature, whether it be in assisting prisoners with job-search or providing them with temporary employment. My results add to this literature by showing that policies that constrain individual behaviour, such as benefit sanctions and mandatory job search, also appear not to affect recidivism.

The remainder of this chapter is structured as follows. In Section 5.2 I describe the data and construction of my dataset. In Section 5.3, I present my methodology, including a description of the identifying assumptions, and the validity tests conducted on them. Section 5.4 presents my results and Section 5.5 concludes.

5.2 Data

5.2.1 Data Sources

In this study, I test the causal effect of exposure to the UC Live Service on prisoner recidivism. To do this, I use offender-case level data provided by Data First, a data-linking programme led by the UK Ministry of Justice (MoJ) and funded by Administrative Data Research UK. These data enable the tracking of individuals' entire journey through the criminal justice system by combining, at the offender level, 4 large-scale administrative MoJ datasets pertaining to Magistrate and Crown Court appearances, prison spells and probation periods. This provides a wealth of information about every offender's offence, sentence, personal characteristics, and outcomes at each stage of the justice system in England and Wales between 1st January 2011 and 31st December 2020. These data also include a longitudinal ID for every offender, thus providing insight into whether an offender reappears in court, prison or probation for subsequent offences within this time period.

To identify whether a prisoner is exposed to the UC Live Service or legacy system upon release, I leverage information within these data on release dates and local authority of residence in tandem with the Live Service rollout timetable published by the DWP. These three factors - release date, local authority of residence, and the date each local authority implemented the UC Live Service - jointly determine which welfare regime a prisoner will be exposed to. While residency information is recorded during a prisoner's trial, and therefore available in the court element of the dataset, there is no guarantee that they will return to that area upon release. Fortunately, their local authority of residency is again recorded as part of their post-prison license agreement, which is available in the probation element of the Data First dataset. However, linking of the prison and probation datasets has been conducted by Data First at the offender level, not offender-case level. Thus, there is not a unique identifier to directly link corresponding prison and probation spells. This necessitates a matching process across the two, which I do using the unique offender ID and any shared offence date or court disposal date listed within both datasets. Once completed, I am able to gain direct insight into the welfare regime prison leavers are exposed to.

The Live Service rollout timetable I use is provided at the Job Centre level.[3] To obtain a timetable at the local authority level, I leverage a schedule for the UC Full Service rollout that matches Job Centre coverage to local authority areas.[4] This enables a matching of Live Service rollout dates to local authorities indirectly. Importantly, it identifies that there are 42 out of 348 local authorities where different Job Centres within the local authority implemented the Live Service on different dates. For these areas, it is difficult to pin down exactly which welfare regime prison leavers are exposed to, so I therefore exclude all releases into these areas from my analysis.

5.2.2 Analytic Sample

To create an offender sample for analysis, I take all prisoners released within 1 year (365 days) before or after the Live Service is introduced in their local authority. Those released up to 1 year before the Live Service form the control group; those released up to 1 year after form the treated group. The length of this window means that some prisoners are released into both welfare regimes at different points in my sample, due to having returned and being re-released from custody. For ease of interpreting my recidivism estimates, I filter out and exclude subsequent releases so there is only one per prisoner. This "baseline" release is not, to be clear, synonymous with isolating prisoners' first *ever* release. It is possible (and many do) to have a custodial history prior to this. Indeed, the baseline release is just the first release within the confines of the 2-year window.

I also exclude two other release types. First, I drop those for prisoners whose demographic characteristics rule them out from claiming the UC Live Service. This includes being married, not aged 18-69 at the time of release or being an immigration detainee. Second, I exclude the releases of individuals remanded in custody while awaiting trial - this ensures my outcomes capture only reoffending (and not also first-time offences).

Figure 5.1 shows the distribution of the 55,089 remaining baseline releases that form the basis of my sample. They span 30th May 2013 and 18th April 2017.[5] The blue histogram shows releases before

[3] Available here Department for Work and Pensions (2015).
[4] Available here Department for Work and Pensions (2018*c*).
[5] I have not shown the full left and right tails for each of these distributions for statistical disclosure reasons.

the Live Service was introduced, and thus prisoners exposed to the legacy regime (control). The red histogram shows releases occurring after the Live Service, and thus prisoners exposed to UC (treated). These distributions overlap due to the staggered nature of the Live Service rollout. My final sample re-centres these release dates so that 0 is the date the Live Service became operational.

To identify instances of reoffending, I use information within my court case data on offence dates and conviction verdicts. This approach is more advantageous than using prison re-entry dates (often used in the literature), since the latter can be a noisy estimate of when offences occurred. However, due to my data only including cases heard on or before December 31, 2020, and it showing an average lag of 109 days from offense to trial within December 2020, I observe relatively few offences being committed within the months of November and December 2020.

To therefore mitigate against the risk of under estimating recidivism rates, I take the 95th percentile of the court waiting times distribution in December 2020, 502 days, and use it as a threshold to determine the latest date an offender can be released from prison and still be included within my sample. Given the maximum recidivism window I explore is 2 years, this means that a prisoner must have been released before 17th August 2017.[6] This restriction does not lead to additional prisoners being excluded from my sample following the previous requirement of being released 365 days before or after Live Service enactment. It does mean, however, that it's difficult to analyse the recidivism effects of the UC Full Service, which was rolled out subsequently. My analysis does not evaluate this version of the programme in part for this reason.[7]

Crime types are aggregated by Data First based on the UK Home Office's offence groupings. These include: violence against the person, sexual offences, robbery, theft, criminal damage and arson, drug offences, possession of weapons, and fraud offences. To obtain a measure of acquisitive crime (i.e. financially motivated) I aggregate the theft, fraud and robbery categories. To obtain a measure of

[6] 31/12/2020 - 365x2 days - 502 days = 17/8/2017.

[7] The other challenge with estimating the impact of the Full Service is that during its rollout the Live Service is already in operation in almost all local authorities. This means that when a prisoner is released, they are very likely to be eligible for UC regardless of which local authority they live. As such, it is difficult to find valid control groups for a Full Service analysis. Only married prisoners would be guaranteed to be ineligible for UC in Live Service areas (assuming they were married before the Live Service rolled). However, analysis into this subgroup would likely suffer additional issues of statistical power, since my data show that 10% of prisoners are married.

violent crimes I use only the category violence against the person. My analysis focuses on total, acquisitive and violent crime; however, I present results for all crime groupings in Appendix C.

5.2.3 Descriptive Statistics

Summary statistics for my sample in Table 5.1 show that reoffending is prolific in both welfare regimes: 57% of prison leavers living in legacy areas and 45% of prison leavers in Live Service areas are found guilty for an offence within 1 year of leaving prison; this rises to 68% and 58% respectively within 2 years. It also highlights that there are clear differences in reoffending outcomes across the two regimes. Surprisingly, it shows that prisoners released after the Live Service is introduced have lower reoffending outcomes, regardless of crime type (total, violent or financial) or window of recidivism. For example, prisoners released into legacy areas are more than twice as likely to have committed an acquisitive crime within the first 5 weeks of release: 7% versus 3% respectively. This is particularly surprising given the 5 week wait for the first payment under UC. Differences in violent crime rates are smaller: the one year reoffending rate for violent crime is 11% for legacy and 9% for the Live Service; for acquisitive crime, the equivalent is 25% and 15%. A closer inspection of Table 5.1 reveals relative differences in reoffending to be declining over time. For example, reoffending rates after 1 quarter are 50% higher in legacy areas (30% versus 20%), yet only 17% higher after 2 years (68% versus 58%). This can be seen clearly in Figure 5.2: the reoffending gap in absolute terms between legacy (blue line) and Live Service (maroon line) areas remaining roughly constant from the second quarter after release through to the eighth quarter after release.

Differences in sample demographics do not appear to be driving this finding. For example, both samples are on average 90% male, 90% British and have an average age of release of 32. Both samples also consist entirely of single prisoners (by construction), and differences in the proportion White, Black and Asian are within 1 percentage point. The same is true for all self-reported religious beliefs except the proportion of Christians in each sample (the difference is within 2 percentage points).

However, Table 5.1 also reveals discrepancies in prisoners' criminal history in the 2 years preceding their baseline prison sentence. On average, prisoners released into legacy districts had 1.47 previous prison spells, compared to 0.94 for those in Live Service areas. This pattern can be seen more broadly for prior court appearances: 3.38 for Live Service areas; 2.28 for legacy areas. The differences are more pronounced for acquisitive crime than for violent crime, which is consistent with the reoffending outcomes reported at the top of the table.

These differences likely reflect a broader trend within the criminal justice system during the Live Service rollout period, namely a steady decrease in the number of cases heard in courts across England and Wales. As shown in Figure 5.3, this trend was particularly acute for acquisitive crimes, which saw a sharp decline of 41.2% between May 2013 and April 2016. Figure 5.3b further reveals that this decline was mainly driven by a reduction in theft cases (-43%). Importantly, the steepest decline in theft occurred between January 2015 and April 2016, the same period during which 94% of local authorities implemented the Live Service. Thus, despite the staggered rollout, the differences in acquisitive crime shown in Table 5.1 likely reflect the fact that fewer cases were heard in court in 2015-16 than in 2014-15.

The question remains however whether the observed decline in court cases reflects an actual drop in theft offences. To explore this, I re-use the monthly police recorded crime data from Chapter 4 and leverage Home Office data on annual police charge rates (the proportion crimes where a suspect is arrested and charged).[8] Figure 5.3c shows that police recorded theft in fact remained relatively stable over the period of the Live Service rollout, whereas charge rates more than have halved between 2011 and 2018 (13.4% to 5.7%). Taking this in conjunction with the concentrated Live Service rollout, it is therefore conceivable that differences in the rates of offending and criminal history between the control and treatment groups reflect changes in charge rates rather than actual changes in criminal behaviour. Fortunately, these differences should not be a threat to identification so long as they develop gradually in the neighbourhood period of the Live Service dates. The following section outlines my RD research design in more detail.

[8] See Home Office (2012, 2018).

5.3 Methodology

5.3.1 RD Research Design

The main empirical challenge to identifying UC's effect on prisoner recidivism is to remove the presence of confounders. Randomly assigning prisoners to UC upon release would solve this issue, since individuals exposed to both benefit regimes would be similar enough on average for a fair comparison of reoffending outcomes. In absence of an RCT, my empirical strategy is to adopt a sharp regression discontinuity (RD) design. This approach exploits the fact that the probability of being assigned to UC is a discontinuous function of prisoners' release date. Specifically, prisoners released before the Live Service is implemented in their local authority are exposed to the legacy system, whereas those released after are exposed to UC. Figure 5.1 provided a visual for these releases (albeit only for the relevant baseline ones in my sample). For my analysis to exploit this, I use the re-centred release dates described in Section 5.2.2 - where 0 is defined as the date that the Live Service became operational in a prison leaver's local authority - as the "running variable" that determines treatment exposure.

Intuitively, the idea with an RD design is that prisoners released marginally before and after the date the Live Service was introduced should be comparable in both observable and unobservable ways. Note, for the remainder of this chapter I often refer to this date as the "cut off". In the following section I discuss the identifying assumptions underlying my RD design and why in principle they would be expected to hold in this setting.

5.3.2 Identifying Assumptions

For the RD design to establish causality, prisoners released in the neighbourhood of the cut-off (either side) must have the same pre-treatment propensity to reoffend on average. A priori this requirement becomes more plausible because prisoners cannot easily manipulate which benefit regime they become exposed to. Indeed, because they are prohibited from claiming benefits while incarcerated, they are subject to whichever welfare regime is operational in their community once released. This rules out the potential for the types of anticipatory behaviour possibly occurring among other demographic groups. For example, individuals may make a claim to the benefit system earlier than they would have otherwise to avoid claiming UC. Similarly, recipients of the legacy benefits may choose to restrict their behaviour so as to avoid triggering a new circumstance that migrates them to UC. Prisoners, by contrast, have no freedom to do so. Their only scope for strategic manipulation is to delay their claim, which they could do, for example, by committing a further transgression so as to push back their release date. This would be an empirical concern only if they did so in order to wait for UC. While possible, in practice this seems unlikely given the disincentives prisoners face. Setting aside the 5 week waiting period, most prisoners will be worse off under UC (some substantially so) because of the way the programme treats single jobseekers (Brewer, et al, 2019). In Section 5.3.3 I test the validity of this assumption.

The second identifying assumption is that if prison leavers choose to claim benefits, they do so on the day they are released. This is needed for the RD design to be sharp. If prison leavers released into a legacy area hold off on claiming, the control group will be contaminated with people on UC, and thus there will be no (or a diminished) discontinuity to identify the treatment effect. Fortunately, for this study, prisoners are assisted and actively encouraged to claim immediately upon release by prison "work coaches". These are DWP employees based in every UK prison whose job is to prepare offenders approaching release for the transition on to welfare or employment. Official government guidance maintains that their role involves checking which benefits prisoners will be entitled to and working with resettlements teams to ensure they have the necessary documents - an ID and a bank account - to make a claim once released (Department for Work and Pensions, 2023). Importantly, they are also tasked with booking a telephone appointment with the Job Centre for the prisoner to take place "on the day of release or earliest available date", and to "ensure the prison leaver understands the importance of any Job Centre telephone appointment that have been booked for them" (Department for Work and Pensions, 2023). This appointment is to assist offenders making their claim that day.

5.3.3 Checking Validity of the RD Design

I test the validity of my first identifying assumption in two ways. First, following Doleac (2017), I predict prisoners' risk of reoffending within 12 months, based on their characteristics, and check whether it is continuous in the running variable around the Live Service introduction. Second, I conduct a McCrary (2008) density test to check for evidence of prisoners manipulating their release date so as to sort into a particular welfare regime. If there is no "jump" in predicted in risk and no evidence of sorting, it suggests prisoners released just before and just after the policy are likely to have had the same pre-treatment propensity to reoffend.

Turning first to predicted risk, the key here is to gain a measure of offenders' propensity to reoffend in the absence of UC. To do so, I estimate using OLS the following linear probability model:

$$Recidivism_{jdc} = \beta_1 Demo_{jcd} + \beta_2 Crim_{jcd} + \epsilon_{jcd} \tag{5.1}$$

Where $Recidivism_{jcd}$ equals 1 if prisoner j living in local authority district d reoffended within 12 months following case c, and 0 otherwise. $Demo_{jcd}$ is a vector of prisoner-case level characteristics: sex, age at release, nationality, ethnicity and religious beliefs. $Crim_{jcd}$ is a vector of prisoner-specific criminal history in the 2 years before case c, including the number of prison spells and court appearances for violent, acquisitive and any crime in general. To measure recidivism risk absent of UC, I use observations from outside my analytic sample to estimate β_1 and β_2 . Specifically, I specify that prisoner j must have been released 12 months *before* the Live Service was introduced in district d, as well as after 1st January 2013 to ensure a 2 year window to identify B_2. This leaves 103,222 baseline releases for estimation. Predicted risk for prisoner i in my analytic sample is computed by the following linear equation:

$$\widehat{Recidivism}_{idc} = \hat{\beta}_1 Demo_{idc} + \hat{\beta}_2 Crim_{idc} \tag{5.2}$$

Figure 5.4a plots $\widehat{Recidivism}_{idc}$ against the running variable after binning prison releases into two-week intervals. Each dot contains approximately 1,000 observations and its position, relative to the y axis, represents the proportion of that prison cohort who reoffended within 12 months of release. The plot shows a clear and persistent downward trend but no discontinuity around the Live Service cut-off. Figure 5.4b shows that the downward trend in risk score becomes more gradual when theft reoffences are excluded from $Recidivism_{jcd}$, as expected. In Appendix C I plot equivalent figures for all offender demographics and criminal history. These similarly show no significant discontinuity around the cut-off.

I next conduct the McCrary density test. This is a more powerful test for the validity of my first identifying assumption. In contrast to the measure of recidivism risk, it can reveal whether there are differences in unobservables across both control and treatment groups. The reason for this is as follows. First, unobservables play a role in formulating prisoners' beliefs about their potential outcomes under each regime. Second, if a prisoner believes that their potential outcome under one regime exceeds the other, theory suggests that they might manipulate their release date - where possible - so as to be exposed to that particular regime. As such, differences in unobservables can be revealed by the physical action of sorting. The McCrary test checks for this sorting by looking for a break in the density of prison releases around the cut-off. The null hypothesis is that there is no break i.e. no manipulation. In this setting it is not possible to sort into the legacy regime given the constraints imposed while incarcerated. The test is therefore only relevant to check for sorting into UC. It computes a log difference in height of -0.0047 and a standard error of 0.0432, which means I cannot reject the null hypothesis of no manipulation at any reasonable significance level. Figure 5.5a shows this result visually: there is no jump or drop in the number of offenders released on either side of the cut-off. Figure 5.5b illustrates the result more clearly by zooming in on the frequency of releases 20 weeks before and after the cut-off. Overall, I conclude there is no indication of sorting behaviour and thus the RD design is credible.

5.3.4 Estimating the Recidivism Effects of UC

With evidence of the main identifying assumption being valid, a natural next step is to plot my main outcomes of interest – (total) recidivism, acquisitive recidivism and violent recidivism – against the running variable. If prisoners' propensity to reoffend is affected by UC, and they claim benefits on the day they are released, there should be a discontinuity on the day the Live Service was launched. Figure 5.6 plots this for reoffending outcomes within 12 months. As before, each dot contains approximately 1,000 prisoners and represents the average reoffending rate for a given cohort released within a 14-day period.

Figure 5.6 shows no clear change in recidivism rates for total or acquisitive crime around the Live Service introduction date. Prisoners released around either side of the cut-off have an average reoffending rate of 18% for acquisitive crime within a year of release. For violent crime, a quadratic polynomial shows a positive discontinuity, but this looks to have been driven by one outlier observation. Overall, the policy looks to have had little to no effect on either type of reoffending outcome at this first pass.

Although there appears to be no visual effect, it is important to quantify my estimate in case the effect sizes are small. This gives rise to a key challenge in an RD setting, namely, correctly specifying the functional forms for $\mathbb{E}[Y_i(1)|X_i = x]$ and $\mathbb{E}[Y_i(0)|X_i = x]$, the unknown regression functions when treated and untreated. Failing to correctly specify these functions will result in the RD design yielding a biased estimate of the treatment effect, even if the identifying assumptions are satisfied. Looking at Figure 5.6, the quadratic polynomial fitting these data appears to approximate its trend well, and therefore looks like a natural candidate for the functional forms. However, polynomials using all observations such as these have been shown to be problematic: the weights on observations on either side of the cut-off can have unattractive properties, and thus can result in treatment effect estimates being sensitive to the order of the polynomial specified - this arises in part because global polynomial often gives poor approximations at boundary points (e.g. at the cut off) (Gelman and Imbens, 2019).

I therefore estimate the RD treatment effect using non-parametric methods, which has become standard in the literature. This approach lets the data decide the appropriate functional forms. Its main advantage is that it is less sensitive to boundary problems because only observations local to the cut off are used to estimate (using least squares) the polynomial fit for $\mathbb{E}[Y_i(1)|X_i = x]$ and $\mathbb{E}[Y_i(0)|X_i = x]$ (Calonico et al., 2014; Porter, 2003; Fan and Gijbels, 1992). It also appears appropriate in this setting given my control and treatment group may not be similar, at least with respect to criminal history, over a 1 year window either side of the Live Service (shown in Table 5.1). Figure 5.4 supports the notion of comparisons being valid local to the cut off. While one possible concern is that non-parametric estimation may not be suitable with a discrete running variable (Lee and Card, 2008), my running variable - the number of days between a release and the Live Service launch - takes on many distinct values which Calonico et al. (2014) suggests should mitigate these concerns.

The question that naturally follows is how local is local. In other words, what is the width of the neighbourhood around the cut-off – the so-called "bandwidth" – to be used for estimation. The choice of bandwidth is fundamental to the analysis as it directly affects the fit of the estimated polynomial, in turn influencing the treatment estimate. Smaller bandwidths induce less misspecification error – i.e. provide a better approximation of the unknown regression functions – since approximations are more accurate the more local they are. The cost, however, is that they lead to greater variance in the estimate since they use fewer observations. To avoid arbitrariness, I choose the bandwidth in a data-driven way. In particular, I use the Calonico et al. (2014) optimal bandwidth selection method, which amounts to picking the local polynomial RD point estimator with a minimum mean squared error (MSE). As the MSE can be decomposed into a bias and variance term, this bandwidth optimally balances this bias-variance tradeoff (Cattaneo et al., 2019).

The last two factors to be considered are the kernel function and polynomial order. The kernel function determines the weight the RD estimator places on observations within the bandwidth (the weight is 0 for observations outside the bandwidth). For this analysis I use a triangular kernel, which increases the weight on observations the closer they are to the cut of. This is an attractive property theoretically given how RD seeks to identify treatment effects. It is also recommended in practice when combined with the MSE optimal bandwidth (Calonico et al., 2014). For the choice of polynomial order, Gelman and Imbens (2019) recommend a linear or quadratic specification. For my main analysis I use the former.[9]

I add demographic and criminal history related covariates to test the robustness of my treatment ef-

[9]Using a quadratic specification does not change my results.

fect estimates. This can be done without fear of inducing bias since these covariates are predetermined and, as shown in Appendix C, continuous around the introduction of the Live Service. Although the latter suggests the legacy and Live Service sample will be similar locally, and thus the RD design will be unbiased, controlling for offender characteristics will help ensure that small differences that do exist do not contaminate my treatment effect estimates. In addition, by helping to explain variation in offending propensity, these controls should improve the efficiency of the RD point estimator. The goal here is to improve this precision without affecting the magnitudes of the estimates. The next section showing my results will test this.

