The Work Programme

A quantitative impact assessment

November 2020

Summary

This report presents an analysis of the impact of Work Programme participation on employment and benefit outcomes, and a cost-benefit analysis based on these findings.

Contents

List of Tables

List of Figures

Acknowledgements

Special thanks go to Professor Richard Dorsett from the University of Westminster for his peer review of the drafts of the impact analysis. Richard's review was confined to the methodology of the impact analysis; Richard did not have access to the actual data used and did not attempt to replicate the results, or review the cost-benefit analysis (which was quality assured internally).

Also thanks to Stephen Aldridge and colleagues at MHCLG for further review and commenting which improved some of the drafting.

And thank you for the following colleagues at DWP for their help towards producing this report including:

Mike Daly; Robin Wood; Stuart Prince; Michael Roussos-Poponis; and Ali McAuley.

The Authors

Jeraldine Kay, Operational Research Analyst at the Department for Work and Pensions.
Simon Marlow, Operational Research Analyst at the Department for Work and Pensions.

Executive Summary

Background

This report presents an impact assessment and accompanying cost-benefit analysis of the Work Programme, which was live between 2011 and 2017 in England, Scotland and Wales.

The programme offered support to a wide variety of people out of work, however this study focuses on individuals aged 25 years and older who had been in receipt of Jobseeker's Allowance or Universal Credit in the Searching for Work conditionality group for 12 months. This group formed Work Programme Payment Group 2, the largest group of Work Programme participants.

Work Programme providers were free to design their own services based on their assessments of individual and local needs and were paid primarily for supporting claimants into employment and helping them stay there.

Methodology

The basis of the analysis is the gap between when referrals to the Work Programme closed in March 2017 and the start of the new Work and Health Programme (from April 2018 the programme started to accept low numbers of referrals for those who had been in receipt of an out of work benefit for 24 months). Those reaching 12 months of claiming before the end of March 2017 can be compared to those reaching 12 months claiming from April 2017, with the difference reflecting the impact of the Work Programme. The focus is on consecutive three month cohorts before and after the programme end but other cohorts are also considered for exploring sensitivity.

To eliminate possible selection biases from the referral process not present after referrals to the programme had closed, an Intention to Treat (ITT) approach was employed, where individuals that were eligible for the Work Programme before the end of March 2017 were compared with individuals that would have been eligible after March 2017. The impacts for the ITT groups are then used to provide an estimate for the impact of participation – the Average Treatment Effect on the treated.

The two cohorts are very similar on observable characteristics but propensity score matching is used to adjust for the small differences between the two groups. However, there are a number of other important issues of internal validity that might bias the results. By drawing groups of cohorts from different, but consecutive, time periods it is possible that changes in the state of the labour market, or other factors varying over time, would also affect the difference between the outcomes of the two groups. Although the period of the study occurred at a time when the overall labour market was gradually improving (which might, all things equal, might cause some underestimation to the estimates) additional tests, using cohorts from different times, attempt to quantify where appropriate, the potential size of these differences.

The selected cohorts are tracked in terms of their employment and benefit status for two years following their referral to the programme. The cost benefit analysis follows

the DWP Social Cost-Benefit Analysis Framework (Fujiwara 2010) methodology. For the purposes of the Cost Benefit Analysis, a further one and two years of impact have been extrapolated from this.

Key findings

Across all the analysis performed, individuals aged 25 years or over and in receipt of Jobseeker's Allowance or Universal Credit in the Searching for Work conditionality group for 12 months who were eligible for the Work Programme in 2017, on average, spent less time in receipt of out of work benefits and more time in employment.

Over the two-year period, following referral to the programme, the main estimate shows that that on average the Intention to Treat (ITT) group spent 17 more days in employment, 26 fewer days in receipt of out of work benefits and 9 extra days on neither out of work benefits nor employment. Since the ITT group consists of 38% participants, and if the impact is only from those participants then it is reasonable to expect that the impact on participation would be around 38% of that observed for the group. With this assumption, the impact of participating on the programme is 46 additional days in employment and not in receipt of an out of work benefit, and 70 fewer days in receipt of an out of work benefit over the two years following their eligibility to the programme.

As indicated in the methodology there are a number of caveats. Although sensitivity analyses have been performed and a range of different cohorts have been used, it cannot be guaranteed that all the differences that might affect outcomes, between the groups, have been captured. These include differences in composition and differences arising from drawing groups from the two consecutive time periods.

Comparisons of individuals from during and after the programme suggest that there are some time effects but they are smaller than the main estimates: the maximum observed fluctuation which is due to time and seasonality effects on ITT groups is 6 days in employment and 9 days in receipt of benefit, which represent an adjustment to the impacts of participation of 16 days in employment and 22 days in receipt of benefit. It is not absolutely certain that the time effects for period between the ITT groups would be greater or smaller than the time effects observed from these other consecutive cohorts, however this does give some assurance around the main estimates, while bearing in mind that there are still some uncertainties.

Another point is that the analysis was during a period when UC was being rolled out across the country: since the comparison group is drawn three months later and it is difficult to map legacy benefits onto UC benefits (our definitions will both under-capture some outcomes or over-capture some outcomes that may not have been captured in the legacy system), this introduces complexity and therefore some additional uncertainty around the impacts – however, we think that effect cannot be large given that both referral cohorts had the same low proportion (a fifth) of UC claims.

In addition, by looking at off-flow rates from before eligibility there was no apparent evidence of anticipation effects which may bias the analysis. Also, by looking at proportion referred to other provision a higher proportion (20%) of the comparison group was observed to be referred to other provision compared to the treatment

group (10%); if the other provision had any impact then this would imply that the estimates understate the impact slightly, all other things equal.

There are a number of factors that limit the external validity, or generalisability, of these results, which need to be borne in mind. Firstly, the impacts were only drawn from the end of the programme for a particular payment group in a particular labour market. It is impossible to know whether the more important influence was whether the programme was at its most mature and effective or whether indeed the programme was starting to wind down - which would mean the estimates would be conservative. Other factors, notably that the period of interest covered the time when decreasing proportions of people who were eligible went on the programme and when the benefit system was changing as Universal Credit was rolling out, also mean that the estimates cannot be safely extrapolated backwards to earlier in the programme; equally some care needs to be taken in generalising these results too far for other times.

Using the observed two year impacts for participation are already, even at this point, net positive for participant, the Exchequer and society:

- For the participant: £1.25, with a range of £1.22 to £1.27
- For DWP benefits: £0.83, with a range of £0.56 to £1.10
- For the Exchequer: £1.50, with a range of £1.00 to £2.01
- For society: £1.75, with a range of £1.16 to £2.31

An important point here is these estimates do not rely on the assumption that the impact from the ITT groups is all due to the Work Programme participants; if some of the ‘impact’ came from the non-participants then the net benefits would be the same and this cost-benefit analysis would give the same results.

It is also clear that the trajectory of the impacts indicates they will extend beyond the tracking period, which is likely to give a fairer representation of potential returns. It is difficult to determine how long they will continue but it is reasonable to suggest that impacts could be extrapolated for a further one or two years. For an extrapolation of a further two years (i.e. a total of four years) the analysis shows that returns would be net positive for participant, exchequer and society:

- For the participant: £1.26, with a range of £1.23 to £1.27
- For DWP benefits: £1.76, with a range of £1.21 to £2.32
- For the Exchequer: £3.21, with a range of £2.17 to £4.25
- For society: £3.51, with a range of £2.42 to £4.51

Some care needs to be taken in making comparisons between these results and those from other labour market interventions. However, it is reasonable to suggest these results, that have been observed for this end part of the programme, would lie in the upper end of results for labour market interventions. It is difficult to bring this study together with previous older evaluation evidence drawn from earlier in the programme to understand why we observe these positive returns but we can

tentatively suggest that the provision had a number of features which align with the overall evidence base for providing effective employment provision e.g. importance of the adviser support, focus on individual client need and continuity. On the other hand, we bear in mind that older programme evidence and other research suggested that the Work Programme did not work as well for some people (e.g. older and those with health and disabilities) – although many of those will have not been within this particular payment group.

1. Introduction

1.1. What is the Work Programme?

The Coalition Government launched the Work Programme in 2011 throughout Great Britain. It was the Department for Work and Pensions' (DWP) flagship scheme, amongst a number of welfare to work initiatives, aimed at helping people in receipt of out of work benefits find sustained work.

Private and public companies were contracted as providers to deliver up to two years support to suit individuals' needs, to find work. Providers were also incentivised to offer in work support to participants in employment, up to the end of the two-year period, i.e. for as much as two years after first entering employment. Participants were referred to the providers from Jobcentre Plus (JCP) and providers were paid on a 'payment by results' (PbR) basis according to participant's achievement of defined employment outcomes.

The Work Programme has been the largest employment programme in the UK with 2 million referrals between 2011 and 2017 across different participant groups, organised into 10 'Payment Groups'. The programme had an overall cost of around £2.9 billion pounds and saw about 630 thousand spells of employment lasting at least 13 or 26 weeks.

1.2. Purpose of the analysis

This report provides a quantitative assessment of the impact Work Programme had upon a specific group of participants' benefit receipt and employment levels. The analysis focusses on the largest of the programme's Payment Groups (PG2) which saw mandatory referrals for individuals aged 25 years and over, who had been in receipt of Jobseeker's Allowance or Universal Credit in the Searching for Work conditionality group for 12 months. The impact is also supplemented with a cost benefit analysis.

Importantly, no previous quantitative assessment of Work Programme's impact has been published. The lack of an appropriate control and its national rollout has produced limited opportunity to develop experimental or quasi-experimental estimates.

The policy was that referrals to this payment group would be made at a specific point in the individual's spell of unemployment - 12 months; this provides an opportunity to restrict a potential counterfactual group in a way other voluntary payment groups do not. It is hoped that some other Payment Groups can be studied in the future.

The current analysis exploits the discontinuity brought about by programme referrals ending in March 2017 which meant that individuals towards the end of the programme could be compared with a group of individuals who would have been eligible for support in the months after referrals ceased. Such an opportunity to isolate the impact of the programme has been unavailable to researchers throughout the programme's life and was not meaningful at the start of the programme as individuals prior to Work Programme were referred to other support in the form of

Flexible New Deal and the labour market was changing rapidly. The analysis seeks to identify and address sources of bias that remain within this approach.

This analysis utilises a discontinuity in programme availability for eligible individuals, when the programme closed for new referrals. However, the programme accepted referrals over a six-year period, from 2011 to 2017, a period over which the economic conditions changed markedly. The analysis estimates the impact for participants of Work Programme Payment Group 2, referred in the last two years of the programme when it was beginning to wind down, therefore caution must be applied in seeking to generalise to those referred earlier in the life of the programme.

The original intention was to compare referrals from the end of the Work Programme with claimants who would have been eligible, after the end of the Programme. However, the referral process meant that a sizeable proportion of eligible people were not referred, which introduces significant selection bias between groups that might have affected the outcomes: the results of this analysis are shown in Appendix H. Instead, the main results come from an Intention to Treat (ITT) analysis, where individuals that were eligible for the Work Programme before the end of March 2017 were compared with individuals that would have been eligible after March. As the treatment group is not completely composed of programme participants, the estimates are used to estimate the impact of participation alone.

This study will contribute to the understanding of contracted provision effectiveness more broadly and provide vital evidence to inform the direction of current and future employment and welfare strategies.

The plan for this report is as follows: Section 2 covers what the Work Programme was and how it was delivered followed by a brief summary of the literature on employment programmes including previous evaluation evidence gathered on this programme; Section 3 describes the analytical approach covering potential sources of bias, limitations and cohort selection, Section 4 explains the impacts findings and main sensitivity analyses, followed by a cost benefit analysis in Section 5. Section 6 finishes with conclusions.

2. Work Programme Policy and Design

2.1. Policy background

The Work Programme started accepting referrals in 2011 and replaced a number of previous welfare to work schemes. The programme was designed to support unemployed and economically inactive individuals into sustained work, during a period of high unemployment as a policy response to the 'Great Recession' in 2008-09. It replaced a number of predecessor initiatives (e.g. Pathways to Work and Flexible New Deal) and it operated throughout England, Scotland and Wales.

Through the Work Programme the Department aimed to encourage effectiveness and efficiency through incentives that went beyond the design of previous employment programmes. Much of the information that influenced the design came from Departmental experience and lessons from the full evaluations of earlier programmes that were applied to the design, either by DWP in the overall design or by providers to whom we highlighted the evidence base. Broadly there were four main features:

- Greater emphasis on outcomes through its payment by results (PbR) model where the majority of provider payments were linked to job outcomes. This is, to date, the largest delivery of UK public services via such a PbR model. Also differential payments were made for groups that were harder to help with the aim of discouraging providers from concentrating effort and resources on those participants for whom they could achieve an employment outcome most quickly or cheaply.
- Provision was 'black box' in the sense that significant freedom was given to providers to decide how best to design provision to deliver personalised and tailored interventions suited to the needs of their clients over a two-year period, with minimal specification and intervention from the department.
- Provision was designed in a way deliver to encourage competition between providers.
- The programme operated through larger, longer contracts with the aim that greater market stability would facilitate the development of provider capacity and expertise and encourage investment to support service delivery innovation.

In addition, as with some other previous programmes, participation was mainly mandatory, particularly where eligibility was by claim duration. In the following sections we describe the how these features were designed in greater detail.

2.2. Work Programme structure

The DWP let the Work Programme across 18 geographical Contract Package Areas (CPAs) across Great Britain in a competitive process. Each area consisted of two or three separate contracts, bringing the total number of contract arrangements to 40.

The contracts were won by 18 unique organisations, with several providers delivering the Work Programme in multiple CPAs. These 'Prime' providers committed to either deliver their services directly or managed them via a supply chain of contractors ('Sub-prime' providers) for five years, until March 2016. This was later extended for a further year, with the programme closing for new referrals in March 2017.

When individuals were referred to the Work Programme, they were randomly assigned to one of the two or three third-party providers operating in their CPA. At the start of the programme, providers in each CPA received an equal share of referrals. To incentivise provider performance, providers outperforming others operating in the same area received an increased proportion of referrals, in a process known as Market Share Shift. Some previous analysis has used this feature to compare the relative effectiveness of providers[1].

The Department also retained the right to withdraw these contracts throughout their life, should the minimum specification not be attained. It enforced this option once within six years of the programme when the original contract holder replaced following a secondary competitive exercise.

Once an individual was referred to the Work Programme, providers were expected to deliver services to them for two years, regardless of whether participants moved into employment or changed benefits. This is in contrast to other provision where a provider is only obliged to support an individual following a 'start' on the programme. Although starts did not feature in the Work Programme, an 'attachment' to the programme was the equivalent used to monitor participant engagement with providers. Across the life of the programme, 98% of referrals had a recorded attachment date.

The Work Programme was made up of a total ten Payment Groups, each with different entry criteria – as described in Appendix A. This analysis focuses on Payment Group 2, a group which included people aged 25 years and older who had been in receipt of either Jobseeker's Allowance or Universal Credit in the Searching for Work conditionality group for a period of 12 months.

Provision was mandatory for those in Payment Group 2 and for most of the other Payment Groups. This meant that people needed to participate in the programme once they had been referred and undertake certain activities, determined by the provider. Non-compliance might incur a benefit sanction leading to a withdrawal of benefit for increasing periods of time: two weeks for an initial sanction, followed by four weeks and then for 26 weeks.

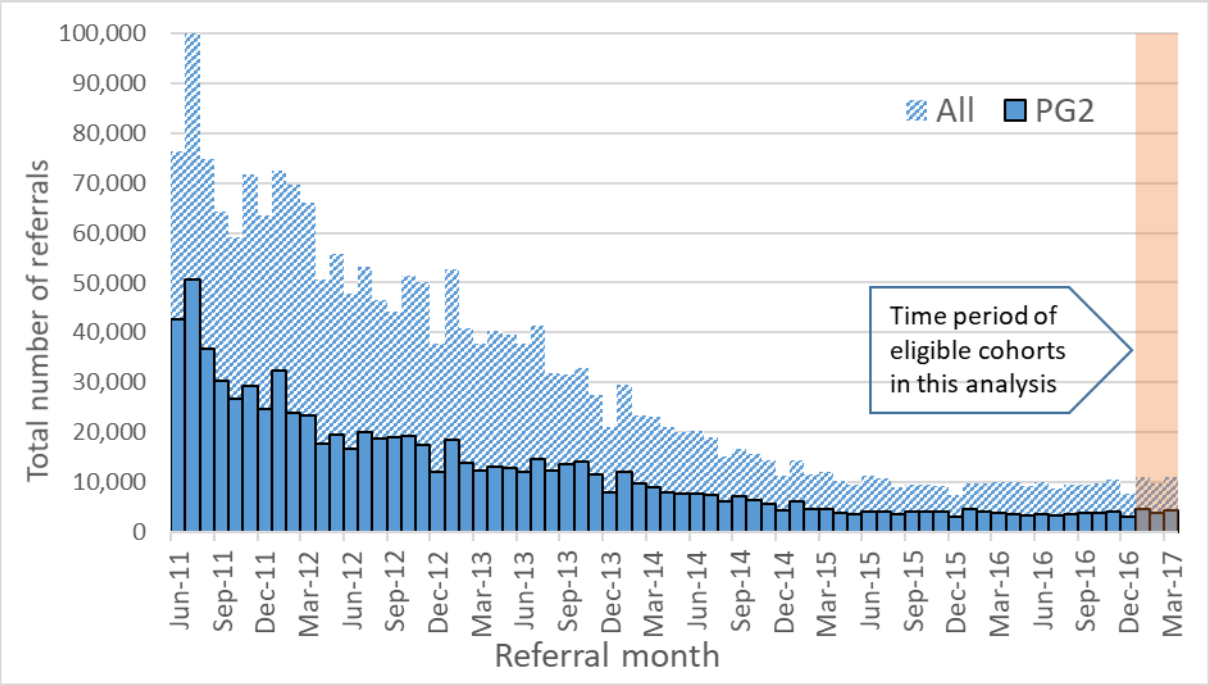
[1] https://www.gov.uk/government/publications/work-programme-factors-associated-with-differences-in-the-relative-effectiveness-of-prime-providers

Participants could end the programme early if the final outcome payment was claimed by the provider or they moved to another employment programme, such as Work Choice. Aside from participants exiting the programme early, the support period lasted two years. In certain cases, if participants entered employment towards the end of this two-year period, providers were incentivised by the payment model to provide in-work support beyond the two-year period.

2.3. Take up of Work Programme

Between June 2011 and March 2017, 1.95 million referrals to the Work Programme were made. Referrals were highest in the early months of the programme when unemployment levels were higher but reduced to around 10,000 each month from early 2015. Figure 2.1 illustrates this and indicates that Payment Group 2 (PG2) was largely responsible for this trend.

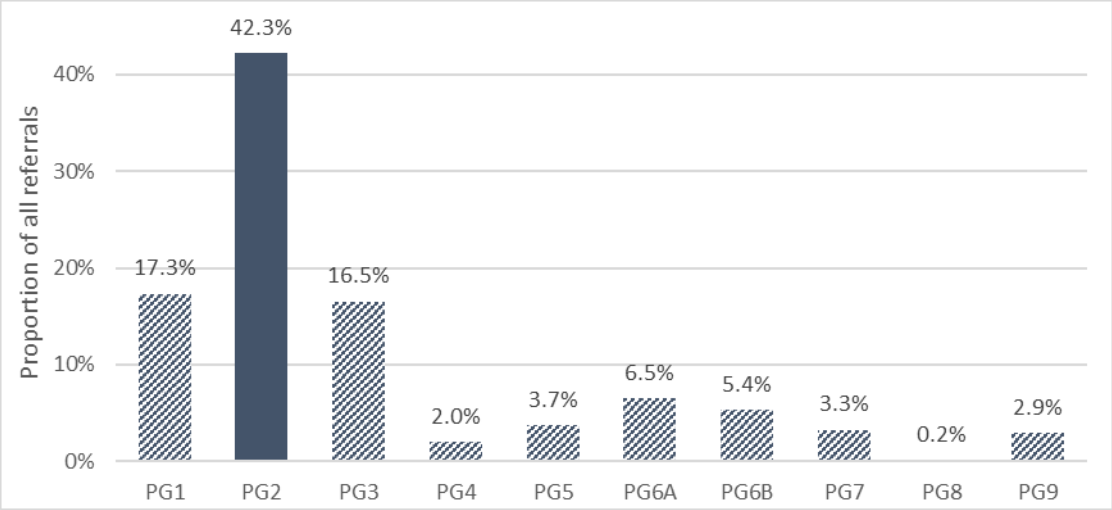
Figure 2.1 Work Programme referral volumes by month



Source: DWP Work Programme Management Information September 2019

Referrals were made across all of the payment groups with Payment Group 2 making up the largest proportion of referrals at 42%.

Figure 2.2 Work Programme referrals - Payment Group proportions



Source: DWP Work Programme Management Information September 2019

2.4. Payment model

The Work Programme saw an almost[2] exclusively Payment by Results (PbR) approach implemented. In the period under consideration for this analysis, all payments were made purely through a PbR approach.

The majority of funding came via outcome payments. Within the tender process, prospective providers were invited to set the prices they were willing to receive for two types of employment outcome:

1. A Job Outcome (JO) – a period of 13 or 26 weeks (Payment Group dependant) in employment and not in receipt of benefit; and
2. Sustainment Outcomes (SO) - multiple 4 week periods of consecutive employment (and not in receipt of benefit) after the JO.

A single Job Outcome could be achieved for each participant within the two years of 'allotted time' on the programme or within a continuous period after programme end, should the individual be in employment when the period of support ends. Multiple consecutive Sustainment Outcomes could be achieved for time in employment following that which merited a JO. The number of SOs available to the provider varied depending on the Payment Group of the individuals and were in place to encourage in-work support and sustainment of employment following core programme support. A pictorial description of how, and when, these outcomes could be achieved within the programme is presented in Appendix B.

A Minimum Performance Level (MPL) for Job Outcomes across the payment groups was agreed with providers. Strong performance against this was rewarded with bonus 'Incentive Payments'[3], and strong performance relative to the other

[2] In the early years of the programme relatively small 'attachment fees' were paid to providers for each referrals made but this ended in July 2014; these made up 20% of the total spend over the life of the programme.
[3] An additional £1,000 was paid for every outcome achieved in PGs 1,2 & 6a that exceeded 118% of MPL

provider(s) in the CPA saw a greater share of the CPA's participants referred to the provider, enabling the opportunity for further revenue through Market Share Shift.

Poorer performance resulted in a reduction of referrals if performance was above MPL but had the potential for contractual penalties including contract removal should performance fall below MPL.

Mid-programme contractual changes brought in an amendment to the MPL and additional Unclaimed Outcome payments being made in relation to sustained employment that was unable to be evidenced by the provider but was evidenced by the Department using administrative employment and benefit data. These were paid at 25% of the JO and SO value.

2.5. Work Programme commercial outcomes and contractual performance

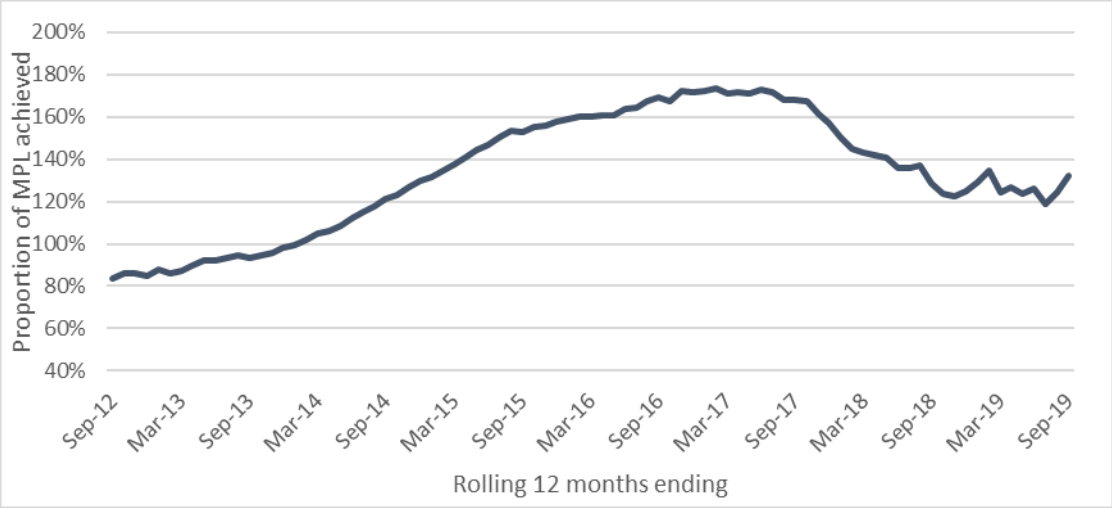
The payment model design and ultimately commercial management of the programme were both centred on a national performance expectation – the Minimum Performance Level (MPL)[4]. This acted as a lower bound for acceptable performance for each contract holder and was based upon a Departmental expectation of the number of individuals within any given cohort that would go on to achieve a Job Outcome.

Performance against this has been strong over the total life of the programme but performance was far below expectations in the early years of the programme as reported in the 2014 NAO report 'The Work Programme'[5]. This was, in part, due to unrealistic expectations of the speed in which employment spells would manifest as recognised performance within the programme and the MPL was adjusted in 2014 to recognise this. The number of outcomes expected remained unchanged but the speed with which they were expected was decreased.

Figure 2.3 below presents for the whole programme, the rolling 12 month performance against this revised metric. It shows that performance was still below MPL in the initial years, 87% in the programme's second year. Performance increased to a peak of 141% in year 4, April 2014 – March 2015, before falling back to 124% in the most recent year available, April 2018 – March 2019. As of September 2019, the programme had achieved 119% of the MPL.

[4] The MPL was derived following an estimation of the Non-Intervention Level (NIL) for each WP payment group. This level is indicative of the number of outcomes participants would receive should the investment via WP not be made. The MPL was set to be a 10% increase on the NIL, i.e. there was an expectation that WP providers would deliver 10% additional employment outcomes.
[5] NAO 'The Work Programme' (2014) https://www.nao.org.uk/wp-content/uploads/2014/07/The-work-programme.pdf

Figure 2.3 Rolling 12 month performance against Minimum Performance Level



Source: DWP Work Programme Management Information September 2019

The commercial outcomes for PG2 were a Job Outcome of six months in employment and off-benefit, and a series of up to 12 smaller Sustainment Outcomes for each four-week spell of continuous employment thereafter.

It is important to note that this presentation of performance is a measure of success against a commercial expectation; it is not in itself, a measure of the performance and effectiveness of the programme. To obtain an understanding of effectiveness, the current analysis compares the cumulative time in days spent in employment and off-benefit from an Intention to Treat group with an inferred counterfactual group in order to assess the impact of participation.

2.6. Brief Literature Review

Over the last few decades, significant public expenditure has been used across the world (e.g. OECD Database on Labour Market Programmes[6] on labour market programmes as a way to reduce numbers of unemployed by trying to improve peoples' labour market outcomes. There have been increasing number of evaluation and evidence overviews – for context of our current study we give a very brief indication of the literature.

Labour Market programmes are generally categorised e.g. Crepon *et al.* (2016) into three broad groups: job search assistance measures, training programs, and subsidized jobs. The study also explained that some individuals are disadvantaged in different ways, so are further away from the labour market and therefore other programmes are needed to overcome barriers – such as building self-esteem, support for health conditions, mentoring and support with re-entering society. These elements could be thought of as fitting across job search assistance and training, or as a separate category – however as Crepon highlights - some traditional approaches may be ill designed and may require more innovative approaches. The Work Programme was designed to be black box so rather than taking a single approach, providers were given the opportunity to decide which approach, or

[6] www.oecd.org/employment/database

combination of approaches, was appropriate on a case by case basis. Therefore, it is difficult to categorise as it contains a number of elements tailored towards a diversity of individuals – but the evaluation evidence, summarised in Appendix C, suggests that the provision consist mainly of job search, a smaller proportion of training but also interventions to help with confidence building and basic mentoring.

There have been an increasing number of large evidence syntheses – possibly the most recent and systematic has been from Card *et al.* (2018) who presented a meta-analysis of impact estimates from over 200 recent econometric evaluations of these programs from around the world. Their main four broad conclusions that we reproduce here are: (1) impacts are typically close to zero in the short run, but become more positive 2-3 years after completion of the program; (2) the time profile of impacts varies by type of program, with larger gains for programs that emphasize human capital accumulation; (3) there is systematic heterogeneity across participant groups, with larger impacts for females and participants who enter from long term unemployment; (4) active labour market programs are more likely to show positive impacts in a recession.

Other older syntheses include Greenberg *et al.* (2004) focussed on 64 US welfare to work programmes. The authors defined programmes emphasising jobsearch and de-emphasising job training as 'work first' and programmes not emphasising jobsearch as training programmes. They found that although the outcomes of training programmes did catch up to 'work first' programmes, the effect of both fell towards zero after 24 months and 'work first' dominated training. The authors point out however, in most instances the training provided only modest support lasting 'no longer than a month or two when it did occur'. And if the programmes directed towards training did not provide useful training then the difference between the two types of support may represent the difference between pressure to find employment and the absence of that pressure for a couple of months.

Evidence from Germany (Lechner *et al.* (2011)) explored the impacts from short (less than 6 months), long (2 years) and retraining programmes over the long term - tracking outcomes over an eight-year period. The evidence suggested that the returns from all training programmes can be considerable: approximately 10 percentage points for short or long training programs and approximately 20 percentage points for retraining. The positive effects take a while to materialise: within 6 months for the short and 12 months for the long training periods, and much later (24 months) for retraining programs.