5.4 Results

In this section I estimate the effect of exposure to UC on recidivism using the date local authorities implemented the Live Service as a sharp cut-off. It is worth highlighting that, as suggested by the word "exposure", the effects estimated in this chapter should not be interpreted as local average treatment effects of *claiming* UC. Indeed, since I do not have access to UC up-take data, I cannot scale my estimates in accordance with offender "compliance". My estimates should instead be interpreted in one of two ways. First, as intent to treat effects - that is, the causal effect of being assigned to treatment. Second, if the Live Service has effects on those not claiming the benefit - i.e. indirect effects - then estimates should be interpreted as local average treatment effects of the programme being launched in a local authority. Indirect effects may arise through a peer effect mechanism, namely, where ex-offenders claiming UC interact with other non-UC claiming ex-offenders - past research has shown peer effects outside of prison to be an important determinant of recidivism (Billings and Schnepel, 2022; Corno, 2017; Damm and Dustmann, 2014).

5.4.1 12 Month Recidivism Rates

Table 5.2 displays the results of my main analysis: the effect of exposure to the Live Service on recidivism rates for total, acquisitive and violent crime in the 12 months following release from prison.

Controls for each specification are shown at the bottom of the table. Standard errors are clustered at the local authority level and adjusted to be bias-corrected, as recommended by Cattaneo et al. (2019).[10]

Turning first to total recidivism in Panel A, Column (1), the baseline model estimates an increase in recidivism of 0.018 percentage points, though it is statistically insignificant at any conventional level. Column (2) controls for sex and age at release, which shrinks the point estimate to 0.013 (again, statistically insignificant). Further controlling for foreign nationality and ethnicity in Column (3) further reduces the estimate to 0.009.[11] Column (4) and (5) include the number of previous court appearances and whether a prison spell took place within the 2 years prior to the baseline prison spell. Table 5.1 showed that these dimensions differed between prisoners released 365 days before and after the Live Service was introduced. Their inclusion therefore attempts to capture any small differences that arise even locally (at the cut-off). Ideally, they should leave the magnitude of the estimate unchanged and explain greater variation in the error term, in turn improving precision. Column (4) and (5) show the estimates do in fact remain stable, at 0.008 for both models. However, they do not improve estimation efficiency, somewhat surprisingly: the standard errors only shrink from 0.021 to 0.019 in Column (4) and (5). Statistical power is in fact low in general across the 5 specifications. This is likely a result of the small bandwidths chosen by the Calonico et al. (2014) selection method (between 87-94 days). Column (5), the model with all controls, is thus only able rule out increases in recidivism by 0.045 percentage points with 95% confidence - that is, an increase of 9.5% relative to the control group mean.

Panel B shows the results for acquisitive crime, the crime type that Becker (1968) would predict UC to influence rates of. Estimates in this panel are positive yet statistically insignificant, thus not providing evidence to support his hypothesis. Estimation efficiency is again an issue with these models, however. The bandwidths chosen by the Calonico et al. (2014) selection method (between 82-91 days) are narrower than that for total crime. It is likely that this reduction is due to the decline

[10]Cattaneo et al. (2019) discusses how using conventional OLS standard errors can lead to invalid inference when non-parametric estimation methods are used. This arises out of the fact that the goal of such methods is to approximate the unknown regression function, rather than assume its exact function form (as with parametric methods). This can lead to misspecification bias that needs to be adjusted for.

[11]Ethnicity includes White, Black or Asian. A dummy for mixed race is excluded as the benchmark.

in convictions for acquisitive offences over the course of my sample period - a narrower bandwidth is needed for rates to be comparable across the control and treatment group. As a consequence of fewer observations being used for estimation, Column (5), the model with all controls, can only rule out increases in acquisitive crime by 18.0%.

Estimates in Panel C similarly show no statistically significant effect for violent crime. Despite the bandwidth being larger (106-120 days) than those for total or acquisitive crime, these estimates are too imprecise to be informative about UC's impact. Column (5), for example, can only rule out increases in violent crime greater than 38% relative to the control group mean.

5.4.2 Impact of UC's 5 week wait

I now explore whether UC's controversial 5 week wait until initial payment led to an increase in crime within this period. The assumption here, as I make throughout my analysis, is that if prisoners make a claim to the benefit system, they do so on the day they are released. This is because I do not observe the date of when prisoners actually claim. Under this assumption, the first 5 weeks of release map directly to the 5-week waiting period under UC. Table 5.3 presents the results. It follows the same structure of Table 5.2: estimates for total, acquisitive and violent crime are displayed in subsequent panels, and controls are shown at the foot of the table for each model. Standard errors are again clustered at the local authority level and adjusted to be bias-corrected Cattaneo et al. (2019).

Panel A suggests the 5-week waiting period does not increase total crime. In fact, surprisingly, all estimates in panel A are negatively signed; though they are statistically insignificant at any reasonable level. Column (1), the baseline model, produces an estimate of -0.017. Columns (2)-(5) shows that adding prisoner demographic and offending history controls does not help improve estimation efficiency. The estimate in Column (5), the model with all controls, computes the same estimate as the baseline model (-0.017). Given its standard error (0.012), it rules out increases in recidivism above 5.34% of the mean.

Estimates in Panel B for acquisitive crime are positive, in contrast to panel A, though again statistically insignificant. The baseline model in Column (1) produces an estimate of 0.002, which corresponds to a 5.56% increase in acquisitive crime relative to the control group mean. Adding prisoner demographic and offending history controls in Column (2)-(5) does not help improve estimation efficiency: standard errors are equal to 0.008 throughout. Panel C tells a similar story for violent crimes. The estimates in each model are positive yet too imprecise to be informative about UC's impact on this offence type.

5.4.3 Recidivism Windows

I now turn to consider multiple recidivism windows. This serves as a useful robustness check to see if my main results are sensitive to the window under consideration. It also enables me to investigate whether UC effects take time to manifest - perhaps, for example, as a result of claimants gradually accumulating debt. As discussed in Section 5.2.2, the reoffending outcomes for all 55,089 prisoners within my sample are observed for up to 2 years. This means that, up to this period, estimation power should not be unduly affected for long recidivism horizons versus short. Further, any differences in treatment estimates across the recidivism windows under consideration will not be a function of a changing sample composition. Standard errors are again clustered at the local authority level and bias-corrected Cattaneo et al. (2019). Bandwidths are chosen using the (Calonico et al., 2014) selection method.

Figure 5.7 shows UC's estimated impact on total, violent and acquisitive crime for all recidivism windows up to 104 weeks following release (2 years). The black line for each figure connects 104 separate RD point estimates, each corresponding to an incrementally longer recidivism window (1 week, 2 weeks,..., 104 weeks). 95% confidence intervals for the estimates are presented as dashed lines. Figure 5.7a) and Figure 5.7c) suggest there is no effect on acquisitive or total crime, consistent with the findings in Section 5.4.1 and Section 5.4.2. Figure 5.7b) suggests violent reoffending steadily increases in the first 33 weeks (approximately 8 months) following release in local authorities with UC, relative to legacy. At its peak in week 33, the point estimate is 0.02 which translates to a 28% increase relative to the average violent re-offending rate within this period (7.13%). This treatment effect is large, yet the low estimation power means that only estimates from around 20-33 weeks are statistically significant at the 5% significance level. Therefore, it is difficult to interpret this as robust evidence for UC causing an increase in (short-run) violent re-offending. Overall, the analysis does not point to any strong evidence for an increase in crime. In Appendix C I show this is also the case for criminal damage, possession of weapons, drug offences, antisocial behaviour, theft and total crime excluding thefts.

5.5 Conclusion

Prison leavers often face significant barriers to employment, making them reliant on welfare support. Yet little is known about how welfare reform affects their likelihood of reoffending. Prior quasi-experimental evidence on this matter pertains to the withdrawal of SNAP benefits for drug offenders in the US. However, this reform is nearly three decades old and only applies to a specific type of offender. Moreover, evidence on its impacts may not accurately reflect those of modern social security systems, which feature a number of policy tools aimed at transitioning individuals into employment. This study aimed to provide updated and policy-relevant evidence on these impacts by studying UC.

Leveraging rich administrative data on court cases, prison and probation spells, I evaluated UC's effect on reoffending within ex-offenders' first 2 years of release from incarceration. To establish causality, I exploited the Live Service staggered rollout across local authorities as discontinuities within a sharp regression discontinuity (RD) design. Results from a McCrary density test and predicted reoffending propensity indicated that the research design was valid.

My results suggest that exposure to the UC Live Service does not lead to an increase in reoffending among prison leavers. My main analysis ruled this out for total and acquisitive forms of crime, for all recidivism windows from 1 week post-release to 2 years. Graphical analysis in Figure 5.7 provided some evidence that violent crime may have been increasing in the first 33 weeks post-release (approximately 8 months). However, while there is a clear upward trend during this period, only estimates from week 20-33 were statistically significant at the 5% significance level. It is therefore difficult to consider this as robust evidence of a treatment effect.

A key limitation of my RD analysis is that my estimates are imprecise. Estimates from my main analysis, for example, could only rule out increases in total crime by 9.5% relative to the control group mean. This was at least in part due to the short bandwidth (around 90 days) chosen by the Calonico et al. (2014) selection method. While in theory increasing its length would improve estimation power, doing so would likely undermine the validity of the design, given Table 5.1 highlighted control and treatment group differences in prisoners' characteristics, specifically criminal history, further away from the cut-off.

Future UC research may be able to overcome this challenge with access to benefit data linked to these criminal justice data. Data linked in this fashion would enable researchers to scale up their estimates based on UC take-up - i.e. the first stage effect. In theory, this should help mitigate the fact that small bandwidths likely have to be used in this setting. A second, related advantage of having benefit-linked data is that it allows researchers to relax the "claim on release date" assumption this analysis hinged on. Together, these factors help strengthen the credibility and statistical power of RD designs seeking to identify UC's causal effect on reoffending.

Table 5.1: Summary Statistics

	Legacy			Live Service		
	Mean	Std Dev	N	Mean	Std Dev	N
Recidivism type and window						
Reoffended 5 weeks	0.165	0.371	29,891	0.100	0.300	25,198
Reoffended 1Q	0.304	0.460	29,891	0.203	0.402	25,198
Reoffended 2Q	0.435	0.496	29,891	0.318	0.466	25,198
Reoffended 3Q	0.513	0.500	29,891	0.397	0.489	25,198
Reoffended 4Q	0.566	0.496	29,891	0.452	0.498	25,198
Reoffended 8Q	0.681	0.466	29,891	0.582	0.493	25,198
Acquisitive reoffence 5 weeks	0.069	0.254	29,891	0.029	0.168	25,198
Acquisitive reoffence 1Q	0.131	0.337	29,891	0.062	0.242	25,198
Acquisitive reoffence 2Q	0.189	0.391	29,891	0.101	0.301	25,198
Acquisitive reoffence 3Q	0.224	0.417	29,891	0.128	0.334	25,198
Acquisitive reoffence 4Q	0.250	0.433	29,891	0.149	0.356	25,198
Acquisitive reoffence 8Q	0.312	0.463	29,891	0.199	0.399	25,198.
Violent reoffence 5 weeks	0.018	0.134	29,891	0.014	0.118	25,198
Violent reoffence 1Q	0.040	0.196	29,891	0.029	0.167	25,198
Violent reoffence 2Q	0.067	0.250	29,891	0.050	0.217	25,198
Violent reoffence 3Q	0.092	0.289	29,891	0.069	0.253	25,198
Violent reoffence 4Q	0.112	0.316	29,891	0.085	0.279	25,198
Violent reoffence 8Q	0.173	0.378	29,891	0.138	0.345	25,198
Demographics						
British	0.904	0.295	29,697	0.898	0.303	25,070
Male	0.897	0.304	29,696	0.905	0.294	25,069
Age at release	31.9	9.608	29,697	32.3	9.862	25,070
Asian	0.053	0.224	29,697	0.061	0.239	25,070
White	0.799	0.401	29,697	0.792	0.406	25,070
Black	0.091	0.287	29,697	0.094	0.292	25,070
Christian	0.431	0.495	29,697	0.417	0.493	25,070
Atheist	0.398	0.490	29,697	0.407	0.491	25,070
Muslim	0.089	0.285	29,697	0.098	0.297	25,070
Jewish	0.002	0.043	29,697	0.002	0.046	25,070
Hindu	0.003	0.056	29,697	0.004	0.064	25,070
Criminal history and predicted reoffending						
Prison spells	1.471	1.242	29,891	0.944	0.536	25,198
Previous court app.	3.378	3.428	29,891	2.281	2.308	25,198
Previous court app. (acquisitive)	1.664	2.597	29,891	0.919	1.600	25,198
Previous court app. (violent)	0.520	0.886	29,891	0.415	0.724	25,198
Predicted reoffend 4Q	0.547	0.201	29,696	0.478	0.143	25,069

Notes: Descriptive statistics for prisoners released up to 365 days after (before) the Live Service is introduced in their local authority of residence are shown on the left (right) hand side, under the heading "Live Service" ("Legacy"). At the bottom of the tables, "app." refers to appearances and "Predicted reoffend 4Q" refers to the predicted 1 year reoffending values given by the linear probability model outlined in Section 5.3.3.

Table 5.2: Effect of UC on Recidivism (12 months)

	(1)	(2)	(3)	(4)	(5)
Panel A: Recidivism					
UC Live Service	0.018	0.013	0.009	0.008	0.008
	(0.022)	(0.021)	(0.021)	(0.019)	(0.019)
Control group mean	0.471	0.472	0.476	0.476	0.476
Observations	55,089	54,765	54,765	54,765	54,765
Bandwidth (days)	87	89	93	94	93
Panel B: Acquisitive Recidivism					
UC Live Service	0.008	0.008	0.007	0.006	0.003
	(0.016)	(0.016)	(0.016)	(0.015)	(0.014)
Control group mean	0.167	0.167	0.167	0.167	0.169
Observations	55,089	54,765	54,765	54,765	54,765
Bandwidth (days)	84	83	82	83	91
Panel C: Violent Recidivism					
UC Live Service	0.010	0.010	0.016	0.016	0.015
	(0.011)	(0.011)	(0.012)	(0.011)	(0.012)
Control group mean	0.100	0.100	0.098	0.100	0.100
Observations	55,089	54,765	54,765	54,765	54,765
Bandwidth (days)	118	120	108	106	101
Sex		Yes	Yes	Yes	Yes
Age at release		Yes	Yes	Yes	Yes
British			Yes	Yes	Yes
Ethnicity			Yes	Yes	Yes
Previous court app.				Yes	Yes
Ex-prisoner					Yes

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.10$. Heteroscedasticity-robust standard errors adjusted for clustering at the local authority level are reported in parentheses. A polynomial of order 1 is used for estimation. Bandwidth size is chosen by the Calonico et al. (2014) selection method.

This table shows the estimated effect on reoffending within 12 months of prison release from exposure to the UC Live Service. Panel A shows results for recidivism (i.e. any form of crime); panel B shows results for acquisitive crime and panel C shows results for violent crime. Controls for each model are shown at the foot of the table.

Column (1) is the baseline model with no controls. Column (2) controls for prisoners' sex and age at release and Column (3) controls for whether the prisoner is of British nationality and whether they are White, Black or Asian (mixed race is the excluded benchmark). Column (4) controls for the number of previous court appearances within the 2 years preceding the baseline prison spell. Column (5) includes a dummy variable for whether the prisoner had been to prison in the 2 years prior to the baseline prison spell.

Table 5.3: Effect of UC 5 Week Wait

	Panel A: Recidivism				
	(1)	(2)	(3)	(4)	(5)
UC Live Service	-0.017	-0.014	-0.012	-0.019	-0.017
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Control group mean	0.122	0.117	0.117	0.122	0.122
Observations	55,089	54,765	54,765	54,765	54,765
Bandwidth (days)	98	120	111	98	101

	Panel B: Acquisitive Recidivism				
	(1)	(2)	(3)	(4)	(5)
UC Live Service	0.002	0.001	0.001	0.000	0.002
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Control group mean	0.036	0.035	0.035	0.036	0.036
Observations	55,089	54,765	54,765	54,765	54,765
Bandwidth (days)	70	69	69	69	71

	Panel C: Violent Recidivism				
	(1)	(2)	(3)	(4)	(5)
UC Live Service	0.002	0.002	0.003	0.003	0.003
	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)
Control group mean	0.016	0.016	0.015	0.016	0.016
Observations	55,089	54,765	54,765	54,765	54,765
Bandwidth (days)	117	113	109	107	104

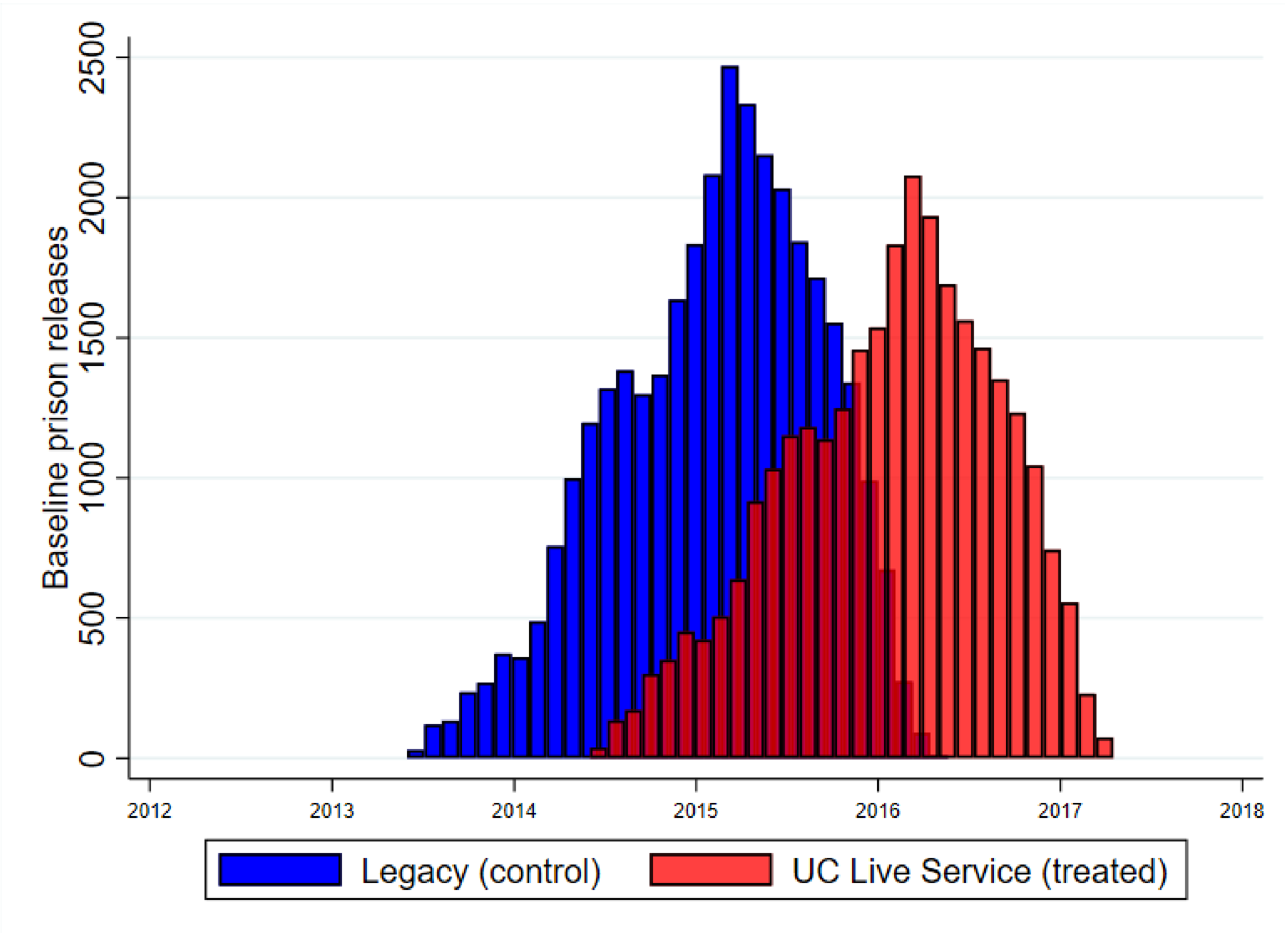
Sex		Yes	Yes	Yes	Yes
Age at release		Yes	Yes	Yes	Yes
British			Yes	Yes	Yes
Ethnicity			Yes	Yes	Yes
Previous court app.				Yes	Yes
Ex-prisoner					Yes

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.10$. Heteroscedasticity-robust standard errors adjusted for clustering at the local authority level are reported in parentheses. A polynomial of order 1 is used for estimation. Bandwidth size is chosen by the Calonico et al. (2014) selection method.

This table shows the estimated effect on reoffending within 5 weeks of prison release from exposure to the UC Live Service. Panel A shows results for recidivism (i.e. any form of crime); panel B shows results for acquisitive crime and panel C shows results for violent crime. Controls for each model are shown at the foot of the table.

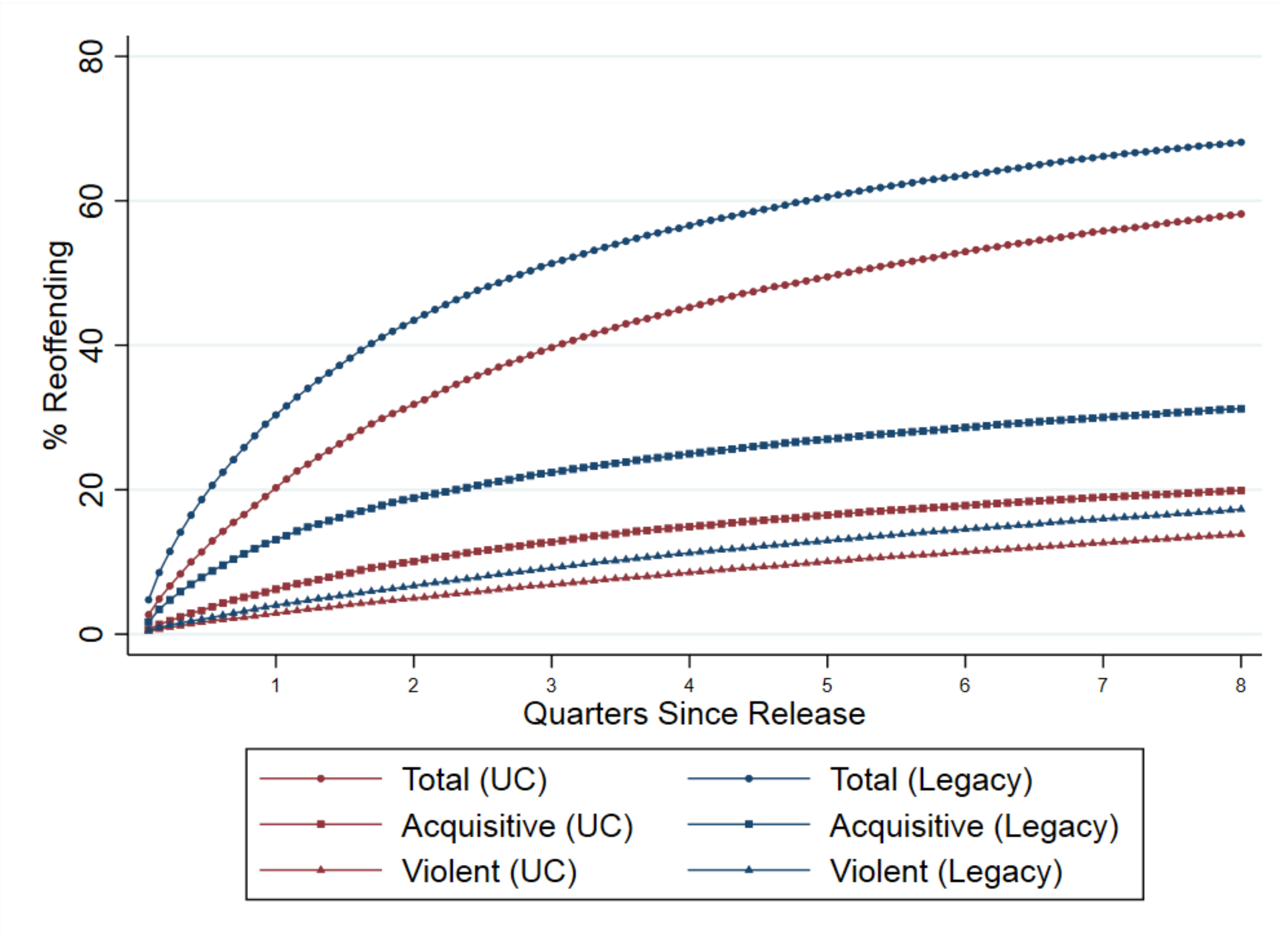
Column (1) is the baseline model with no controls. Column (2) controls for prisoners' sex and age at release and Column (3) controls for whether the prisoner is of British nationality and whether they are White, Black or Asian (mixed race is the excluded benchmark). Column (4) controls for the number of previous court appearances within the 2 years preceding the baseline prison spell. Column (5) includes a dummy variable for whether the prisoner had been to prison in the 2 years prior to the baseline prison spell.

Figure 5.1: Baseline Prison Releases



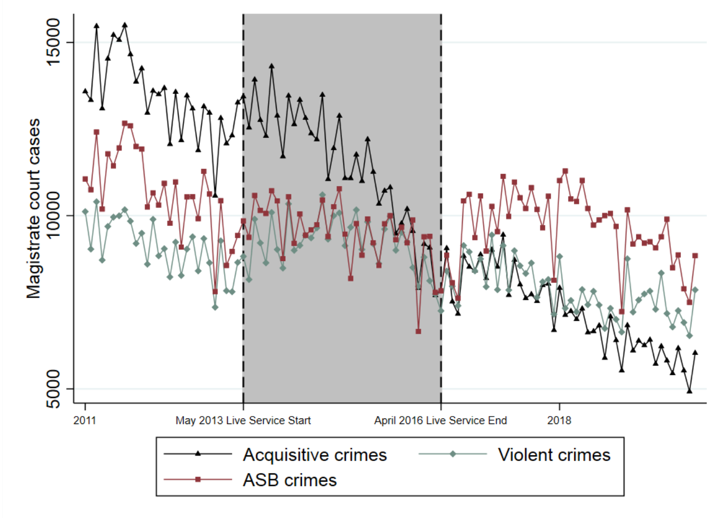
Notes: This figure shows the 55,089 "baseline" prison releases I measure reoffending against. Prison leavers released into a local authority where the legacy benefit system was operating at the time of release are shown in blue. The equivalent for the UC Live Service is shown in red. A baseline release refers to a prisoner's first release within a 365 day window either side of the date when the UC Live Service was implemented in their local authority. The distributions overlap because the UC Live Service was rolled out gradually across local authorities (over 3 years).

Figure 5.2: Reoffending within Analytic Sample

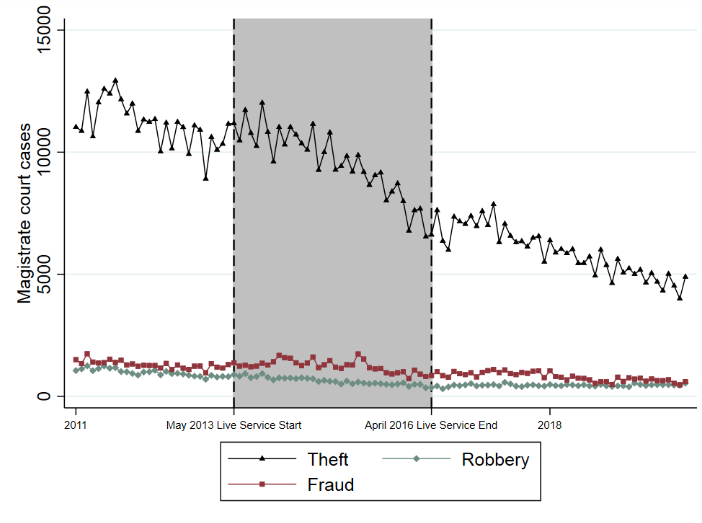


Notes: This figure shows the reoffending rates for prisoners within my analytic sample - that is, reoffending rates following prisoners' "baseline" prison release. The blue lines show the percentage of prison leavers who reside in Legacy local authorities and the red lines show that for Live Service local authorities. "Total" refers to any offence committed, "Acquisitive" refers to a financially motivated crime (robbery, theft or fraud) and "Violent" refers to any violent offence against another person.

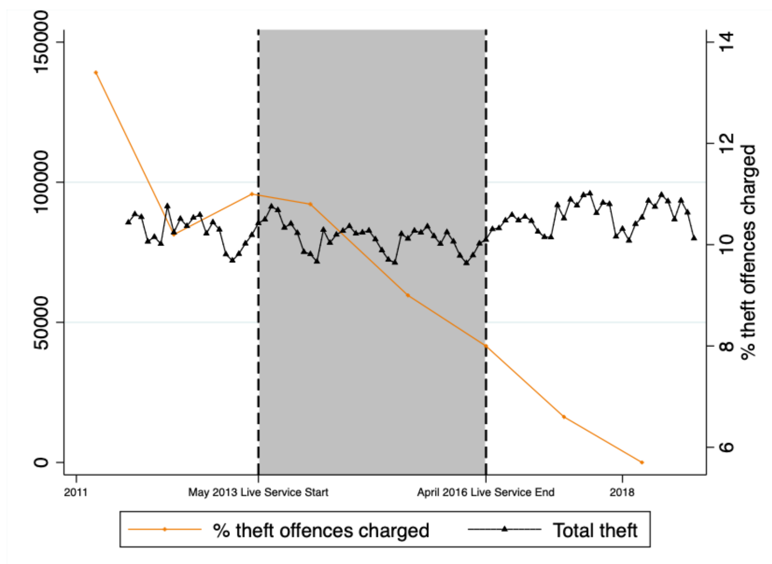
Figure 5.3: Court Appearances



(a) Aggregated crime types



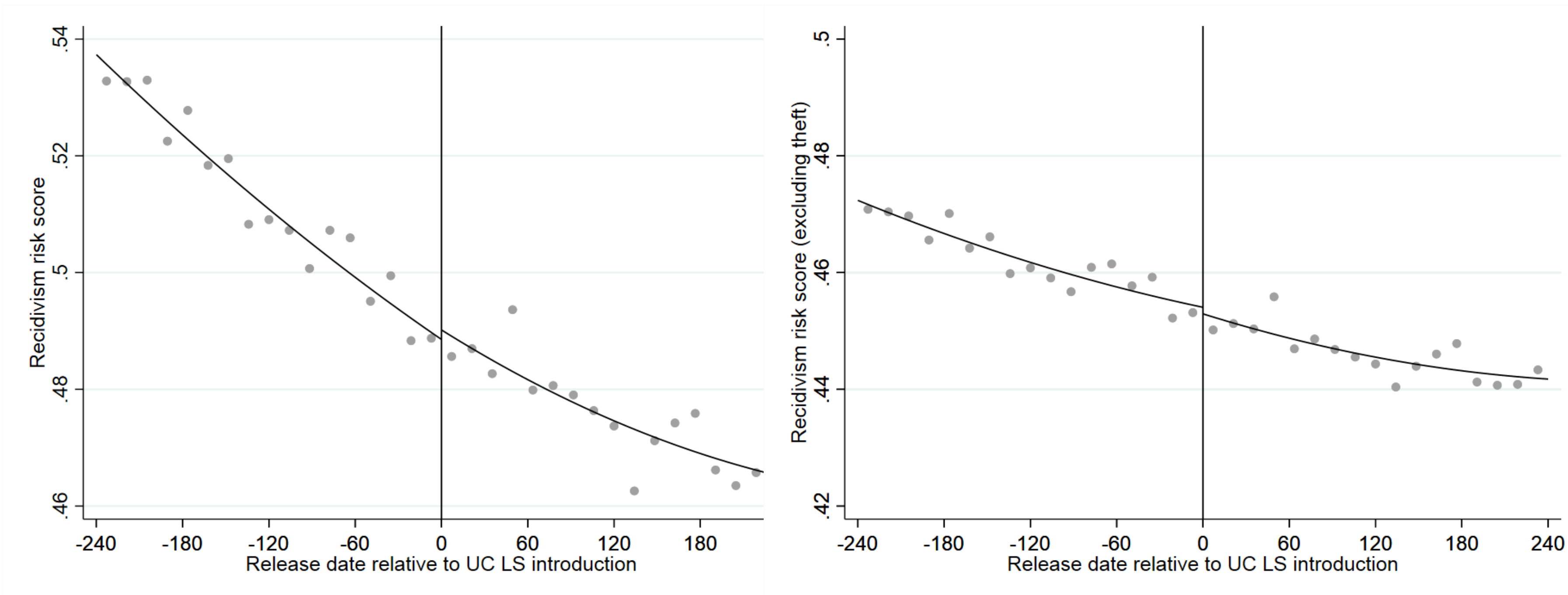
(b) Acquisitive crime breakdown



(c) Police recorded theft

Notes: The top left figure shows the number of cases being held in magistrates courts for acquisitive types of crime, antisocial behaviour (ASB) and violent crimes. It shows a sharp decline in court cases for acquisitive crime during the period of the Live Service rollout. The top right figure disaggregates acquisitive types of crime into its Home Office categories: theft, robbery and fraud. It shows that it is the decline in theft cases that is driving the reduction in acquisitive crime cases overall. The bottom figure shows police recorded theft and Home Office prosecution rates during the period of my sample. It shows that the number of thefts has been fairly stable over the course of my sample, however there has been a decline prosecution rates.

Figure 5.4: Recidivism Risk

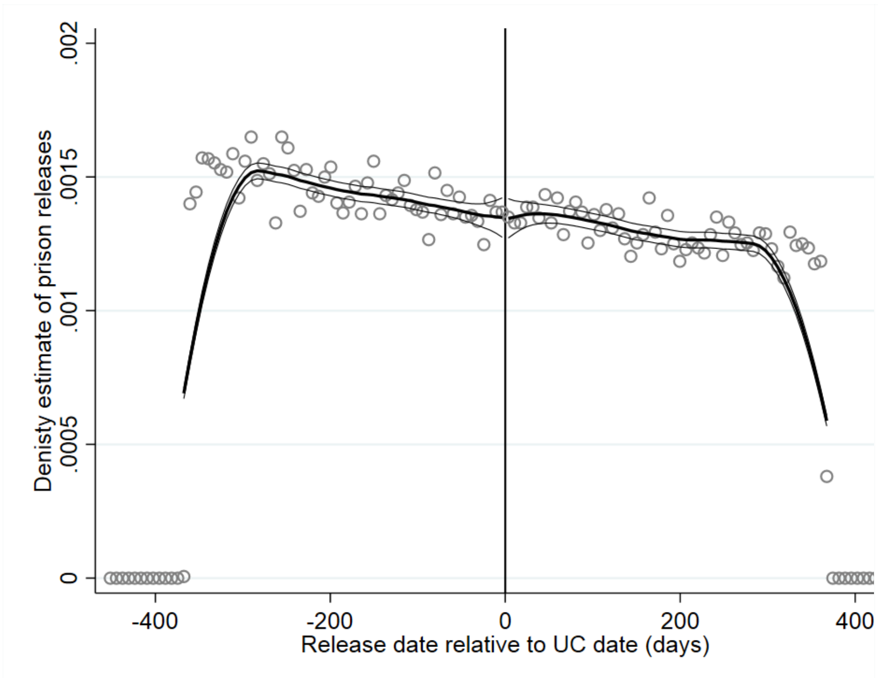


(a) Total crime

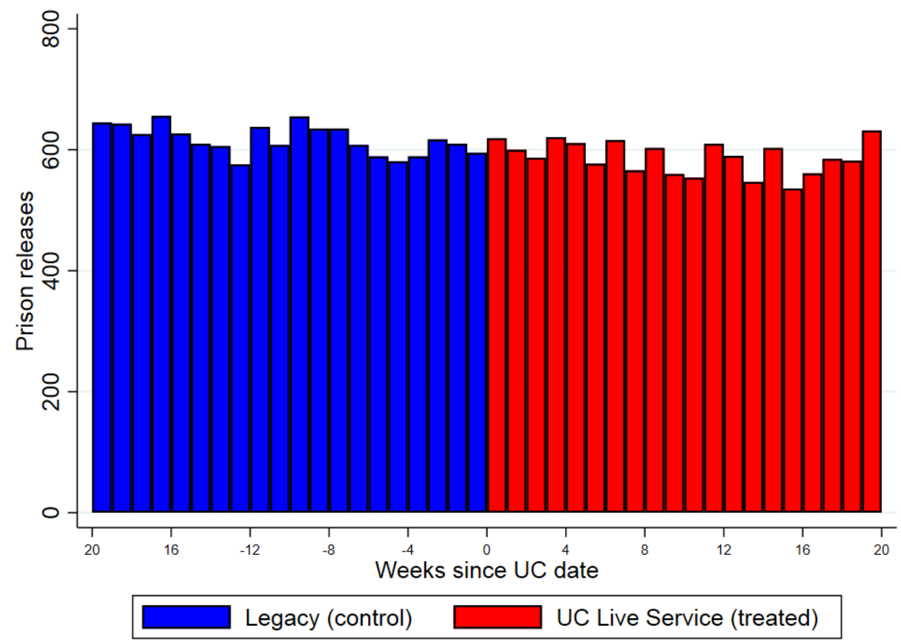
(b) Total crime excluding theft

Notes: The left figure plots average predicted recidivism risk against the running variable after binning prison releases into two week intervals. Each dot contains approximately 1,000 observations on average and its position, relative to the y axis, represents the proportion of that prison cohort who reoffended within 12 months of release. The right figure shows the same graph having removed theft offences.

Figure 5.5: Check for Manipulation around Cut Off



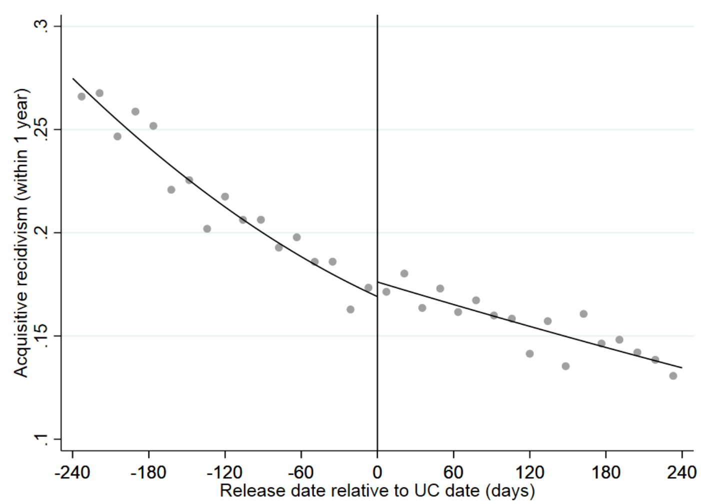
(a) McCrary density test



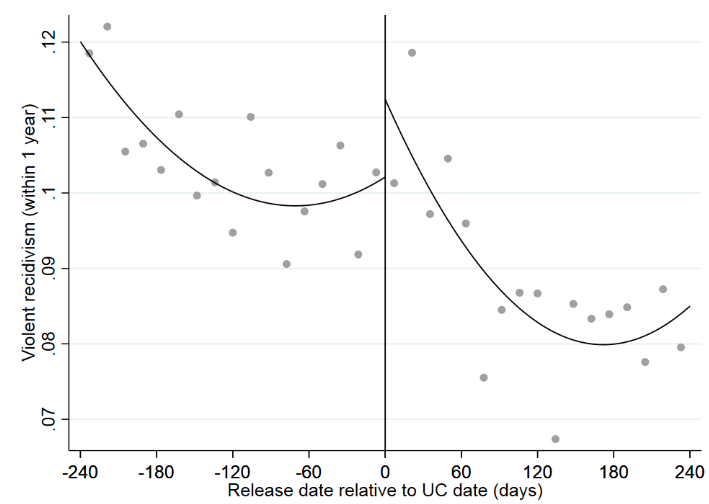
(b) Frequency of releases

Notes: These two figures check for whether prisoners attempt to sort into either welfare regime by manipulating their release dates (for example, by committing a further transgression in prison). The figure on the left provides a visual of the McCrary density test. It shows that there is no break in density around the cut off. The figure on the right shows the number of prison releases each week before and after the cut off. Again, there appears to be no sorting.

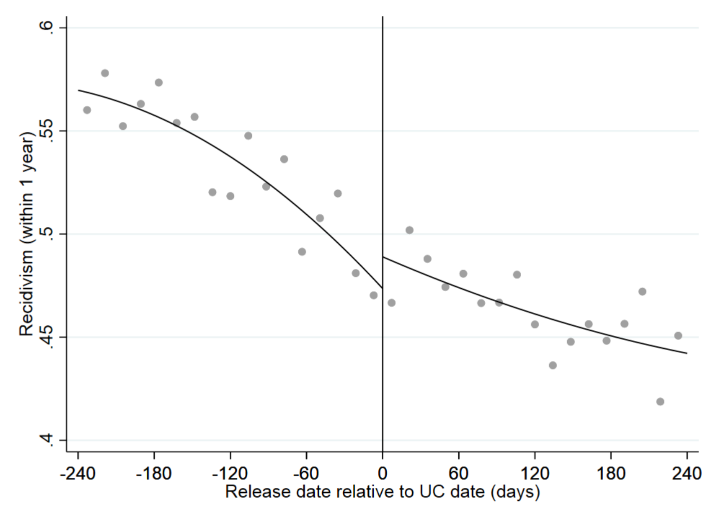
Figure 5.6: RD Plots



(a) Acquisitive crime



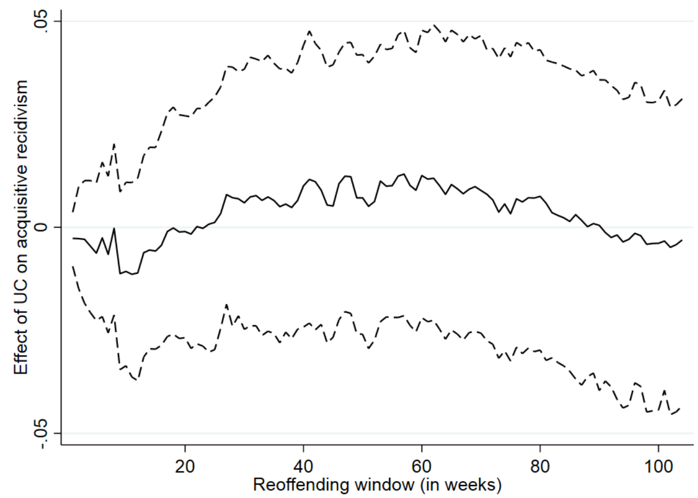
(b) Violent crime



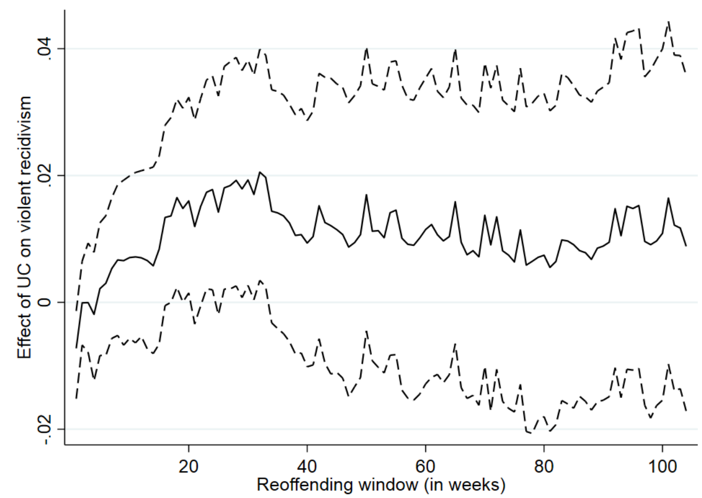
(c) Total crime

Notes: The top left figure plots the average proportion of prison leavers who commit an acquisitive offence within 12 months of release against the running variable. Each dot represents the average outcome of approximately 1,000 prison leavers within a given two week interval. The figure on the top right and the figure at the bottom show the same plot but for violent and total (i.e. any) offences, respectively.