When we turn to labour market programmes in the UK the programmes most closely related to the Work Programme (though there are many differences) are the New Deals (1998-2010) – specifically New Deal for Young people, mandatory at 6m, and New Deal 25+ for those aged 25+ and mandatory at 18m. Beale et al. (2008) showed that over a four-year period, participants spent, on average, 64 fewer days claiming out of work benefits than the comparison group – though the impact does decline. To give a wider context, some other employment programmes that have more differences that have been evaluated, include:

1. Work Experience (2011-): Haigh and Woods (2016) showed that participation in a 2-8 experience placement with an employer gave an average impact of 10 days off benefit, and 47 days in employment over two years, with no evidence of decline; at that point fiscal return was £1.50 for each £1 spent.
2. Sector-based work academies (2011-): Ward et al. (2016) also showed that in work training and offer of guaranteed interview gave an average impact of 29 days off benefit and 50 days in employment, also with no evidence of decline; at that point a fiscal return of around £1.20 for each £1 spent.
3. Mandatory Work Activity (2011-15): Prince et al. (2012) showed that up to a month of mandated work gave no evidence of impact.
4. Future Jobs Fund (2009-2011): Hillmore et al. (2012) showed that a 6 months paid job costing £6,500 generated a high benefit and employment impact (11 percentage points) after two years, that was not declining, giving a fiscal return of around £1, were it to continue for a further 4 years. Although the evaluation could not measure the substitution or displacement effects that may have been present.

As discussed in Section 2.1 some of the lessons drawn from the wider evaluations of these programmes and other DWP programmes were applied to the Work Programme design, the contractual changes mentioned in Section 2.4, and throughout the lifetime of the programme.

Finally, Appendix C summarises some of the previous evidence for the Work Programme, which we refer to later when we discuss the results in Section 4.5.

3. Analytical approach

3.1. Overview of the methodology used

This evaluation aims to estimate the impact of participation on the Work Programme for people aged 25 years and older who had been in receipt of either Jobseeker's Allowance or Universal Credit in the Searching for Work conditionality group for a period of 12 months. This group who were referred to the Work Programme was known as Payment Group 2. The main focus of the assessment is by comparing the subsequent labour market outcomes on those who were eligible for the Work Programme between January 2017 and March 2017 with those who would have been eligible between April 2017 and June 2017 had the Work Programme not ended. Then by taking into account the proportion of participants, these estimates are then used to develop an estimate of the impact of participation.

Impacts are defined as 'the outcomes that occurred under the Work Programme' minus 'the outcomes that would have occurred anyway'. The first part of this equation is straightforward to calculate, as the outcomes of those who participated in the programme can be directly observed. A comprehensive view of paid outcomes are described by the National Statistics[7]; in this study they are measured from data on benefit receipt held by the Department for Work and Pensions (DWP) and employment information based on earnings recorded through HMRC's Real Time Information (RTI). However, the second figure, commonly referred to as the 'counterfactual', is more difficult to determine because it is impossible to know what the outcomes would have been for the participant group in the absence of the programme.

However, until now there has not been a published quantitative impact assessment of the Work Programme. This was because the programme was rolled out simultaneously across all areas and groups, and because cohorts of claimants reaching the 12-month threshold, before the Work Programme was introduced, were eligible for support from predecessor programmes, giving no opportunity to define a comparison group – either by randomly allocating some claimants to a control group, or identifying a comparable group not eligible for the programme. For the group analysed here, Payment Group 2, participation was mandatory, so there were no contemporaneous non-participants from whom a comparison group could be drawn[8].

The closure of the programme to new referrals offered an opportunity for a quasi-experimental assessment of impact through the discontinuity it created. During the following year, there is no evidence that there were any differences in normal Jobcentre Plus support and there was no similar alternative large scale provision in place at that time: the most similar successor programme, the Work and Health Programme, became available in April 2018 for low numbers of people who had

[7] Work Programme National Statistics https://www.gov.uk/government/collections/work-programme-statistics--2

[8] It may be possible to identify comparable groups for other Payment Groups, and this is being explored separately to this analysis

been in receipt of either Jobseeker's Allowance or Universal Credit in the Searching for Work conditionality group for a period of 24 months[9].

The difference between the actual observed outcomes and the estimated counterfactual outcomes gives the true impact of the programme if the only relevant difference between the treatment and comparison group, that affects the outcomes, is that the participant group took part in the Work Programme.

The original intention was to compare groups of participants and non-participants, however this led to some concerns of selection bias in the participant group. After investigations, it was decided that an Intention to Treat (ITT) approach would be more appropriate and select groups based on their eligibility to the programme. The reasons for this are explained in Section 3.2.1. This leaves two remaining potential sources of bias that are discussed in Sections 3.2.2 and 3.2.3: time trends (from drawing cohorts from consecutive time periods) and compositional differences respectively.

Addressing these sources of bias tackle the internal validity of this study. The end result of applying this methodology is a quantitative estimate of the impact of the intervention on each outcome of interest. Additional limitations and challenges to the external validity are discussed in Section 3.3.

The analysis focussed on comparing consecutive three monthly cohorts - this was chosen as the best trade-off between choosing a sufficiently large cohort against a cohort that did not extend too far in time away from the end of the programme. Shorter cohorts would mean smaller numbers and lower proportions of participants; longer cohorts might be more confused by time effects and compromise the plausibility of inferring the counterfactual. Nonetheless we note that as part of sensitivity analysis in Section 4.4 results were very similar to those using monthly cohorts before and after the end of the programme, a three month cohort spanning October to December 2016 compared with April to June 2017, and also for a longer 6 month cohort.

The sources for the outcome measures are spells of out-of-work benefits as recorded in DWP administrative datasets and evidence of employment through the HMRC Real Time Information (RTI) feed. Section D.4 has more detail of the HMRC RTI feed. Since for all individuals, whether participating or not, full information is available on their benefit claims and on their earnings from employment, these impact estimates can be expressed in a number of different ways. For example, in terms of the average additional days each participant spent in work or off benefit as a result of participating in the Work Programme; or alternatively, as a change in the likelihood of participants being in work or on benefit at a specific point following the intervention. This study has focused on the duration of employment and has not considered earnings as an outcome, however this is an area of exploration for future studies.

[9] Work Choice was available for some individuals with health barriers during all of the period of the Work Programme up until November 2017 - when the Work and Health Programme was available for a similar group.

3.2. Addressing potential sources of bias and internal validity

The success of the evaluation is dependent on identifying and addressing potential sources of bias that might affect the outcomes not due to participation in the Work Programme satisfactorily.

3.2.1. Reasons for adopting an Intention to Treat (ITT) approach

The original intention was to compare the outcomes of two cohorts of individuals: one group who were referred to the programme between January 2017 and March 2017 and a second group of individuals who would have met the eligibility conditions between April 2017 and June 2017, but because the referrals were closed they were not referred to the programme. The outcomes of this second group of non-participants would then allow an estimate of the counterfactual to be made, i.e. outcomes that would have occurred for the first cohort in the absence of the programme.

However, while the comparison group includes all those who would have been eligible, not all those who became eligible were ultimately referred to the programme to became part of the participant group and many were referred later than the point of eligibility. Referrals to the programme were not made automatically when an individual satisfied all the eligibility criteria. Instead, the referral process was manual and initiated by Work Coaches, so referrals did not occur immediately on the date an individual became eligible for the programme, or in many cases the referral never happened. For example, Figure E.1 in Appendix E shows that 45% of individuals meeting the eligibility criteria between October 2016 and December 2016 were not referred to the programme. Furthermore, over 75% of referrals made to Payment Group 2 occurred within a window from between 30 days before, to 90 days after the individual became eligible for the programme. There could have been many reasons for this, such as if a referral to the programme was deferred in the event the individual was taking part in some other work-related activity.

Therefore, there is a risk that the participant group could be defined partly on the basis of unobserved characteristics which the non-participant group is not. So it is likely that this might introduce some selection bias into the participant group, which does not exist in the non-participant group. The analysis was performed - matching was carried out on observable characteristics (as described in section 3.2) to try to eliminate some bias, analysis of these groups is described in Appendix H and the results are discussed in Section 4.5 alongside the main ITT results. The point though is that even if the participant/non-participant group compares well on observable characteristics after matching, there is no way of knowing how well they would match on unobserved characteristics. However, the fact that even after matching there are some differences in the observable characteristics suggests that some bias remains related to some unobserved factors (in contrast, the ITT groups that we describe next are almost identical looking after matching).

Therefore, the main focus of the analysis is on a set of cohorts, defined as individuals meeting the eligibility criteria for the programme regardless of the referral status, i.e. on an Intention to Treat (ITT) basis. These ITT cohorts were selected from the same time periods of the participant and non-participant groups, i.e. individuals meeting the eligibility criteria between January 2017 and March 2017 and a second group meeting the criteria between April 2017 and June 2017. Since the

groups are selected on an ITT basis, the January 2017 to March 2017 groups contain eligible non-participants alongside participants, which dilutes the level of participation seen in the programme and dilute any impact that there might be.

Although the ITT groups avoid the selection biases, they are still selected from distinct time periods. Therefore, it is still crucial to understand time effects, such as labour market changes, changes to the benefit system and seasonal effects, as outlined in the next section.

3.2.2. Time trend and seasonal changes between cohorts

The main challenge for all the analysis stems from the need to draw two groups from different time periods. The groups are drawn from consecutive time periods either side of the referral discontinuity to minimise any differences in circumstances, e.g. labour market changes, changes to the benefit system and seasonal effects, but it cannot be assumed that the changes are negligible without investigation.

Extra cohorts are assessed in an attempt to quantify the scale of any effects across similar time periods, although this does not guarantee that the effects are the same as the main period of interest:

- To assess the effect of seasonality and other time effects from across two consecutive quarters from a year earlier, cohorts are selected of eligible individuals between January 2016 to March 2016 and between April 2016 to June 2016.
- To assess the change observed between two consecutive quarters after referrals to the programme had closed, cohorts of individuals are selected that would have been eligible for the Work Programme, had it continued, between April 2017 to June 2017 and between July 2017 to September 2017
- To further assess the change observed between the two consecutive quarters after referrals to the programme had closed, cohorts of eligible individuals from a year earlier are selected between April 2016 to June 2016 and between July 2016 to September 2016.

Alongside this, descriptive analysis of the labour market has been collated to identify whether any trends might be considered and should be borne in mind: full details can be found in Appendix G.1. In summary, the time period of interest was a time when labour market conditions were good, with gradually improving employment and unemployment rates. The expectation therefore is that the differences between the groups would not be sizeable, given labour market improvements were gradual and small. And given that the comparison group may have faced a slightly more favourable labour market, theoretically we might expect this to lead to an underestimation of the impact, all else being equal.

As well as possible changes to the labour market, the benefit system itself changed over the period of this study with the introduction of Universal Credit (UC) which we describe briefly here.

The rollout of UC started in 2013 and was staggered across geographical areas, with UC available in all Jobcentre Plus (JCP) offices by December 2018. The majority of the rollout took place over the period of this study, and was accelerated in 2018. In January 2017, less than 10% of JCP offices were offering UC to claimants and by December 2017 this had increased to around one third of all JCP offices. Over 50% of JCPs were offering UC by June 2018, and UC was available in all JCPs by the end of December 2018[10].

UC was designed to replace six legacy benefits; however, the two systems do not map directly onto one another. Furthermore, the first phases of UC rollout did not offer functionality to replicate the six intended legacy benefits from the start, and this rollout was also phased. For this analysis, the two systems have been broadly mapped together and are considered in combination, to allow for individuals transitioning at different times. It is not possible to completely reproduce where people would have been if only one system was available - it can only be approximated. Therefore, some caution should be exercised when comparing benefit outcomes over two different time periods.

Additionally, by design, UC allows claimants to claim benefit whilst in employment, tapering according to earning amounts. Outcome states are considered that retain mutually exclusive benefit and employment states and also allow for an overlap of these two states.

Overall, because UC is not the same as JSA and roll out was happening quickly at the time of cohorts of interest then these interactions do potentially add some extra complication to the analysis, which we will describe alongside the findings in Section 4.3.

3.2.3. Compositional differences between cohorts

A further possible source of bias in the impact estimation would be if the characteristics of claimants in the two periods were different. A statistical technique known as propensity score matching (PSM) was used to attempt to address this by selecting a comparison group of individuals from the later period which matched the treatment cohort on all relevant characteristics. A full description of propensity score matching and how it was implemented can be found in Appendix D.

A key consideration of PSM is the ‘Conditional Independence Assumption’. That is that all relevant differences between the treatment and comparison groups are controlled for, and that the groups are the same in all regards except for their participation in the programme. The theory then is that none of the difference in outcomes between the groups can be attributed to differences in the make-up of the two groups. However, this matching can only be done on observed characteristics and there is no guarantee that all unobserved characteristics are captured.

Propensity Scores are determined by a logistic regression model where the inputs are a selection of characteristic variables and benefit and employment history. The

[10] https://www.gov.uk/government/publications/universal-credit-transition-to-full-service/universal-credit-transition-rollout-schedule-march-2018-to-december-2018#annex-a--sites-currently-live-with-universal-credit-full-service-november-2015--february-2018

sensitivity of the matching carried out is tested by creating several models using various permutations of characteristic, benefit and employment information. A well-scored model should show few differences between the groups prior to the point of an intervention, i.e. for the main analysis this is the point of eligibility for the programme in the groups. This indicates that observational differences have been accounted for and the constructed comparison group is characteristically similar to the treatment group.

The observed characteristics employed for the matching include age, gender, geographical location, receipt of benefits and time in employment. Appendix D details the data and sources used, along with further information on the observed characteristics. Appendix F shows how similar the groups were on these observed characteristics.

It is also possible that the prospect of referral to mandatory provision, such as Work Programme Payment Group 2, could change the way customers interact with the labour market or other work-related initiatives prior to referral. It could be that potential participants may make extra efforts to find employment in anticipation of the Work Programme to avoid mandated referral, or delay such efforts in order to secure the support that the provision might provide. Once the programme has stopped receiving referrals this incentive no longer exists, which could change the composition of the comparison group.

In order to investigate whether there are any of these anticipation effects, further labour market analysis in Appendix G.2 attempts to identify any changes in benefit off-flows, where a change in labour market interactions in a period prior to programme referral would reasonably be expected to manifest as either an increase or decrease in benefit off-flows. This analysis showed some evidence of an uplift in off-flow rates at around 6 months after claim start, which is likely to be attributable to the automatic cessation of Contribution Based benefit claims. However, there was no observable evidence of a change in off-flows prior to benefit claims reaching 12 months in duration for cohorts of claims during the period of interest.

3.3. Limitations of this methodology and external validity

Referrals were open to the programme from June 2011 to March 2017 and there were significant economic changes during this period, particularly with the recovery from a global recession. Therefore, an ideal evaluation would consider a programme in a steady state. However, the discontinuity being used for this analysis only occurs at the end of Work Programme.

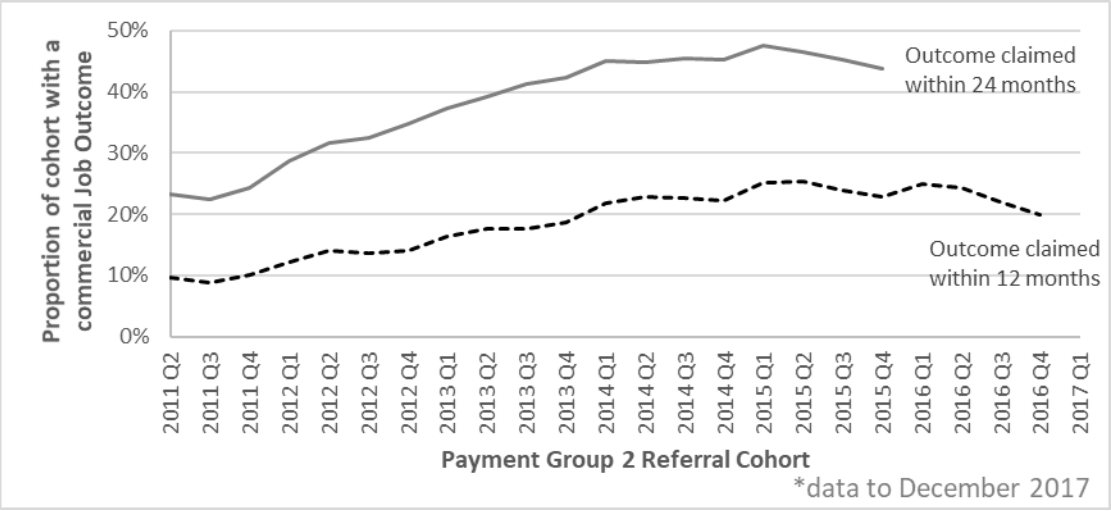
Individuals referred earlier on in the programme are unlikely to be represented by an inferred counterfactual drawn from a comparison group after the end of programme as they are likely to be individuals in different circumstances, in a different labour market, resulting in estimated impacts that would include effects other than impacts of the Work Programme. In addition, across the life of the Work Programme, different levels of referrals and volumes of participants were observed, as described in Section 2.2 and specifically in Figure 2.1. Referrals over the last 3 months of the programme, which are considered in the following analysis, accounted for 1.5% of total programme referrals to Payment Group 2. Therefore, the time period for which

the results of this study can be reasonably generalised is uncertain, as it is not possible to derive impacts for the group relating to early years of the programme.

As well as changes in labour market conditions over the six-year period of referrals to the programme, as the economy recovered from a global recession there will have been changes in programme delivery. It is plausible that the programme became more effective as it matured and providers gained experience; but it is also possible that as the programme was winding down it became less effective.

Figure 3.1 shows that the Commercial performance in terms of paid job outcomes of Payment Group 2 in the Work Programme showed steady improvement over the early years of the programme, reaching a peak around year 4 before starting to tail off in later years as the programme wound down. This evidence, alongside anecdotal evidence from DWP performance managers that the programme started to wind down from 2017, suggests that any estimate of impact from the end of the programme might produce a lower estimate than the impact when the programme was operating in a steady state. However, since impacts from earlier on in the programme cannot be estimated there is no way of knowing that.

Figure 3.1 Commercial performance of Work Programme Payment Group 2 and payment of job outcomes



*Source: DWP Work Programme Management Information September 2019

The Work Programme consisted of ten distinct participant types, organised in the form of Payment Groups. Since these groups were distinct, any impacts estimated for Payment Group 2 cannot also be reasonably generalised to another group.

In this analysis, detailed information on benefit and employment history has been used to match together cohorts, which has been shown[11] to be an effective proxy for fixed or historic unobserved attributes such as general motivation, though not contemporaneous unobserved factors e.g. getting sick. However, it is impossible to prove that the available data is sufficient to account for all the relevant variation between the participant and comparison groups to meet the conditional independence assumption.

[11] Caliendo, Mahlstedt and Mitnik, 2014

The selected characteristics are ones for which DWP has access to data for and which have been found in previous evaluations to be important controlling variables. Evaluation of the success of the matching between groups only demonstrates that they match on observed characteristics. There is no way of testing that the matching has controlled for unobserved relevant differences between the two populations. However, this analysis considers a group where participation was mandatory and so considers entire cohorts. Therefore, the matching is attempting to control for any differences between these whole cohorts. Had this study focused on a group where participation was voluntary, a greater emphasis would need to be placed on controlling for characteristic differences between groups.

Previous evaluations of programmes have used data to estimate spells of employment provided to HMRC through P45 and P46 returns. Since 2013, HMRC has collected Real Time Information (RTI) as a timelier reporting mechanism of employment spells. As well as improvements to employment start and end dates, RTI also captures the amount paid to an individual so can be used to determine earnings for a given spell of employment. However, to date no analysis has been performed on earnings as an outcome and neither the P45 and P46 data nor RTI capture information on self-employment.

3.4. Cohort selection

As explained in Section 3.2, in recognition of possible selection biases arising from not being able to replicate selection of the non-participant group in the same way as the participant group, the main analysis is based on drawing groups selected on an Intention to Treat (ITT) basis. That is individuals meeting the eligibility criteria in the specified time periods regardless of their referral status to the programme. For Work Programme Payment Group 2, these eligibility criteria are:

1. The individual is aged 25 years or older;
2. The individual has been claiming any combination of Jobseeker's Allowance or Universal Credit with Searching for Work conditionality for the last 12 months, or has linked claims totalling 12 months; and
3. The individual has not previously participated in the Work Programme in any Payment Group.

Appendix E explains the selection of these ITT groups in further detail and other potential cohorts selected on the same basis, in addition to information about the proportions of each cohort which were referred to the Work Programme.

Although the groups selected on an ITT basis remove some of the selection effects described above, the analysis leads to an estimate of the impact on the whole cohort who became eligible between January 2017 and March 2017 which means that it will understate the impact on those who actually participated in the programme. This is a material issue, since only around 38% of the ITT treatment cohort were Work Programme participants. As explained in Section 3.2.1, there was a gap between an individual becoming eligible for the programme and a referral actually occurring – for 75% of referrals the referral occurred within 3 months of the individual becoming

eligible. Therefore, it is likely that referrals to the programme did not proceed in the time between an individual becoming eligible and when the referral would have been made, particularly for the later months.

Section 4.4 shows three other ITT treatment groups that were used to explore sensitivity – shorter cohorts from March 2017 and April 2017, a treatment group for those who became eligible between October 2016 and December 2016 compared with April 2017 to June 2017 and a six-month treatment group of those who become eligible between October 2016 and March 2016 compared with April 2017 and September 2017. The short cohorts have a short time differential between the groups but a low proportion of programme participants. The latter treatment groups have higher proportions of programme participants (Oct 2016 - Dec 2016 is 55% programme participants), but introduce a greater time differential between itself and the ITT comparison group of individuals becoming eligible. Other groups with higher levels of participation might also have been selected but time effects from cohorts from further apart in time would have introduced greater uncertainty.

Finally, it is noted that a small proportion, around 1.5%, of the ITT comparison group, were also Work Programme participants. This is because they were referred to the programme earlier than the eligibility criteria determined, which again is likely for operational reasons.

4. Impact assessment findings

4.1. Outcomes of propensity score matching

This initial section shows the results of the propensity score matching to determine the extent to which the groups for comparison can be matched well on observable characteristics and whose outcomes provide a valid estimate of the counterfactual outcomes for the relevant treatment group. The main focus is on the ITT analysis but results from the participant and non-participant groups are also mentioned briefly.

The first aspect examined is the extent to which the comparison sample provided common support for individuals in the treatment cohort. For an individual in the treatment group to be 'on support', there must be at least one individual in the comparison group with a propensity score within the matching bandwidth of the treated individual's score, i.e. individuals are sought from the comparison group who have similar characteristics to individuals in the treatment group. See Appendix D for more detail on the matching approach.

For propensity score matching to be successful, it is important that as many as possible of the treated cohort are on support following the matching, since the impact estimates generated by comparing the matched groups will only be valid for those treated for whom common support is available.

Table 4.1 below shows the number of individuals in the treatment and comparison groups before and after matching, and the resulting proportion of the treatment group for whom common support was found. The results show that the proportion of the treatment group on support was 99.9%, indicating that this aspect of the matching has been successful – and this was slightly more than for the participant/non-participant group analysis, as shown in Table F.3.

Observing the propensity scores for the ITT treatment and comparison groups, indicates that in addition to 99.9% of the treatment group being on support, there does not appear to be an over-reliance on a small number of comparators.

Figure 4.1 Propensity scores for ITT treatment and comparison groups

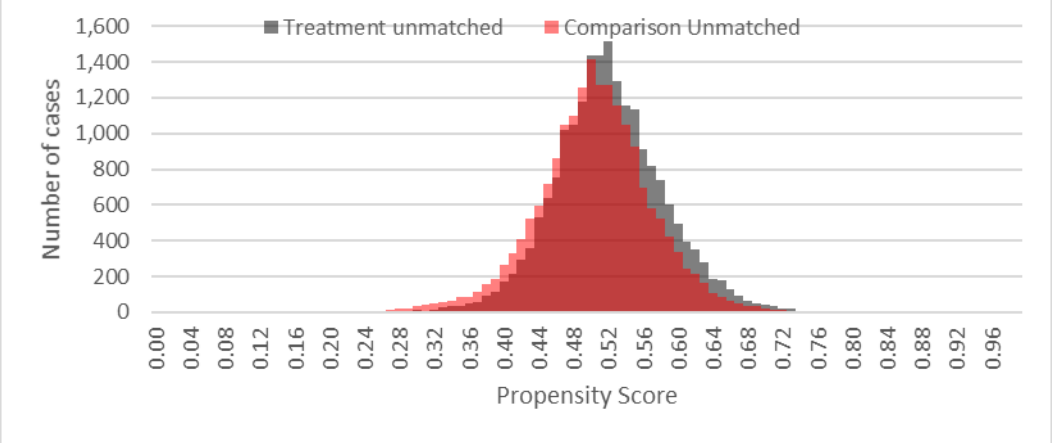
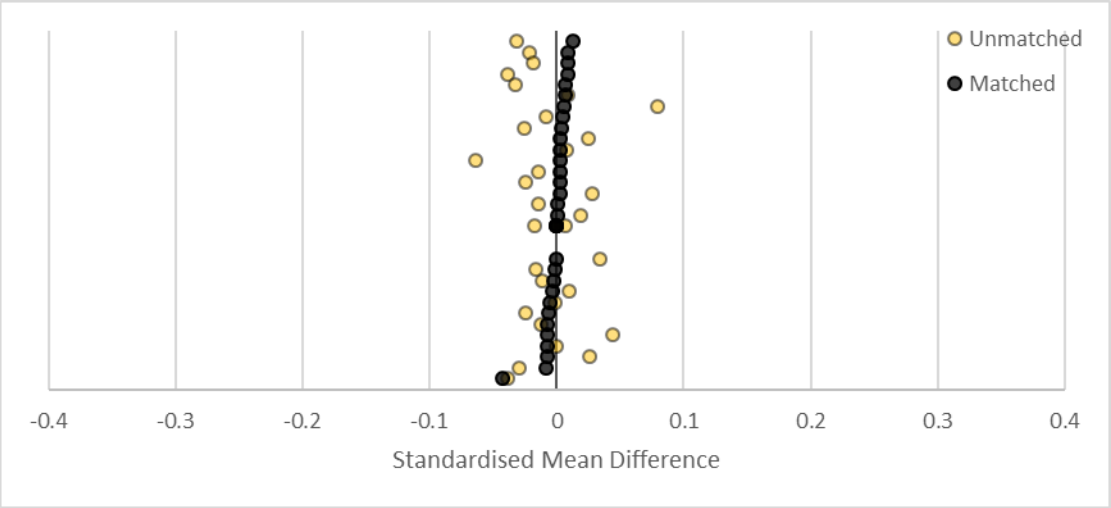


Table 4.1 Intention to Treat group sizes before and after matching and the proportion of participants on support

Cohort name	Size before matching	Size after matching	Lost through matching	Proportion on support
ITT Treatment group	20,129	20,113	16	99.9%
ITT Comparison group	18,785	20,113	N/A	N/A

Figure 4.2 shows the Standardised Mean Differences (SMD) to describe differences between the treatment and comparison cohorts before and after matching - further details for each variable are in Appendix F.1. These are calculated for each characteristic as the difference between the means of the two groups, divided by the pooled standard deviation of the two groups. SMD scores of 0 indicate no difference between the groups for a particular characteristic. A commonly used boundary[12] to indicate a difference between the two groups of some significance is an SMD of ±0.2.

Figure 4.2 Standardised Mean Differences between characteristics for Intention to Treat cohorts



Overall, the results show that before matching the differences between ITT treatment and comparison groups were quite small: all observed characteristics have SMDs between -0.1 and 0.1. After matching the differences are even smaller: all variables after matching show a SMD of less than 0.015 with only one, (mean time spent with a claim to either Jobseeker's Allowance (JSA) or Universal Credit (UC) with Searching for Work conditionality in the 52 weeks prior to programme eligibility) that has a greater difference, at 0.045.

In addition, statistical significance testing has been carried out between these variables. Before matching there were only four variables with statistically significant differences out of 23 variables at the 5% level between the treatment group and the

[12] Cohen J., 1977

matched comparison group - as seen in Table F.1 and denoted by an asterisk (*). After matching there were no statistically differences across these observed characteristics.

4.2. Comparison of groups and measuring impact

For each of the cohorts considered, the proportions of each group in various outcome states have been monitored. Three primary mutually exclusive outcome states have been defined as:

- Claiming an Out of Work (OW) benefit
- In employment and not claiming an OW benefit
- Neither claiming an OW benefit nor in employment

For the purposes of this study, an OW benefit is one of:

1. Jobseeker's Allowance (JSA)
2. Income Support (IS)
3. Employment and Support Allowance (ESA)
4. Incapacity Benefit (IB)
5. Universal Credit (UC): Searching for Work conditionality
6. Universal Credit (UC): Working with Requirements conditionality
7. Universal Credit (UC): Preparing for Work conditionality
8. Universal Credit (UC): Planning for Work conditionality

These groups specifically exclude Universal Credit conditionality groups "Working – no requirements" and "No work requirements" to match as closely as possible to the legacy benefit system. Figure D.1 illustrates how the two systems have been mapped together.

The state 'in employment' refers to having an employment spell as submitted by an employer to HMRC through Real Time Information (RTI).

The state 'Neither claiming an OW benefit nor in employment' occurs when no OW benefit or RTI employment spells have been identified for a given week. Two particular types of individual who may make up a substantial proportion of this group are: self-employed individuals and people in full-time education. These individuals would be counted in this outcome state, unless they also had earnings as an employee. Self-employment earnings are not captured by RTI.

The three primary outcome states are mutually exclusive in that each individual can only be in one state in any given week, but individuals can move between states from one week to the next. To support the analysis of these three primary states, additional states have been defined and monitored. These additional states are not mutually exclusive and overlaps may exist between them. These additional states are as follows:

- Claiming either JSA or UC with Searching for Work conditionality

- Claiming the remaining OW benefits: ESA, IS, IB and UC with Working with Requirements, Preparing for Work, or Planning for Work conditionality
- Claiming OW benefit **and** in employment
- Claiming Disability Living Allowance (DLA) or Personal Independence Payment (PIP)

Outcome states 4 and 5 indicate the split between individuals actively searching for work and those who are out of work for other reasons such as health conditions or parental responsibilities. Although some minor overlaps can exist between these two states, they are an effective disaggregation of the key outcome state of claiming an OW benefit.