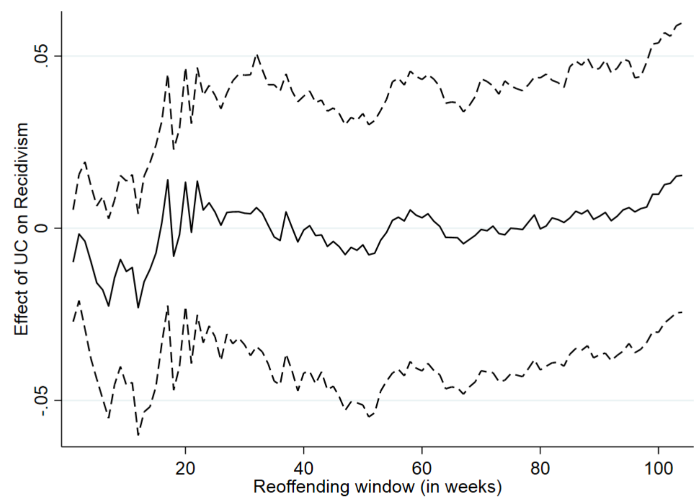
Figure 5.7: Effect of UC on Recidivism over a Period of 2 years since Release



(a) Acquisitive crime



(b) Violent crime



(c) Total crime

Notes: These figures show the estimated effect of the UC Live Service on different recidivism windows up to 2 years (104 weeks) post prison release. The black line in each figure is made up 104 separate RD treatment effect estimates, with each one incrementally increasing the recidivism period in question by 1 week. Dashed lines represent 95% confidence intervals. The figure on the top left shows results for acquisitive crime; the figure on the top right shows results for violent crime; and the figure at the bottom shows results for total crime. All figures show estimates for UC's impact on recidivism at the extensive margin.

Chapter 6

Conclusion

Universal Credit (UC) is a monumental reform to the UK's social security system. It is replacing six major working-age benefits into a single means-test payment for individuals out of work or in employment on low wages. Beyond simplification, UC recasts the central tenet of the benefit system to one that prioritises the transition into employment. In practice, this has meant reducing benefit expenditure by £2 billion per year whilst simultaneously attempting to ensure work is always rewarded over inactivity. UC also aims to achieve through its introduction of a number of sub-polices. These include moving to a monthly payments schedule; introducing a more punitive sanction regime, and pairing it with more stringent work-related conditionality; paying housing benefit to tenants rather than landlords; and moving to a digital system.

This thesis aimed to improve our understanding of UC by providing quantitative evidence on its impacts on mental health, local crime rates, and reoffending among prisoners. I explored each of these questions within a quasi-experimental framework, which enabled my analysis to provide credibly causal estimates.

Chapter 2

In Chapter 2, I provided an overview of UC: I described its purpose, the changes to the welfare state it brings, and the manner in which it has been implemented across the UK.

Chapter 3

In Chapter 3, I studied UC's impact on mental health in England using administrative data from NHS Digital and Public Health England. My analysis expanded on previous research in three main ways. First, I explored a broader range of mental health outcomes. Second, my estimates rely on weaker identifying assumptions, specifically, those pertaining to treatment effect homogeneity and UC's labour market impact. Third, I focus on objectively measurable health outcomes, which in turn enabled my analysis to capture plausibly distinct variation in mild, moderate and severe levels of clinical depression caused by the programme. My empirical strategy was to exploit the UC's Full Service rollout across English local authorities between June 2015 and December 2018 within a difference-in-difference framework.

I found no robust evidence of UC leading to an increase in mild, moderate or severe levels of clinical depression in communities following its Full Service rollout. Specifically, my estimates suggest the programme had no discernible impact on the GP prescribing of mental health related treatments, mental heath related hospitalisations, referrals to the main National Health Service (NHS) talking therapy programme, or suicidality. One possible exception pertained to the number of open referrals to NHS second care mental health services; estimates for this analysis found suggestive evidence of UC leading to a 2.4% increase. However, caution is needed when interpreting this finding.

Chapter 4

In Chapter 4, I documented the impact of UC on local authority level crime rates in England and Wales using monthly police recorded crime data from UKCrimeStats. I improved on previous UC-crime studies by employing new difference-in-differences estimators that are robust to treatment effect heterogeneity, a particularly important assumption to relax in this setting given that, by design, UC's caseload accumulates once the programme is launched. Additionally, my main analysis used an estimator that is uniquely tailored to handle settings with multiple staggered treatments, such as UC's Live Service and Full Service. This further improves upon the credibility of my estimates and, separately, enabled my analysis to provide the first criminological estimates of the Full Service - the version of UC in operation to date.

Contradicting the existing literature, I found no robust evidence of the Live Service or Full Service leading to an increase in crime. While some estimates within my analysis linked the Live Service to a rise in acquisitive crime, they were not robust to subsequent checks.

Chapter 5

In Chapter 5, I estimated UC's intent-to-treat effect on prison leavers propensity to reoffend. To the best of my knowledge, this represents the first welfare reforms estimates pertaining to recidivism among ex-prisoners outside of a U.S. context. To explore these effects I leveraged a novel dataset covering the universe of court cases, prison sentences and probation periods within England and Wales between 1st January 2011 and 31st December 2020. These data contained a unique offender ID that enabled me to track ex-offenders through the criminal justice system and observe their reoffending outcomes post release. I disentangled UC's effect from potential confounders by exploiting the fact that small, plausibly exogenous, differences in prisoners' release dates saw them exposed to UC or not, depending on when their local authority of residence implemented the programme. This gave rise to a sharp regression discontinuity design.

I found no evidence to suggest that UC affected prisoners' recidivism rates upon release from incarceration. This was regardless of both offence type or recidivism window in question (up to 2 years post release). Of particular note was the lack of effect within the first several weeks of release, as this indicated that UC's 5 week until initial payment did not have criminological consequences. That said, estimates from this chapter were imprecise, and thus only able to rule out large recidivism effects from the programme. I note that access to benefit data linked to these criminal justice data would enable future research to improve upon this analysis.

Bibliography

Agan, A. and Starr, S. (2017), 'The effect of criminal records on access to employment', *American Economic Review* **107**(5), 560–564.

Agnew, R. (1992), 'Foundation for a general strain theory of crime and delinquency', *Criminology* **30**(1), 47–88.

Aizer, A. and Doyle Jr, J. J. (2015), 'Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges', *The Quarterly Journal of Economics* **130**(2), 759–803.

Alonso, J., Petukhova, M., Vilagut, G., Chatterji, S., Heeringa, S., Üstün, T. B., Alhamzawi, A. O., Viana, M., Angermeyer, M., Bromet, E. et al. (2011), 'Days out of role due to common physical and mental conditions: results from the who world mental health surveys', *Molecular psychiatry* **16**(12), 1234–1246.

Andersen, K. (2020), 'Universal credit, gender and unpaid childcare: Mothers' accounts of the new welfare conditionality regime', *Critical Social Policy* **40**(3), 430–449.

Andersen, L. H., Dustmann, C., Landersø, R. K. et al. (2019), *Lowering welfare benefits: Intended and unintended consequences for migrants and their families*, Centre for Research and Analysis of Migration, Department of Economics . . . .

Angrist, J. D. and Pischke, J.-S. (2009), *Mostly harmless econometrics: An empiricist's companion*, Princeton university press.

Arie, S. (2018), 'Doctors' concerns over universal credit are mounting', *BMJ: British Medical Journal (Online)* **363**.

Barr, B., Kinderman, P. and Whitehead, M. (2015), 'Trends in mental health inequalities in england during a period of recession, austerity and welfare reform 2004 to 2013', *Social Science & Medicine* **147**, 324–331.

Basu, S., Rehkopf, D. H., Siddiqi, A., Glymour, M. M. and Kawachi, I. (2016), 'Health behaviors, mental health, and health care utilization among single mothers after welfare reforms in the 1990s', *American journal of epidemiology* **183**(6), 531–538.

Becker, G. S. (1968), 'Crime and punishment: An economic approach', *Journal of Political Economy* **76**(2), 169–217.

Bennett, P. and Ouazad, A. (2020), 'Job displacement, unemployment, and crime: Evidence from danish microdata and reforms', *Journal of the European Economic Association* **18**(5), 2182–2220.

Billings, S. B. and Schnepel, K. T. (2022), 'Hanging out with the usual suspects neighborhood peer effects and recidivism', *Journal of Human Resources* **57**(5), 1758–1788.

Bitler, M. P., Gelbach, J. B. and Hoynes, H. W. (2005), 'Welfare reform and health', *Journal of Human resources* **40**(2), 309–334.

Björklund, A. (1985), 'Unemployment and mental health: some evidence from panel data', *Journal of Human Resources* pp. 469–483.

Blank, R. M. (2002), 'Evaluating welfare reform in the united states', *Journal of Economic Literature* **40**(4), 1105–1166.

Borjas, G. J. (2003), 'Welfare reform, labor supply, and health insurance in the immigrant population', *Journal of Health Economics* **22**(6), 933–958.

Borusyak, K., Jaravel, X. and Spiess, J. (2022), 'Revisiting event study designs: Robust and efficient estimation', *Available at SSRN 2826228* .

Brewer, M., Dang, T. and Tominey, E. (2022), 'Universal credit: Welfare reform and mental health'.

Brewer, M., Joyce, R., Waters, T. and Woods, J. (2019), 'Universal credit and its impact on household incomes: the long and the short of it', *Institute for Fiscal Studies* .

Britto, D. G., Pinotti, P. and Sampaio, B. (2022), 'The effect of job loss and unemployment insurance on crime in brazil', *Econometrica* **90**(4), 1393–1423.

Bubonya, M., Cobb-Clark, D. A. and Wooden, M. (2017), 'Mental health and productivity at work: Does what you do matter?', *Labour economics* **46**, 150–165.

Bushway, S. D. and Sweeten, G. (2007), 'Abolish lifetime bans for ex-felons', *Criminology & Pub. Pol'y* **6**, 697.

Butcher, K. F. and LaLonde, R. (2006), 'Female offenders' use of social welfare programs before and after jail and prison: Does prison cause welfare dependency?'.

Caliendo, M. (2019), 'Health effects of labor market policies: Evidence from drug prescriptions'.

Callaway, B. and Sant'Anna, P. H. (2021), 'Difference-in-differences with multiple time periods', *Journal of Econometrics* **225**(2), 200–230.

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

Card, D., Chetty, R. and Weber, A. (2007), 'The spike at benefit exhaustion: Leaving the unemployment system or starting a new job?', *American Economic Review* **97**(2), 113–118.

Card, D., Johnston, A., Leung, P., Mas, A. and Pei, Z. (2015), 'The effect of unemployment benefits on the duration of unemployment insurance receipt: New evidence from a regression kink design in missouri, 2003–2013', *American Economic Review* **105**(5), 126–130.

Carr, J. B. and Packham, A. (2019), 'Snap benefits and crime: Evidence from changing disbursement schedules', *Review of Economics and Statistics* **101**(2), 310–325.

Cattaneo, M. D., Idrobo, N. and Titiunik, R. (2019), *A practical introduction to regression discontinuity designs: Foundations*, Cambridge University Press.

Centre for Social Justice (2009), 'Dynamic benefits: towards welfare that works', *London: Centre for Social Justice* .

Cheetham, M., Moffatt, S., Addison, M. and Wiseman, A. (2019), 'Impact of universal credit in north east england: a qualitative study of claimants and support staff', *BMJ open* **9**(7), e029611.

Chioda, L., De Mello, J. M. and Soares, R. R. (2016), 'Spillovers from conditional cash transfer programs: Bolsa família and crime in urban brazil', *Economics of Education Review* **54**, 306–320.

Citizens Advice (2017), 'Universal credit and debt', `https://www.citizensadvice.org.uk/Global/CitizensAdvice/welfare%20publications/Universal%20Credit%20and%20Debt%20-%20final.pdf`. Accessed: 2023-02-15.

Citizens Advice (2018), 'Making a universal credit claim', `https://www.citizensadvice.org.uk/Global/CitizensAdvice/welfare%20publications/Making%20a%20Universal%20Credit%20claim%202018%20-%20final.pdf`. Accessed: 2023-02-15.

Cook, P. J., Kang, S., Braga, A. A., Ludwig, J. and O'Brien, M. E. (2015), 'An experimental evaluation of a comprehensive employment-oriented prisoner re-entry program', *Journal of Quantitative Criminology* **31**, 355–382.

Corman, H., Dave, D. M., Das, D. and Reichman, N. E. (2013), 'Effects of welfare reform on illicit drug use of adult women', *Economic inquiry* **51**(1), 653–674.

Cornaglia, F., Crivellaro, E. and McNally, S. (2015), 'Mental health and education decisions', *Labour Economics* **33**, 1–12.

Corno, L. (2017), 'Homelessness and crime: do your friends matter?', *The Economic Journal* **127**(602), 959–995.

Currie, J. and Grogger, J. (2002), 'Medicaid expansions and welfare contractions: offsetting effects on prenatal care and infant health?', *Journal of health economics* **21**(2), 313–335.

Czabanski, J. (2008), *Estimates of cost of crime: history, methodologies, and implications*, Springer Science & Business Media.

Damm, A. P. and Dustmann, C. (2014), 'Does growing up in a high crime neighborhood affect youth criminal behavior?', *American Economic Review* **104**(6), 1806–1832.

Danziger, Mary Corcoran, S. and Heflin, C. (2000), 'Barriers to the employment of welfare recipients', *Prosperity for all?: The economic boom and African Americans* **245**.

Danziger, S. K., Kalil, A. and Anderson, N. J. (2000), 'Human capital, physical health, and mental health of welfare recipients: Co-occurrence and correlates', *Journal of Social Issues* **56**(4), 635–654.

Davis, O. (2019), 'What is the relationship between benefit conditionality and mental health? evidence from the united states on tanf policies', *Journal of Social Policy* **48**(2), 249–269.

de Chaisemartin, C. and d'Haultfœuille, X. (2020), 'Two-way fixed effects regressions with several treatments', *arXiv preprint arXiv:2012.10077* .

De Chaisemartin, C. and d'Haultfoeuille, X. (2022), Difference-in-differences estimators of intertemporal treatment effects, Technical report, National Bureau of Economic Research.

De Chaisemartin, C. and d'Haultfoeuille, X. (2020), 'Two-way fixed effects estimators with heterogeneous treatment effects', *American Economic Review* **110**(9), 2964–2996.

DeLeire, T., Levine, J. A. and Levy, H. (2003), 'Welfare reform and the uninsured', *Typescript, University of Chicago* .

DeLeire, T., Levine, J. A. and Levy, H. (2006), 'Is welfare reform responsible for low-skilled women's declining health insurance coverage in the 1990s?', *Journal of Human resources* **41**(3), 495–528.

Dell, M., Feigenberg, B. and Teshima, K. (2019), 'The violent consequences of trade-induced worker displacement in mexico', *American Economic Review: Insights* **1**(1), 43–58.

Department for Work and Pensions (2010), 'Linking data on offenders with benefit, employment and income data', https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/48897/universal-credit-full-document.pdf. Accessed: 2022-11-10.

Department for Work and Pensions (2011), 'Iain duncan smith sets out next steps for moving claimants onto universal credit', https://www.gov.uk/government/news/iain-duncan-smith-sets-out-next-steps-for-moving-claimants-onto-universal-credit. Accessed: 2023-02-18.

Department for Work and Pensions (2015), 'Universal credit: 29 apr 2013 to 3 dec 2015', https://www.gov.uk/government/statistics/universal-credit-29-apr-2013-to-3-dec-2015. Accessed: 2022-09-17.

Department for Work and Pensions (2018*a*), 'Understanding the impact of universal credit on the labour market', https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/714858/understanding-the-impact-of-universal-credit-on-the-labour-market.pdf. Accessed: 2022-11-10.

Department for Work and Pensions (2018*b*), 'Universal credit full service claimant survey', https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/714842/universal-credit-full-service-claimant-survey.pdf. Accessed: 2022-11-10.

Department for Work and Pensions (2018*c*), 'Universal credit roll out schedule', https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/693928/universal-credit-transition-rollout-schedule.pdf. Accessed: 2022-09-17.

Department for Work and Pensions (2021), 'Official statistics benefit cap: number of households capped to february 2021', ://www.gov.uk/government/statistics/benefit-cap-number-of-households-capped-to-february-2021/benefit-cap-number-of-households-capped-to-february-2021: :text=When%20the%20benefit%20cap%20was,single%20adults%20with%20no%20children.

Department for Work and Pensions (2023), 'Supporting prison leavers: a guide to universal credit', `https://www.gov.uk/government/publications/universal-credit-and-prison-leavers/supporting-prison-leavers-a-guide-to-universal-credit`. Accessed: 2022-11-10.

Department of Health (2014), 'International comparisons of health and wellbeing', `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/277595/International_comparisons.pdf`. Accessed: 2023-02-18.

Deshpande, M. and Mueller-Smith, M. (2022), 'Does welfare prevent crime? the criminal justice outcomes of youth removed from ssi', *The Quarterly Journal of Economics* **137**(4), 2263–2307.

Dix-Carneiro, R., Soares, R. R. and Ulyssea, G. (2018), 'Economic shocks and crime: Evidence from the brazilian trade liberalization', *American Economic Journal: Applied Economics* **10**(4), 158–195.

Dobkin, C. and Puller, S. L. (2007), 'The effects of government transfers on monthly cycles in drug abuse, hospitalization and mortality', *Journal of Public Economics* **91**(11-12), 2137–2157.

Doleac, J. L. (2017), 'The effects of dna databases on crime', *American Economic Journal: Applied Economics* **9**(1), 165–201.

Draca, M. and Machin, S. (2015), 'Crime and economic incentives', *Annual Review of Economics* **7**(1), 389–408.

Dwyer, P. (2018), 'Final findings: overview', *Welfare conditionality: sanctions, support and behaviour change. Economic and Social Research Council. Available at: www. welfare conditionality. ac. uk/wp-content/uploads/2018/05/40414_Overview-HR4. pdf (accessed 20 June 2019)* .

d'Este, R. and Harvey, A. (2022), 'The unintended consequences of welfare reforms: Universal credit, financial insecurity, and crime', *The Journal of Law, Economics, and Organization* .

Ehrlich, I. (1973), 'Participation in illegitimate activities: A theoretical and empirical investigation', *Journal of Political Economy* **81**(3), 521–565.

Ettner, S. L., Frank, R. G. and Kessler, R. C. (1997), 'The impact of psychiatric disorders on labor market outcomes', *ILR Review* **51**(1), 64–81.

Fallesen, P., Geerdsen, L. P., Imai, S. and Tranæs, T. (2018), 'The effect of active labor market policies on crime: Incapacitation and program effects', *Labour Economics* **52**, 263–286.

Fan, J. and Gijbels, I. (1992), 'Variable bandwidth and local linear regression smoothers', *The Annals of Statistics* pp. 2008–2036.

Fishback, P. V., Johnson, R. S. and Kantor, S. (2010), 'Striking at the roots of crime: The impact of welfare spending on crime during the great depression', *The Journal of Law and Economics* **53**(4), 715–740.

Foley, B. (2017), 'Delivering on universal credit', *Citizens Advice.[Online]. Accessed on* **23**, 18.

Foley, C. F. (2011), 'Welfare payments and crime', *The Review of Economics and Statistics* **93**(1), 97–112.

Foster, R., Adams, L., Svabaes, S., O'Driscoll, C. and Thomson, D. (2018), 'Universal credit full service survey'.

Fougère, D., Kramarz, F. and Pouget, J. (2009), 'Youth unemployment and crime in france', *Journal of the European Economic Association* **7**(5), 909–938.

Gelman, A. and Imbens, G. (2019), 'Why high-order polynomials should not be used in regression discontinuity designs', *Journal of Business & Economic Statistics* **37**(3), 447–456.

Goodman-Bacon, A. (2021), 'Difference-in-differences with variation in treatment timing', *Journal of Econometrics* **225**(2), 254–277.

Gould, E. D., Weinberg, B. A. and Mustard, D. B. (2002), 'Crime rates and local labor market opportunities in the united states: 1979–1997', *Review of Economics and Statistics* **84**(1), 45–61.

Griffiths, R., Wood, M., Bennett, F. and Millar, J. (2020), 'Uncharted territory: Universal credit, couples and money'.

Grogger, J., Grogger, J. T., Karoly, L. A. et al. (2005), *Welfare reform: Effects of a decade of change*, Harvard University Press.

Gyani, A., Pumphrey, N., Parker, H., Shafran, R. and Rose, S. (2012), 'Investigating the use of nice guidelines and iapt services in the treatment of depression', *Mental health in family medicine* **9**(3), 149.