The design of UC explicitly allows claimants to engage in employment activity whilst still being in receipt of the benefit. As the rollout of UC was phased, the Work Programme was largely operating either prior to UC rollout or whilst UC was in its infancy so the overlap between receipt of OW benefits and employment is expected to be relatively low. Outcome state 6 allows this to be observed and accounted for. Finally, outcome state 7 which refers to spells of DLA or PIP are not mutually exclusive with OW benefit claims or employment, as recipients can also be in or out of work or claiming an OW benefit.

The proportions of each group in each outcome state were monitored for 104 weeks prior to reaching eligibility to the programme and for 104 weeks following that point. The point of eligibility is shown as week 0 on all outcome and impact charts. In figures throughout, grey lines indicate unmatched groups for reference. Impact is defined as the percentage point difference between the proportions of each group in an outcome state at any given time. For the weeks prior to the point at which eligibility starts, the difference between compared groups should be small, since the propensity score matching is expected to ensure that benefit and employment histories of the groups are similar. After this point the groups will diverge, if the Work Programme has an impact for the outcome state in question. If the Work Programme is successful the expected differences would be positive for employment outcomes, and negative for the proportion on OW benefits

All graphs show the 95% confidence intervals to show the uncertainty around the impact estimate in each week. This uncertainty reflects the size of the samples used to produce the impact estimate.

The impact estimate can also be aggregated to provide estimates of the impact of the programme on the overall time an individual spent in any given outcome state following their eligibility to the programme. The additional number of days that participants spent in a given outcome state in a given week is calculated by multiplying the percentage-point impact for that week by seven, i.e. the number of days in a week. Summing across all weeks in the tracking period gives an estimate of the total number of additional days spent by participants in the given outcome state.

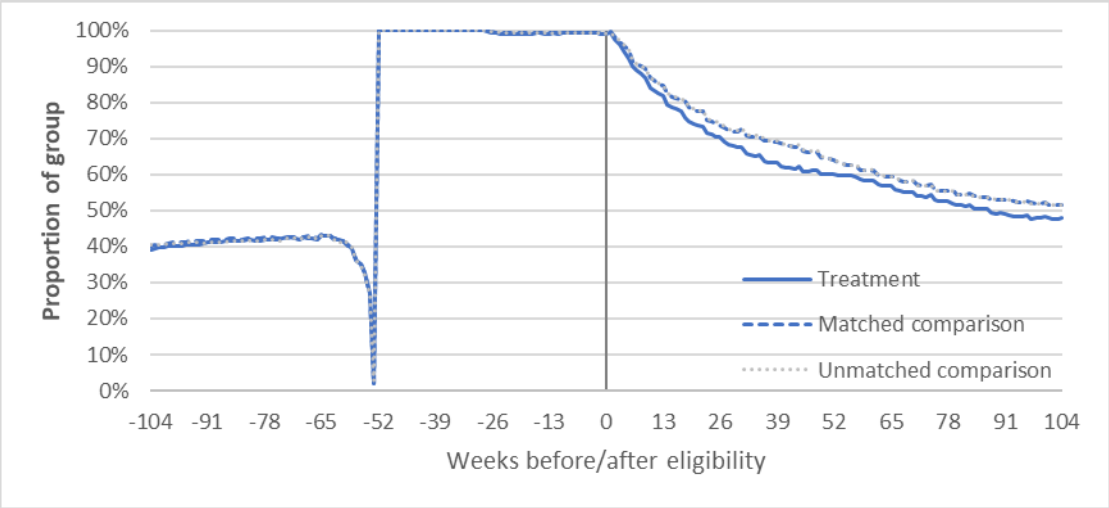
4.3. Findings from Intention to Treat analysis

The Intention to Treat groups for treatment and comparison are assessed across the three primary outcome states defined in section 4.2. The three primary outcome states are monitored and impacts estimated, followed by the calculation of additional days spent by the treatment group in each outcome state. These results are then followed by the four additional outcome states, including impact estimates, and then by the calculation of additional days in each state. Finally, engagement with other provision is tracked.

4.3.1. Primary outcome states

The following graphs show the proportions of each cohort in each of the three primary mutually exclusive outcome states. These are followed by the estimated impacts which are calculated by subtracting the proportion of the matched ITT comparison group from the proportion of the ITT treatment group for each outcome state for each of the 208 monitored weeks. It should be noted that in order to see the impacts clearly, the following charts are presented on slightly different scales and care should be taken when comparing any given pair of charts.

Figure 4.3 Proportion of ITT cohorts in receipt of an OW benefit



Since the ITT groups have been defined by strictly applying the Payment Group 2 eligibility criteria, the unusual, but expected, pattern of proportions in receipt of an OW benefit prior to eligibility is seen in Figure 4.3. Individuals become eligible once they have been in receipt of Jobseeker's Allowance (JSA) or Universal Credit (UC) with Searching for Work conditionality for 12 months, or 52 weeks. This means that the proportion in receipt of one of these benefits at 53 weeks prior to eligibility is almost zero, followed by the proportion in receipt of one of these benefits at 52 weeks prior to eligibility being 100%. Linking rules[13] have been taken into account when determining eligibility, which means that these proportions are not exactly zero and 100%, respectively.

[13] for the purposes of eligibility claims are considered continuous if there are gaps of less than 1 month between them

Figure 4.4 Impact on the likelihood of being in receipt of an OW benefit

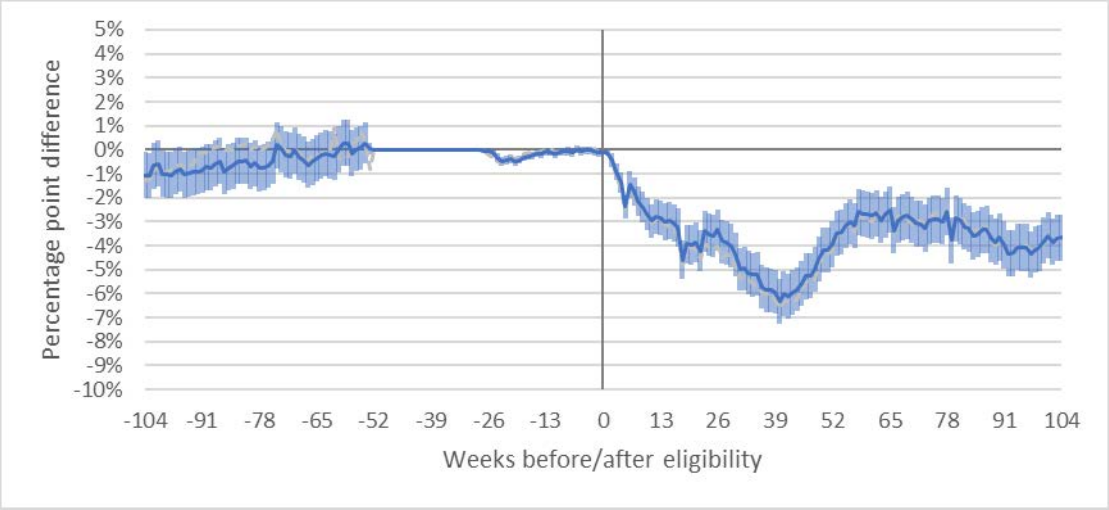


Figure 4.5 Proportion of ITT cohorts in employment and not claiming OW benefit

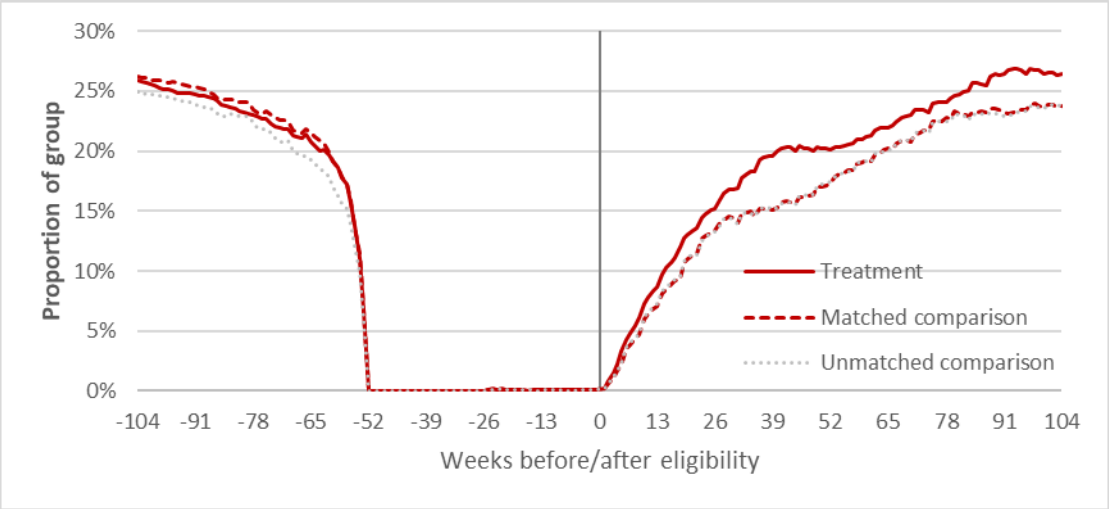


Figure 4.6 Impact on the likelihood of being in employment and not claiming OW benefit

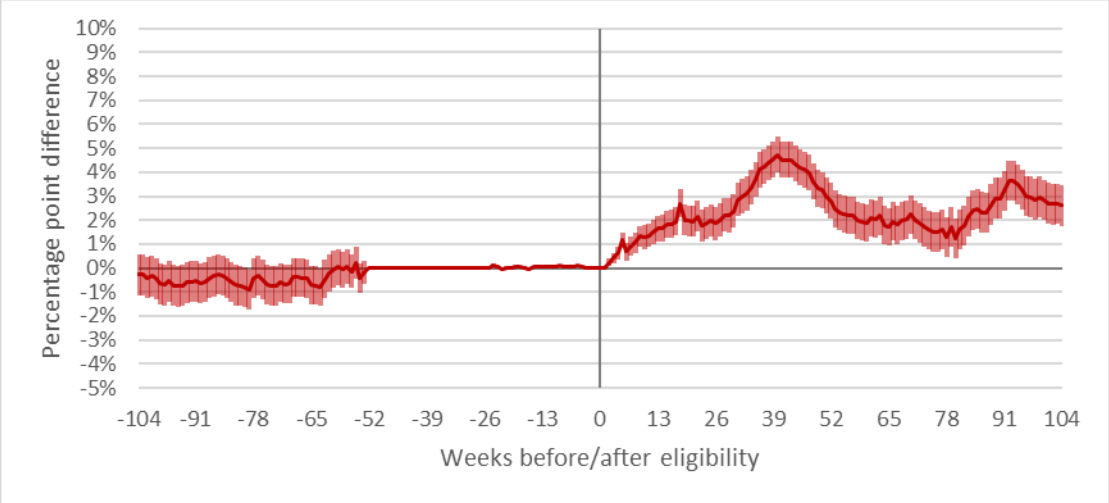


Figure 4.7 Proportion of ITT cohorts neither claiming OW benefit nor in employment

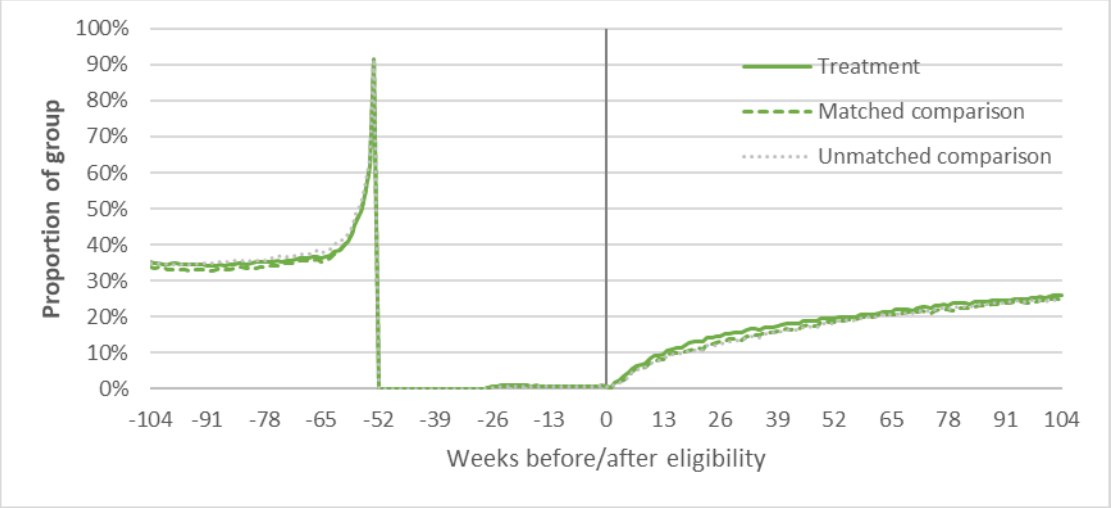
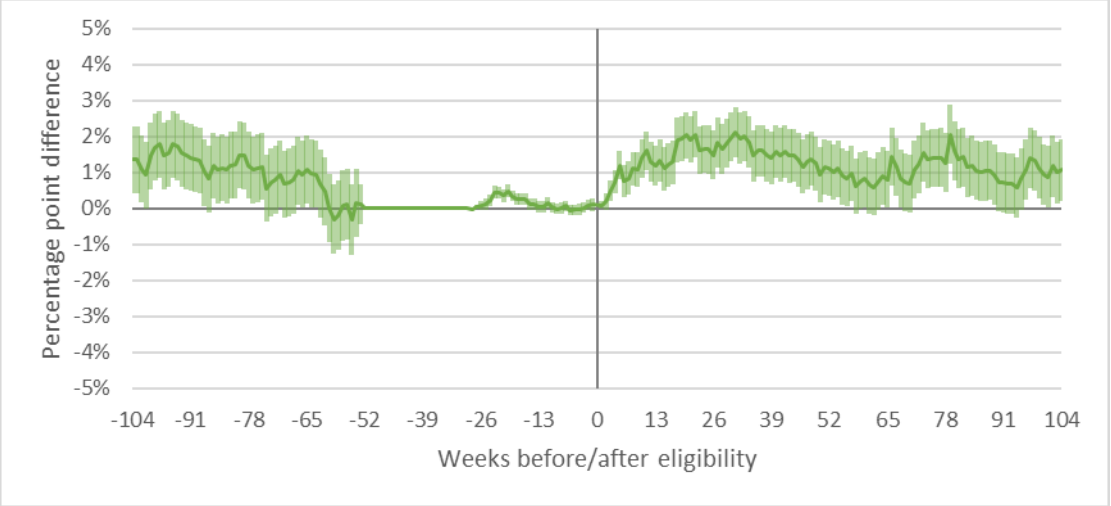


Figure 4.8 Impact on the likelihood of neither claiming OW benefit nor in employment



The series of charts, in Figure 4.3 to Figure 4.8, suggest that prior to the eligibility date, the ITT groups were very evenly matched across outcome types, with only some small significant differences observed for the neither claiming OW benefit nor in employment in weeks 104 to 78 prior to eligibility. This increases confidence that the matching has been effective, particularly for time spent on OW benefits and time spent in employment.

Overall, the results show that the ITT treatment group spent more time in employment and less time in receipt of OW benefits than the matched ITT comparison group.

We note that earlier in Section 3.2 we explained that because UC is not the same as the legacy benefits (Intensive Work Search Conditionality Group is not exactly the same as JSA and our definition of out-of-work benefits will both under-capture some outcomes or over-capture some outcomes that may not have been captured in the legacy system), and UC roll out was happening quickly at the time of cohorts of

interest, then these interactions add some extra complication to the analysis. In the later comparison cohort, there was a higher proportion, before matching of those with a UC history. This potentially matters since UC provides greater incentives to move into work, and therefore, all other things being equal, it would lead to the comparison cohort having more favourable employment outcomes. We observe that this would, if uncorrected for, lead to an *understatement* of the programme impact, all else being equal. But more importantly, we observe that while this difference was present in the unmatched data, after matching the difference was negligible. It was not possible to do subgroup analysis as the numbers of UC participants in the ITT were very low, however the analysis for the participant/non-participant comparison in Appendix H, where numbers were higher, suggested no difference between the two groups (but caution needs to be taken about this analysis for all the reasons that have been mentioned).

As described in section 4.2, the following analysis shows the additional days the ITT treatment group estimated for each of the defined outcome states over the 2-year tracking period.

Figure 4.9 Impact on additional days in the three primary outcome states over 104 weeks

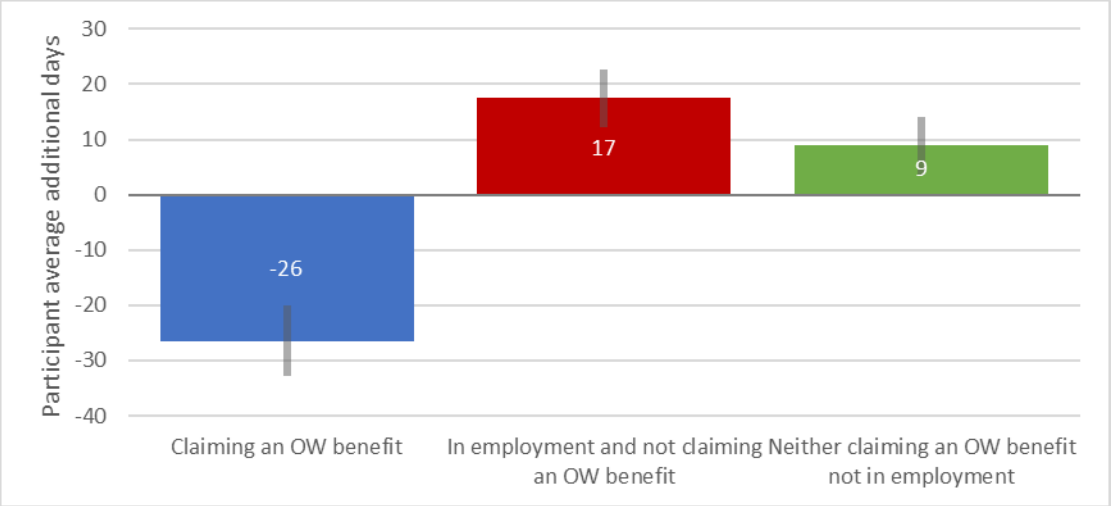


Figure 4.9 shows that, compared to the matched ITT comparison group, over 2 years the ITT treatment group spent:

- Less time on OW benefits – 26 fewer days
- More time in employment – 17 more days
- More time neither on OW benefits nor in employment – 9 more days

Since the ITT treatment group consists of 38% participants, it is reasonable to expect that the estimated impact for the eligible group would be around 38% of that for the participant group if all the impact can be attributed to participants. In the next section we show how these results are used to estimate the impact of participation.

Finally, we note that the ITT treatment group spent slightly more time than its comparator neither in receipt of an OW benefit nor in employment. Some of the outcomes in this state are likely to be self-employment or individuals in education.

Neither of these activities are captured by the other two primary outcome states. From DWP Management Information, around 15% of Job Outcomes claimed were for individuals in self-employment, so it is likely that a considerable portion of the additional days neither in receipt of an OW benefit nor in employment might be explained by self-employment (which would have the effect of pushing up the impact on overall employment, all else being equal). When self-employment data becomes available for 2018-20 in the future it might be possible to test this hypothesis.

4.3.2. Additional outcome states

Figure 4.10 to Figure 4.13 cover the additional five outcome states defined in section 4.2. These outcome states are not mutually exclusive and individuals can be counted in multiple outcome states in any given week.

Figure 4.10 Proportion of ITT cohorts in receipt of JSA or Universal Credit with Searching for Work conditionality

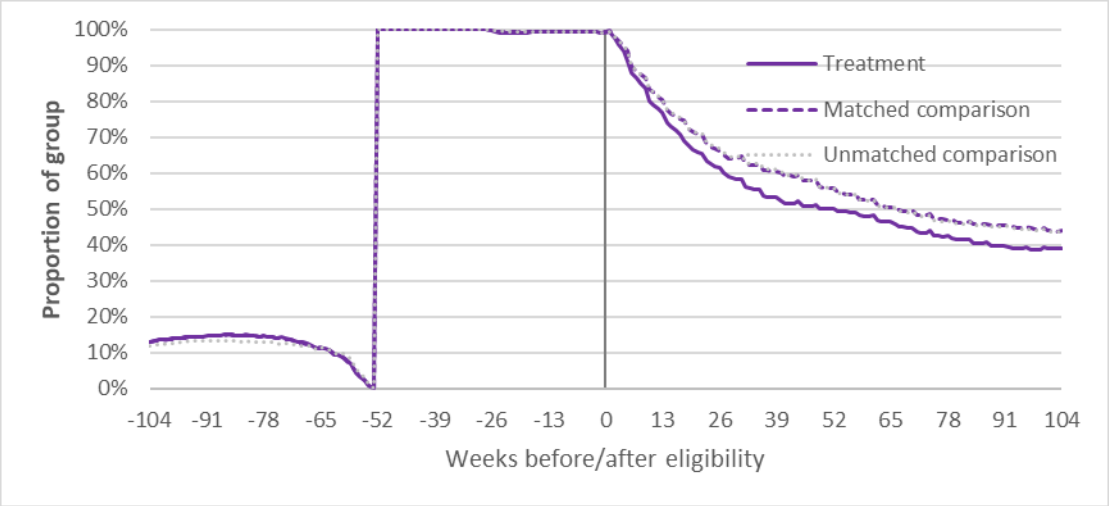


Figure 4.11 Impact on the likelihood of being in receipt of JSA or Universal Credit with Searching for Work conditionality

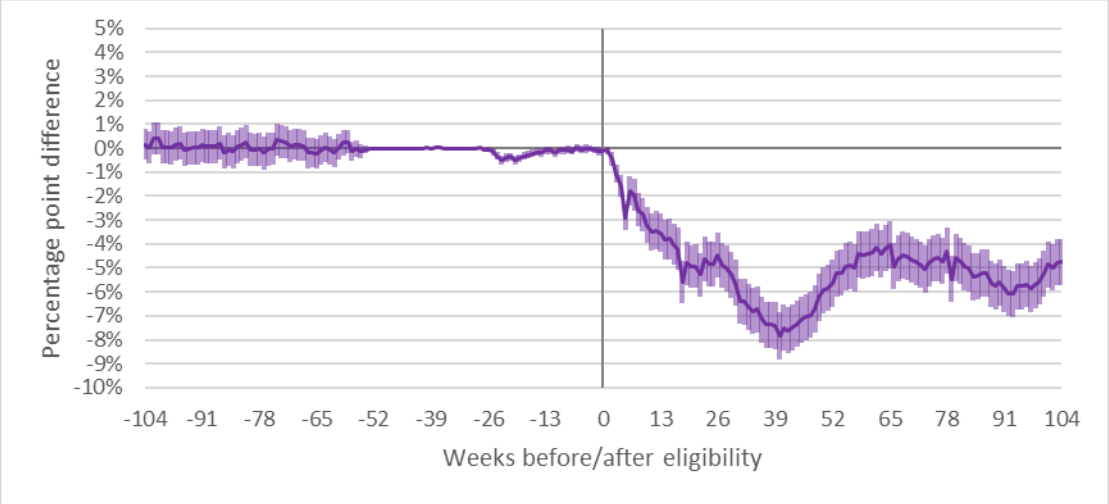


Figure 4.12 Proportion of ITT cohorts in receipt of the remaining OW benefits

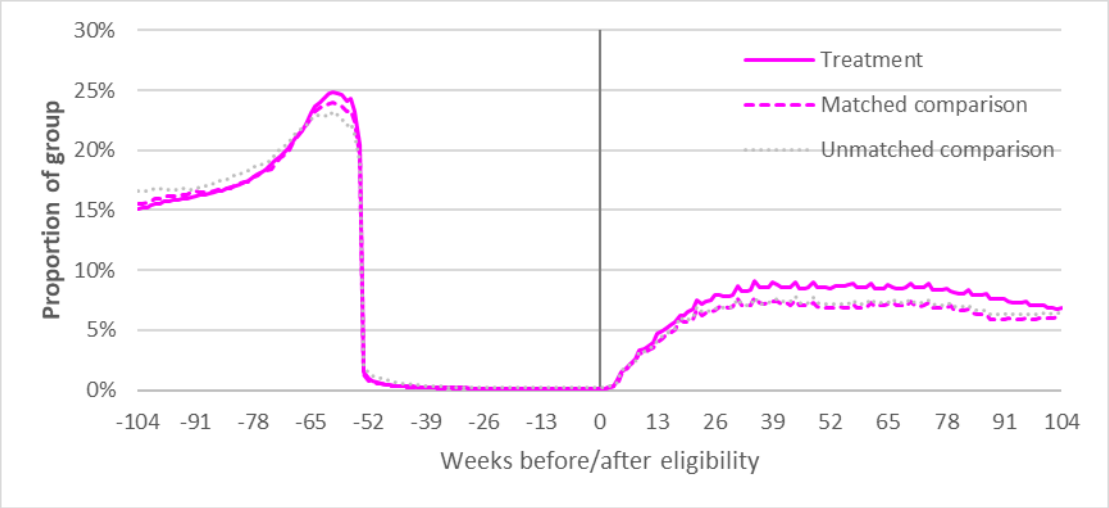
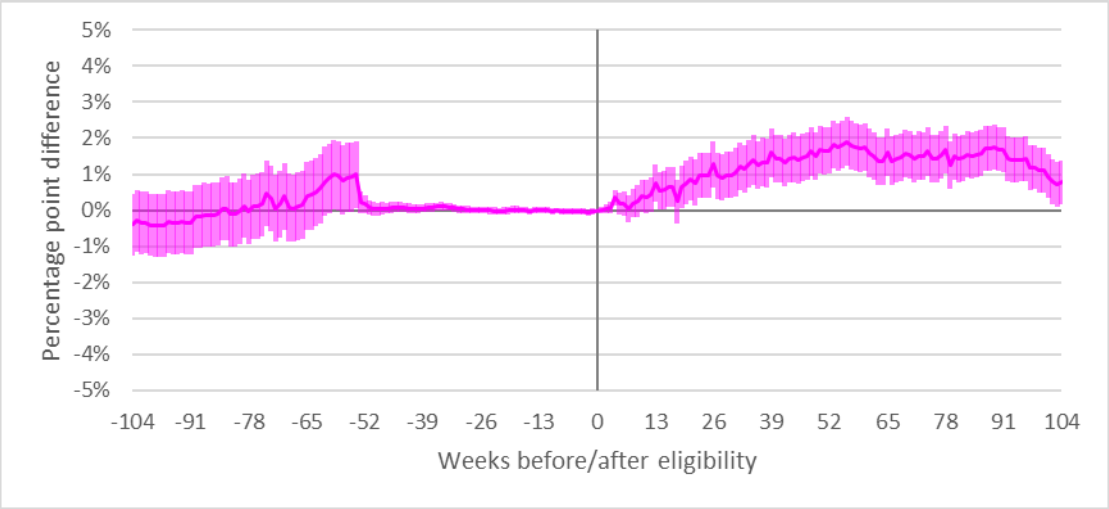


Figure 4.13 Impact on the likelihood of being in receipt of the remaining OW benefits



The unusual pattern in Figure 4.3 for receipt of any OW benefit is also present for receipt of either JSA or UC in the Searching for Work conditionality group, but not for receipt of other OW benefits. Instead, the other OW benefits show a similar pattern to that seen in Figure 4.7, where there is a gradual increase prior to the start of the eligibility period. This is because the eligibility is linked explicitly to the receipt of JSA or UC with Searching for Work conditionality, rather than receipt of other OW benefits. Again the results show that these groups are balanced very well on these further outcome states prior to eligibility with no significant differences in any one week. Figure 4.11 in particular demonstrates how well the treatment and comparison groups are matched, as the period prior to eligibility shows no statistical difference between the two groups.

Claims to Employment and Support Allowance (ESA) make up a significant proportion of the other OW benefits. ESA cannot be claimed at the same time as JSA so if an individual ends a claim to ESA and starts a claim to JSA in the same week they will be counted in both Figure 4.10 and Figure 4.12.

Overall the results show that the ITT treatment group spend less time in receipt of JSA or UC with Searching for Work conditionality than the ITT comparison group in the two years following reaching their eligibility. However, the group spends more time in receipt of other OW benefits. The net effect is that the ITT treatment group spend less time on OW benefits than the ITT comparison group in the two years following eligibility.

It is not clear from this study why this group spend more time on other OW benefits. A plausible reason could be that the treatment group were more engaged with the benefit system or the third-party providers delivering the Work Programme, which led them to better identify when they were eligible for other OW benefits. One untested hypothesis here is that a participant changing their circumstances significantly and switching benefits may have an effect on the provider by possibly exempting them from programme participation or moving them into a Payment Group with increased financial incentives.

While the primary outcome types are defined to be mutually exclusive, and OW benefit claims take precedence over employment, it is still possible that there are circumstances where an overlap of benefit claim and employment could occur. For example, an OW benefit claim ends at the start of a week and employment begins at the end of the same week, or individual claiming UC is earning but earnings are insufficient to exceed certain conditionality thresholds. Figure 4.14 and Figure 4.15 show the proportion and the impact of the overlap which may occur when an individual is in receipt of an OW benefit and also in employment recorded by RTI in the same weeks. This is more prevalent for individuals in receipt of UC.