Haider, S. J., Jacknowitz, A. and Schoeni, R. F. (2003), 'Welfare work requirements and child well-being: Evidence from the effects on breast-feeding', *Demography* **40**(3), 479–497.

Hardie, I. (2021), 'The impact of universal credit rollout on housing security: an analysis of landlord repossession rates in english local authorities', *Journal of Social Policy* **50**(2), 225–246.

Hartfree, Y. (2014), 'Universal credit: the impact of monthly payments on low income households', *Journal of Poverty and Social Justice* **22**(1), 15–26.

Herbst, C. M. (2013), 'Welfare reform and the subjective well-being of single mothers', *Journal of Population Economics* **26**, 203–238.

Hirsch, A. E., Dietrich, S. M., Landau, R., Schneider, P. D., Ackelsberg, I., Bernstein-Baker, J. and Hohenstein, J. (2002), 'Every door closed: Barriers facing parents with criminal records.'.

HMRC (2022), 'Child and working tax credits statistics: Provisional awards', `https://www.gov.uk/government/statistics/child-and-working-tax-credits-statistics-provisional-awards-april-202` Accessed: 2023-04-15.

Holtfreter, K., Reisig, M. D. and Morash, M. (2004), 'Poverty, state capital, and recidivism among women offenders', *Criminology & Public Policy* **3**(2), 185–208.

Holzer, H. J., Raphael, S. and Stoll, M. A. (2003), 'Employment barriers facing ex-offenders', *Urban Institute Reentry Roundtable* pp. 1–23.

Holzer, H. J., Raphael, S. and Stoll, M. A. (2006), 'Perceived criminality, criminal background checks, and the racial hiring practices of employers', *The Journal of Law and Economics* **49**(2), 451–480.

Holzer, H. J., Raphael, S. and Stoll, M. A. (2007), 'The effect of an applicant's criminal history on employer hiring decisions and screening practices: Evidence from los angeles', *Barriers to reentry* **4**(15), 117–150.

Home Office (2012), 'Crimes detected in england and wales 2011/12'.
URL: *https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment{_}data/file/116435/hosb0812.pdf*

Home Office (2018), 'Crime outcomes in england and wales: year ending march 2018'.
URL: *https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/729127/crime-outcomes-hosb1018.pdf*

House of Commons Library (2016), 'Population estimates gp registers: why the difference?', https://commonslibrary.parliament.uk/population-estimates-gp-registers-why-the-difference/. Accessed: 2023-02-15.

House of Commons Library (2018), 'Mental health statistics for england: prevalence, services and funding', http://allcatsrgrey.org.uk/wp/download/public_health/mental_health/SN06988-1.pdf. Accessed: 2023-01-15.

House of Lords (2020), 'The economics of universal credit: call for evidence', https://publications.parliament.uk/pa/ld5801/ldselect/downloads/Call_for_evidence-the_economics_of_Universal_Credit.pdf. Accessed: 2022-11-10.

Immervoll, H. and Knotz, C. (2018), 'How demanding are activation requirements for jobseekers?'.

Jacob, B. A. and Lefgren, L. (2003), 'Are idle hands the devil's workshop? incapacitation, concentration, and juvenile crime', *American Economic Review* **93**(5), 1560–1577.

Jayakody, R., Danziger, S. and Pollack, H. (2000), 'Welfare reform, substance use, and mental health', *Journal of Health Politics, Policy and Law* **25**(4), 623–652.

Johnsen, S. and Blenkinsopp, J. (2018), 'Final findings: Lone parents', *Welfare Conditionality: Sanctions, Support and Behaviour Change* .

Joseph Rowntree Foundation (2018), 'Uk poverty', `https://www.jrf.org.uk/report/uk-poverty-2018`. Accessed: 2023-01-15.

Kaestner, R. and Kaushal, N. (2003), 'Welfare reform and health insurance coverage of low-income families', *Journal of health economics* **22**(6), 959–981.

Kaestner, R. and Tarlov, E. (2006), 'Changes in the welfare caseload and the health of low-educated mothers', *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management* **25**(3), 623–643.

Kaplan, G. A., Siefert, K., Ranjit, N., Raghunathan, T. E., Young, E. A., Tran, D., Danziger, S., Hudson, S., Lynch, J. W. and Tolman, R. (2005), 'The health of poor women under welfare reform', *American Journal of Public Health* **95**(7), 1252–1258.

Katikireddi, S. V., Molaodi, O. R., Gibson, M., Dundas, R. and Craig, P. (2018), 'Effects of restrictions to income support on health of lone mothers in the uk: a natural experiment study', *The Lancet Public Health* **3**(7), e333–e340.

Katz, L. F. and Meyer, B. D. (1990), 'Unemployment insurance, recall expectations, and unemployment outcomes', *The Quarterly Journal of Economics* **105**(4), 973–1002.

Khanna, G., Medina, C., Nyshadham, A., Posso, C. and Tamayo, J. (2021), 'Job loss, credit, and crime in colombia', *American Economic Review: Insights* **3**(1), 97–114.

Knab, J., Garfinkel, I. and McLanahan, S. (2006), 'The effects of welfare and child support policies on maternal health (working paper# 2006–04-ff)', *Center for Research on Child Wellbeing. Princeton University* .

Knab, J., Garfinkel, I. and McLanahan, S. (2008), 'The effects of welfare and child support policies on maternal health and wellbeing', *Making Americans healthier: Social and economic policy as health policy* pp. 281–305.

Lee, D. S. and Card, D. (2008), 'Regression discontinuity inference with specification error', *Journal of Econometrics* **142**(2), 655–674.

Lim, D., Sanderson, K. and Andrews, G. (2000), 'Lost productivity among full-time workers with mental disorders', *The journal of mental health policy and economics* **3**(3), 139–146.

Lim, K.-L., Jacobs, P., Ohinmaa, A., Schopflocher, D. and Dewa, C. S. (2008), 'A new population-based measure of the economic burden of mental illness in canada', *Chronic Dis Can* **28**(3), 92–98.

Lim, K. Y., Pickering, R. et al. (2020), 'Crime heterogeneity and welfare spending: theory and empirical evidence based on the universal credit system', *Nottingham Business School Discussion Paper* **4**.

Luallen, J., Edgerton, J. and Rabideau, D. (2018), 'A quasi-experimental evaluation of the impact of public assistance on prisoner recidivism', *Journal of Quantitative Criminology* **34**, 741–773.

Machin, S. and Marie, O. (2006), 'Crime and benefit sanctions', *Portuguese Economic Journal* **5**, 149–165.

Mallar, C. D. and Thornton, C. V. (1978), 'Transitional aid for released prisoners: Evidence from the life experiment', *Journal of Human Resources* pp. 208–236.

Mathers, C. D. and Loncar, D. (2006), 'Projections of global mortality and burden of disease from 2002 to 2030', *PLoS medicine* **3**(11), e442.

McCollister, K. E., French, M. T. and Fang, H. (2010), 'The cost of crime to society: New crime-specific estimates for policy and program evaluation', *Drug and alcohol dependence* **108**(1-2), 98–109.

McCrary, J. (2008), 'Manipulation of the running variable in the regression discontinuity design: A density test', *Journal of Econometrics* **142**(2), 698–714.

McDaid, D., Park, A.-L., Davidson, G., John, A., Knifton, L., McDaid, S., Morton, A., Thorpe, L. and Wilson, N. (2022), 'The economic case for investing in the prevention of mental health conditions in the uk'.

McManus, S., Bebbington, P. E., Jenkins, R. and Brugha, T. (2016), *Mental health and wellbeing in England: the adult psychiatric morbidity survey 2014*, NHS digital.

Meara, E. and Frank, R. G. (2006), 'Welfare reform, work requirements, and employment barriers'.

Merton, R. K. (1938), 'Social structure and anomie', *American Sociological Review* **3**(5), 672–682.

Mind (2018), '40 per cent of all gp appointments about mental health', https://www.mind.org.uk/news-campaigns/news/40-per-cent-of-all-gp-appointments-about-mental-health/. Accessed: 2022-11-10.

Mind (2023), 'Health and social care rights'.
URL: *https://www.mind.org.uk/information-support/legal-rights/health-and-social-care-rights/about-healthcare/*

Ministry of Justice (2014), 'Linking data on offenders with benefit, employment and income data', https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/304411/experimental-statistics.pdf. Accessed: 2022-11-10.

Moffitt, R. (1985), 'Unemployment insurance and the distribution of unemployment spells', *Journal of Econometrics* **28**(1), 85–101.

Morris, P. A. (2008), 'Welfare program implementation and parents' depression', *Social Service Review* **82**(4), 579–614.

Munyo, I. and Rossi, M. A. (2015), 'First-day criminal recidivism', *Journal of Public Economics* **124**, 81–90.

Murphy, G. C. and Athanasou, J. A. (1999), 'The effect of unemployment on mental health', *Journal of Occupational and Organizational Psychology* **72**(1), 83–99.

NAO (2005), 'Dealing with the complexity of the benefits system', *Report by the Comptroller and Auditor General* .

National Audit Office (2013), 'Universal credit: early progress', https://www.nao.org.uk/wp-content/uploads/2013/09/10132-001-Universal-credit.pdf. Accessed: 2023-02-15.

National Audit Office (2018), 'Rollout universal credit', https://www.nao.org.uk/wp-content/uploads/2018/06/Rolling-out-Universal-Credit.pdf. Accessed: 2023-02-15.

National Audit Office (2020), 'Universal credit: getting to first payment', https://www.nao.org.uk/wp-content/uploads/2020/07/Universal-Credit-getting-to-first-payment.pdf. Accessed: 2023-02-18.

National Health Service (2015), 'Choice of gp practice guidance on the new out of area patient registration arrangements'.
URL: *https://www.england.nhs.uk/wp-content/uploads/2017/10/Guide-out-area-reg-1214.pdf*

National Health Service (2022), 'Information for secondary, community and mental health service staff', https://digital.nhs.uk/services/nhs-app/nhs-app-guidance-for-gp-practices/guidance-on-nhs-app-features/accelerating-patient-access-to-their-record/information-for-health-and-care-providers-working-in-secondary-care-community-care-and-ment
Accessed: 2023-02-18.

National Health Service (2023), 'How to register with a gp surgery', https://www.nhs.uk/nhs-services/gps/how-to-register-with-a-gp-surgery/. Accessed: 2023-01-15.

National Institutional for Health and Care Excellence (2017), 'Depression in adults: treatment and management', https://www.nice.org.uk/guidance/ng222/documents/guideline-appendix-22. Accessed: 2022-11-10.

National Institutional for Health and Care Excellence (2021), 'Scenario: Choice of analgesic', https://cks.nice.org.uk/topics/analgesia-mild-to-moderate-pain/management/choice-of-analgesic/. Accessed: 2022-11-10.

OECD (2021), 'Scaling up policies that connect people with jobs in the recovery from covid-19'.
URL: *https://www.oecd-ilibrary.org/content/paper/a91d2087-en*

Öster, A. and Agell, J. (2007), 'Crime and unemployment in turbulent times', *Journal of the European Economic Association* **5**(4), 752–775.

Pager, D. (2003), 'The mark of a criminal record', *American Journal of Sociology* **108**(5), 937–975.

Petersilia, J. (2003), *When prisoners come home: Parole and prisoner reentry*, Oxford University Press.

Porter, J. (2003), 'Estimation in the regression discontinuity model', *Unpublished Manuscript, Department of Economics, University of Wisconsin at Madison* **2003**, 5–19.

Public Accounts Committee (2013), 'Universal credit: early progress', https://publications.parliament.uk/pa/cm201314/cmselect/cmpubacc/619/130911.htm. Accessed: 2022-11-10.

Rabindrakumar, S. and Dewar, L. (2018), 'How benefit sanctions push single parents further from work'.

Raphael, S. and Winter-Ebmer, R. (2001), 'Identifying the effect of unemployment on crime', *The Journal of Law and Economics* **44**(1), 259–283.

Redcross, C., Millenky, M., Rudd, T. and Levshin, V. (2011), 'More than a job: Final results from the evaluation of the center for employment opportunities (ceo) transitional jobs program', *OPRE Report* **18**.

Reeves, A. and Loopstra, R. (2021), 'The continuing effects of welfare reform on food bank use in the uk: the roll-out of universal credit', *Journal of Social Policy* **50**(4), 788–808.

Rege, M., Skardhamar, T., Telle, K. and Votruba, M. (2019), 'Job displacement and crime: Evidence from norwegian register data', *Labour Economics* **61**, 101761.

Reichman, N. E., Teitler, J. O. and Curtis, M. A. (2005), 'Tanf sanctioning and hardship', *Social Service Review* **79**(2), 215–236.

Riddell, C. and Riddell, R. (2006), 'Welfare checks, drug consumption, and health evidence from vancouver injection drug users', *Journal of Human Resources* **41**(1), 138–161.

Rose, E. K. (2018), 'The effects of job loss on crime: evidence from administrative data', *Available at SSRN 2991317* .

Roth, J. and Sant'Anna, P. H. (2021), 'Efficient estimation for staggered rollout designs', *arXiv preprint arXiv:2102.01291* .

Roth, J., Sant'Anna, P. H., Bilinski, A. and Poe, J. (2022), 'What's trending in difference-in-differences? a synthesis of the recent econometrics literature', *arXiv preprint arXiv:2201.01194* .

Schmitt, J. and Warner, K. (2011), 'Ex-offenders and the labor market', *WorkingUSA* **14**(1), 87 – 109.
URL: *https://brill.com/view/journals/wusa/14/1/article-p87_7.xml*

Schmitt, J., Warner, K., Gupta, S. et al. (2010), 'The high budgetary cost of incarceration'.

Shapiro, J. M. (2005), 'Is there a daily discount rate? evidence from the food stamp nutrition cycle', *Journal of public Economics* **89**(2-3), 303–325.

Sheely, A. and Kneipp, S. M. (2015), 'The effects of collateral consequences of criminal involvement on employment, use of temporary assistance for needy families, and health', *Women & health* **55**(5), 548–565.

Stephens Jr, M. (2003), '"3rd of tha month": do social security recipients smooth consumption between checks?', *American Economic Review* **93**(1), 406–422.

Stephens Jr, M. and Unayama, T. (2008), 'Can governments help households smooth consumption? evidence from japanese public pension benefits', *Work. Pap., Carnegie Mellon Univ* .

Sun, L. and Abraham, S. (2021), 'Estimating dynamic treatment effects in event studies with heterogeneous treatment effects', *Journal of Econometrics* **225**(2), 175–199.

Timmins, N. (2016), 'Universal credit', *From Disaster to Recovery* .

Tiratelli, M., Bradford, B. and Yesberg, J. (2022), 'The political economy of crime: Did universal credit increase crime rates?', *The British Journal of Criminology* .

Travis, J., Solomon, A. L. and Waul, M. (2001), 'From prison to home: The dimensions and consequences of prisoner reentry'.

Trussell Trust (2019), 'Why we need to end the wait for universal credit', `https://www.trusselltrust.org/wp-content/uploads/sites/2/2019/09/PolicyReport_Final_ForWeb.pdf`. Accessed: 2023-01-15.

Tuttle, C. (2019), 'Snapping back: Food stamp bans and criminal recidivism', *American Economic Journal: Economic Policy* **11**(2), 301–327.

Uggen, C. (2000), 'Work as a turning point in the life course of criminals: A duration model of age, employment, and recidivism', *American Sociological Review* pp. 529–546.

Van den Berg, G. J., Van der Klaauw, B. and Van Ours, J. C. (2004), 'Punitive sanctions and the transition rate from welfare to work', *Journal of Labor Economics* **22**(1), 211–241.

Visher, C., Debus, S. and Yahner, J. (2008), 'Employment after prison: A longitudinal study of releasees in three states'.

Watson, B., Guettabi, M. and Reimer, M. (2020), 'Universal cash and crime', *Review of Economics and Statistics* **102**(4), 678–689.

Webster, D. (2018), 'Benefit sanctions statistics', https://cpag.org.uk/sites/default/files/files/18-11%20Sanctions%20Stats%20Briefing%20-%20D.Webster.docx. Accessed: 2021-4-10.

Western, B., Braga, A. A., Davis, J. and Sirois, C. (2015), 'Stress and hardship after prison', *American Journal of Sociology* **120**(5), 1512–1547.

Whitty, P. and Gilbody, S. (2005), 'Nice, but will they help people with depression? the new national institute for clinical excellence depression guidelines', *The British Journal of Psychiatry* **186**(3), 177–178.

Wickham, S., Bentley, L., Rose, T., Whitehead, M., Taylor-Robinson, D. and Barr, B. (2020), 'Effects on mental health of a uk welfare reform, universal credit: a longitudinal controlled study', *The Lancet Public Health* **5**(3), e157–e164.

Wilde, P. E. and Ranney, C. K. (2000), 'The monthly food stamp cycle: shopping frequency and food intake decisions in an endogenous switching regression framework', *American Journal of Agricultural Economics* **82**(1), 200–213.

Williams, E. (2021), 'Unemployment, sanctions and mental health: the relationship between benefit sanctions and antidepressant prescribing', *Journal of Social Policy* **50**(1), 1–20.

World Health Organisation (2023), 'Global burden of disease (gbd)', https://www.healthdata.org/gbd#:~:text=The%20Global%20Burden%20of%20Disease,be%20improved%20and%20disparities%20eliminated. Accessed: 2022-11-10.

Wright, S., Robertson, L. and Stewart, A. B. (2022), 'Universal credit and the invalidation of mental health problems: claimant and jobcentre plus staff experiences', *Journal of Poverty and Social Justice* **30**(2), 151–170.

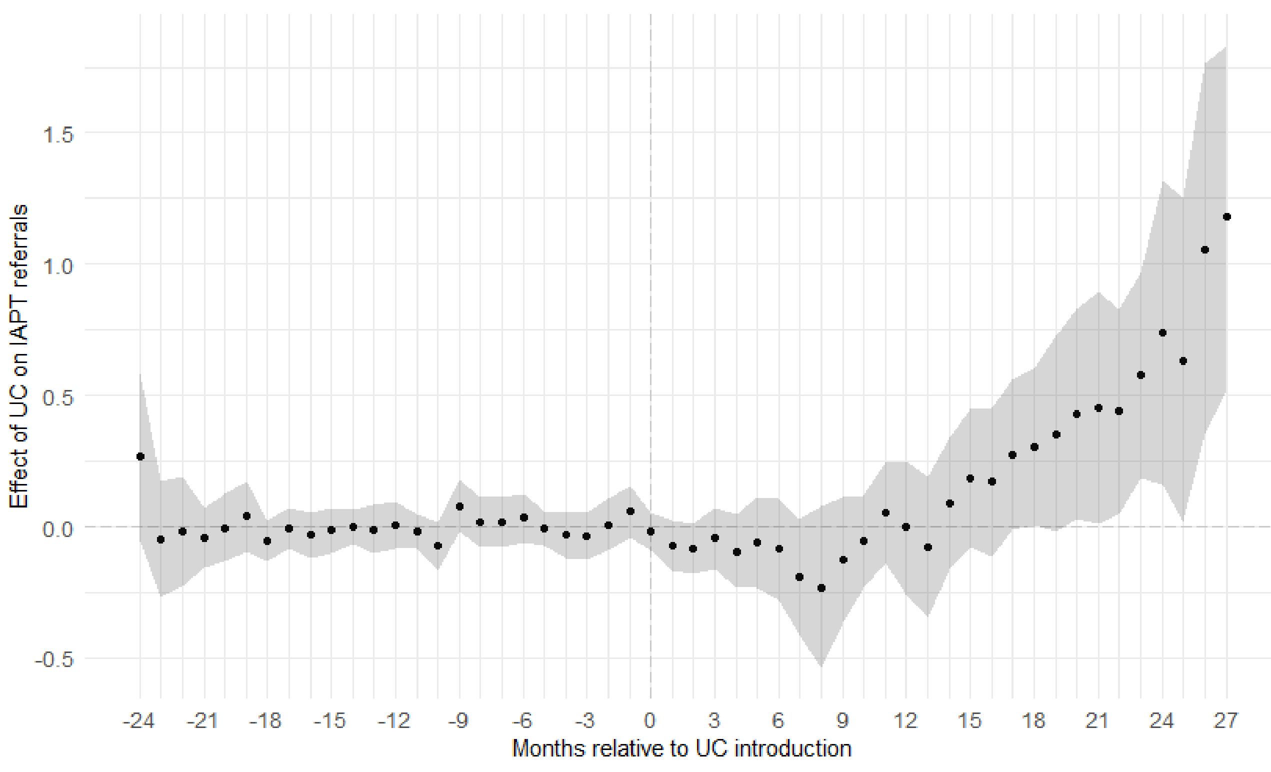
Yang, C. S. (2017*a*), 'Does public assistance reduce recidivism?', *American Economic Review* **107**(5), 551–555.

Yang, C. S. (2017*b*), 'Local labor markets and criminal recidivism', *Journal of Public Economics* **147**, 16–29.