Figure 4.14 Proportion of ITT cohorts in receipt of an OW benefit and in employment

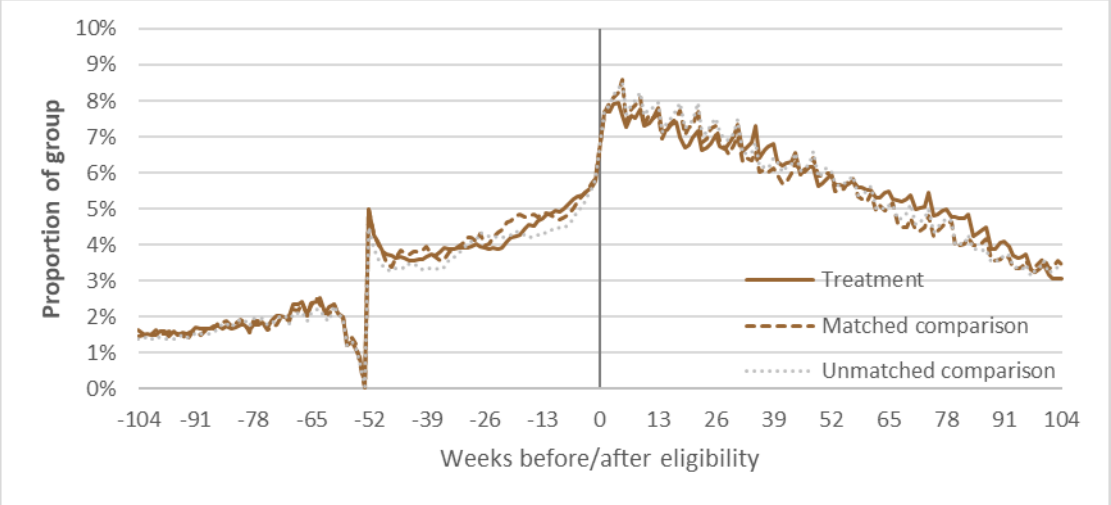
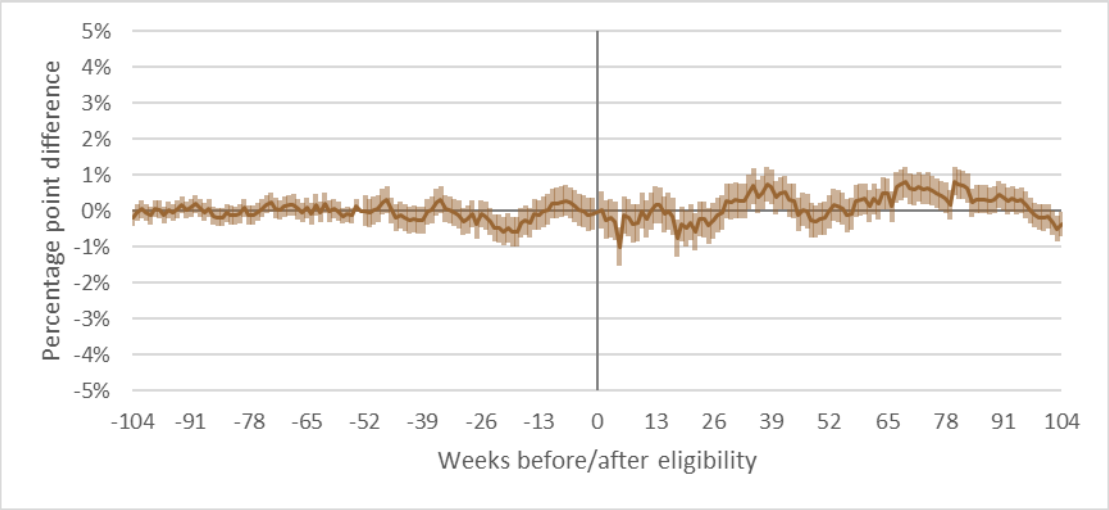


Figure 4.15 Impact on the likelihood of being in receipt of an OW benefit and in employment



A discontinuity, at around 52 weeks before referral, is observed in Figure 4.14, corresponding with the pattern seen in the proportion of individuals in receipt of an OW benefit in Figure 4.3. This is reflective of the application of the eligibility criteria and is observed in both the ITT treatment and ITT comparison groups. Nonetheless, the results again show there are no significant differences between the groups for the overlap between time in receipt of OW benefits and being in employment. Similarly this can also be observed in the additional outcome state of in employment irrespective of OW benefit stats, as shown in Figure 4.16 and Figure 4.17 below.

Figure 4.16 Proportion of ITT cohorts in employment irrespective of OW benefit status

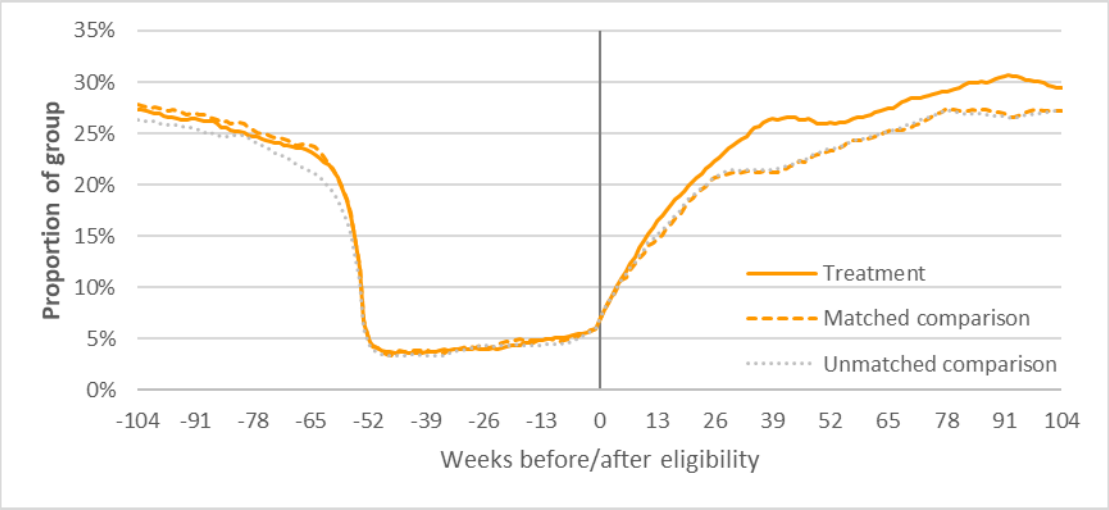
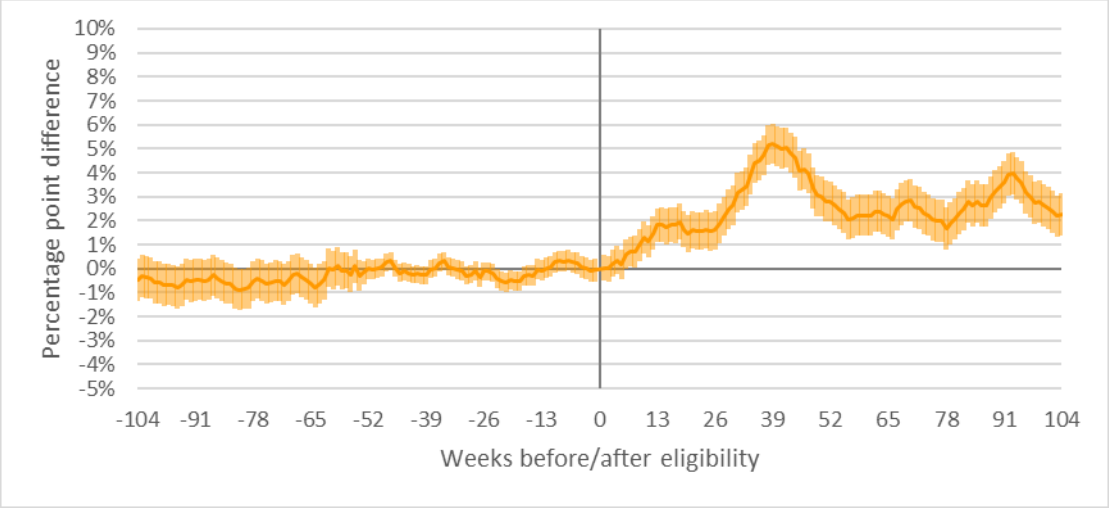


Figure 4.17 Impact on the likelihood of being in employment irrespective of OW benefit status



As expected, since the impact of being in employment and on OW benefits was shown to be minimal in Figure 4.15, Figure 4.17 shows similar levels of impact as the main outcome state of in employment and not in receipt of OW benefit in Figure 4.6.

Although the receipt of Disability Living Allowance (DLA) or Personal Independence Payment (PIP) are not dependent on the receipt of any other benefits or employment, it has previously been observed that there is an association between claiming ESA and claiming DLA or PIP. So it is possible that if the programme had a significant impact on receipt of ESA, that there might also be an impact on the receipt of DLA or PIP. The following graphs explore this outcome state.

Figure 4.18 Proportion of ITT cohorts (post-matching) in receipt of DLA or PIP

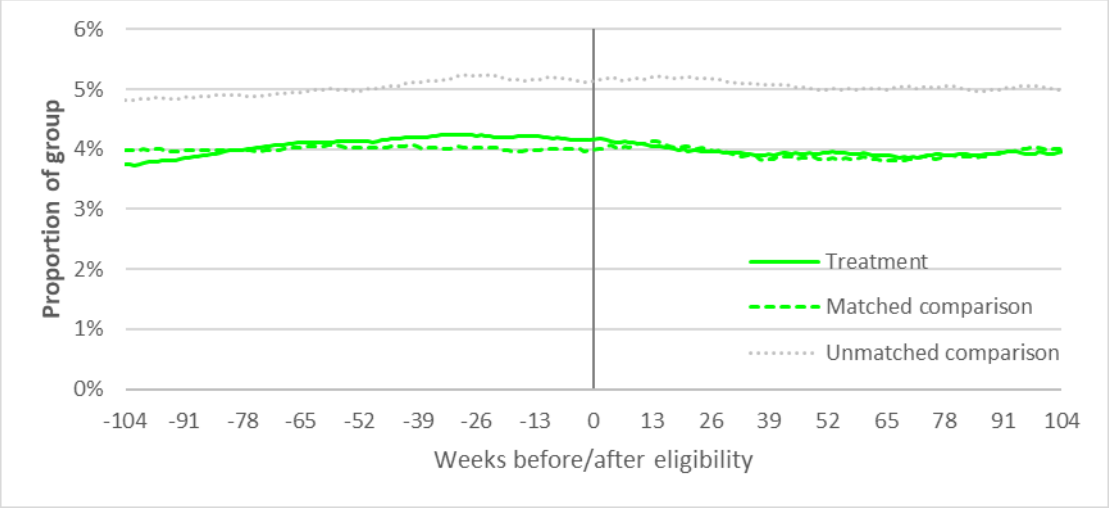


Figure 4.19 Impact on the likelihood of being in receipt of DLA or PIP

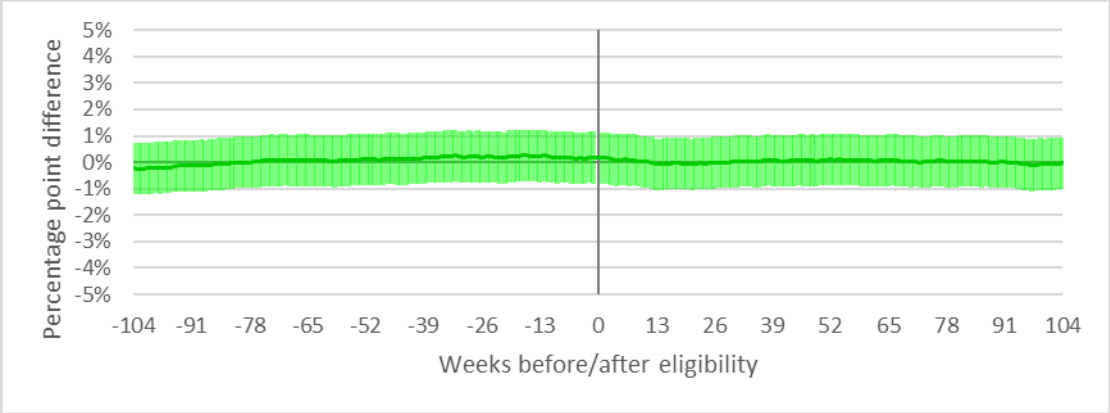


Figure 4.18 and Figure 4.19 show that there are no differences in the likelihood of being in receipt of either DLA or PIP between the ITT treatment and ITT comparison groups before and after eligibility point, after matching. This is a different result from the participant/non-participant analysis (Appendix H.2) where this variable was not balanced prior to eligibility and showed a slight divergence after reaching eligibility. This analysis suggest that use of the ITT groups has avoided some of the bias that might be in the participant/non-participant groups, albeit with a much smaller proportion of programme participants.

These five additional outcome states can also be measured in terms of the additional days the ITT treatment group spends in each state beyond the ITT comparison group.

Figure 4.20 Additional days in the five additional outcome states over 104 weeks

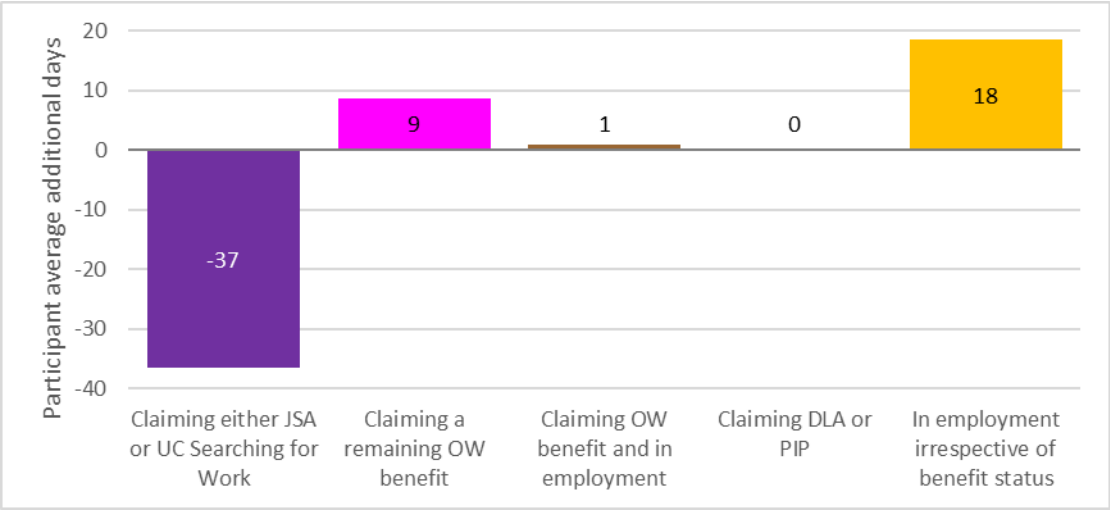


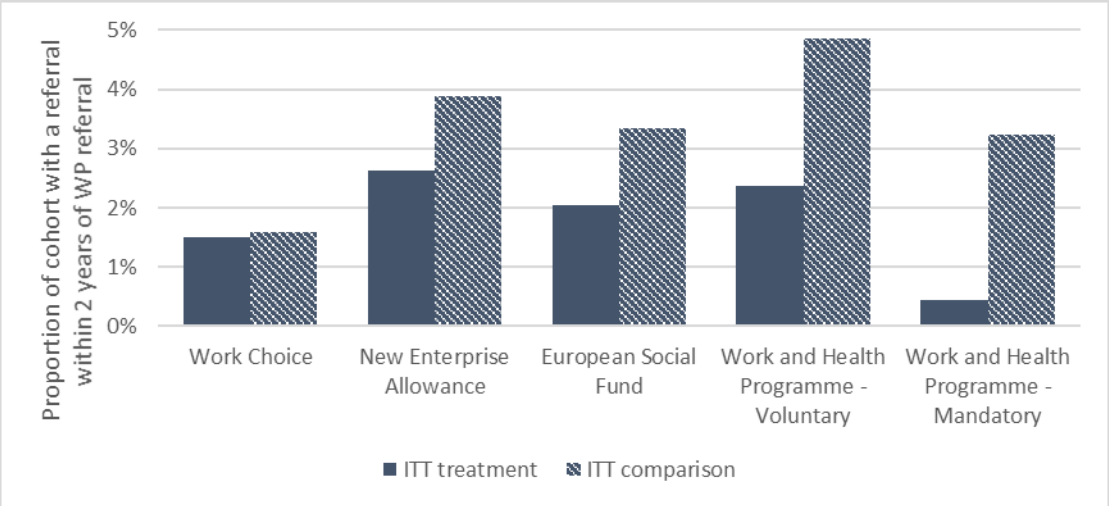
Figure 4.20 shows that the ITT treatment group spends less time in receipt of JSA or UC in Searching for Work conditionality, but more time in receipt of other OW benefits than the ITT comparison group. In contrast to the participant and non-participant groups in Figure H.13, the ITT treatment group spends almost the same time claiming OW benefits and in employment as the ITT comparison group, and the groups also spend the same amount of time in receipt of DLA or PIP.

4.3.3. Engagement with other provision

Some of the ITT treatment and comparison group may have been referred to other provision, either on a mandatory or voluntary basis, in the period following their eligibility to the Work Programme. This may have an impact on their outcomes in the two years following becoming potentially eligible to the Work Programme. We can track whether individuals in the groups go on DWP contracted provision, but they may also go on non-DWP provision that is not tracked. Other DWP programmes are designed to offer employment support to help tackle different barriers to employment than Work Programme, or in the case of New Enterprise Allowance, to support an individual into becoming self-employed. These referrals are not mutually exclusive, and it is possible that an individual could have referrals to multiple programmes.

This analysis only looks at whether someone has been referred in the two years following reaching eligibility; it does not attempt to determine when the referral occurred in relation to their eligibility to the Work Programme or whether the individual progressed further than a referral. Not all programmes were available during the two years tracked in this analysis. For example, the Work and Health Programme had a phased rollout starting in late November 2017 for voluntary participants in some areas in England and Wales and then from April 2018, low numbers of individuals reaching 24 months of unemployed were referred. Figure 4.21 below shows the proportions of each ITT cohort that had a referral to selected employment provision in the two years following reaching eligibility.

Figure 4.21 Referrals to other contracted provision within 2 years of Work Programme eligibility for the ITT cohorts



The results show that about 10% of the ITT treatment group were referred to other contracted provision, compared to about 20% of the ITT comparison group.

If other provision does have a positive impact, the fact that a higher proportion of the comparison group participate in other provision would lead to their outcomes being higher than if they were receiving only standard Jobcentre Plus support, and hence to the understatement of the net impact. Past evaluation have generally found positive impacts for these programmes; but given the differences in the proportions participating, the consequent understatement of the programme impact will be minor.

4.4. The impact of participation and time effects

4.4.1. Impact of participation

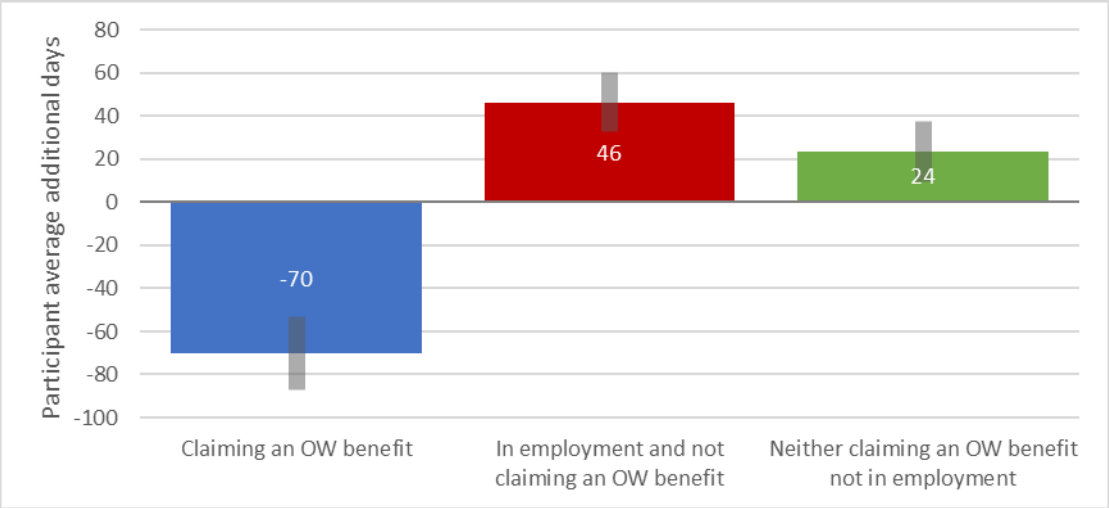
The previous section sets out how the basis for impact estimates for those eligible to join the Work Programme in Jan-March 2017, compared with those who would have become eligible in April-June 2017, had the programme still been running. We have used propensity score matching to adjust for the (relatively small) differences in the characteristics of claimants in those two quarters, and have carried out a series of sensitivity checks to assure ourselves that little if any of this difference can be attributed to changes in labour market conditions; these additional 17 days in employment and 26 fewer days on out of work benefits are our best estimate of the net impact of the Work Programme on this Intention to Treat cohort of claimants.

As set out in Section 3 however, only a minority of the cohort – 38% - actually participated in the programme. It is natural therefore to ask what the impact was on those who actually participated; technically this would be the Average Treatment Effect on the Treated, which for the rest of the report we describe this as the impact of participation.

To estimate the impact of participation on employment, we first observe that, with a total of 20,113 claimants in the cohort, the aggregate impact was 3.4 million days of employment. If we accept that the ITT comparison is a plausible counterfactual for the ITT treatment group, then it follows that there is a subset of the ITT comparison cohort which is a valid counterfactual for participants (although we don't know who they would be). So if we assume that the programme had no impact on those who did not register that would mean that the 3.4 million extra days of employment were shared amongst the 7,591 individuals who actually participated in the WP – that is, an average of an additional 45 days per participant. Another way of looking at this is that the average of 17 days more in employment for the overall cohort is the combined effect of 45 days for the 38% who participated, and zero for the 62% who did not. The implicit and reasonable assumption here is that the difference between the participant counterfactual outcomes of the group with the ITT group within the comparison cohort and the rest of the comparison cohort is the same as the difference in counterfactual outcomes between participants and non-participants in the treatment cohort. We note that although we have estimated that average impact on participation is 45 days, we cannot know for sure that the impact be the same for the whole cohort.

This analysis depends as stated on the assumption that the programme only has an impact on those who actually participate – an assumption which needs to be tested. The most likely way in which the programme could have had an effect on non-participants is through *anticipation effects*. That is, that those who are approaching the eligibility threshold either intensify their jobsearch efforts in order to avoid being referred to the programme, or reduce them because they want to make sure that they are referred. The checks that we made in Appendix G.2 found no evidence of this. Accordingly, Figure 4.22 show the results of the estimated impacts on participants only for the three different outcome measures, in terms of additional days, with confidence intervals calculated in the same way.

Figure 4.22 Additional days in the three primary outcome states in terms of impact on participants



4.4.2. Time effects

As described in Section 3.2 several additional cohorts, drawn from other consecutive time periods, have also been compared to understand the magnitude of any time based effects. These cohorts are:

1. After the programme finished, April 2017 to June 2017 and July 2017 to September 2017
2. One year prior to the lead comparison, January 2016 to March 2016 and April 2016 to June 2016
3. One year prior to the groups in (1) above, April 2016 to June 2016 and July 2016 to September 2016

The details of the analysis are in Appendix G.3. The differences between cohorts selected from 2016 suggest that changes between consecutive cohorts in that period were minor – Figure G.16 shows there was no change in receipt of out of work benefits and the earlier cohort had 3 fewer days in employment. If this effect had been present during the time period for the main analysis, then it would have meant that time effects would have contributed 3 days to the main analysis – overestimating the benefit impact (i.e. the benefit impact would be slightly less negative).

However, there is a greater difference for the cohorts after the end of the programme – the earlier cohort had 6 fewer days in employment and 9 extra days on out of work benefits, as shown in Figure G.8. If this effect had been the present for the main analysis, then it would have meant that the main impact estimates had underestimated the employment impact by 6 days and the out of work benefit impact by 9 days (i.e. more negative).

These differences suggest that there are some time effects for the ITT groups which can be as large as 6 days in employment and 9 days on benefit. We do not know what exactly is the extent of the time effect for the main ITT groups of interest but this analysis suggests a conservative view is that they can lead to an overestimation 6 days in employment and 9 days on benefit; a best case view is that they might lead to an underestimation the other way. Similarly the results of the participant/non-

participant comparison in Appendix H suggest that the fluctuations are no greater 3 days in employment and 6 days on benefit, as seen in Table H.1.

These time effects can be used to adjust the main impacts to account for these time differences. Table 4.2 below shows the impact from the ITT analysis and impact on participants with ranges derived from the potential time effects.

Table 4.2 Summary of ITT and impact on participants results, with ranges of time effects

Cohorts	In receipt of OW benefit	In employment and not in receipt of OW benefit	Neither in receipt of OW benefit nor in employment
Impact from ITT analysis (without time effects)	-26	17	9
Range of impact on ITT (with time effects)	-18 to -35	11 to 24	7 to 11
Impact on participants (without time effects)	-70	46	24
Range of impact on participants (with time effects)	-47 to -93	30 to 63	17 to 30

Overall this suggests, that even when taking into account potential time effects, the programme had a positive impact on individuals participating in it, resulting in fewer days in receipt of an OW benefit and more days in employment. As explained in Section 4.3 self-employment is not recorded through RTI and is instead captured as a 'neither in receipt of OW benefit nor in employment' outcome, but it is unclear how much, if any of this outcome type consists of individuals who have become self-employed.

4.4.3. Further sensitivity analysis

Further sensitivity analyses of other cohorts are also presented in Appendix G.3. These cohorts are:

1. Shorter 1 month cohorts drawn closer to the end of the programme: March 2017 and April 2017
2. Wider range between cohorts: October to December 2016 and April to June 2017
3. Longer 6 month cohorts: October 2016 to March 2017 and April to September 2017

As described in Section 3, there is a trade-off between shorter cohorts where time effects will be smaller but there will be fewer participants against longer cohorts where time effects will be more uncertain but there will be more participants. In 1. the proportion of participants in the treatment group is only 26% - so it's reassuring to see that the results are lower than in the main analysis. In 2. and 3. the proportion of participants increases, but time effects are more uncertain: results are slightly higher in 2. and are very similar in 3. Overall, although we believe that this further analysis

is not as informative as the main analysis, it does provide additional reassurance around the main results.

4.5. Summary of results

The success of the evaluation is dependent on potential sources of bias that might affect the outcomes being identified and addressed satisfactorily. The previous sections have considered an Intention to Treat analysis to avoid selection bias arising from comparing participants and non-participants and several sensitivity analyses on additional groups and sub-groups in order to quantify potential effects from two key sources:

- Time trend and seasonal changes between cohorts; and
- Compositional differences between cohorts.

Propensity score matching was used to try to control for differences in composition – it showed that the groups were very well balanced prior to reaching eligibility (particularly when compared with the participant/non-participant comparison). However, more potentially significant issues arise from drawing cohorts from different time periods and also that not all eligible people were referred. These biases have been explored, for example by:

- Comparing cohorts from different time periods to look at time effects; and
- Looking for evidence of any anticipation effects.

The outcome of these analyses suggest that although there are some time effects they are unlikely to affect the findings substantively and therefore offers assurance for the main estimates, while bearing in mind that are still some uncertainties.

Table 4.3 below summarises the results of the analysis carried out across these groups, including results from the participant/non-participant comparison described in Appendix H, and the impact on participation.

Table 4.3 Summary of results

Cohorts	Days in receipt of OW benefit	Days in employment and not in receipt of OW benefit	Days neither in receipt of OW benefit nor in employment
ITT impact from comparing January 2017 to March 2017 and April 2017 to June 2017	-26	17	9
Time effects from comparing April 2017-June 2017 and July 2017-September 2017	9	-6	-2
Time effects from comparing January 2016-March 2016 and April 2016-June 2016	0	1	-1

Time effects from comparing April 2016 -June 2016 and July 2016-September 2016	0	3	-2
ITT impact from comparing October 2016 – March 2017 to April 2017 – September 2017	-24	16	8
ITT impact from comparing March 2017 with April 2017 sub-groups	-14	8	6
Impact from comparing participants January 2017 to March 2017 with non-participants April 2017 to June 2017	-25	37	-12
Impact of participation (using January 2017 to March 2017 and April 2017 to June 2017 ITT groups)	-70	46	24
Range of impact on participants	-47 to -93	30 to 63	17 to 30

Finally, we note that when comparing the ITT results with those from the participant/non-participant group, it can be seen that:

- the impact on participation ITT-derived estimates suggest greater effects than the participant/non-participant analysis.
- the outcome for which the participant/non-participant and impact on participation ITT-derived estimates compare most closely is employment; this is also the outcome for which the matching worked best in the participant/non-participant group analysis.
- the out of work nor in employment results are of a positive sign for the ITT analysis rather than a negative sign for the participant/non-participant analysis.

The main reason offered for the differences between the ITT analysis and participant/non-participant analysis, as seen in Appendix H, is that the ITT analysis has avoided the selection bias arising from the participant group. However, the similarity around the analysis on the employment state does suggest that the participant/non-participant is not wildly affected and gives some tentative reassurance to the ITT estimates for the impact on participation.

5. Cost benefit analysis

This section presents a cost benefit analysis (CBA) for participation on Work Programme Payment Group 2 based on the impact results in Section 4 for a two-year period following referral to the programme, plus additional extrapolated estimates of a further one and two years – giving a total of three and four years following reaching individuals reaching eligibility.

5.1. Cost benefit analysis methodology

The methodology underpinning the cost benefit analysis (CBA) is based on the Department for Work and Pensions (DWP) Social Cost Benefit Analysis (SCBA) framework as set out by Fujiwara (2010) as has been used in previous quantitative impact analyses, such as for Sector–based Work Academies. This framework aims to provide a thorough, consistent and evidence-based approach to the CBA of employment programmes. The application of this approach is outlined below in terms of:

- whose perspective is considered – section 5.2;
- which costs and benefits are estimated – section 5.3; and
- the estimated scale of the costs and benefits under consideration – section 5.4.

This framework does, however, exclude a number of costs and benefits where it was not possible to obtain robust evidence, for example, the cost of hiring and training as incurred by employers. Further detail on the limitations of the adopted methodology is discussed in section 5.5.

5.2. Perspectives under consideration

The costs and benefits are considered from the perspectives of:

- The Work Programme participant;
- The Department for Work and Pensions;
- The Exchequer, i.e. the government budget perspective; and
- Society.

For each of these perspectives, cost benefit ratios can be calculated. These show for each £1 that participation in the programme cost, how much value is returned.

The accuracy of the CBA estimates is dependent on the robustness of the impacts from which they are derived. To mitigate some of this uncertainty a series of sensitivity tests have taken place, as already described in Chapter 4.

For the purposes of this analysis, ‘society’ represents an aggregate of all British citizens. Therefore, a cost or benefit to participants, their employers or the Exchequer can also represent a cost or benefit to society. However, it should be noted that many of the gross impacts are essentially ‘transfer payments’. Transfer payments represent a cost to one group of citizens but a benefit to another. For example, the wages earned during additional employment as a result of the Work Programme represent a benefit to participants but a cost to their employers. Such

transfer payments cancel out when estimating the net benefits of a policy from society's perspective.

An example of a net benefit to society is the increase in output that occurs when a policy leads to resources being used more efficiently. In the case of the Work Programme, this occurs when participants, who were previously producing no output, produce output during additional time spent in unsubsidised employment as a result of the policy. This additional output represents a net benefit to employers and society.

5.3. Costs and benefits under consideration

Table 5.1 summarises the impacts which have been translated into monetised costs and benefits for inclusion in this CBA. These impacts and the associated costs and benefits are discussed separately below.

Table 5.1 Monetised costs and benefits of the Work Programme

	Perspective			
Impact	**Participants**	**DWP**	**Exchequer**	**Society**
Increase in output	0	0	0	+
Increase in wages	+	0	0	0
Programme costs	0	-	-	-
Reduction in operational costs	0	+	+	+
Reduction in benefit payments	-	+	+	0
Increase in taxes	-	0	+	0
Increases in travel and childcare costs	-	0	0	-
Reduction in healthcare costs	0	0	+	+
Redistributive costs and benefits	+	0	0	+
Social cost of Exchequer finance	0	0	0	-

Key: '+' denotes a net benefit; '-' denotes a net cost; '0' denotes neither cost nor a benefit.

5.3.1. Increase in output

This refers to the economic output produced by participants as a result of additional time spent in employment. This output represents a benefit to employers (who sell it) and society (who consume it). The DWP does not have information on the value of this output so it is necessary to make a number of simplifying assumptions, discussed below.

The labour market is assumed to be perfectly competitive. This implies that employers will hire workers up to the point where the value of an additional unit of output is equal to the associated marginal cost of production. The cost of production, and therefore the value of the output produced during additional spells in employment, is assumed to equal the commensurate gross wage payments and employers' National Insurance contributions.

5.3.2. Increase in wages

This refers to the gross wages received by participants during additional time spent in employment. Wages represent a benefit to participants but a cost to their employers. This means they do not represent a net cost or benefit to society as a whole, except via redistributive effects described below.

With the introduction of Real Time Information (RTI), which includes earnings information, it may be possible in the future for further analysis to take place on earnings received by participants. However, this is not the case for this study, and previously assumed wage averages of former benefit recipients have been used[14].

5.3.3. Programme costs

For DWP, the Exchequer and Society the costs of programme participation are the amounts paid to third-party providers to run the programme. This analysis does not take into account other smaller costs which will include administrative costs for running the programme, although these will be reduced by having less contact with Work Coaches, and also subsidised courses that providers might use. For participants, the costs are reductions in benefits received and increases in income tax and National Insurance Contributions.

The costs comprised attachment fees, paid in the early years of the programme, and outcome fees paid on a Payment by Results (PbR) approach. Further information on the payment model can be found in Appendix B.

As outcome fees were paid on a PbR basis, they did not occur at fixed points during a participants' time on the programme. Through analysis of payments made over the life of the programme, the £1417 spent per person was estimated to occur over three years following programme referral, as follows:

- £667 in year 1 following referral;
- £417 in year 2 following referral; and
- £333 in year 3 following referral.

These costs are estimated in relation to programme referral, however key cohorts are defined on their eligibility to the programme. As shown in section E.3, over 75% of referrals made occurred between 30 days before to 90 days after eligibility. Due to possible biases in these delays, it is not clear how these costs should be adjusted to be used in reference to eligibility dates rather than referral dates. However, as the costs are aggregated to an annual frequency, and delays from eligibility to referral largely lie in a 4-month range, it is reasonable to assume the costs would be aggregated similarly.

These payments represent a cost to the Exchequer and society, as this diverts economic resources from alternative uses.

[14] Estimates of the average weekly wage received are based on the findings of Adams and others (2012)

5.3.4. Reduction in operational costs

Work Programme participants are less likely to receive support from Jobcentre Plus advisers following a placement because they are more likely to be working and less likely to be claiming benefit. As a result, this also means participants are less likely to participate in other DWP employment programmes. This translates into operational savings which represent a benefit to the Exchequer and society, as economic resources can be reallocated to alternative uses.

5.3.5. Reduction in benefits

This refers to the net reduction in benefit entitlement and take-up that occurs when participants spend additional time in employment as a result of participation on the Work Programme[15]. This represents a cost to participants but a benefit to the Exchequer, which means there is no net cost or benefit to society as a whole, except via redistributive effects explained below. Changes in benefit entitlement and take-up are estimated using the DWP Policy Simulation Model[16].

5.3.6. Increase in taxes

This refers to the increase in income tax, National Insurance and indirect tax revenue that occurs when participants spend additional time in employment as a result of participation in the Work Programme. This represents a benefit to the Exchequer but a cost to participants and employers, which means there is no net cost or benefit to society as a whole, except via redistributive effects below. Increases in tax revenue are estimated using the DWP Policy Simulation Model[17].

5.3.7. Increase in travel and childcare costs

This refers to the additional travel and childcare costs that are incurred by participants during additional employment as a result of participation in the Work Programme. This also represents a cost to society as the provision of additional travel and childcare services diverts economic resources from alternative uses.

5.3.8. Reduction in healthcare costs

This refers to the reduction in National Health Service (NHS) costs which is expected to occur when participants spend additional time in unsubsidised employment as a result of their participation in the Work Programme[18]. This represents a benefit to the Exchequer, via reductions in NHS expenditure, and society, as economic resources which had been allocated to healthcare provision can be reallocated to alternative uses.

[15] Increases in Tax Credit entitlement and take-up are more than offset by reductions in Jobseeker's Allowance, Housing Benefit and Council Tax Benefit

[16] The DWP Policy Simulation Model is a microsimulation model which combines data from the Family Resources Survey with information on the UK tax and benefit systems. This allows users to estimate the changes in benefit payments and tax revenue that occur when unemployed individuals with a given set of characteristics move into work

[17] In order to estimate increases in indirect tax revenue, Office for National Statistics estimates of indirect tax burdens were applied to estimates of participants' disposable income obtained from the DWP Policy Simulation Model

[18] Fujiwara (2010) presents evidence of a causal relationship between individuals' employment status and NHS usage

5.3.9. Redistributive costs and benefits

This refers to the redistributive costs and benefits associated with monetary transfers between participants, employers and the Exchequer. In line with the methodology prescribed in the HM Treasury Green Book[19], participants, who have relatively low incomes, are assumed to value each additional pound more highly than employers and the average taxpayer, who both have a relatively high income compared to participants. This implies, for example, that monetary transfers from the Exchequer to participants represent a benefit to society as a whole. In line with the recommendations of Fujiwara (2010), redistributive costs and benefits are estimated by applying a 'welfare weight' of 2.5 to monetary transfers made to and from programme participants.

5.3.10. Social cost of Exchequer finance

This refers to the cost of raising the tax revenue that was required to finance the Work Programme. This 'deadweight loss' arises because taxation creates market distortions which have an adverse effect on economic efficiency. The distribution of the social cost of Exchequer finance amongst members of society depends on the specific details of taxation policy.

For the purposes of this analysis, it is assumed to accrue to society as a whole. In line with the recommendations of Fujiwara (2010), the social cost of Exchequer finance is assumed to equate to 20 per cent of the net cost of the programme to the Exchequer. However, this estimate is subject to considerable uncertainty.

5.4. Estimating the scale of the benefits under consideration

The scale of the costs and benefits of the Work Programme depends on the magnitude and the duration of its impacts. Specifically, it depends on the programme's impacts on the amount of time that participants spend:

- in employment; and
- neither in receipt of Out of Work benefits nor in employment.

These impacts have been derived from the two of the three mutually exclusive outcome states relating to being "off benefit", as presented in section 4.2.

The tracking period in this study covers two years from the point of referral. It is clear that impacts will not cease immediately after the two year tracking period, so it is reasonable to extrapolate estimated impacts into the future. However, it is a judgement as to how long impacts might continue and whether the observed impacts would be sustained at the same level, or decay at some rate.

For this judgement, we should bear in mind that the period of support on the Work Programme can be for a long time. For payment Group 2 participants support was offered for a period of up to two years following referral. Then, for participants finding and remaining in employment for at least 6 months, up to an additional 12 months of support was on offer to help participants sustain their time in employment, while

[19] See HM Treasury (2003)

providers received sustainment payments. So, total support for these participants could amount to a maximum of 42 months, or around 3 and a half years.

Therefore, given the trajectory of the impacts and that participants could be supported on the programme for up to three and a half years in some cases, a one-year extrapolation would seem pessimistic while extrapolating for three years or more is, while certainly possible, difficult to justify from the tracked data.

So in this analysis we show the estimated costs and benefits of the programme in relation to three possible scenarios:

- The impacts of the programme do not persist beyond the tracking period -the most conservative estimate;
- The impacts of the programme persist for one year beyond the tracking period; and
- The impacts of the programme persist for two years beyond the tracking period.

To calculate the latter two scenarios, the programme impacts have been extrapolated using additive seasonal exponential smoothing with three-month seasonality. This broadly assumes that the level of impact is maintained beyond the tracking period and does not decay. It should be borne in mind that the further the impacts are assumed to persist beyond the tracking period, the less reliable the resultant cost and benefit estimates become.

It is also necessary to mention the possibility that a proportion of the positive employment impacts experienced by programme participants are obtained at the expense of other people. If this is the case, then the overall benefits of the programme will be overstated. However, this 'substitution effect' is unlikely to be large for a supply side programme.

5.5. Limitations of this approach

The CBA estimates presented are subject to two main caveats. Firstly, the accuracy of the estimates depends on the robustness of the impact estimates from which they are derived as described in Section 4 and the validity of the assumptions upon which they are base, as outlined in Sections 5.3 and 5.4.

Secondly, the CBA estimates exclude a number of potentially significant costs and benefits due to a lack of robust evidence[20]. For example, it has not been possible to obtain robust estimates relating to:

- the additional leisure time which participants forego – this represents a potential cost to participants and therefore society;

[20] For a thorough discussion of the non-monetised costs and benefits of employment programmes, see Fujiwara (2010).

- the non-pecuniary benefits associated with additional time in unsubsidised employment – these represent a potential benefit to participants and therefore society;
- the cost of hiring and training incurred by employers – this represents a potential cost to employers and therefore society;
- the reduction in crime[21] which may result from the programme – this represents a potential benefit to society; and
- the economic multiplier effect which may result from the programme – this represents a potential benefit to society.

These non-monetised costs and benefits should be borne in mind when interpreting the CBA estimates presented in section 5.7.

5.6. Findings of cost benefit analysis

By using the departmental Social Cost Benefit Analysis (SCBA) model, estimates can be made on the cost benefit ratios from each of the selected perspectives.

As inputs, this model uses:

- Programme costs per year
- Additional days in employment
- Additional days not in receipt of an OW benefit

As discussed in Section 2.4, the cost of the programme per participant varied over the life of the programme. This is largely due to attachment fees that were only payable during the early years of the programme, and were not paid for later referrals. Whilst this study focuses on participants engaging with the provision in 2017, using the cost-per-capita for these customers alone would ignore capital investment made in the provision in earlier years, thus underestimating the costs of the programme. Overall, the average cost per participant on the programme over the lifetime of the programme was around £1400 excluding VAT. These costs are estimated as being incurred across a three-year period as described in Section 5.3.3.

Appendix I also gives results from a secondary outcome state – in employment irrespective of their OW benefit status. Ranges for these outcome states are formed using the ranges produced by consideration of potential time effects.

Figure 5.1, Figure 5.2 and Figure 5.3 show the central estimate broken down into the observed two-year impact and the extrapolated years three and four, for each of the three outcomes states. The ranges for additional days in each outcome state have been calculated using the largest estimated time effects, as shown in Table 4.2, and

[21] Fujiwara (2010) presents evidence of a causal relationship between individuals' income levels and their propensities to commit acquisitive crime.

projected forward to four years using additive seasonal exponential smoothing with three-month seasonality.

Figure 5.1 Cumulative average additional days for participants in receipt of an OW benefit

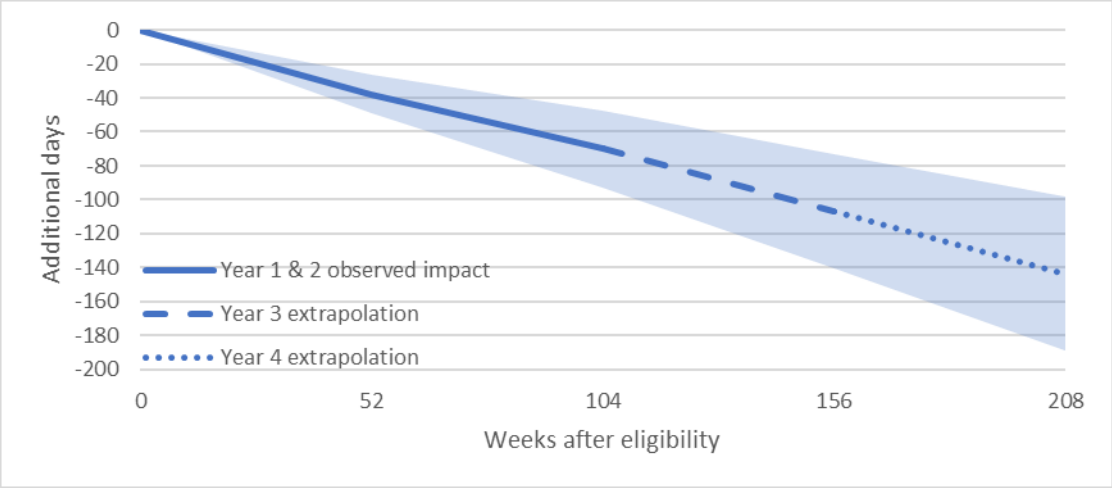


Figure 5.2 Cumulative average additional days for participants in employment and not in receipt of an OW benefit

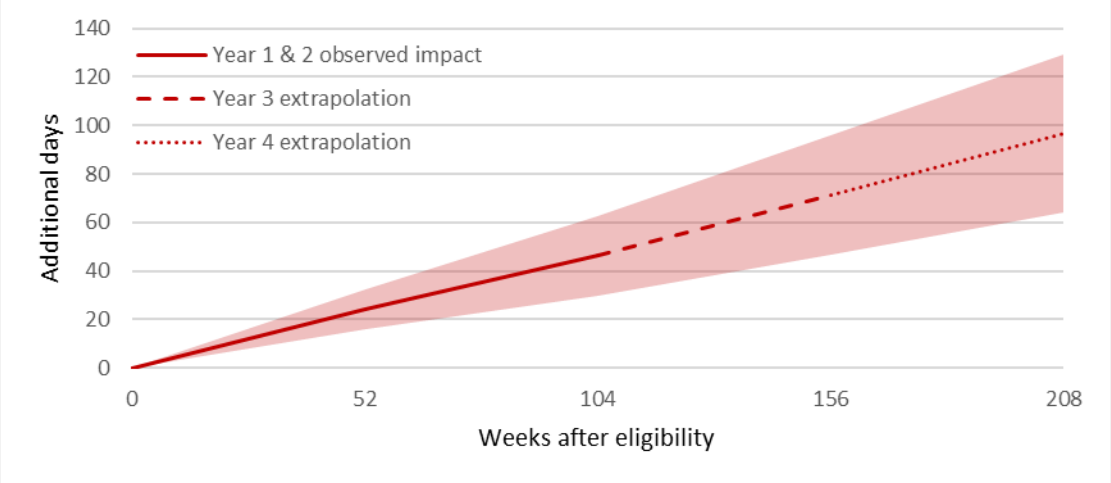


Figure 5.3 Cumulative average additional days for participants in employment irrespective of OW benefit status

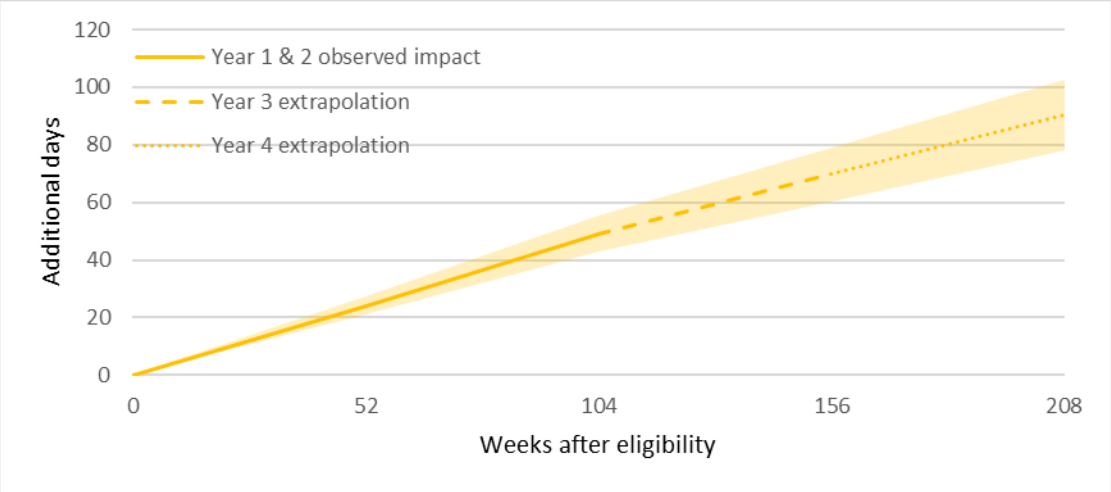


Table 5.2 shows the estimates for the total additional days in each outcome state for each year following eligibility to the programme.

Table 5.2 Estimated additional days extrapolated over four years

	Years 1 and 2 - observed	Year 3 - extrapolated	Year 4 - extrapolated
In receipt of an OW benefit			
Impact from ITT analysis	-26	-14	-14
Impact of participation	-70	-37	-37
Range of impact of participation (from potential time effects)	-46 to -94	-26 to -48	-26 to -48
In employment and not in receipt of an OW benefit			
Impact from ITT analysis	17	10	10
Impact of participation	46	25	25
Range of impact of participation (from potential time effects)	30 to 63	17 to 33	17 to 33
In employment, irrespective of OW benefit status			
Impact from ITT analysis	19	8	8
Impact of participation	49	21	21
Range of impact of participation (from potential time effects)	43 to 55	18 to 24	18 to 24

5.7. Findings of cost benefit analysis

By using the estimated impacts over a four-year period and assuming the costs of a participant are incurred over a three-year period, Table 5.3 shows the impacts using additional days for a participant in employment and not in receipt of an OW benefit using ITT estimates for the impact of participation for the three scenarios – 2 year observed, 3 year extrapolated and 4 year extrapolated. Appendix I shows the

estimated additional days in employment irrespective of OW benefit status and shows similar cost benefit ratios.

Table 5.3 Cost Benefit ratio estimates using in employment and not in receipt of an OW benefit

		Participant	DWP	Exchequer	Society
Impact of participation	2 year observed	£ 1.25	£ 0.83	£ 1.50	£ 1.75
	3 year extrapolated	£ 1.26	£ 1.31	£ 2.38	£ 2.68
	4 year extrapolated	£ 1.26	£ 1.76	£ 3.21	£ 3.51
Range of impact	2 year observed	£1.22 to £1.27	£0.56 to £1.10	£1.00 to £2.01	£1.16 to £2.31
	3 year extrapolated	£1.23 to £1.27	£0.89 to £1.72	£1.60 to £3.15	£1.82 to £3.48
	4 year extrapolated	£1.23 to £1.27	£1.21 to £2.32	£2.17 to £4.25	£2.42 to £4.51

The accuracy of the CBA estimates is dependent on the robustness of the impacts from which they are derived. As explained in the previous section these estimates have been calculated by extrapolating an additional one and two-years of impact from the observed two-year impact.

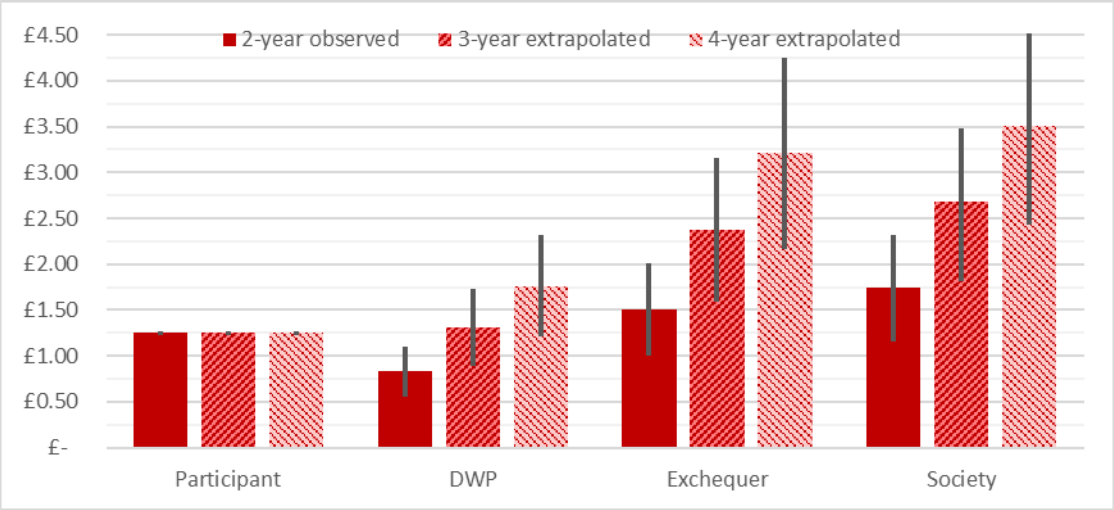
Participation in the programme is on average a net benefit to the participant. For every £1 less received in benefits, paid in taxes, such as NIC, or additional costs incurred, such as travel and childcare, the participant earns around £1.20 to £1.30 in wages and increased tax credits.

And for each £1 invested in the programme:

- the return to DWP is estimated to be £0.83 (a range of £0.56 to £1.10) over the 2-year observed period, rising to £1.76 (a range of £1.21 to £2.32) when impacts are extrapolated over 4 years.
- the return to Exchequer is estimated to be £1.50 (a range of £1.00 to £2.01) over the 2-year observed period, rising to £3.21 (a range of £2.17 to £4.25) when impacts are extrapolated over 4 years.
- the return from a broader societal cost benefit perspective, is estimated to be £1.75 (a range of £1.16 to £2.31) over the 2-year observed period, rising to £3.51 (a range of £2.42 to £4.51) when impacts are extrapolated over 4 years.

Similar observations can be made for estimates using in employment irrespective of OW benefit status, and can be seen in Appendix I. A graphical representation of the cost benefit ratios are presented in Figure 5.4 below.

Figure 5.4 Cost benefit ratios for impacts estimated over four years using in employment and not in receipt of an OW benefit



5.8. Conclusions of cost benefit analysis

The observed impact estimates from Chapter 4 have been used in a cost benefit analysis to quantify the return on investment. The impacts are unlikely to decay quickly, therefore a cost benefit analysis has also been performed on one and two year extrapolations.

For the observed time periods, the main estimates suggest that for each £1 spent on the programme, the return is £0.83 on average in DWP benefits, £1.50 back to the Exchequer in reduced benefit spending and increases in income tax – i.e. fiscal savings and £1.75 on the wider society measure. In addition, participants on average, are better off following their participation in the programme. When possible time effects are considered, then the return to the Exchequer could range from £1.00 to £2.01 for each £1 spent over the observed two-year period.

Under the assumption that the impact can be extrapolated for an additional two years then for each £1 spend is estimated to return £1.76 in DWP benefits with a range of £1.21 to £2.32 to account for time effects, £3.21 back to the Exchequer with a range of £2.17 to £4.25 or £3.51 on a wider society measure with a range of £2.42 to £4.51.

It is difficult to make comparisons between results of this programme and previous programmes given the different designs, participant groups, and contexts under which they operated. However, when we consider the overall evidence that we described in Section 2.6 we can make the general statement, that the results suggest that the effectiveness of part of the programme that we can observe (the end of the programme), as measured by return on investment, can be broadly said to be in the upper end for labour market programmes.

The natural question is then to ask whether we have any evidence to suggest what features in the design might have led to the results that we observe. This is a very challenging question with no definitive answer mainly because the provision was designed to be personalised and tailored to the needs of the participant, with multiple

and different interventions. However, we can briefly reflect on some of the wider evidence base described in Section 2.6 alongside the relevant findings from qualitative and survey research with providers (Foster et al. 2014) and participants (Newton et al. 2014) that we described in Appendix C to understand better the nature of the personalisation and interventions used. However, we need to bear in mind that the research was performed earlier in the programme when the programme was less mature, caseloads were higher and before the time that the results of the study had been drawn.

Overall when we consider the evidence it aligns broadly with some of the elements that appear to be important for providing effective support (e.g. Hasluck and Green, 2007). These include: a strong focus on support to finding a job through a work focused approach, provision personalised to needs, the importance of the role of the adviser to build a good relationship with the client – meeting regularly and giving continuity over a period of time, and some in-work support. Most participants reported that that had received enough support to find work and of those who didn't feel that they said they would have benefited from more support more effective, personalised advice and support. Of those who had not found work, two in three felt that the programme had made them 'a little' or 'a lot' more likely to find work. At the same time, some groups (particularly older and disabled) were less likely to report that provision was adequate and those that did not get a job offered mixed views about the provision. This suggests, as Carter and Whitworth (2015) and Scholz and Ingold (2020) highlight, that there was some evidence of paying less attention and therefore less focus on those people further away from the labour market which needs to be addressed better in the design of future programmes.

6. Conclusions

This report presents the results of an impact evaluation and subsequent Cost Benefit Analysis (CBA) for participants in Work Programme Payment Group 2 which was for individuals aged 25 years and older who had been in receipt of Jobseeker's Allowance or Universal Credit in the Searching for Work conditionality group for 12 months.

6.1. Impact analysis

The main analysis uses an Intention to Treat (ITT) approach to compare the employment and benefit status of individuals who became eligible for Work Programme Payment Group 2 between January 2017 and March 2017 with a matched group of individuals who would have been eligible the programme in the three months after it ceased between April 2017 and June 2017.

The analysis exploits the fact that individuals who would have been eligible for the Work Programme, just after the end of the Work Programme, but did not receive an equivalent offer and so form a plausible comparison group. The period of the study occurred in a time when the labour market was gradually improving, which might, all things equal, underestimate the estimates.

The main estimates suggest that within two years following eligibility, the Intention to Treat group spent 17 more days in employment and 26 fewer days in receipt of out of work benefits and 9 more days in neither in employment nor out of work benefits. Since the Intention to Treat groups contained a proportion of individuals that were not referred to the programme, we can estimate the impact of participation if it can be assumed that all of the observed impact is due to participation in the programme. In this case, the results suggest that programme participants experienced 46 additional days in employment and not in receipt of an Out of Work benefit, 70 fewer days in receipt of an Out of Work benefit and 24 more days neither in employment nor out of work benefits, over the two years following their eligibility to the programme.

However, there are a number of uncertainties: although a number of sensitivity analyses have been performed and a range of cohorts have been used, it cannot be guaranteed that all the differences that might affect outcomes, between the groups, have been captured.

Propensity score matching was used to control for observable differences in composition and this led to groups that were very well balanced prior to reaching eligibility. However, more potentially significant issues arise from drawing cohorts from different time periods. Comparisons of individuals from during and after the programme suggest that there are some time effects but they are relatively small compared to the main estimates: the maximum observed fluctuation due to time and seasonality effects for the impacts of participation is 16 days in employment and 22 days in receipt on benefit. It is not absolutely certain that the time effects for the period from which we infer the impact would be greater or smaller than the time effects observed from these other consecutive cohorts, however this does give some

assurances around the main estimates, while bearing in mind that are still some uncertainties.

Another point is that the analysis was during a period when UC was being rolled out across the country: since the comparison group is drawn three months later and it is difficult to map legacy benefits onto UC benefits (our definitions will both under-capture some outcomes or over-capture some outcomes that may not have been captured in the legacy system), this introduces complexity and therefore some additional uncertainty around the impacts – however, we think that effect cannot be large given that both referral cohorts had the same low proportion (a fifth) of UC claims.

In addition, by looking at off-flow rates from before eligibility there was no apparent evidence of any anticipation effects which may bias the analysis. Also, by looking at proportion referred to other provision it can be observed that a higher proportion (20%) of the comparison group were referred to other provision compared to the treatment group (10%). If the other programmes have any impact, then this would push the estimates up slightly.