Appendix A

Appendix - Chapter 3

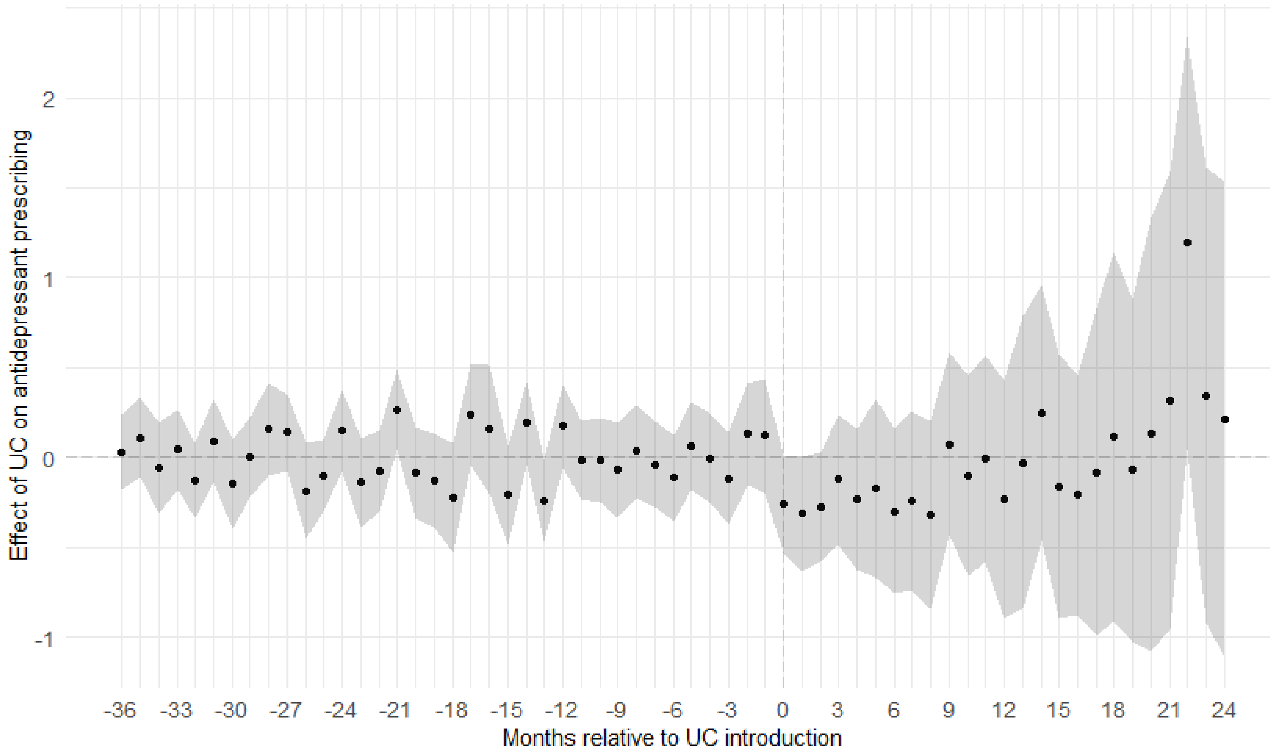
Figure A1: Event Study Plot - IAPT



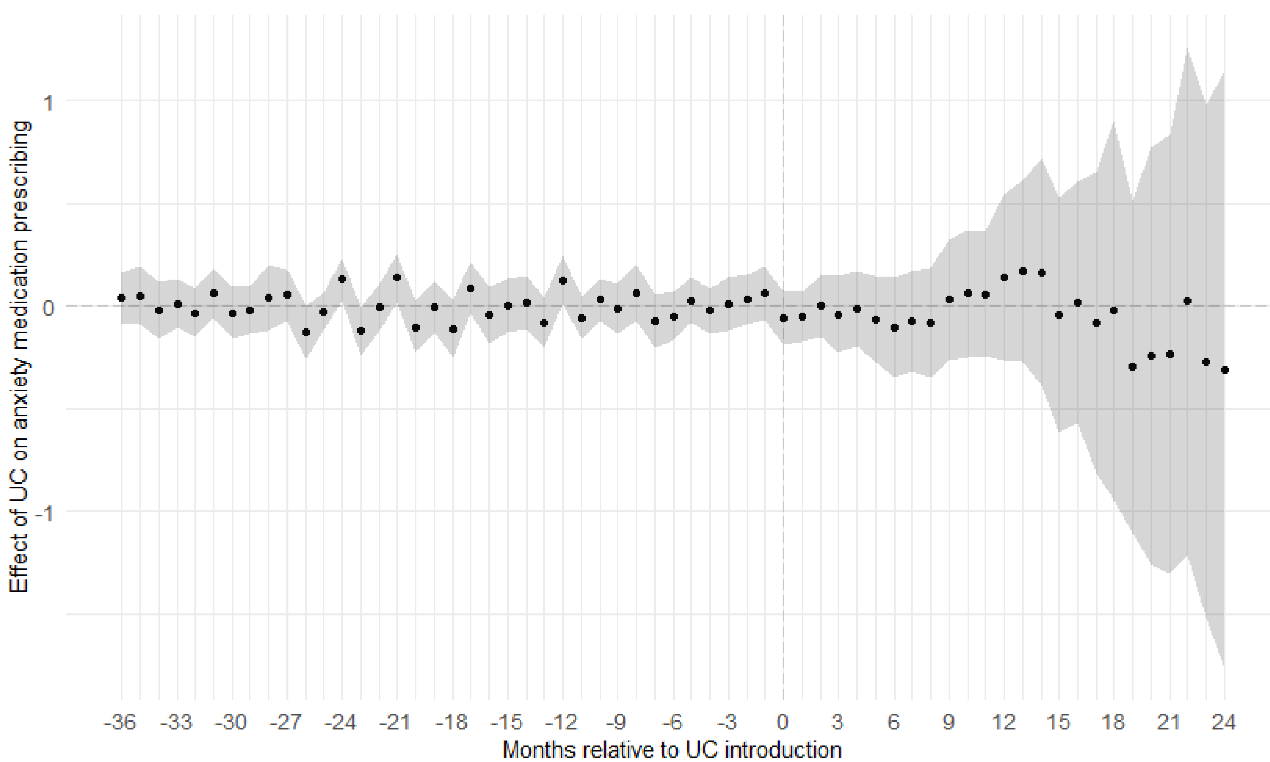
Notes: This figure shows the estimated impact of UC on IAPT referrals per 1,000 working-age population, using the Callaway and Sant'Anna (2021) estimator. Each dot represents a point estimate. Shaded area represents 95% confidence intervals.

Figure A2: Event Study Plot - Antidepressants and Anxiety Medication

(a) Antidepressants



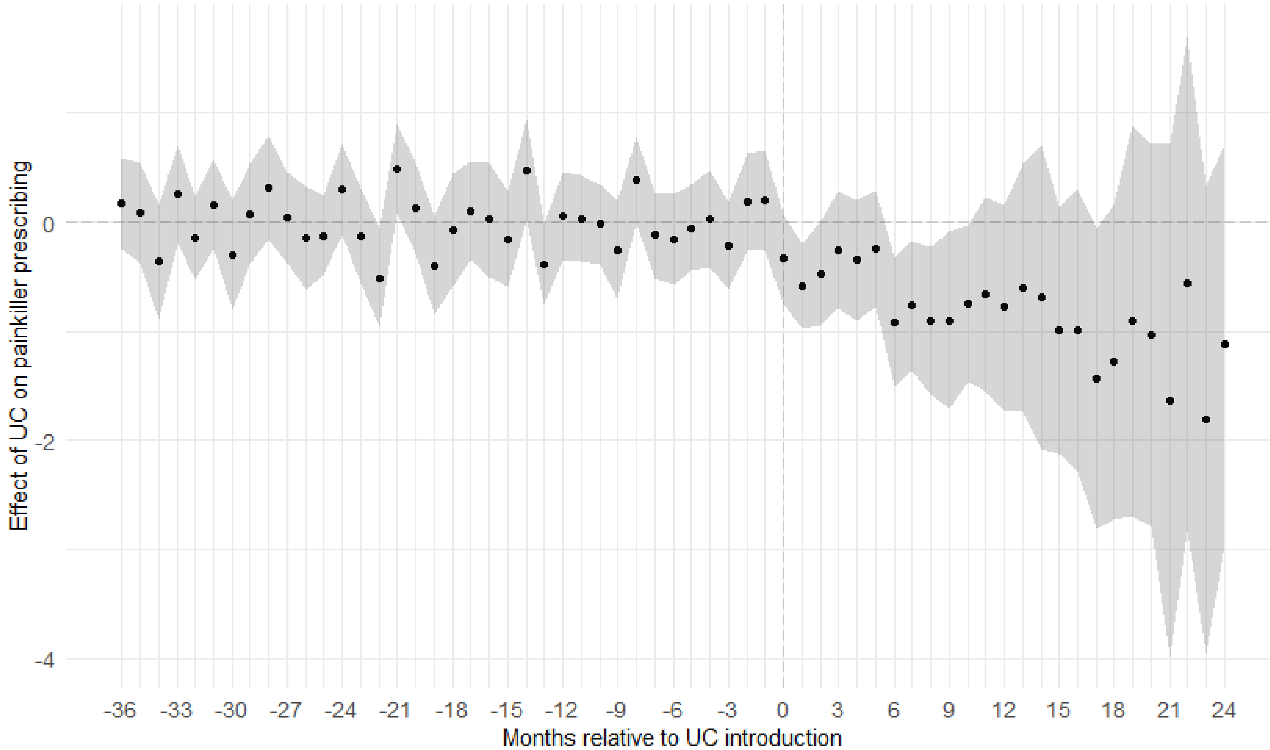
(b) Anxiety medication



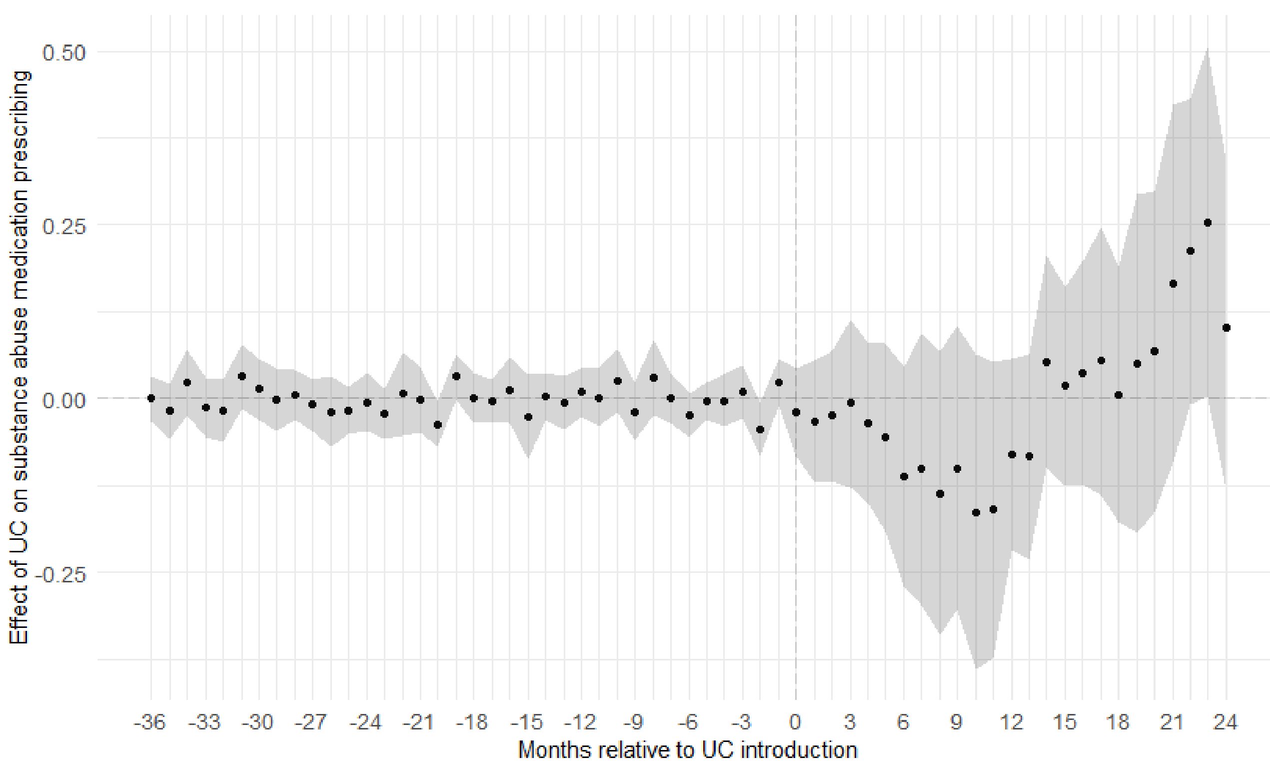
Notes: These figures show the estimated impact of UC on antidepressant and anxiety medication prescribing per 1,000 patients, using the Callaway and Sant'Anna (2021) estimator. Each dot represents a point estimate. Shaded area represents 95% confidence intervals.

Figure A3: Event Study Plot - Painkillers and Substance Abuse Medication

(a) Painkiller medication



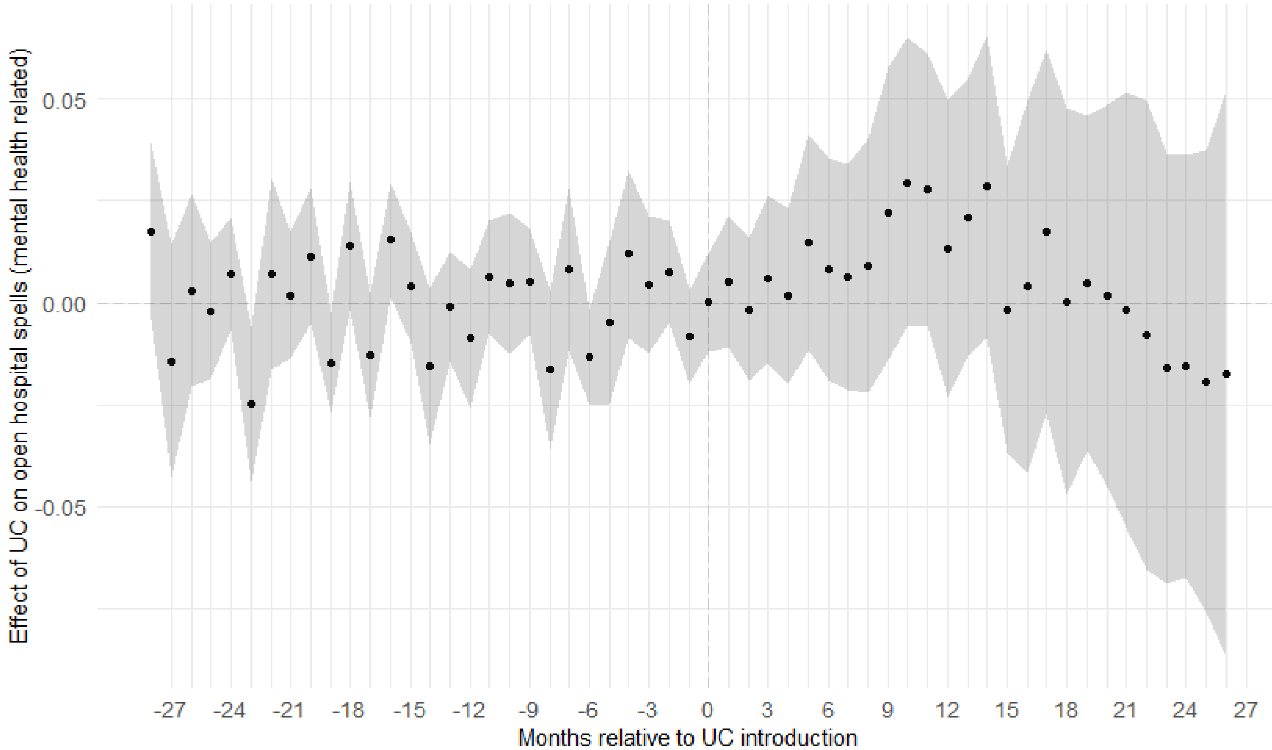
(b) Substance abuse medication



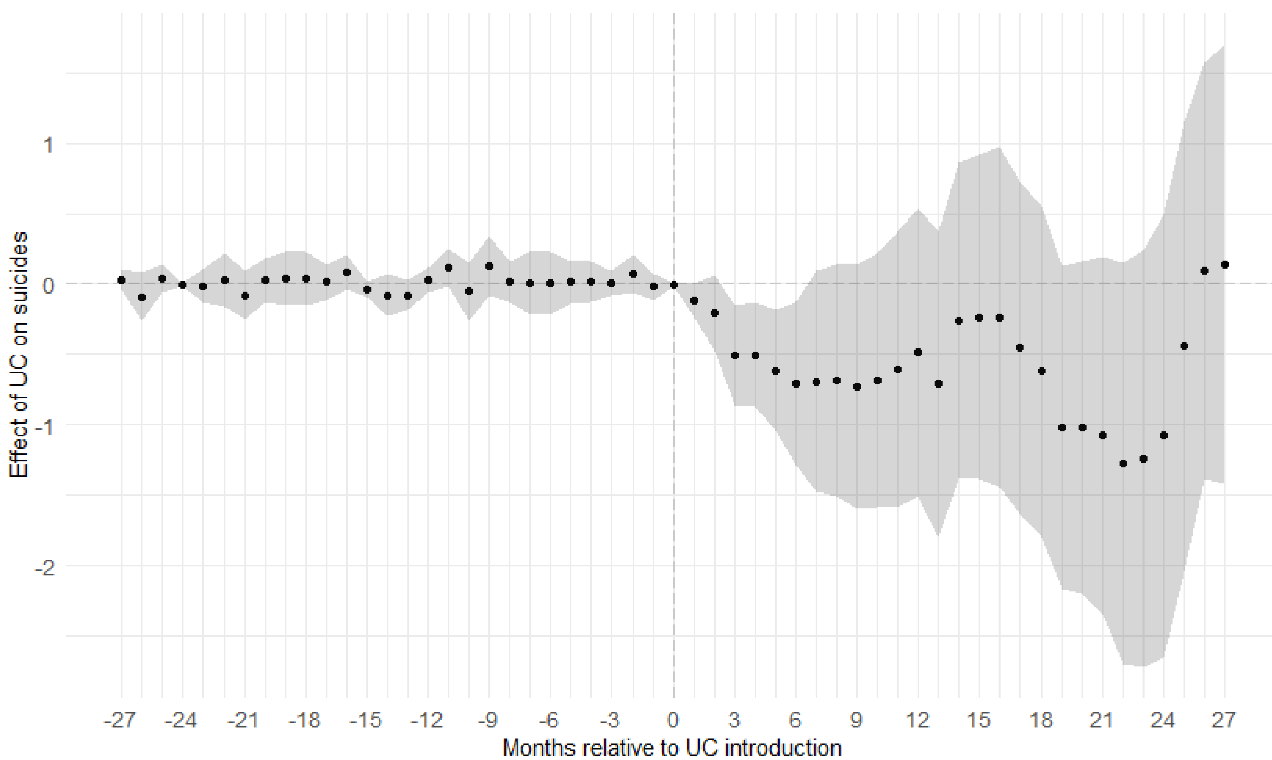
Notes: These figures show the estimated impact of UC on painkiller and substance abuse GP prescribing per 1,000 patients, using the Callaway and Sant'Anna (2021) estimator. Each dot represents a point estimate. Shaded area represents 95% confidence intervals.

Figure A4: Event Study Plot - Open Hospital Spells and Suicides

(a) Open hospital spells



(b) Suicides

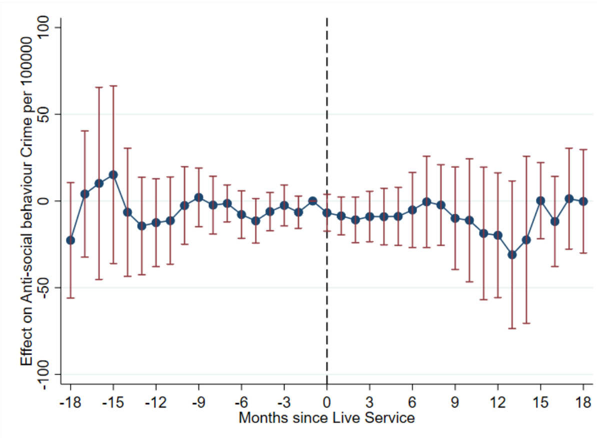


Notes: These figures show the estimated impact of UC on the number of open hospital spells per 1,000 working-age population and suicides per 100,000 population, using the Callaway and Sant'Anna (2021) estimator. Each dot represents a point estimate. Shaded area represents 95% confidence intervals.