It is clear that the impacts will continue for longer than two years following a referral and also participants will still be receiving support for over two years. This provides a rationale for an extrapolation of these impacts which are shown for a further one and two years.

There are a number of factors that limit the external validity, or generalisability, of these results, which need to be borne in mind. Firstly, the impacts were only drawn from the end of the programme for a particular payment group in a particular labour market. It is impossible to know whether the more important influence was whether the programme was at its most mature and effective or whether indeed the programme was starting to wind down - which would mean the estimates would be conservative. Other factors, notably that the period of interest covered the time when decreasing proportions of people who were eligible went on the programme and when the benefit system was changing as Universal Credit was rolling out, also mean that the estimates cannot be safely extrapolated backwards to earlier in the programme; equally some care needs to be taken in generalising these results too far for other times.

6.2. Cost benefit analysis

A Cost Benefit Analysis (CBA) of these impacts have been carried out using the Department for Work and Pensions' Social Cost Benefit Analysis model.

Using the observed two year impacts for each of the perspectives shows that the main estimates are already, even at this point, net positive for participant, the Exchequer and society (figures in brackets are the ranges as given from the potential time effects):

- For the participant: £1.25, with a range of £1.22 to £1.27
- For DWP benefits: £0.83, with a range of £0.56 to £1.10
- For the Exchequer: £1.50, with a range of £1.00 to £2.01

- For society: £1.75, with a range of £1.16 to £2.31

An important point here is these estimates do not rely on the assumption that the impact from the ITT groups is all due to the Work Programme participants; if some of the 'impact' came from the non-participants then the net benefits would be the same and this cost-benefit analysis would give the same results.

It is clear that the trajectory of the impacts indicates they will extend beyond the tracking period, which is likely to give a fairer representation of potential returns. It is difficult to determine how long they will continue but it is reasonable to suggest that impacts could be extrapolated for a further one or two years. For an extrapolation of a further two years (i.e. a total of four years) the analysis shows that returns would be net positive for participant, exchequer and society:

- For the participant: £1.26, with a range of £1.23 to £1.27
- For DWP benefits: £1.76, with a range of £1.21 to £2.32
- For the Exchequer: £3.21, with a range of £2.17 to £4.25
- For society: £3.51, with a range of £2.42 to £4.51

Given that the main assumptions for the CBA are the assumptions for the impacts, the same caveats also apply to these estimates as for the impacts.

Care is needed in making comparisons with other labour market programmes but it is reasonable to suggest these results that have been observed for this end part of the programme would lie in the upper end of results for labour market programmes. It is very difficult to bring this study together with previous older evaluation evidence drawn from earlier in the programme to understand why we observe these results but we can tentatively suggest that the provision had a number of features which align with the overall evidence base for providing effective employment provision, e.g. importance of the adviser support and continuity. On the other hand, we bear in mind that previous programme evidence also suggested that the provision did not work as well for some people (e.g. older and those with health and disabilities) – although many of those will have not been within this particular payment group.

Appendix A. Work Programme Payment Groups

As the Work Programme was such a large programme, it provided support to a broad range of individuals. It was acknowledged that they would have different support needs, so referrals were made into one of several Payment Groups. These Payment Groups evolved over the life of the programme, most notably Payment Group 6 split into Payment Groups 6A and 6B to separate individuals in receipt of Employment and Support Allowance (ESA) according to their prognosis, and the introduction of Payment Group 9 for prison leavers in receipt of Jobseeker's Allowance (JSA) which was not open to referrals from the start of the Work Programme.

People were referred to the Work Programme on both a mandatory and voluntary basis, dependent on their circumstances, and all referrals were made via Work Coaches in Jobcentre Plus (JCP). The Payment Groups were largely driven by the benefit that the individual was in receipt of:

1. Jobseeker's Allowance (JSA)
2. Universal Credit (UC) in the Searching for Work conditionality group
3. Employment and Support Allowance (ESA)
4. Incapacity Benefit (IB)
5. Income Support (IS)

Table A.1 below shows each of the ten Work Programme Payment Groups.

Table A.1 Work Programme Payment Groups

Payment Group	Descriptor
1	Individuals aged 18 to 24 years old in receipt of JSA or UC with Searching for Work conditionality for 9 months
2	Individuals aged 25 years and over in receipt of JSA or UC with Searching for Work conditionality for 12 months
3	Individuals in receipt of JSA or UC with Searching for Work conditionality meeting 'early access' conditions
4	Individuals in receipt of JSA or UC with Searching for Work conditionality, previously in receipt of IB
5	Individuals in receipt of ESA or a UC equivalent volunteering for the programme
6A	Individuals in receipt of ESA or a UC equivalent with a prognosis of 3 or 6 months
6B	Individuals in receipt of ESA or a UC equivalent with a prognosis of 12 months or over
7	Individuals in receipt of ESA or a UC equivalent, previously in receipt of IB
8	Individuals in receipt of IS or IB in England only
9	Individuals in receipt of JSA or UC with Searching for Work conditionality, from day one of release from prison

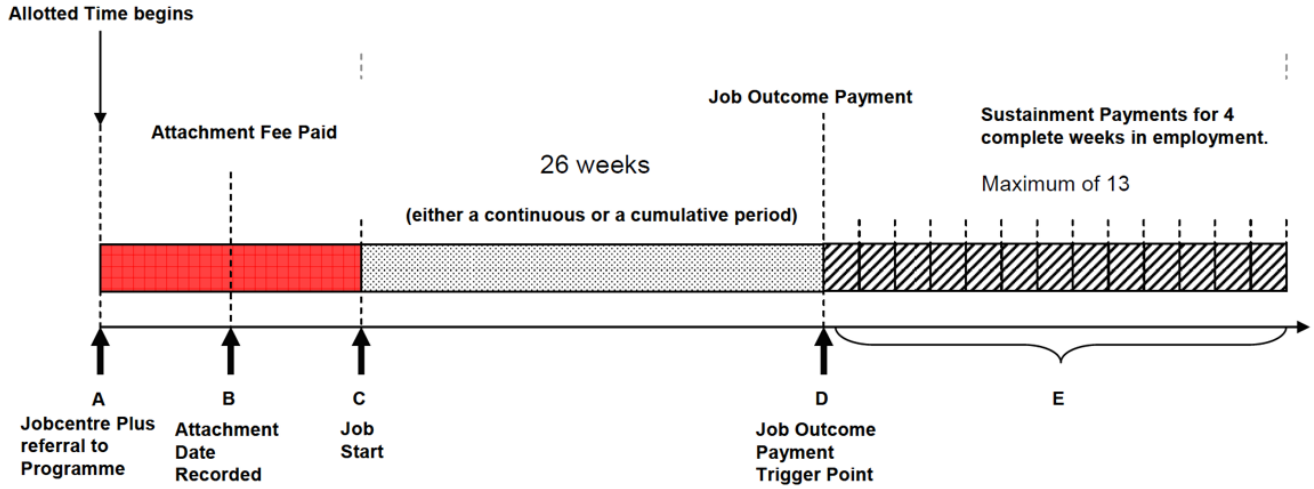
Appendix B. Work Programme Payment Model

The Work Programme operated almost exclusively on a Payment by Results (PbR) approach. Outcome payments included a single Job Outcome (JO) payment, Sustainment Outcome (SO) payments and Incentive payments. In the early years of the programme, providers were also paid “attachment fees”, which made up 20% of the total spend over the life of the programme.

Attachment fees and outcome payments were paid at different rates depending upon the Payment Group to reflect the different levels of support provided. For Payment Group 2, the focus of the analysis, a single JO could be claimed by a provider after a participant had been in work for a continuous or cumulative period of 26 weeks. Once a JO had been claimed, providers could claim SO payments. These were paid when participants remained in employment for a period of 4 complete weeks in. For PG2, a total of 13 SOs could be claimed for each 4-week period of employment.

Figure B.1 below presents the Payment Model diagram to illustrate some of the payment elements.

Figure B.1 The Work Programme payment model



Appendix C. Summary of other Work Programme evaluation evidence

A brief summary of the evidence on the Work Programme, drawn, and in some places reproduced, mainly from Meager (2014):

- Providers reported that they normally deliver support through personal advisers, usually face-to-face, and aim to offer continuity of adviser support. Participants, however, suggested that such continuity was less commonly experienced by some groups (for example, older participants) than others. Fortnightly meetings were most common and most participants were happy with the frequency of contact.
- In line with previous schemes, there was more evidence of 'work-first' approach (job search support to get people quickly into work) rather than on human-capital based approaches (for example, training programmes). Most participants received help with CVs, job search and interview techniques while few were referred to training provision or to support designed to address specific barriers to employment (for example, health conditions, accommodation problems or caring responsibilities). Evidence suggested limited use of subcontractors (especially specialist providers) in supply chains to deliver support interventions, but rather support being delivered through generalist, in-house staff.
- Most participants who cited difficulties finding work reported that the interventions received were helpful in overcoming their barriers and moving closer to work. However, some groups (older, disabled and better-qualified participants in particular) were less likely to report the interventions as helpful.
- Overall, most participants thought the support they received was adequate, although disabled people and people with health conditions, and highly qualified participants were significantly more likely to feel that they had not received enough support. Participants with health conditions and disabilities often did not feel ready to progress towards work – they were much more likely to be looking for support related specifically to medical or disability matters and they were also less likely than participants as a whole to wish for more meetings or contact with advisers.
- Half the participants in work while on the programme reported that they had received in-work support (especially participants with caring responsibilities, or those with a long period since they had last worked). Most felt the amount of in-work support they received was about right and had not felt pressurised by providers to stay in work.

Qualitative research on those who had completed the programme, but not found a job, offered some additional insight, such as:

- Some, who had a good relationship with providers, wanted to remain on the programme, looking for work; others, less satisfied with their contact with the provider, were keen to leave the programme.
- Some (especially older participants, and with health conditions) believed they were too ill to work, and reported having little support from providers (often because their conditions inhibited regular contact). Others completing their time on the programme, mainly JSA claimants, remained optimistic about their employment prospects, and a further group were planning entry to further education or training on leaving the programme.
- As with other participants, this group had mixed views on whether the programme had made a difference to them. Some appreciated positive and supportive adviser contact, but this did not always lead them to feel that the programme had made a difference. Others highlighted benefits such as an improved CV or greater confidence as a result of the programme. Some of those completing the programme criticised it for not delivering the promised personalised support, and some highlighted a need for more contact time with advisers, and more access to training linked to labour market opportunities.

There were a number of findings about mandation which include:

- Most participants were aware of the mandatory nature of the programme and the implications of not engaging with it, and a general acceptance that such an approach was ‘reasonable’ in principle.
- 40% of participants said that awareness of the threat of sanctions made them more likely to comply with provider requests, but slightly more than half felt the sanctions regime had made no difference to compliance.
- From participants’ accounts there was little to indicate that they believed that the threat and operation of sanctions had changed their job search behaviour or had increased their likelihood of entering work.

On the issue of personalisation, findings included:

- Providers were seen by participants as delivering a high level of ‘procedural’ personalisation, creating friendly, mutually respectful relationships with participants, and using assessment and action-planning tools which incorporated a degree of ‘procedural’ personalisation in their operation.
- Although there was less evidence of ‘substantive’ personalisation in the sense of delivering customised support services to individual participants, tailored to specific needs, the majority of participants said they received support that matched their needs either very or fairly well. For these participants, a standardised service was deemed sufficient and appropriate because the interaction with the adviser provided the individualised support that many appreciated. Some participants benefited from frequent meetings while for

others (for example, those waiting for external interventions such as health services), meetings spaced months apart were welcome and appropriate.

- A minority (particularly older and more highly-qualified participants) felt their needs weren't met because of insufficient personalisation.
- The way in which providers engaged with participants (particularly early on), the style of engagement adopted by personal advisers and the extent to which interventions were seen by participants as 'appropriate', were important influences both on participant job search motivation and on their commitment and willingness to engage with the programme.

Finally, findings about variation of provision included:

- Evidence suggested that participants' readiness to work and other characteristics (but not payment group) were used by providers to vary the frequency and intensity of support they received.
- Participants confirmed that continuity of adviser contact was the norm: at the two-year point 70% reported seeing the same adviser always or almost always, indicating a high level of adviser continuity (although older participants reported less continuity).
- One in ten participants did not receive any additional support beyond adviser meetings (more prevalent amongst women, the youngest and oldest participants and those with health conditions/disabilities). There was little evidence that providers had offered specialised and targeted support to help participants address particular barriers to work;
- Participants with health conditions and disabilities often reported a different experience from those in other groups, although many felt this was appropriate to their circumstances.
- Some of the variation in support apparent in the early stages of the programme had diminished 18 months on.
- Participants with health conditions and disabled people reported different experiences from other participants, though most were content with the level of support received. Sometimes these participants were offered less frequent, but longer appointments, and/or a frequency of appointments that they saw as appropriate for their needs or their ability to work. It also seemed that some of these differences may have moderated over time as the programme developed.
- Other evidence, notably Whitworth and Carter (2015, 2020), particularly draws attention to the evaluation evidence and incentives for focussing on resources towards to those closest to the labour market ('creaming') and paying less attention towards those further from the labour market ('parking'); Ingold (2020) in an in depth qualitative study showed that the employment

preparation mechanism pushed disabled participants further away from paid employment, rather than towards workplace inclusion.

Appendix D. Propensity Score Matching and data

D.1. Why use propensity score matching?

Propensity score matching is often employed in evaluations of this type in order to more accurately estimate the counterfactual, i.e. the outcomes that would have occurred in the absence of the programme. It is impossible to measure the counterfactual directly, since it is impossible to know what outcomes participants would have attained without the programme, so it is necessary to estimate this figure using outcomes for individuals who did not participate.

Referral to Payment Group 2 was mandatory for all eligible individuals, meaning that a suitable non-participant group does not exist for the time the Work Programme was accepting referrals. It is possible that differences may occur between the groups. Any relevant differences between the participant and non-participant groups would introduce bias into the impact estimate if this was calculated using the non-participant group as a whole to estimate the counterfactual.

The aim of employing propensity score matching is to eliminate this potential source of bias by creating a comparison group which is identical to the participant group in terms of all relevant observable characteristics. A testing plan, as detailed in section 3, was employed to test whether differences between the groups were relevant and potentially affecting any observed impact.

The assumption that the matching has successfully controlled for all factors affecting both the likelihood of participation in the programme and the likelihood of attaining the outcomes under evaluation is referred to as the ‘conditional independence assumption’. If this assumption is met, then the two groups being should be indistinguishable from two groups selected at random from the same population and they are the same in all regards except for the actual treatment. The result is that any differences observed after this process can be attributed as the impact of the programme itself.

This PSM approach is broadly consistent with a number of previous Department for Work and Pensions (DWP) impact assessments of labour market initiatives, for example for Sector-based Work Academies by Ward et al. (2016). This analysis introduces some minor refinements to the PSM methodology used in earlier studies, in terms of the variables employed to create the propensity score for each individual. The new variables have been added as a result of new data becoming available, with the aim of improving the accuracy of the matching. A full list of the variables used is provided in section D.3.

D.2. How propensity score matching is carried out

The starting point for propensity score matching is to define an overall sample containing both participants and non-participants in the intervention under consideration. The methodology applied to select the samples used in this evaluation is described in Appendix E. Once the sample has been defined, propensity score matching is carried out as follows:

1. Data on the characteristics of individuals in the sample are used as the input to a logistic regression model, here using a logit approach, to estimate the

probability of each individual participating in the programme. This probability is also known as an individual's 'propensity score'.

2. The propensity scores determined in part (1) are then used to match participants to individuals in the comparison group with a similar likelihood of participating in the intervention. Here, the matching approach used was single nearest neighbour with replacement, meaning one non-participant could be matched to multiple participants, i.e. many-to-one.

Under this approach each participant is matched to the single non-participant with the closest propensity score. To ensure that the proximity of a non-participant was appropriate, and avoid matching between participants and non-participants with very different propensity scores, a caliper was applied. A non-participant would only be matched to a participant if the difference in their propensity scores was less than the width of the caliper. In this case the caliper was set at 5% of the standard deviation of all propensity scores. This caliper width is similar to other ongoing impact analysis, but smaller than previous analyses, which have extended to a width of 20% of the score standard deviation and allowed matches with a relative larger difference between propensity scores. Replacement of the non-participants allows a non-participant to be matched to multiple participants.

At the end of this process, the desired outcome is for the large majority of participant records to be matched with a record in the non-participant group. The records where matching has been successful are described as 'on support'. The proportion of records which are on support is one of the main criteria used to assess the quality of the matching result. The degree of common support achieved in this analysis is described in section 4.1.

The other major success factor is the bias reduction achieved by the matching, i.e. how similar the matched comparison group is to the participant group compared with the unmatched comparison group based on their observed characteristics. Monitoring of characteristics for the participant, non-participant and Intention to Treat cohorts, along with an assessment of Standardised Mean Differences (SMD) and whether differences are statistically different, is discussed in Appendix F.

D.3. Variables used for matching

As described in section D.2, the scoring element of propensity score matching was carried out using a logistic regression model. A range of characteristic variables were selected to generate the scores. The characteristic variables selected for both the participant and non-participant matching and the matching of the Intention to Treat (ITT) groups are described below. Section E.1 details the exclusions made from the participant group before the scoring took place.

D.3.1. Personal characteristics

Characteristic information for individuals included in the participant, non-participant and Intention to Treat groups was gathered across a range of characteristics:

- The individual's age on the date of referral, or equivalent. This was used as a continuous decimal variable.

- The recorded gender of the individual.
- The parental status of the individual, stored across three separate binary variables. These binary variables stored if the individual was recorded as a lone parent, part of a couple with children, having no children.
- The recorded ethnicity of the individual, stored in two separate binary variables: white; all other recorded ethnicities.

D.3.2. Benefit history

Benefit histories were drawn from DWP administrative datasets and aggregated as described in section D.4. For each of the 104 weeks prior to the referral, or equivalent, date the presence of a relevant benefit spell was monitored. A week is classed as having a benefit spell within it if at least one day of the spell occurs within that week.

For the purposes of matching, weeks were selected at a 13-week interval to be included in the scoring model for both spells of Jobseeker's Allowance (JSA) or Universal Credit (UC) with Searching for Work conditionality, and other Out of Work (OW) benefits separately. To improve the quality of the match, following initial testing, variables for additional weeks around the point of eligibility were included. Included in the final model were variables denoting the presence of a benefit claim as follows:

1. JSA or UC Searching for Work conditionality at 1, 2, 3, 4, 13, 26, 39, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 78, 91 and 104 weeks
2. Other OW benefits at 13, 26, 39, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 78, 91 and 104 weeks

These 13-weekly intervals were selected to allow spells to be monitored at a relatively granular level, without limiting the ability to find non-participant matches for participants or being processing intensive. It also allowed for additional data points to be included where necessary.

Additional variables indicating if in the 2 years prior to the referral, or equivalent, date an individual has had:

1. Claims to Disability Living Allowance (DLA) or Personal Independence Payment (PIP)
2. Claims to UC in any conditionality group
3. Referrals to any other Contracted Employment Provision as recorded by the Provider Referrals and Payments (PRaP) system

D.3.3. Employment and earning history

Spells of employment were monitored in a similar way to spells of benefits, i.e. for each of the 104 weeks prior to the referral, or equivalent, a week was counted as having a spell of employment if it contained at least one recorded day of employment. These spells were included in the model at 13-week intervals, at 13-, 26-, 39-, 52-, 65-, 78-, 91- and 104-weeks.

Using Real Time Information (RTI) to monitor employment spells also enabled earnings to be tracked and included in the scoring model. Individual earnings were aggregated for each 3-month period prior to the referral, or equivalent, date, for the 2-year pre-programme period. This resulted in eight separate variables relating to earnings in each of those periods.

D.3.4. Additional characteristics

Geographical information in the form of an individual's recorded postcode mapped to one of the 18 Contract Package Areas (CPAs) for the Work Programme was included in the matching model. Further information on local unemployment rates at given points in time to represent characteristics of the local labour markets were also used in sensitivity tests of the matching model.

D.4. Sources of data

The Department for Work and Pensions (DWP) holds information about individuals claiming benefits and their claims and people participating in employment programmes in the form of administrative datasets. These datasets vary in their scope according to the benefit or programme they relate to. To maximise the use of this data in this study, information from many sources has been drawn together to create a rich composite dataset for all groups that were being compared. The purpose of these sources in this analysis can be broken down into three main categories:

- Benefit and characteristic information;
- Employment programme information; and
- Employment and earnings information

D.4.1. Benefit and characteristic information

Benefits delivered by DWP utilise several administrative platforms, so the information on the benefit spells, such as start and end dates, and claimant characteristics, such as age and postcode, are stored in different administrative datasets. To ensure the fullest coverage possible of benefit claims, the following sources of data were used:

1. National Benefits Database;
2. Universal Credit Reference Sets; and
3. Personal Independence Payment Atomic Data Store.

Additional characteristic information was drawn from:

- Jobseeker's Allowance Payment System Atomic Data Store;
- Client Extract;
- Work Programme Analytical Dataset.

D.4.2. Employment programme information

Since 2009, employment programmes contracted to third-party providers, including the Work Programme, have been administered using the Provider Referrals and

Payments (PRaP) system. The PRaP system records details of all referrals made to provision and commercial outcomes achieved. Most programmes are subdivided into Contract Package Areas (CPAs), which generally relate to contiguous geographical regions, where the programme is delivered by a specific third-party provider. Information gathered in the PRaP system is made available for analysis through the Contracted Employment Provision Management Information dataset, which has been used in this study to obtain information on Work Programme referrals and referrals to other contracted employment programmes.

D.4.3. Employment and earnings information

Previous evaluations of other provisions have used employment data from the Work and Pensions Longitudinal Study (WPLS), which captured details of employment start and end dates taken from employer submissions to HMRC on P45 and P46 forms. The quality of this data was known to be limited, since historically P45 and P46 forms were not required to be submitted for people whose earnings were below the Lower Earnings Limit, and employment start and end dates were often approximate or missing. In addition, the dataset did not capture any information on people who were self-employed.

Starting in 2013, the P45/P46 reporting system to HMRC was phased out in favour of the Real Time Information (RTI) system. Under RTI employers must submit information to HMRC each time an employee is paid, including the amount paid. RTI offers substantial improvements to the P45/P46 system, as employers must now provide information on all of their employees if even one employee of the company is paid above the Lower Earnings Limit. Previously under the P45/P46 system, submissions were not required for employees earning below the Lower Earnings Limit.

Full rollout of RTI was achieved in 2014, and all employers have been required to report in real time with 1.9 million schemes covering 48 million employees[22], meaning that all individuals selected for each of the cohorts studied in this evaluation have RTI employment information for at least 2 years prior to their referral or eligibility to the programme.

As well as improvements to employment start and end dates, RTI also captures the amount paid to an individual so can be used to determine earnings for a given spell of employment. However, RTI does not capture any information about earnings through self-employment.

Through a data sharing agreement, HMRC provides DWP with RTI for claimants of Jobseeker's Allowance (JSA) and Universal Credit. By definition, this includes participants Work Programme Payment Group 2 and therefore non-participants to compare to.

D.4.4. Data not included

Education is a key area not included with the data used, including both past participation, achieved qualification levels and future intention to study. Higher Education qualifications are available via the Department for Education's

[22] 2014 HMRC – RTI: improving the operation of PAYE

Longitudinal Education Outcomes (LEO) data, which commenced in 2016. However, given the age group covered for this study are individuals aged 25 years and older, with a mean age of around 42 years, it unlikely that this data would provide substantial coverage for individuals assessed and was therefore discounted.

D.5. Data cleaning and structures

As outline in the previous section, there were several sources of Department for Work and Pensions (DWP) administrative data employed in this analysis to gather information on the characteristics of individuals selected in any of the cohorts used for matching.

It is not mandatory for all characteristics information to be captured, so by drawing on multiple sources it was possible to gain more complete data, for example where the ethnicity of a person was not recorded as part of one benefit claim but was recorded at a later date. When collating information on a single characteristic from multiple sources, it is possible that there may be some conflicting data. When selecting which version of the characteristic to use, the most commonly occurring non-missing value was selected.

For input into the propensity score model, characteristics were coded into separate binary variables. For example, the categorical variable ethnicity became three separate binary variables: white; Black, Asian, Minority ethnic and other; and missing.

Similarly, spells of benefit and employment were coded into 208 binary variables spanning the 104-week period before and 104-week period after the referral date, pseudo-referral date or eligibility date, as appropriate. Weeks were defined as 7-day periods counted from the referral, or equivalent, date, with negatives denoting weeks prior to referral and positives denoting weeks following referral or equivalent. A week was marked as having a claim to a particular benefit or time in employment if at least one day of a spell was contained within the week. It is therefore possible for a single week to show a claim to an out of work benefit and some employment for an individual. In this case when assessing outcomes, the claim to benefit is given priority. Results in Section 4.2 include estimates for the size of this overlap of benefit claims and employment.

D.5.1. Universal Credit

In 2013, a change to the benefit system started its phased rollout in the form of Universal Credit (UC). The intention of UC is to replace 6 legacy benefits:

1. Jobseeker's Allowance (JSA)
2. Employment and Support Allowance (ESA)
3. Income Support (IS)
4. Child Tax Credit
5. Housing Benefit
6. Working Tax Credit

UC uses a series of conditionality groups, which determine the amount and type of work-related activity that an individual is required to do to remain eligible for the benefit. These conditionality groups are more fluid than the separate legacy benefits, and claimants move between them much more freely dependent on their circumstances. If a claimant is eligible for more than one conditionality group, they are assigned the group with the lowest level of conditionality, where low levels of conditionality mean there are fewer work-related requirements for a claimant to satisfy to be eligible. In order of lowest to highest levels of conditionality, the groups are as follows:

- Working – no requirements
- No work requirements
- Planning for work
- Preparing for work
- Working – with requirements
- Searching for work

Administrative data in the form of reference datasets, summarises claims to UC into month-long Assessment Periods, to reflect a claimant's circumstances and earnings. For analytical purposes, the reference datasets record a single conditionality group for the whole Assessment Period. Consecutive Assessment Periods have been joined together to form continuous spells of conditionality, in a similar way to the recording of spells of legacy benefits.

UC conditionality groups do not map directly on to legacy benefit types, and a feature of UC design brings Legacy benefits and UC conditionality groups do not map directly onto one another, so it is only possible to approximate equivalencies between the two systems. For the purposes of this analysis, the following approximations have been made:

1. Jobseeker's Allowance to Universal Credit Searching for Work
2. Employment and Support Allowance, Income Support, Incapacity Benefit to Universal Credit Planning for work, Preparing for work, Working – with requirements

Claims of UC with conditionality of No work requirements and Working – no requirements have also been used to monitor a more complete view of individuals and if they have accessed any aspects of UC.

Figure D.1 below illustrates how legacy benefits and UC conditionality groups have been aggregated for the purposes of this evaluation.

Figure D.1 Classification of legacy benefits and Universal Credit conditionality groups

	Legacy benefit	Universal Credit	Out of Work benefits	JSA and UC searching for work	Other OW benefits
Jobseeker's Allowance	✓		✓	✓	
Employment and Support Allowance	✓		✓		✓
Incapacity Benefit	✓		✓		✓
Income Support	✓		✓		✓
Searching for Work		✓	✓	✓	
Working - with requirements		✓	✓		✓
Planning for Work		✓	✓		✓
Preparing for Work		✓	✓		✓
Working - no requirements		✓			
No work requirements		✓			

In addition to changes to the type of people claiming benefit, UC uses a new administrative data system and claimants primarily interact with the benefit digitally.

UC also allows for individuals to claim benefit whilst in employment, tapering according to earning amounts. Whilst this was possible to a limited extent with legacy benefits, it is an explicit policy design of UC.

The number of participants claiming UC or not UC prior to being eligible was too small to be able to perform a sub-group analysis, however this analysis was possible for the participant/non-participant group – as described in Appendix H.

Appendix E. Defining comparison cohorts

E.1. Intention to Treat Groups

As explained in Section 3.1, the main focus of the analysis is on based on comparing groups drawn solely from individuals that meet the eligibility criteria, irrespective of whether they actually participated in the programme. By using Department for Work and Pensions (DWP) administrative datasets, individuals can be identified that:

1. Are aged 25 years or older
2. Have been claiming any combination of Jobseeker's Allowance or Universal Credit with Searching for Work conditionality for the last 12 months, or has linked claims totalling 12 months
3. Have not previously participated in the Work Programme in any Payment Group.

The date at which individuals meet all three of these criteria is known as the eligibility date. Cohorts can then be constructed on an Intention to Treat (ITT) basis as groups of individuals with eligibility dates:

1. Between January 2017 and March 2017 for the ITT treatment group
2. Between April 2017 and June 2017 for the ITT comparison group

These groups of individuals are tracked from the date at which they meet the eligibility criteria, which can be applied equally to both groups and should therefore minimise any selection effects. Additional ITT cohorts can be defined similarly, based on the date that individuals meet all of the eligibility criteria.

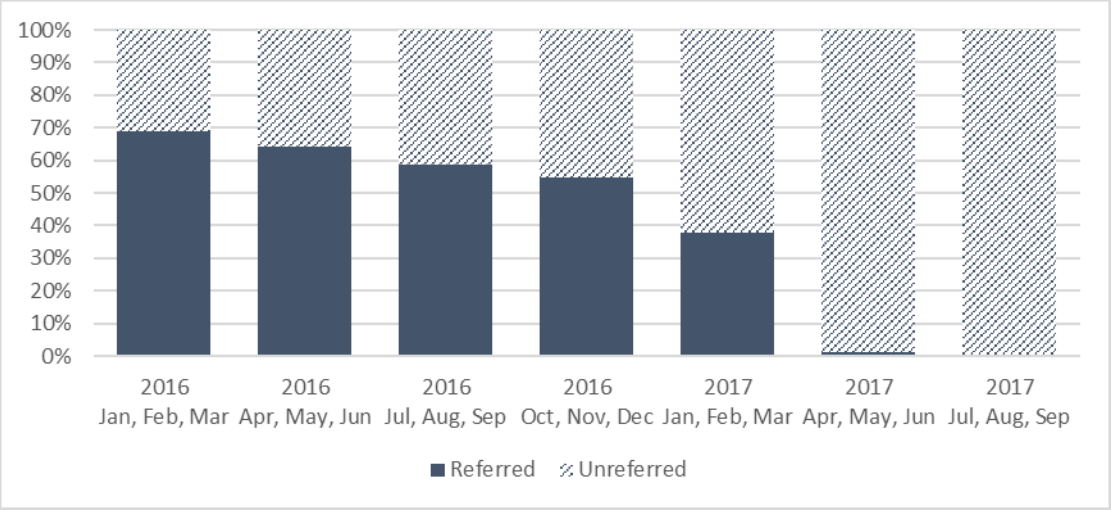
Although this ITT methodology avoids selection biases to the programme, it does introduce other challenges and is still subject to other potential effects.

As the ITT groups are selected irrespective of whether individuals were referred to the Work Programme, the groups are therefore a mixture of participants and non-participants. This means that of the ITT treatment group, drawn from individuals eligible between January 2017 to March 2017, around 38% were participants. Additionally, a small proportion, around 1.5%, of the ITT comparison group, drawn from individuals eligible between April 2017 and June 2017, were also participants.

Further to this, any potential labour market changes or anticipation effects should still be expected as the ITT methodology does not attempt to address these. This means that whilst the proportion of individuals participating in the Work Programme of those that met the eligibility criteria does increase as the group selected is from earlier in the programme, changes to the labour market are likely to become more apparent as the two groups share less time operating under the same conditions.

Figure E.1 below shows the proportions seven distinct cohorts selected on an ITT basis that did receive a Work Programme referral, including the two groups selected for analysis.

Figure E.1 Proportion of ITT cohorts with Work Programme referrals



Whilst Figure E.1 does show that the proportion of each group with a referral to the Work Programme is higher further back in time, this increases the risk of additional time-based biases, particularly labour market changes. As part of investigations into time effects, additional ITT groups have been defined in a similar way for different time periods.

E.2. Participant Group

The original intention of the analysis was to compare a group of the group of referrals to Payment Group 2 between January 2017 and March 2017, known as the 'participant group' with a group of people who would have been eligible, the 'non-participant group'. However, as explained in Section 3 this led to some concerns of selection bias in the participant group which led us to focus on an ITT analysis. Nonetheless, we show here how the participant/non-participant groups are comprised with the results in Appendix H.

To be eligible for participation in Work Programme Payment Group 2, the following criteria should have been met:

1. The individual is aged 25 years or older
2. The individual has been claiming any combination of Jobseeker's Allowance or Universal Credit with Searching for Work conditionality for the last 12 months, or has linked claims totalling 12 months
3. The individual has not previously participated in the Work Programme in any Payment Group.

Some referrals to Work Programme Payment Group 2 during the period January 2017 to March 2017 are excluded from the participant group. These exclusions are made in specific circumstances:

1. The referral was recorded as being inappropriate
2. The individual was referred multiple times, only the first referral was retained

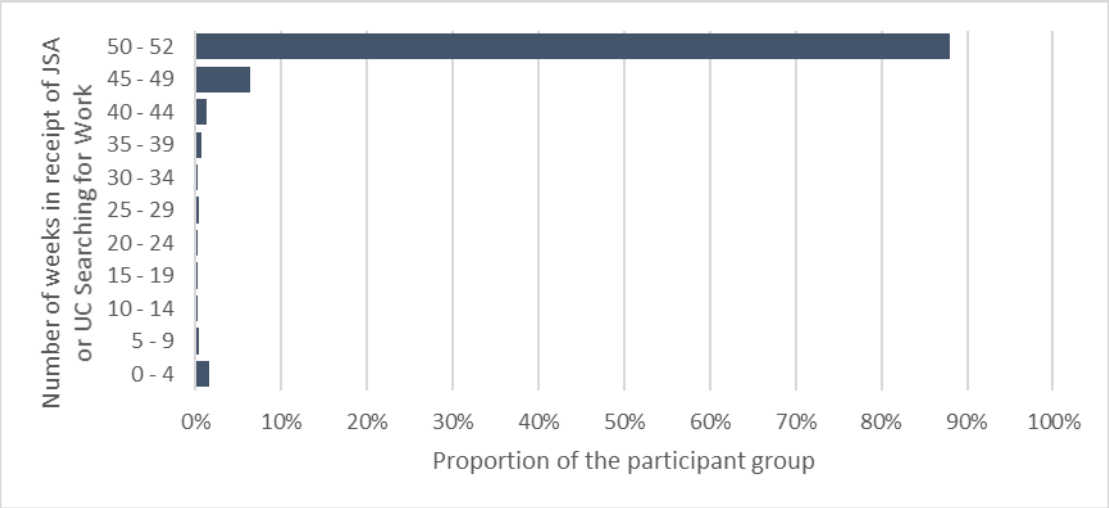
3. The individual had previously participated in the Work Programme

For the first point ensured that the participant group only consists of individuals that contracted providers were expected to engage with as part of the Work Programme: the proportion of referrals recorded as inappropriate was less than 1%. The second point means that repeated and duplicate referrals were excluded to ensure that the participant group was a set of unique individuals and that their engagement with the Work Programme was only evaluated once. Finally, the third point means that previous participants were excluded to ensure that the eligibility criteria were met correctly and allowed for a distinct non-participant group to be drawn. The purpose of this study was to assess the impact of a single instance of participation. Allowing multiple instances of participation in the programme would mean any lasting impact from the first instance could not be disaggregated from any subsequent interactions with the programme. Furthermore, the eligibility criteria state that to have been eligible an individual must not have participated in the programme previously.

Additional exclusions were made where individuals had been referred to the programme, but upon investigation in administrative datasets appeared to have significantly less recorded time on benefit than was necessary to be eligible for referral. It is possible that some of this information was held clerically or that the individual was referred in error. Through linking rules, it is possible that an individual does not have 52 continuous weeks of benefit claims prior to their referral to the programme, but this should result in fewer than 10 out of 52 weeks without a benefit claim attributable.

Figure E.2 below shows the distribution of benefit history coverage for the participant group.

Figure E.2 Referred individuals and the number of weeks with a relevant claim in the 52 weeks prior to referral



This shows a large peak of individuals, around 95%, with 45 or more of the 52 weeks prior to programme referral with at least one recorded benefit claim present. A much smaller peak, around 2%, is also present showing participants with fewer than 5 weeks with a recorded benefit claim. Referred individuals with fewer than 11 weeks (20%) of recorded benefit information were therefore been excluded. This cut-off point was selected to balance the exclusion of individuals with low recorded amounts

and potentially low-quality benefit history information whilst retaining as many individuals for the participant group as possible.

In summary, the participant group comprised of all referrals made to Work Programme Payment Group 2 between January 2017 and March 2017, less the referrals recorded on the PRaP system as inappropriate, less duplicate referrals and referrals made for individuals that have previously participated in the Work Programme, less individuals with fewer than 11 weeks with a relevant benefit claim in the 52 weeks prior to referral. Table E.1 below shows the size of each of these groups and the final size of the participant group.

Table E.1 Number of referrals excluded from the participant group and reasons for exclusion

Group description	Group size
Referrals made to Work Programme Payment Group 2 between January 2017 and March 2017	12,798 (100%)
Referrals recorded as inappropriate	251 (2.0%)
Referrals duplicated or for individuals previously participated in the Work Programme	133 (1.0%)
Referrals with fewer than 11 weeks with a relevant benefit claim in the 52 weeks prior to referral	264 (2.1%)
Participant group	12,150 (94.9%)

E.3. Non-Participant Group

The initial selection criteria used to define a non-participant group tries to mirror as closely as possible the criteria applied in defining the participant cohorts. For Work Programme Payment Group 2, these criteria are:

1. The individual is aged 25 years or older
2. The individual has been claiming any combination of Jobseeker's Allowance or Universal Credit with Searching for Work conditionality for the last 12 months, or has linked claims totalling 12 months
3. The individual has not previously participated in the Work Programme in any Payment Group.

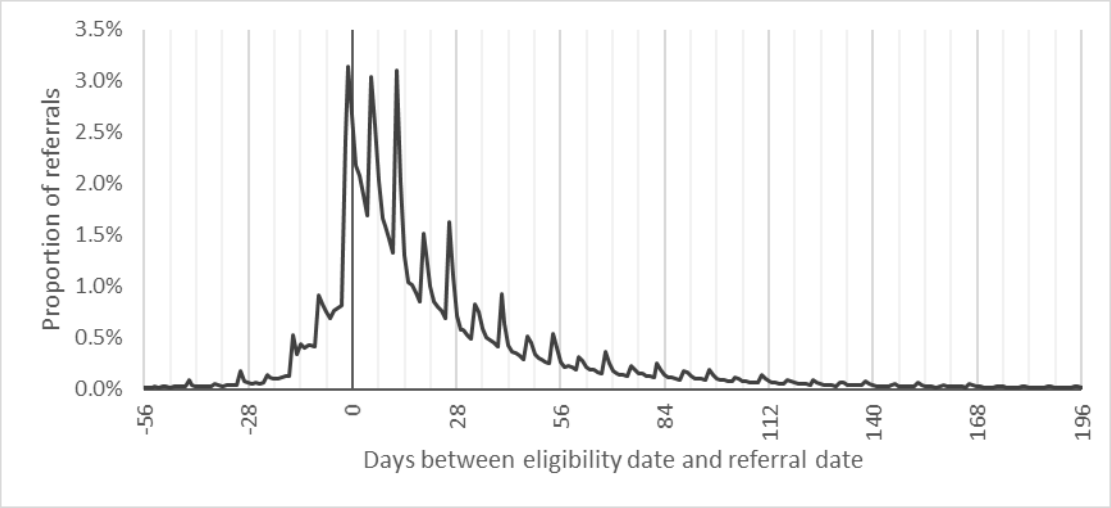
In order to reflect the time taken between an eventual participant meeting all three eligibility criteria and being referred to the programme for the group of non-participants, pseudo-referral dates were assigned to all non-participants meeting the eligibility criteria from administrative datasets. The method employed calculated a distribution of times between meeting eligibility criteria and programme referral for all individuals referred. This distribution was then randomly sampled from for each non-participant and the selected time added on to the data the non-participant met the eligibility criteria to create the pseudo-referral date. This ensured that the distribution of times taken between individuals meeting the eligibility criteria and their corresponding referral or pseudo-referral dates was identical.

Once assigned, pseudo-referral dates for non-participants were subsequently treated as equivalent to actual referral dates for participants. For example, pseudo-referral dates were used to align the time periods over which outcomes were compared between the participant and non-participant groups in order to determine the impact of the programme.

Following the assignment of pseudo-referral dates, the non-participant group to be matched to the participant group was selected as individuals with a pseudo-referral date between 01 April 2017 and 30 June 2017, i.e. the next 3-month cohort of individuals following the end of referrals to the Work Programme.

The distribution of times taken for participants between meeting all three eligibility criteria and actual referral has very long tails, where individuals were either referred much earlier or much later than expected. Around 10% of referrals were made more than 90 days before or more than 180 days after records showed the individual was eligible for the programme. The date of eligibility could not be calculated for around 2.5% of all participants. Figure E.3 below shows the distribution in days between the date a participant met the programme eligibility criteria and when their referral to the programme took place.

Figure E.3 Difference between eligibility date and referral date for Work Programme Payment Group 2 participants



* A negative number of days indicates that the referral took place prior to the eligibility criteria being fully met.

To ensure that non-participants were assigned pseudo-referral dates allowing relevant comparisons to take place, the distributions of times from eligibility to referral was limited to between 30 days prior to eligibility and 90 days after eligibility; this period accounted for over 75% of referrals made to Payment Group 2. Thus individuals meeting the eligibility criteria between 01 January 2017 and 30 July 2017 form the wider population from which the non-participant group is drawn. This means that some individuals who became eligible for the programme whilst referrals were still open, but were ultimately not referred to the programme, will form part of the non-participant group.

The number of individuals retained in the unmatched non-participant group following this process is given in Table F.3.

E.4. Side-by-side comparison of selected groups

Table E.2 shows the comparison of the both ITT and participant/non-participant groups used in the study.

Table E.2 Composition of selected groups for comparison

Participant Group	Non-participant Group	ITT Treatment Group	ITT Comparison Group
Individuals referred to Work Programme Payment Group 2 between January 2017 and March 2017	Individuals with a pseudo-referral to Work Programme Payment Group 2 between April 2017 and June 2017	Individuals meeting Work Programme Payment Group 2 eligibility criteria between January 2017 and March 2017	Individuals meeting Work Programme Payment Group 2 eligibility criteria between April 2017 and June 2017
100% Work Programme participants	0% Work Programme participants	38% Work Programme participants	1.5% Work Programme participants
Excludes individuals with previous referrals to the Work Programme			
Individuals with 12 months of claims to Jobseeker's Allowance or Universal Credit in Searching for Work conditionality			
Linking rules applied to join claims together with gaps of less than 1 month between them			

Appendix F shows a more detailed overview of the characteristics of each of the selected cohorts, see.

Appendix F. Characteristics before and after matching

F.1. Characteristics and balance of Intention to Treat cohorts

Table F.1 shows a selection of characteristics for the matched ITT groups. As well as the two matched groups, also shown are characteristics for the comparison group prior to matching and the group of individuals excluded from the treatment group. Note, 16 individuals were excluded from the treatment group after matching.

Table F.1 Selected characteristics of Intention to Treat cohorts

Characteristic	ITT treatment group	Unmatched ITT comparison group	Matched ITT comparison group
Group Size	20,113	18,785	20,113
% with JSA or UC Searching for Work at referral	99%	99%	99%
% with other OW benefit claims at referral	0%	0%	0%
% with DLA or PIP at referral	4%	5%	4%
Mean weeks of JSA or UC Searching for Work	51.8	51.8	51.8
Mean weeks of other OW benefits	0.1	0.2	0.1
Mean weeks of DLA or PIP	2.2	2.7	2.1
Mean weeks of any OW benefits	51.8	51.8	51.8
Mean weeks in RTI employment	2.2	2.1	2.3
Mean age	42.6	42.5	42.6
% aged 18 to 24 years	0%	0%	0%
% aged 25 to 49 years	70%	70%	70%
% aged 50 years and over	30%	30%	30%
% with any UC claim	35%	39%*	35%
% male	57%	55%*	57%
% white	70%	67%*	70%
% non-white	23%	24%	23%
% Lone Parents	16%	17%	16%
% parents	24%	24%	23%
% in Central England and Wales	29%	27%	29%
% in North of England	29%	28%	29%
% in South of England	32%	35%	32%

% in Scotland	10%	10%	10%
% with referral to other contracted provision	19%	15%*	19%

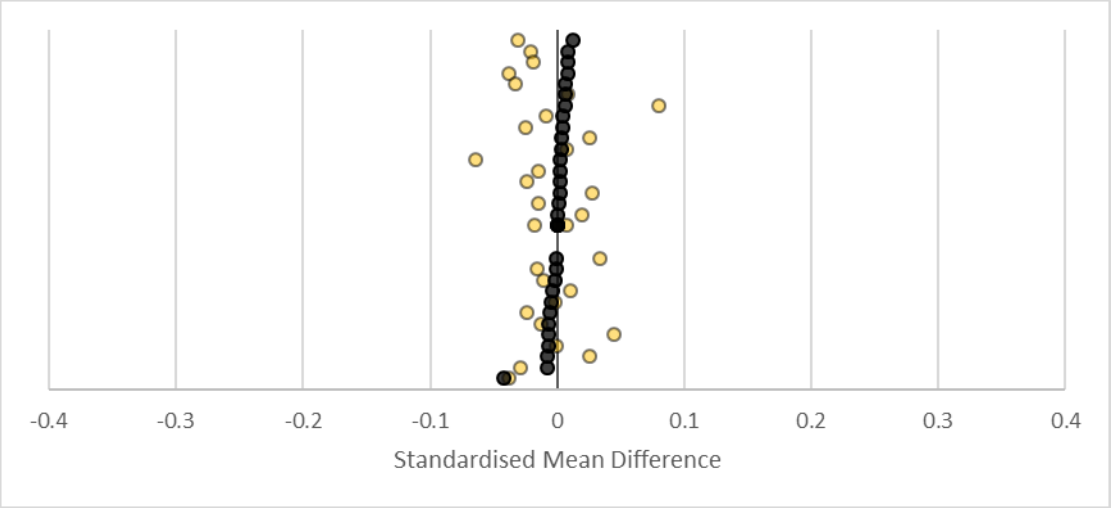
* denotes a statistically significant difference, at the 5% level compared to the treatment group in column 1.

Table F.1 shows that the ITT treatment and comparison groups are very similar across the observed characteristics. Before matching there were only four variables with statistically significant differences as denoted by an asterisk (*) (% male, % white, % Universal Credit (UC) claim prior to becoming eligible and % with referral to other contracted provision) out of 23 variables at the 5% level between the treatment group and the matched comparison group. After matching there were no statistically significant differences across these observed characteristics.

The small number of individuals excluded from the ITT treatment group, 16 in total, exhibit some differences from the ITT treatment group. There are differences across most characteristics, and some of the larger differences include the mean number of weeks out of 52 spent in RTI employment and the proportion referred to another contracted programme, both of these characteristics are much higher for the group of excluded individuals than the ITT treatment and ITT comparison cohorts.

In an alternative way to compare the groups Figure F.1 shows the Standardised Mean Differences (SMD): SMD measures the difference between mean values of two groups in standard deviations. A smaller SMD means the groups are more similar to one another.

Figure F.1 Standardised Mean Differences between characteristics for Intention to Treat cohorts



Overall, the results show that before matching the differences between ITT treatment and comparison groups were quite small: all observed characteristics have SMDs between -0.1 and 0.1. After matching the differences are even smaller: all variables after matching show a Standardised Mean Difference (SMD) of less than 0.015 with only one, (mean time spent with a claim to either Jobseeker's Allowance (JSA) or Universal Credit (UC) with Searching for Work conditionality in the 52 weeks prior to programme eligibility) that has a greater difference, at 0.045.

Table F.2 below shows the characteristics of the ITT treatment and ITT comparison groups broken down by whether they were participants or non-participants.

Table F.2 Selected characteristics of disaggregated Intention to Treat cohorts

Characteristic	ITT treatment non-participants	ITT treatment participants	ITT comparison non-participants	ITT comparison participants
Group Size	12,522	7,591	19,770	343
% with JSA or UC Searching for Work at referral	99%	99%	99%	100%
% with other OW benefit claims at referral	0%	0%	0%	0%
% with DLA or PIP at referral	3%	6%	4%	8%
Mean weeks of JSA or UC Searching for Work	51.7	51.9	51.8	52.0
Mean weeks of other OW benefits	0.2	0.1	0.1	1.6
Mean weeks of DLA or PIP	1.6	3.2	2.1	3.9
Mean weeks of any OW benefits	51.7	51.9	51.8	52.0
Mean weeks in RTI employment	2.5	1.9	2.3	0.7
Mean age	42.1	43.5	42.5	43.8
% aged 18 to 24 years	0%	0%	0%	0%
% aged 25 to 49 years	71%	68%	70%	65%
% aged 50 years and over	29%	32%	30%	35%
% with any UC claim	47%	15%	35%	20%
% male	59%	53%	56%	63%
% white	70%	69%	70%	66%
% non-white	20%	27%	22%	29%
% Lone Parents	14%	20%	16%	14%
% parents	21%	28%	23%	25%
% in Central England and Wales	29%	28%	29%	22%
% in North of England	30%	29%	29%	23%
% in South of England	31%	34%	32%	39%
% in Scotland	10%	10%	10%	15%
% with referral to other contracted provision	16%	26%	18%	100%

Table F.2 shows that the characteristics of the ITT treatment and ITT comparison groups are broadly similar, though some differences exist between participants and

non-participants. Participants from either ITT group were slightly older on average than their non-participant counterparts and were more likely to be lone parents.

Non-participants from either ITT group were more likely to have some evidence of a claim to UC prior to becoming eligible. This is particularly evident for non-participants in the ITT treatment group, who became eligible for referral whilst referrals were still available. Almost half of these have some evidence of a UC claim. As well as being engaged with other labour market activity, it is possible that some of these individuals were eligible for the programme but not referred because operational resource was diverted to transitioning to UC or that claims spanning both UC and legacy benefit platforms were not easily identified in all cases.

F.2. Characteristics and balance of participant and non-participant cohorts

Table F.3 below shows the size of the participant and non-participant groups, as well as the proportion of the participant group remaining 'on support' following the matching.

Table F.3 Participant and non-participant group sizes before and after matching and the proportion of participants on support

Cohort name	Size before matching	Size after matching	Lost through matching	Proportion on support
Participant group	12,150	12,098	52	99.6%
Non-participant group	13,444	12,098	N/A	N/A

Table F.4 shows a selection of characteristics for the matched participant and non-participant groups. As well as the two matched groups, also shown are characteristics for the non-participant group prior to matching and the group of individuals excluded from the participant group, as described in section E.1.

Table F.4 Selected characteristics of participant and non-participant cohorts

Characteristic	Participants	Unmatched Non-participants	Matched Non-participants	Excluded Participants
Group Size	12,098	13,444	12,098	316
% with JSA or UC Searching for Work at referral	97%	99%	96%	34%
% with other OW benefit claims at referral	0%	0%	0%	11%
% with DLA or PIP at referral	6%	5%	5%	5%
Mean weeks of JSA or UC Searching for Work	51.0	51.5	51.5	7.6

Mean weeks of other OW benefits	0.2	0.2	0.1	6.2
Mean weeks of DLA or PIP	3.0	2.7	2.5	2.5
Mean weeks of any OW benefits	51.1	51.5	51.5	14.1
Mean weeks in RTI employment	2.2	2.3	1.9	3.7
Mean age	43.2	42.6	43.5	42.8
% aged 18 to 24 years	0%	0%	0%	1%
% aged 25 to 49 years	68%	70%*	67%	66%
% aged 50 years and over	31%	30%	33%	33%
% with any UC claim	18%	39%*	17%	29%
% male	54%	53%	55%	65%
% white	68%	66%*	70%	66%
% non-white	26%	23%*	25%*	25%
% Lone Parents	20%	18%*	20%	6%
% parents	28%	25%*	27%	13%
% in Central England and Wales	29%	27%	29%	28%
% in North of England	28%	28%	29%	33%
% in South of England	33%	35%	32%	31%
% in Scotland	10%	10%	9%	8%
% with referral to other contracted provision	10%	14%*	9%	5%

* Denotes a statistically significant difference for unmatched and matched non-participant groups in columns 2 and 3 compared to the participant group in column 1.

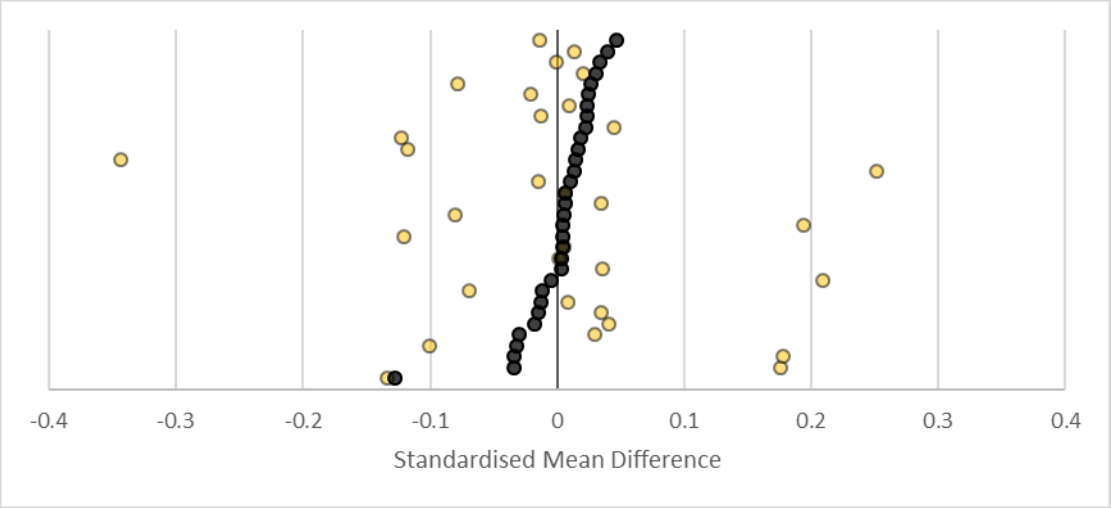
Table F.4 shows that before matching there were a number of statistically significant differences. For example, in the unmatched non-participant group 39% of individuals had some evidence of a Universal Credit (UC) claim prior to their pseudo-referral compared to only 18% of the participant group. In contrast, the matched non-participant group had only 17% of individuals with evidence of a UC claim prior to their pseudo-referral, which is much closer to the participant group. After the matching many of the differences are reduced and the only statistically significant difference that remains at the 5% level is non-white ethnicity.

Table F.4 also shows the characteristics of the group excluded from the participant cohort, as described in section E.1. This group of excluded individuals has an average of 7.6 out of a possible 52 weeks in receipt of JSA or UC in Searching for Work conditionality and only around a third were in receipt of one of these benefits at the time of their referral. This suggests that their exclusion from the participant group is appropriate as they do not appear to fully meet the eligibility criteria. One plausible reason for this is that these individuals were eligible for another payment group but were referred in error to Payment Group 2.

Figure F.2 shows the SMDs for the participant and non-participant groups – the results suggest that the unmatched groups were fairly similar as differences between

most monitored characteristics were between -0.2 and 0.2. Matching the participant and non-participant cohorts brought them into closer to alignment with one another, with only one characteristic having an SMD score outside the range of -0.05 to 0.05. However, it is clear that the balance is not as good as seen for the ITT groups in Figure F.1.

Figure F.2 Standardised Mean Differences between characteristics for participant and matched non-participant cohorts



The largest difference with a SMD of -0.13 is the mean time spent with a claim to either Jobseeker's Allowance (JSA) or Universal Credit (UC) with Searching for Work conditionality in the 52 weeks prior to the referral, or equivalent, date. This is likely due to the strict eligibility criteria used to define the non-participant group by time on JSA or UC. Allowing for linking rules means there may be a some very short gaps in those 52 weeks prior to the eligibility point, but the variance, and therefore standard deviation, of each of the groups will be very small. Consequently, even small differences in the means will appear as relatively large differences in SMD calculations.

F.3. Comparison of characteristics of participant, non-participant and Intention to Treat (ITT) cohorts

For ease of comparison, Table F.5 brings together the participant and matched non-participant groups with the ITT treatment and matched ITT comparison groups.

Table F.5 Selected characteristics of participant/non-participant, & ITT cohorts

Characteristic	Participants	Matched Non-participants	ITT treatment group	Matched ITT comparison group
Group Size	12,098	12,098	20,113	20,113
% with JSA or UC Searching for Work at referral	97%	96%	99%	99%

% with other OW benefit claims at referral	0%	0%	0%	0%
% with DLA or PIP at referral	6%	5%	4%	4%
Mean weeks of JSA or UC Searching for Work	51.0	51.5	51.8	51.8
Mean weeks of other OW benefits	0.2	0.1	0.1	0.1
Mean weeks of DLA or PIP	3.0	2.5	2.2	2.1
Mean weeks of any OW benefits	51.1	51.5	51.8	51.8
Mean weeks in RTI employment	2.2	1.9	2.2	2.3
Mean age	43.2	43.5	42.6	42.6
% aged 18 to 24 years	0%	0%	0%	0%
% aged 25 to 49 years	68%	67%	70%	70%
% aged 50 years and over	31%	33%	30%	30%
% with any UC claim	18%	17%	35%	35%
% male	54%	55%	57%	57%
% white	68%	70%*	70%	70%
% non-white	26%	25%	23%	23%
% Lone Parents	20%	20%	16%	16%
% parents	28%	27%	24%	23%
% in Central England and Wales	29%	29%	29%	29%
% in North of England	28%	29%	29%	29%
% in South of England	33%	32%	32%	32%
% in Scotland	10%	9%	10%	10%
% with referral to other contracted provision	10%	9%	19%	19%

*Denotes a statistically significant difference between participant/non-participant groups in column 1 and 2 and ITT groups in column 3 and 4.

Overall, we can see that these four groups appear to be relatively similar, however the final matching is better for the ITT groups than the participant/non-participant group.

We note that the ITT treatment and ITT comparison cohorts have around twice the proportion of individuals with a claim to Universal Credit than the participant and non-participant groups. This is also the case for referrals to other contracted provision, with the ITT treatment and ITT comparison groups having around twice the proportion as the participant and non-participant groups.