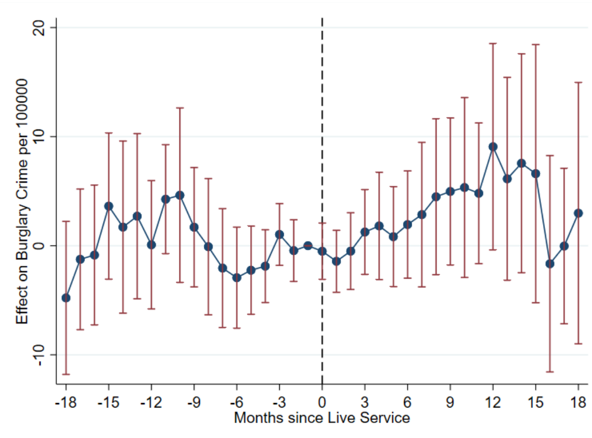
Appendix B

Appendix - Chapter 4

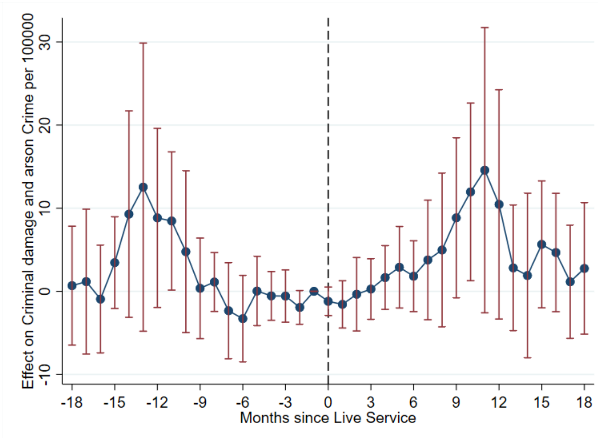
Figure B1: Live Service Event Study Plots - Crime Breakdown



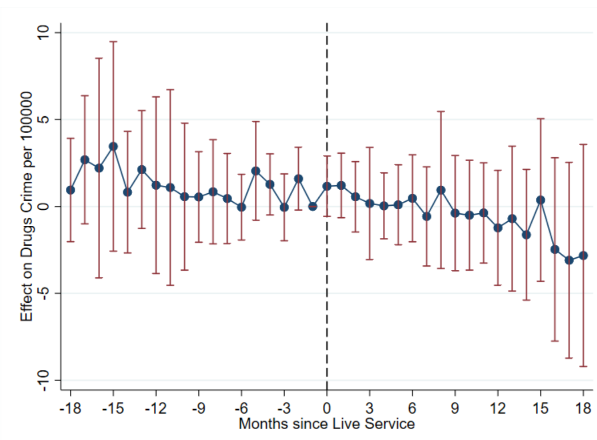
(a) Anti-social behaviour



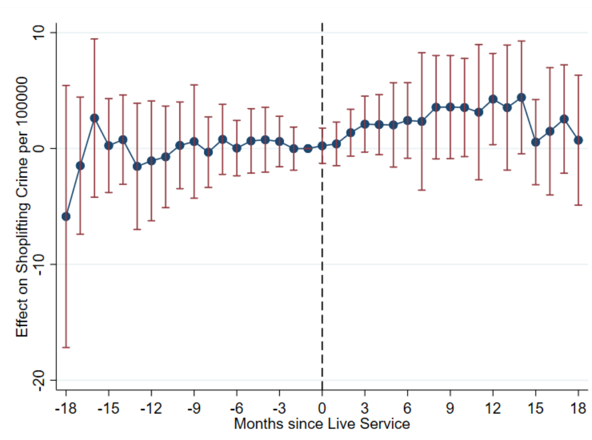
(b) Burglary



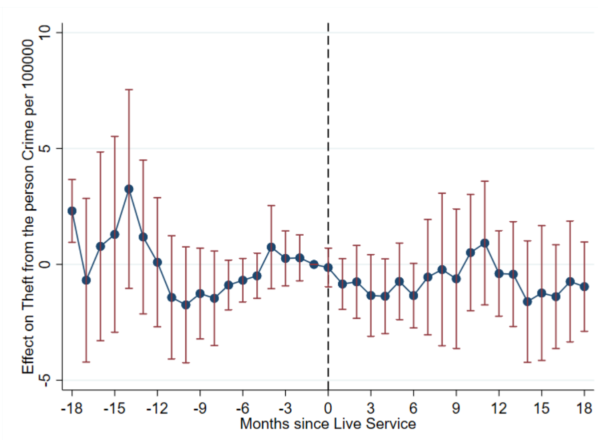
(c) Criminal Damage & Arson



(d) Drugs



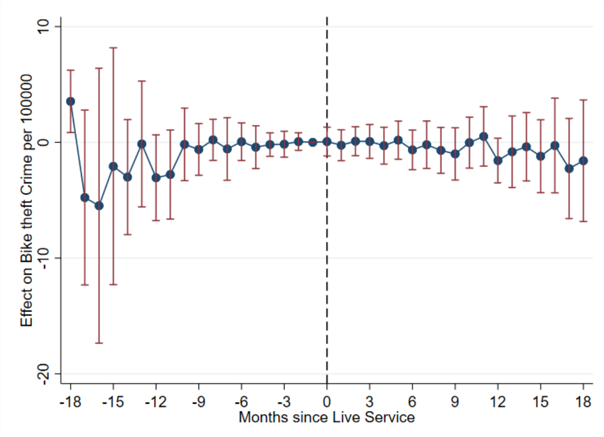
(e) Shoplifting



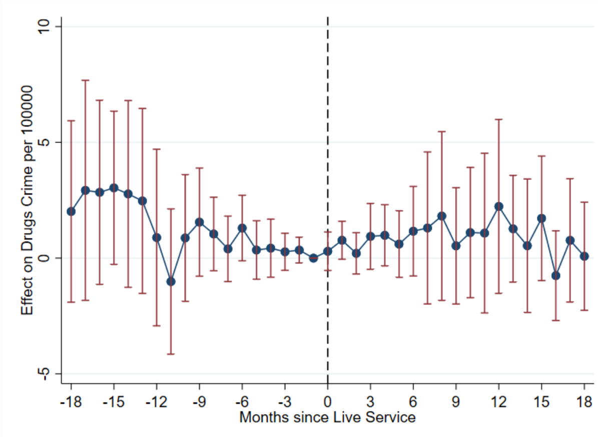
(f) Theft from the person

Notes: These figure show how the average effect of the LS evolved over time for various crime rates across local authorities. 95% confidence intervals are shown as red lines. DiD_l estimates the dynamic effect l by comparing the change in crime rates between local authorities that started receiving the LS l months ago to districts not yet converted from the legacy system. This comparison is over $l+1$ months, from the month before treatment to l month after. Placebo estimates on the left of t=0 test the assumptions of parallel trends and no anticipation. Standard errors are clustered at the local authority level.

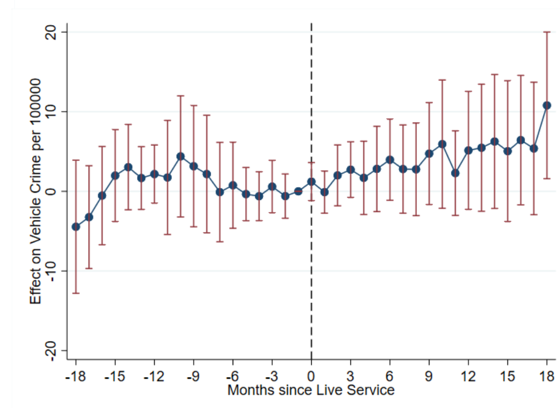
Figure B2: Live Service Event Study Plots - Crime Breakdown



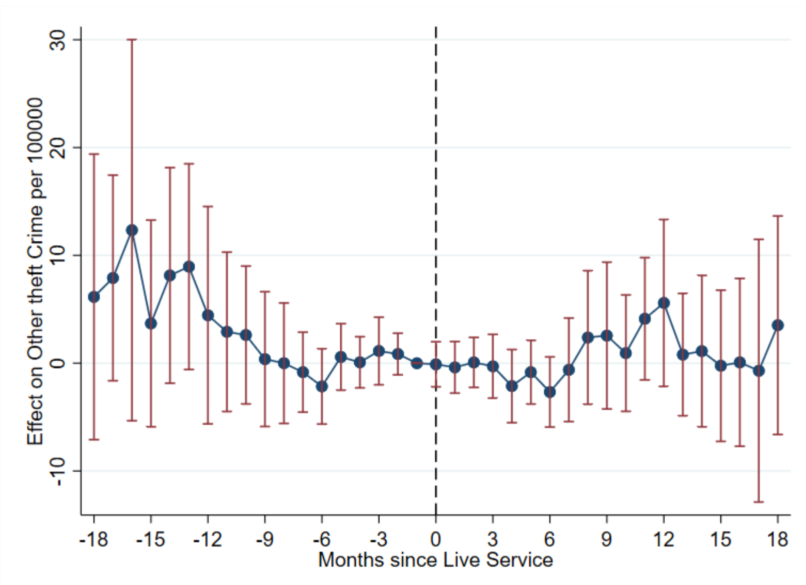
(a) Bike theft



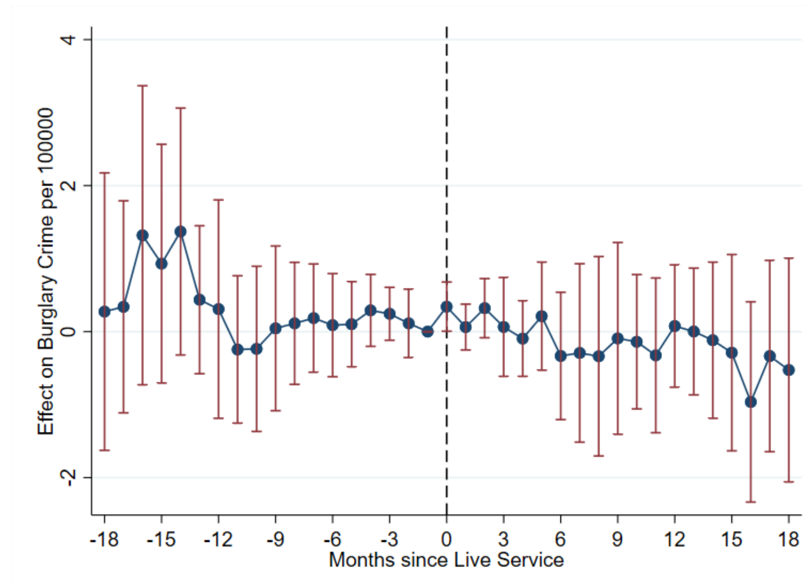
(b) Other crimes



(c) Vehicle



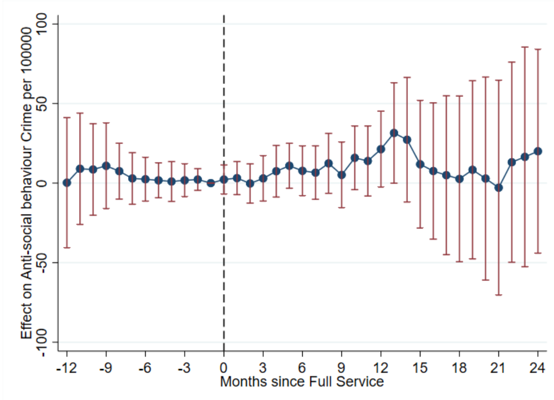
(d) Other theft



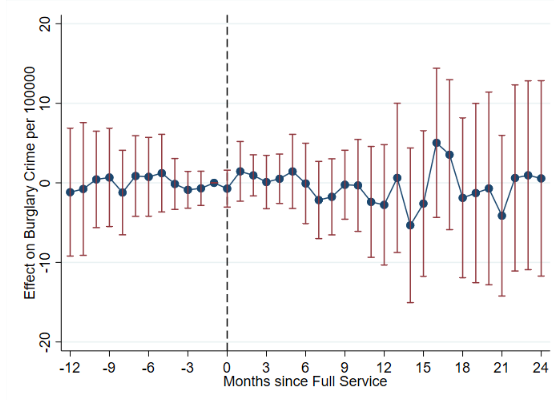
(e) Robbery

Notes: These figure show how the average effect of the LS evolved over time for various crime rates across local authorities. 95% confidence intervals are shown as red lines. DiD_l estimates the dynamic effect l by comparing the change in a given crime type between local authorities that started receiving the LS l months ago to districts not yet converted from the legacy system. This comparison is over $l + 1$ months, from the month before treatment to l month after. Placebo estimates on the left of t=0 test the assumptions of parallel trends and no anticipation. Standard errors are clustered at the local authority level.

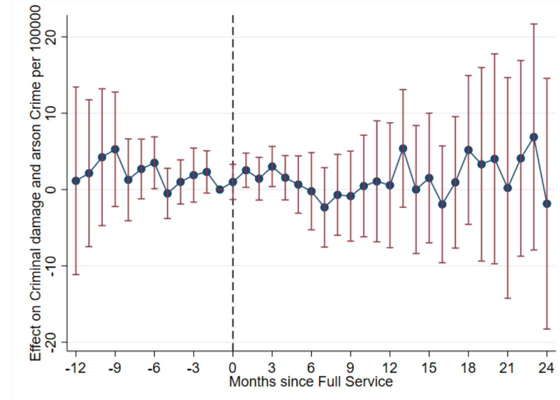
Figure B3: Full Service Event Study Plots - Crime Breakdown



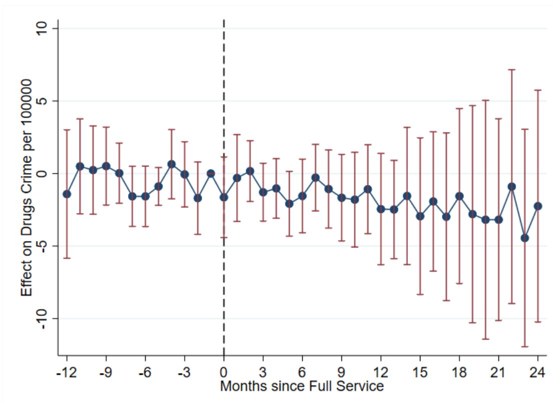
(a) Anti-social behaviour



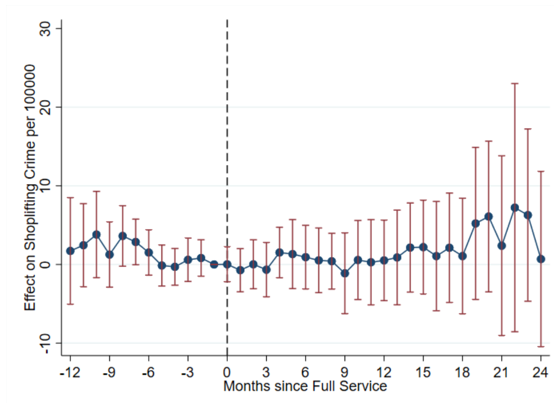
(b) Burglary



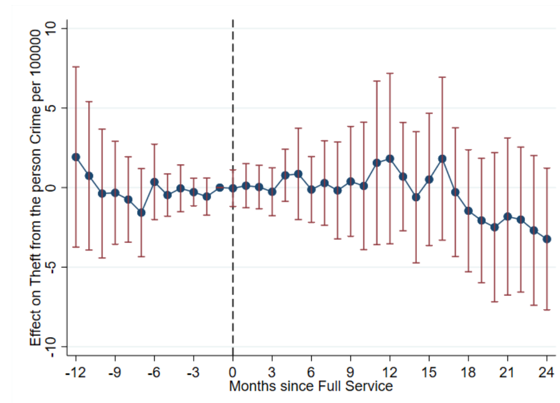
(c) Criminal Damage & Arson



(d) Drugs



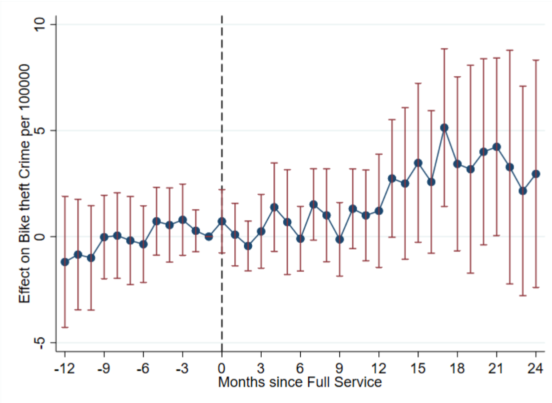
(e) Shoplifting



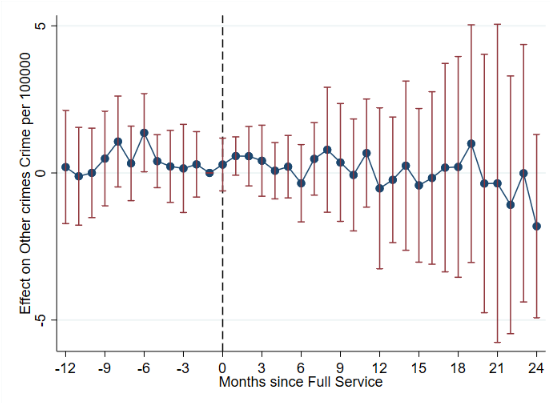
(f) Theft from the person

Notes: These figure show how the average effect of the FS evolved over time for various crime rates across local authorities. 95% confidence intervals are shown as red lines. DiD_l estimates the dynamic effect l by comparing the change in a given crime type between local authorities that started receiving the FS l months ago to districts still operating the LS. This comparison is over $l+1$ months, from the month before treatment to l month after. Only local authorities that started receiving the FS within the same quarter are used for the comparison. Placebo estimates on the left of t=0 test the assumptions of parallel trends and no anticipation. Standard errors are clustered at the local authority level.

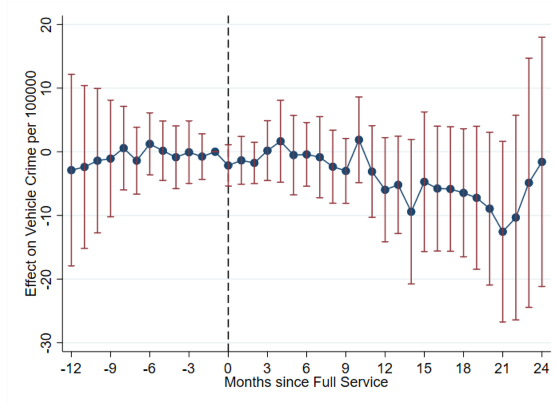
Figure B4: Full Service Event Study Plots - Crime Breakdown



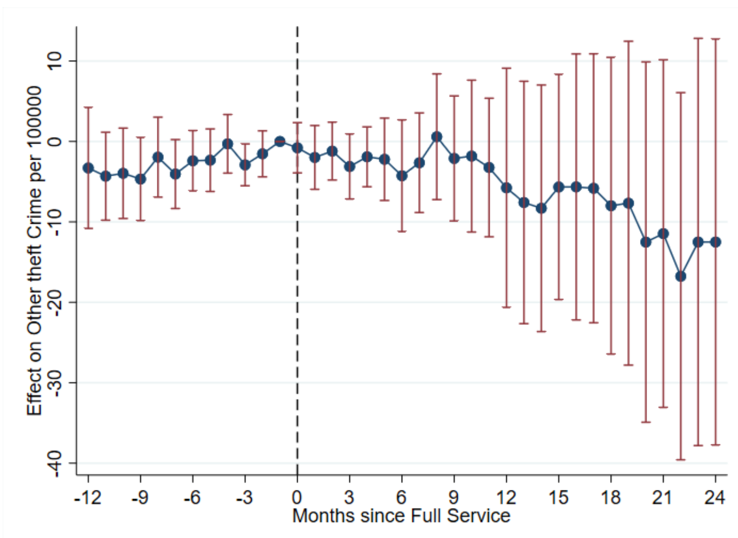
(a) Bike theft



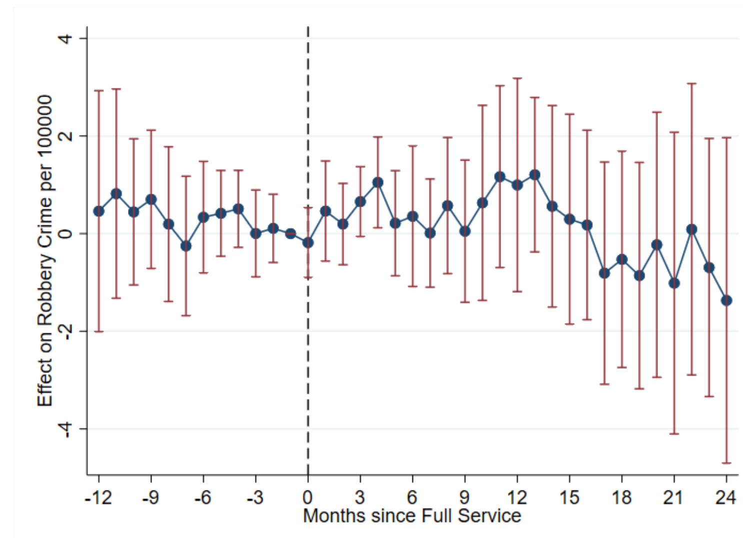
(b) Other crimes



(c) Vehicle



(d) Other theft



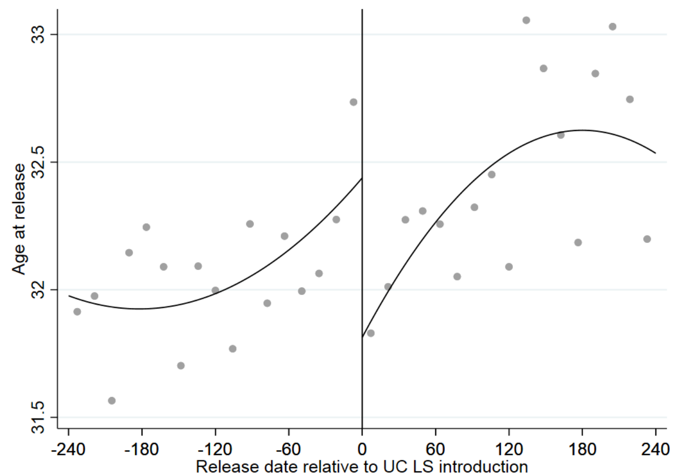
(e) Robbery

Notes: These figure show how the average effect of the FS evolved over time for various crime rates across local authorities. 95% confidence intervals are shown as red lines. DiD_l estimates the dynamic effect l by comparing the change in a given crime type between local authorities that started receiving the FS l months ago to districts still operating the LS. This comparison is over $l+1$ months, from the month before treatment to l month after. Only local authorities that started receiving the FS within the same quarter are used for the comparison. Placebo estimates on the left of t=0 test the assumptions of parallel trends and no anticipation. Standard errors are clustered at the local authority level.

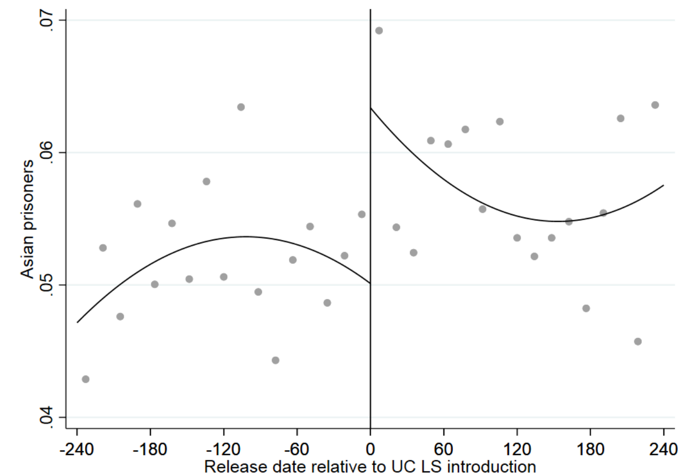
Appendix C

Appendix - Chapter 5

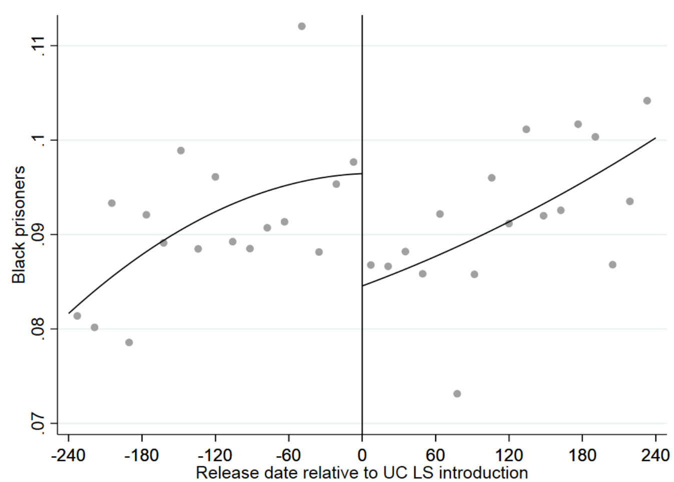
Figure B1: Check for Discontinuity in Demographic Controls



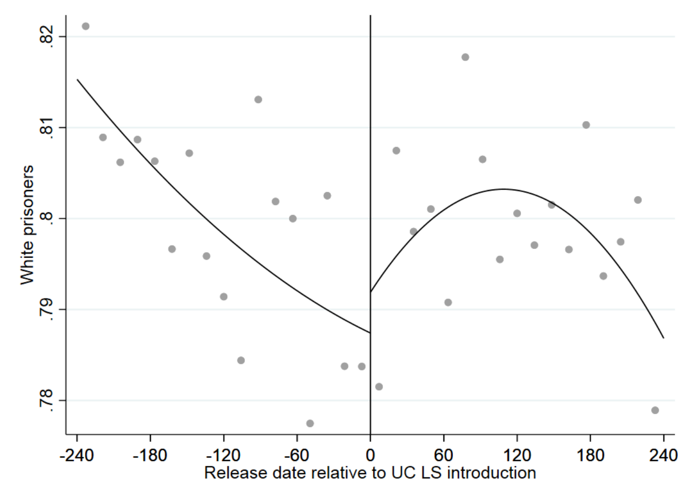
(a) Age at release



(b) Asian



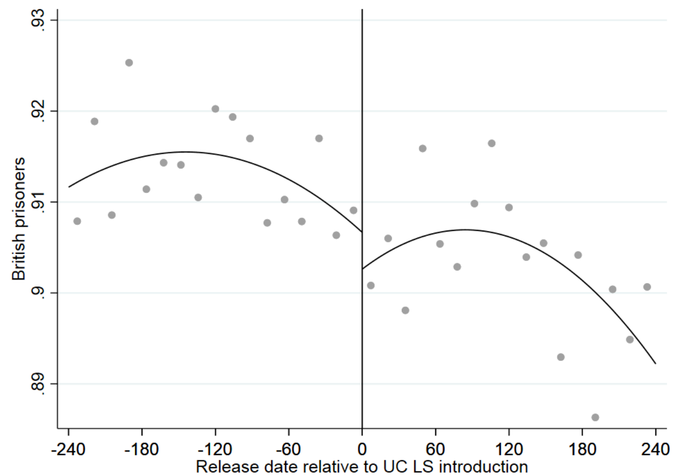
(c) Black



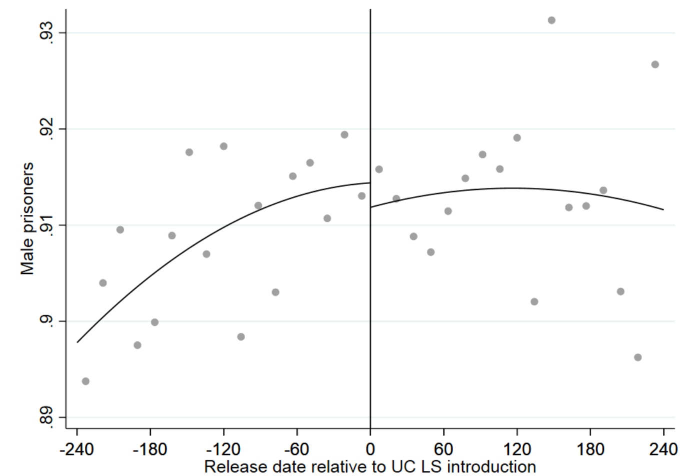
(d) White

Notes: These figures check for discontinuities in prisoners' demographics around the period of the UC Live Service (LS) introduction. The x axis in each figure is the running variable used for the analysis: prisoners' release date relative to the UC Live Service introduction date. The top left figure plots prisoners' average age at release; the top right figure plots the proportion of Asian prison leavers; the bottom left figure plots the proportion of Black prison leavers; and the bottom right figure plots the proportion of white prison leavers. Each dot in the figures represents the average outcome of approximately 1,000 prison releases within a given two-week interval.

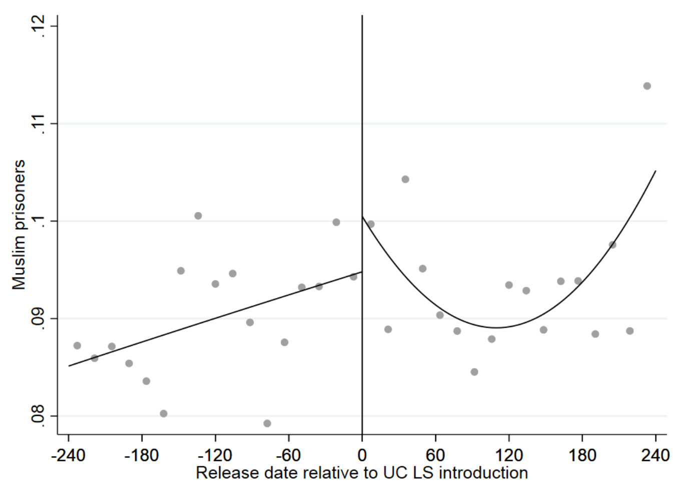
Figure B2: Check for Discontinuity in Demographic Controls



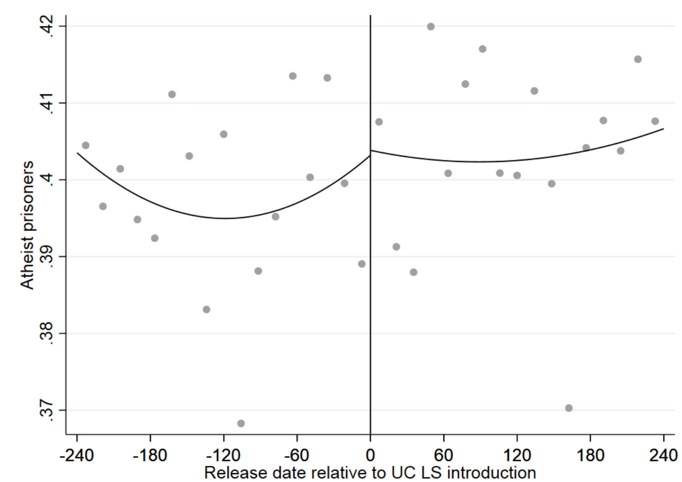
(a) British



(b) Male



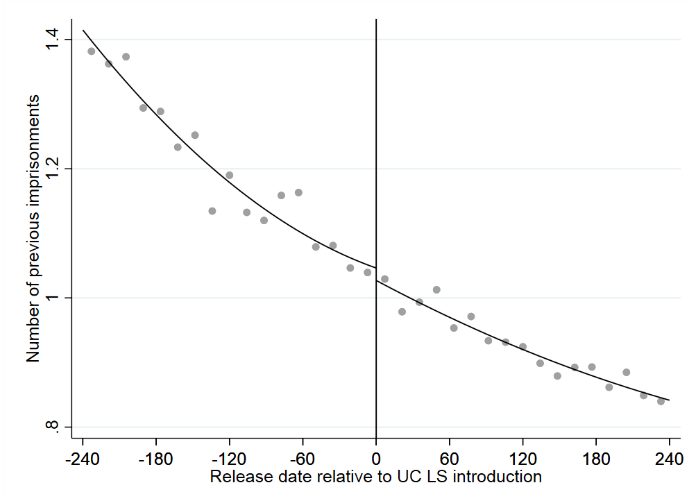
(c) Muslim



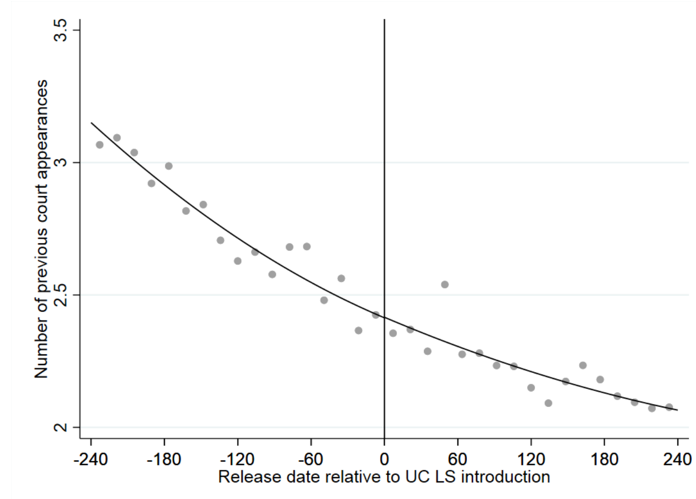
(d) Athiest

Notes: These figures check for discontinuities in prisoners' demographics around the period of the UC Live Service (LS) introduction. The x axis in each figure is the running variable used for the analysis: prisoners' release date relative to the UC Live Service introduction date. The top left figure plots the proportion of British prison leavers; the top right figure plots the proportion of Male prison leavers; the bottom left figure plots the proportion of Muslim prison leavers; and the bottom right figure plots the proportion of Athiest prison leavers. Each dot in the figures represents the average outcome of approximately 1,000 prison releases within a given two-week interval.

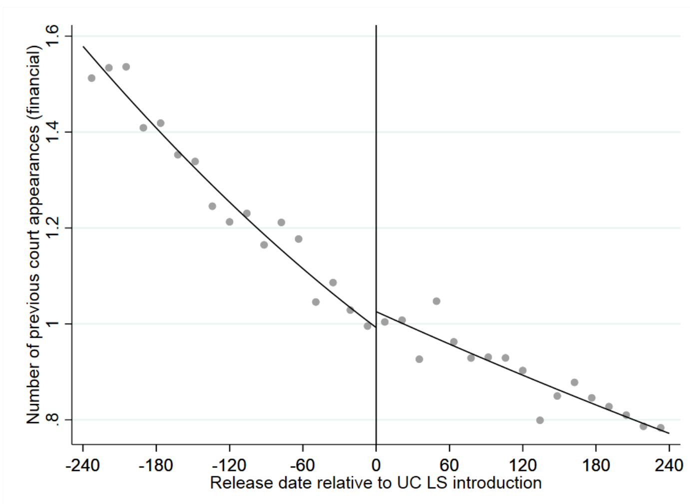
Figure B3: Check for Discontinuity in Criminal History



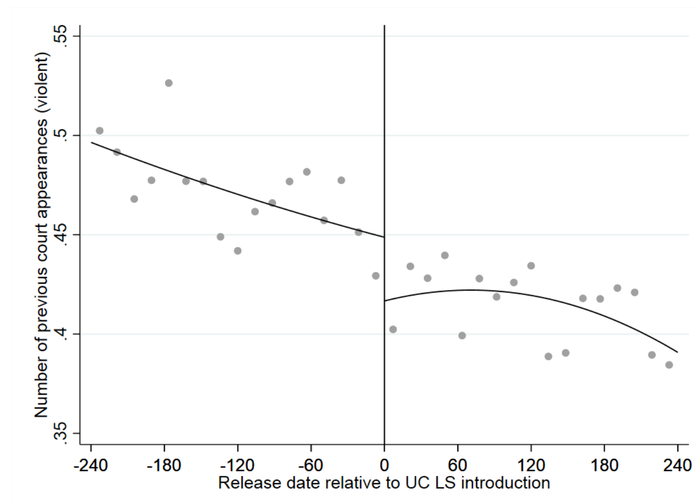
(a) Prison spells



(b) Number of previous court appearances



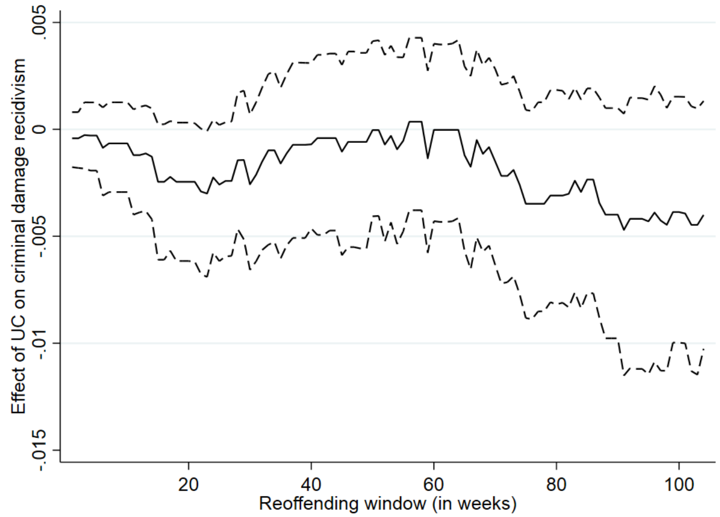
(c) Number of previous court appearances (acquisitive)



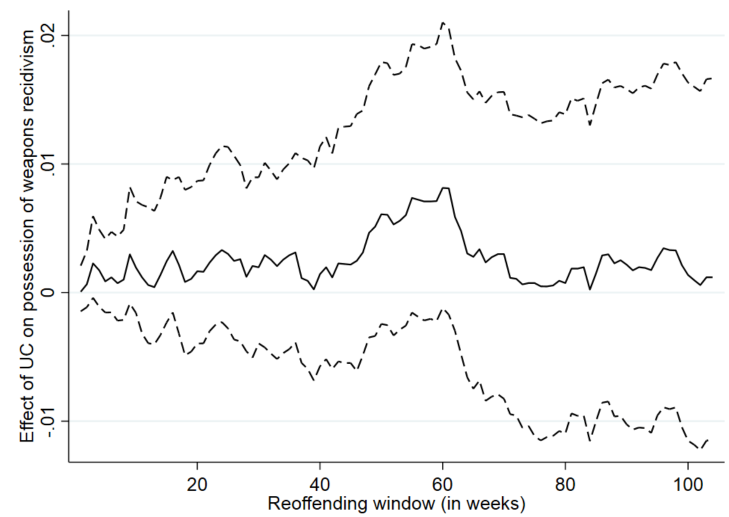
(d) Number of previous court appearances (violent)

Notes: These figures check for discontinuities in prisoners' criminal history around the period of the UC Live Service (LS) introduction. The x axis in each figure is the running variable used for the analysis: prisoners' release date relative to the UC Live Service introduction date. The top left figure plots the average number of previous imprisonments; the top right figure plots the average number of previous court appearances; the bottom left figure plots the average number of acquisitive-crime related previous court appearances; and the bottom right figure plot the average number of previous violent-crime related court appearances. Each dot in the figures represents the average outcome of approximately 1,000 prison releases within a given two-week interval.

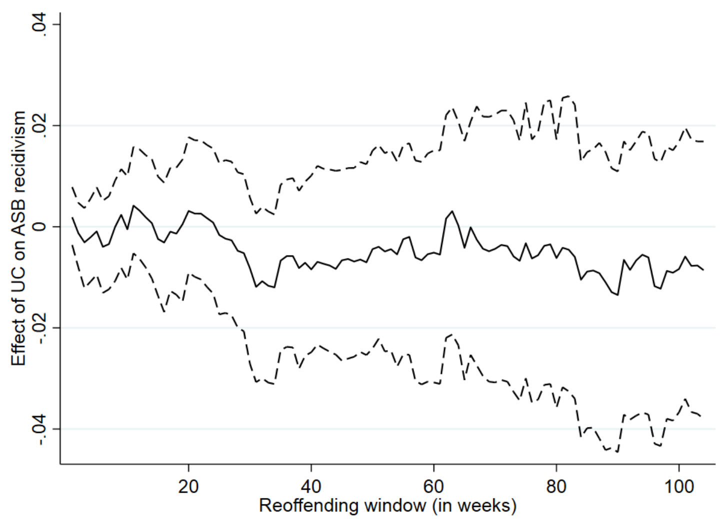
Figure B4: Effect of UC on Recidivism over a Period of 2 Years since Release - Crime Breakdown



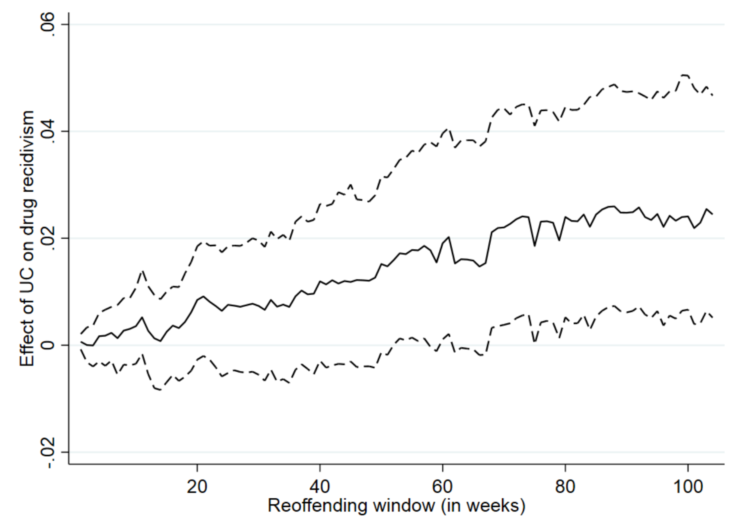
(a) Criminal damage and arson



(b) Possession of weapons



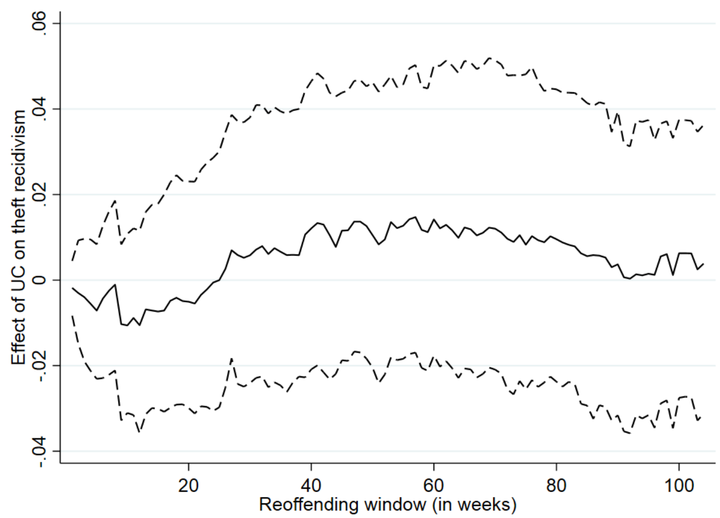
(c) Antisocial behaviour



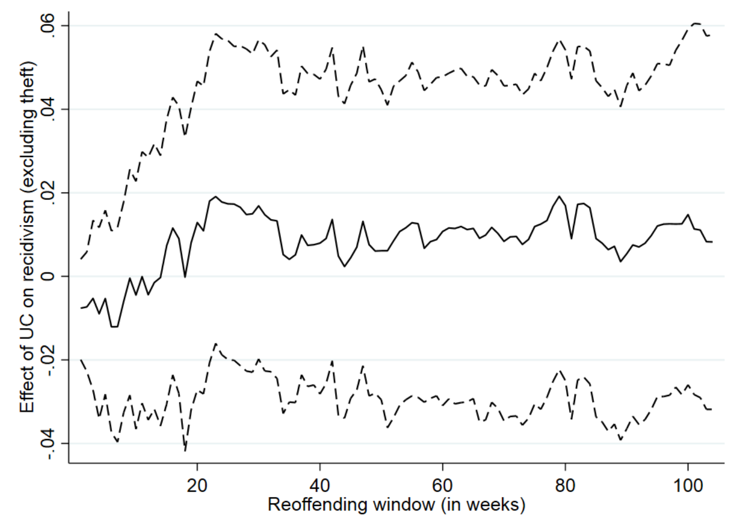
(d) Drug offences

Notes: This figure shows the effect of the UC Live Service on different recidivism windows, up to 2 years (104 weeks). The black line in each figure is made up 104 separate RD treatment effect estimates, with each one incrementally increasing the recidivism period in question by 1 week. Dashed lines represent 95% confidence intervals.

Figure B5: Effect of UC on Recidivism over a Period of 2 Years since Release - Crime Breakdown



(a) Theft



(b) Total crime excluding theft

Notes: This figure shows the effect of the UC Live Service on different recidivism windows, up to 2 years (104 weeks). The black line in each figure is made up 104 separate RD treatment effect estimates, with each one incrementally increasing the recidivism period in question by 1 week. Dashed lines represent 95% confidence intervals.