Appendix G. Other effects and biases

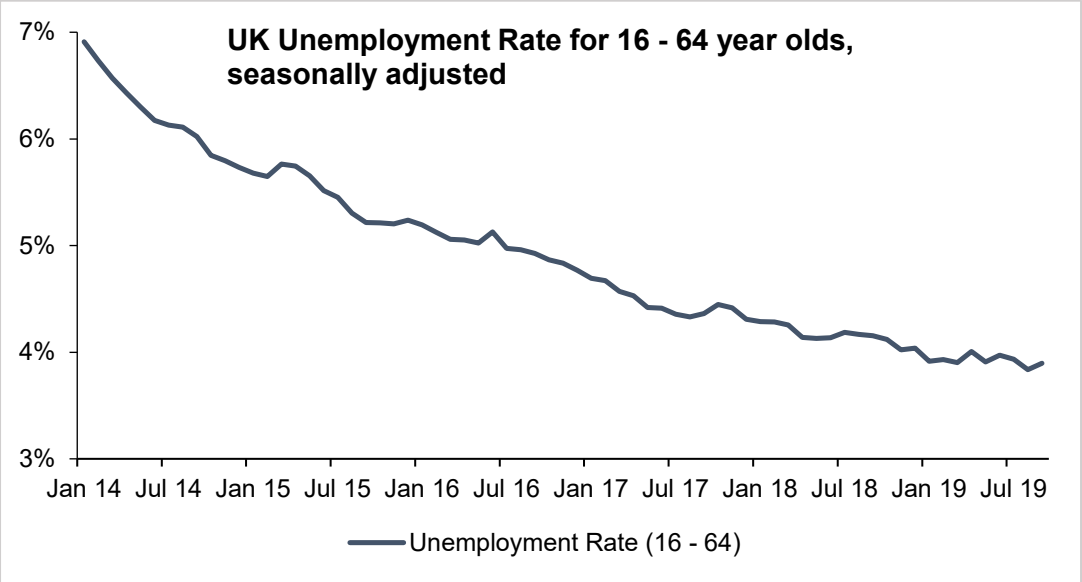
G.1. Labour market changes

In drawing a comparison group of non-participants from a different time period to the participant group, this opens up the possibility of differences in outcomes for the two groups to be attributable to aspects other than their participation in the programme. These differences include the labour market conditions that each group is experiencing. The groups have been selected to minimise the time difference between them, which means they will largely be active under the same labour market conditions. However, there are some differences and the point through an individual's benefit and Work Programme journey that these took place will also be different.

The following analysis provides a high level description of the labour market at a national level for the period up to the end of 2019 to provide some context to the study of the groups in the period of interest – which includes the tracking period. Unemployment rates, number of job vacancies and claims to Jobseeker's Allowance (JSA) and Universal Credit (UC) in the Searching for Work conditionality group have all been considered to give a broad view of the labour market over time.

Figure G.1 shows that unemployment rates dropped quickly following the relatively high levels seen between 2009 and 2012 and slowed in more recent years: the seasonally adjusted UK unemployment rate for people aged 16- to 64-years-old dropped gradually from 6.9% in January 2014 to March 2014 to 3.9% in September 2019 to November 2019. Similarly, employment rates rose from 72.5% to 76.3% in the same period.

Figure G.1 Seasonally adjusted UK unemployment rate for people aged 16- to 64-years-old

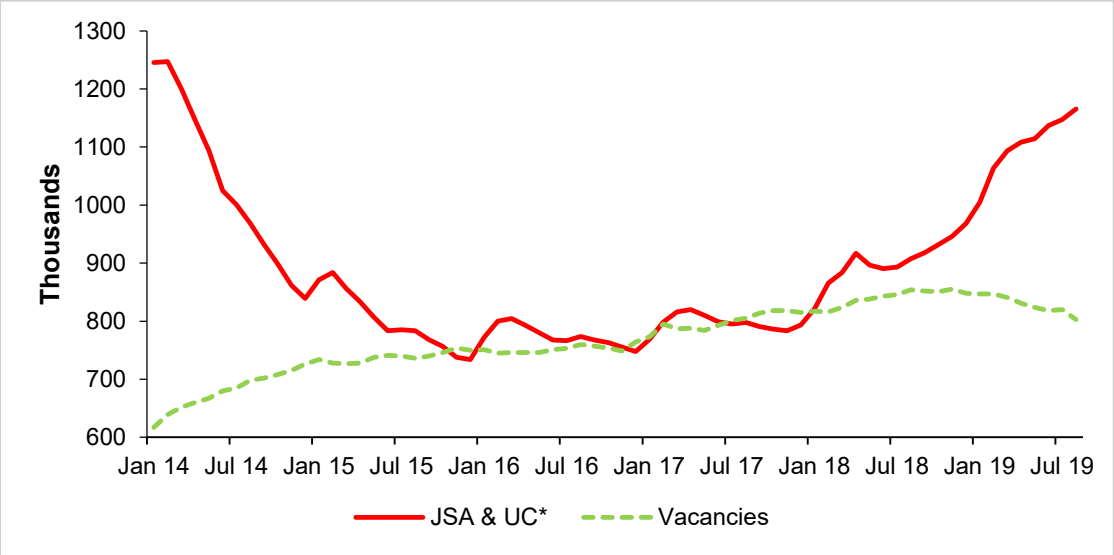


Source: ONS statistics: Labour Market Overview

Figure F.2 shows that since 2015, there had been a gradual rise in job vacancies in the UK, increasing from around 700 thousand vacancies in 2015 to over 800 thousand in 2018-19. This is alongside the number of people on out of work benefits

with requirements to seek employment, either JSA or UC in the Searching for Work conditionality group, moving from around 870 thousand in January 2015 to 820 thousand in January 2018, with low points of around 750 thousand in intervening months. However, since 2018, this increased to nearly 1.2 million by September 2019.

Figure G.2 UK vacancies and claimants in receipt of out of work benefits with requirements to seek employment



Source: ONS statistics Labour Market Overview
*Universal Credit (UC) claims as all conditionality groups with a requirement to be actively seeking employment.

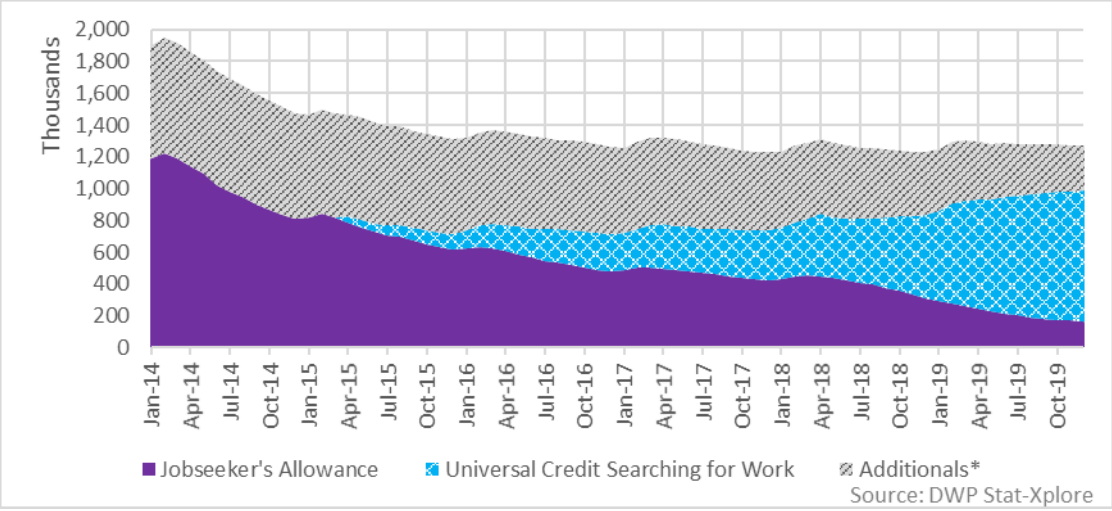
Unemployment and employment rates and vacancy numbers suggest that the labour market had gradually eased since the recession. Therefore, if the participant group is drawn from a time when the labour market was *slightly* more difficult than the time the non-participant group was drawn, then this study might be expected to *slightly* underestimate the impact of the Work Programme, all other things being equal. However, at the same time the number of people on out of work benefits seeking employment has increased which could possibly lead to a slight overestimation of the benefit impact. This coincides with changes to the benefit system in the form of a gradual rollout of UC, which replaces several 'legacy' out of work benefits such as JSA.

A feature of the design of UC brings additional groups into the Searching for Work conditionality group. This means that irrespective of the condition of the labour market, the number of people on out-of-work benefits seeking employment is set to rise[23]. Under UC, people who would have previously claimed JSA are included in the Searching for Work conditionality group. However, this conditionality group contains a broader span of claimants than under JSA definitions, such as partners of claimants, people who previously would have only received Housing Benefit or Child Tax Credits and people who are awaiting a Work Capability Assessment under UC. Therefore, estimates have been made of the proportion of 'additional' people who

[23] Alternative Claimant Count https://www.gov.uk/government/publications/alternative-claimant-count-statistics-background-information-and-methodology/alternative-claimant-count-statistics-background-information-and-methodology

would have been under Searching for Work conditionality had UC existed at the time and been fully rolled out for all areas of the country and all claimant types. Figure G.3 below illustrates the level of people on out of work benefits under this revised definition, known as the Alternative Claimant Count.

Figure G.3 The Alternative Claimant Count



*Additionals are described as people who would have been under Searching for Work conditionality had Universal Credit (UC) existed and been fully rolled out from 2013

This suggests a much more stable labour market than shown in Figure G.2, suggesting that groups of individuals drawn from consecutive quarters in 2017 may have been operating under broadly similar labour market conditions. This suggests that we can have some confidence, at a national level, that when consecutive cohorts in time are compared that effects on benefit and employment from time should be fairly small. Therefore, in theory, the impact on outcomes, as measured between a group eligible for the Work Programme and ones that did not, should be broadly attributable to the programme

G.2. Anticipation effects

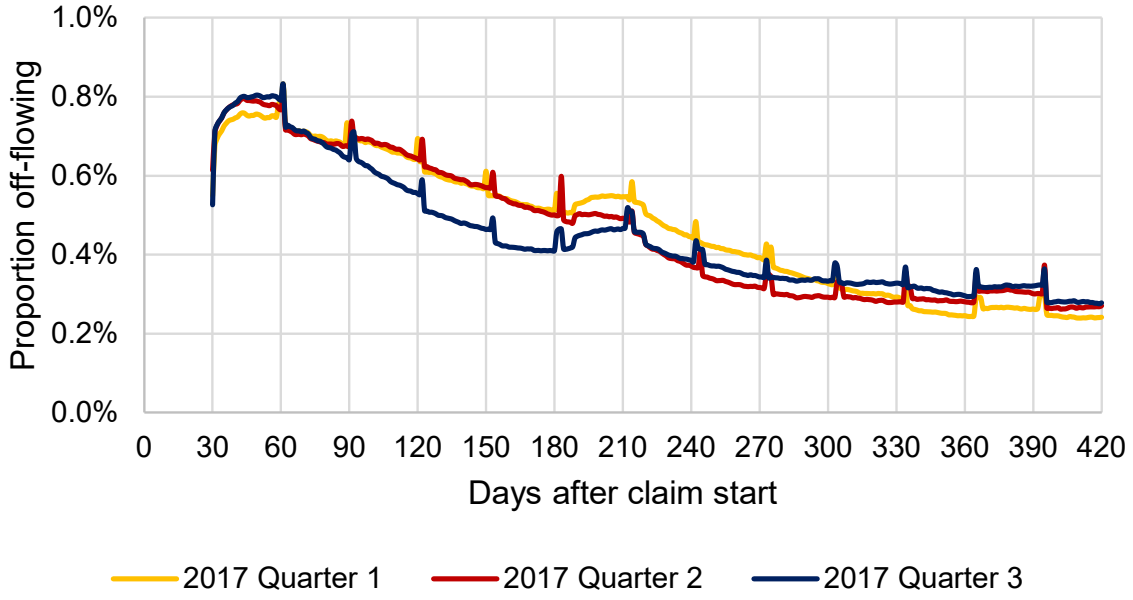
There is a risk that the existence of a provision such as the Work Programme could impact the way individuals interact with the labour market prior to referral. It is possible that individuals may make extra efforts to find employment in order to avoid a mandated referral, or delay such efforts in order to secure the support that the provision might provide, known as the 'anticipation effect'. Were this to happen this could lead to shifts in unobservable differences between treatment and comparison. If such anticipation effects exist, it might be expected that these would manifest as changes in the off-flow rate of qualifying benefits in the period preceding referral whilst the provision is active, that are not present when the provision has ended.

To assess this, details of all Jobseeker's Allowance (JSA) claims and claims to Universal Credit (UC) in the Searching for Work conditionality regime were collated for three cohorts of claim starts. These cohorts were claims starting between: January 2017 and March 2017; April 2017 and June 2017; July 2017 and September 2017. These cohorts were selected as groups of individuals that, should their claims

have continued, would have met the Work Programme Payment Group 2 eligibility criterion for claiming benefit between January 2018 and September 2018.

The rolling 30-day mean conditional daily off-flow rate was calculated for these cohorts and presented below in Figure G.4.

Figure G.4 Rolling 30-day mean conditional daily off-flow rates for claimants of Jobseeker's Allowance and Universal Credit Searching for Work



The data source for spells on UC considers the conditionality group for an individual across a period of one month, resulting in conditionality changes for individual claimants only being observed at monthly intervals. This results in the 'spikes' in monthly off-flow rates observed in Figure G.4. These 'spikes' are therefore not indicative of anticipative effects, and instead are artefacts of the data source used.

Across the selected cohorts, there is an uplift in off-flow rates observed at around 180 to 210 days, approximately 6 months, after claim start. This corresponds with, and is therefore likely to be attributable to, automatic cessation of Contribution Based claims for JSA, which occurs after a period of 6 months in receipt of JSA.

To assess whether any anticipation effects might have been present for referrals that should occur after 12 months of benefit claims, the period preceding this has been considered. In the period 300 to 360 days, approximately 10 to 12 months, after the claim start, there is no observable change in off-flow rates that might suggest that individuals were adjusting their interaction with the labour market to either avoid or ensure participation in the Work Programme.

This suggests that there is no evidence that there are any changes in off-flow rates just prior to or at the point an individual would become eligible for referral to Work Programme Payment Group 2.

G.3. Intention to Treat cohort selection

The ITT groups were selected based on individuals meeting all eligibility criteria for Work Programme Payment Group 2 in two specific consecutive time periods, between January 2017 and March 2017 and between April 2017 and June 2017.

In this section we present results using further cohorts to explore the possibility of time effects and sensitivity in selection of cohorts.

1. After the programme finished, April 2017 to June 2017 and July 2017 to September 2017
2. One year prior to the lead comparison, January 2016 to March 2016 and April 2016 to June 2016
3. One year prior to the groups in (1) above, April 2016 to June 2016 and July 2016 to September 2016
4. Shorter 1 month cohorts drawn closer to the end of the programme: March 2017 and April 2017
5. Wider range between cohorts: October to December 2016 and April to June 2017
6. Longer 6 month cohorts: October 2016 to March 2017 and April to September 2017

G.3.1. Time effects from after the programme

To explore possible time effects in the period after the programme finished an additional ITT cohort was drawn to follow on from the cohort of 18,771 individuals becoming eligible from April 2017 to June 2017, i.e. 20,316 individuals becoming eligible between July 2017 and September 2017. These cohorts of individuals met the eligibility criteria for the programme after referrals had closed at the end of March 2017, so if the time difference between the two does not have an effect on outcomes, the impact is expected to be nil.

Figure G.5 Differences on the likelihood of being receipt of an OW benefit for the ITT cohorts of April 2017 to June 2017 and July 2017 to September 2017

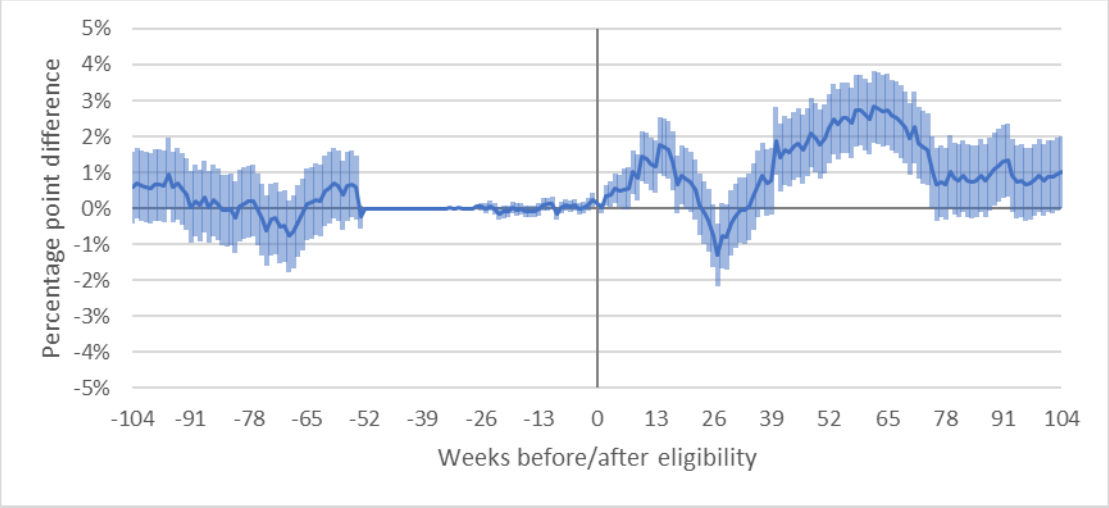


Figure G.6 Differences on the likelihood of being in employment and not claiming OW benefit for the ITT cohorts of April 2017 to June 2017 and July 2017 to September 2017

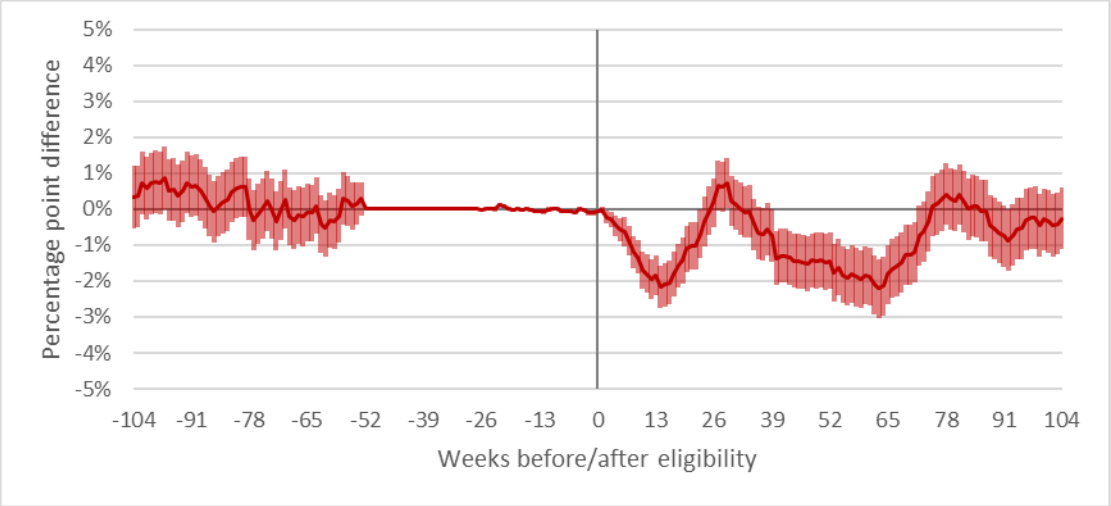


Figure G.7 Differences on the likelihood of neither claiming OW benefit nor in employment for the ITT cohorts of April 2017 to June 2017 and July 2017 to September 2017

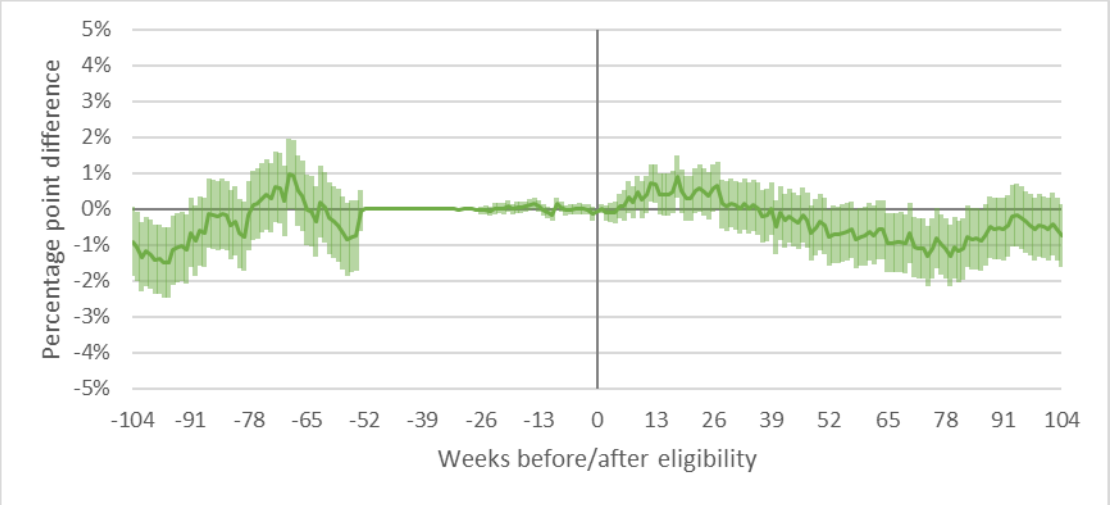
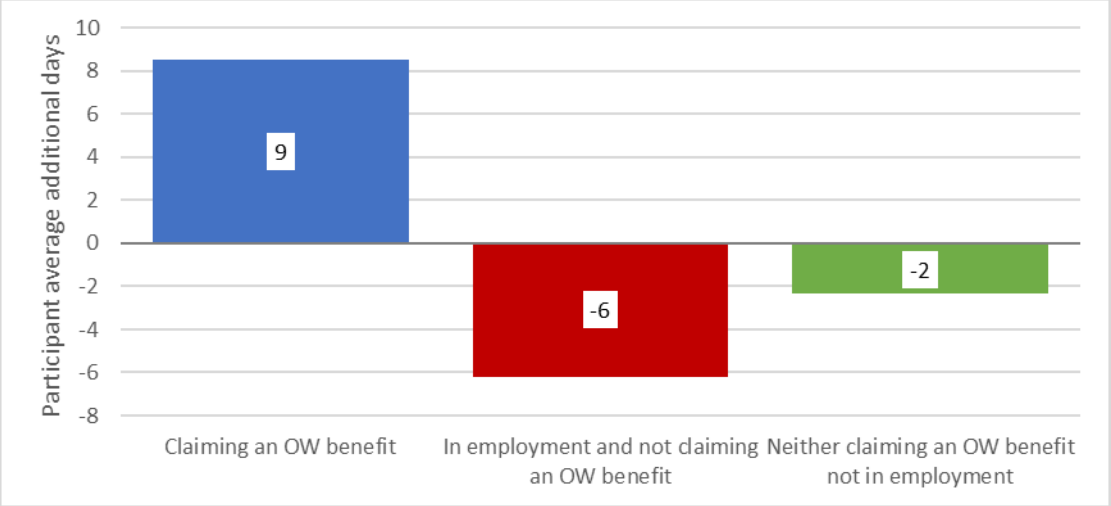


Figure G.8 Additional days in the three primary outcome states over 104 weeks for the ITT cohorts of April 2017 to June 2017 and July 2017 to September 2017



From the impacts shown in Figure G.5 to Figure G.7, and the additional days calculated for Figure G.8, it appears that there are some differences between the cohorts of April 2017 to June 2017 and the cohort of July 2017 to September 2017. Both of these cohorts are taken after the Work Programme closed for new referrals, and some of the tracked time is before an equivalent programme became available. It is possible that some of this observed difference is due to changes in the labour market or other time effects between the cohorts. Some individuals in these cohorts will also have been amongst the first to become eligible for the Long-Term Unemployed group of the Work and Health Programme, a group which opened for referrals in April 2018.

G.3.2. Time effects from ITT groups from during the Work Programme

A further possibility is that the difference between the two quarters is the result of a seasonal effect. To explore this, two separate ITT cohorts were drawn from one year earlier than the tested cohorts, i.e. 16,021 individuals eligible between January 2016 to March 2016 and 15,108 individuals eligible between April 2016 and June 2016.

Referrals to the programme were active at the time of these cohorts and they both experienced similar levels of programme participation, 69% and 64% respectively, so if there are no material changes in delivery and the time difference between the two does not have an effect on outcomes, the difference in outcomes is expected to be nil. As was observed in previous graphs, the presence of seasonality might be expected, but the overall impact would still be zero.

The following charts in Figure G.9 to Figure G.12 show the observed difference in outcomes between these two cohorts of participants for the three key outcome states.

Figure G.9 Differences on the likelihood of being receipt of an OW benefit for the ITT cohorts of January 2016 to March 2016 and April 2016 to June 2016

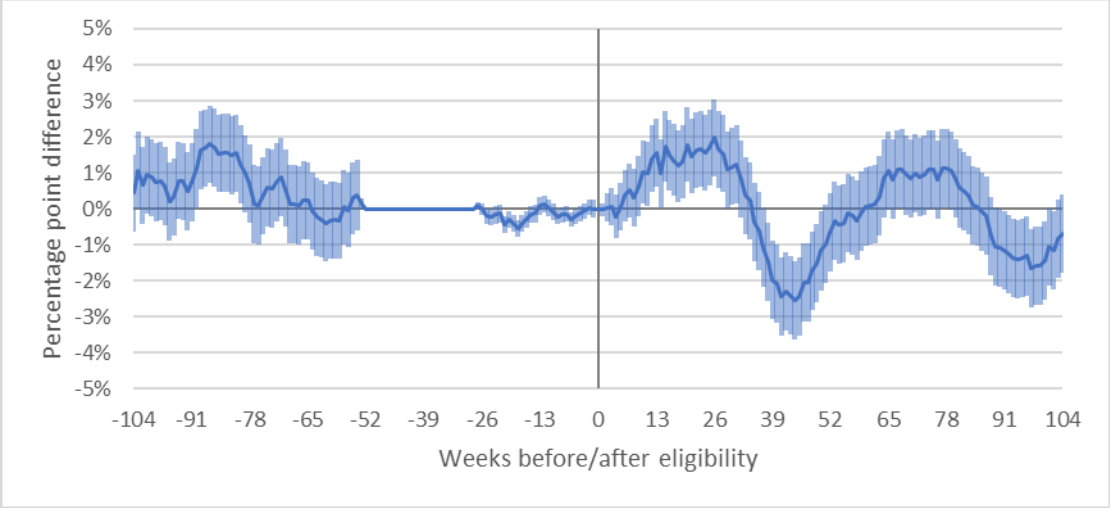


Figure G.10 Differences on the likelihood of being in employment and not claiming OW benefit for the ITT cohorts of January 2016 to March 2016 and April 2016 to June 2016

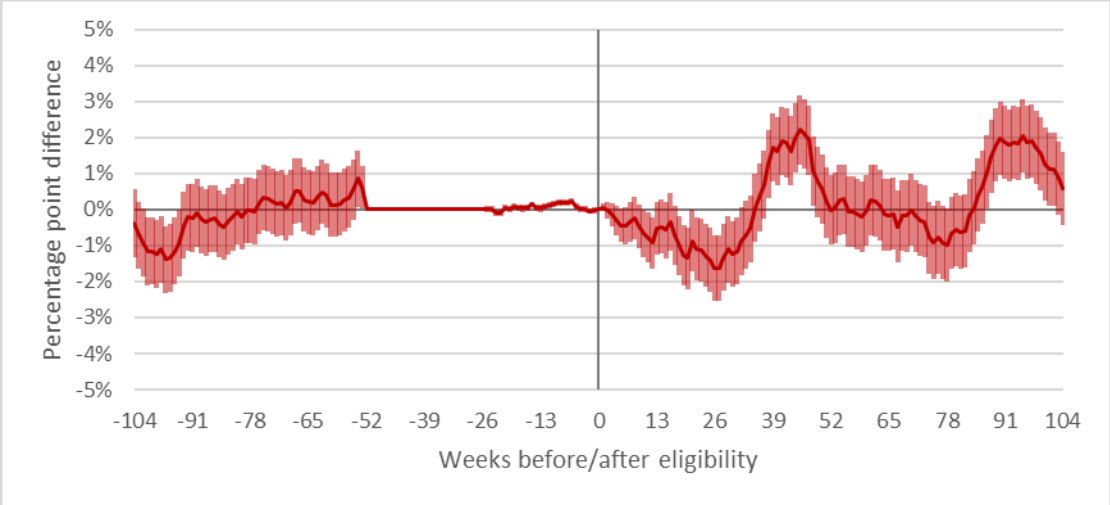


Figure G.11 Differences on the likelihood of neither claiming OW benefit nor in employment for the ITT cohorts of January 2016 to March 2016 and April 2016 to June 2016

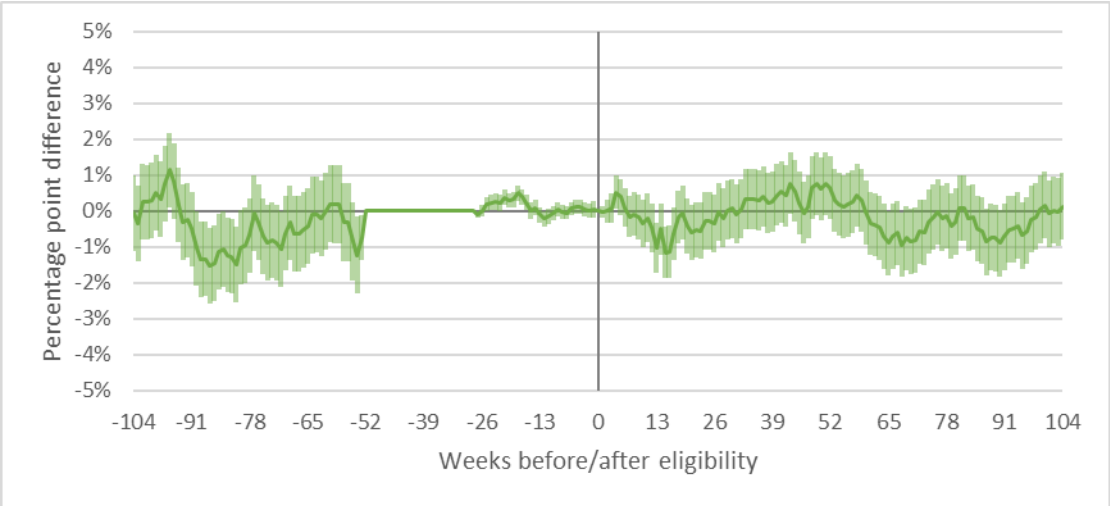
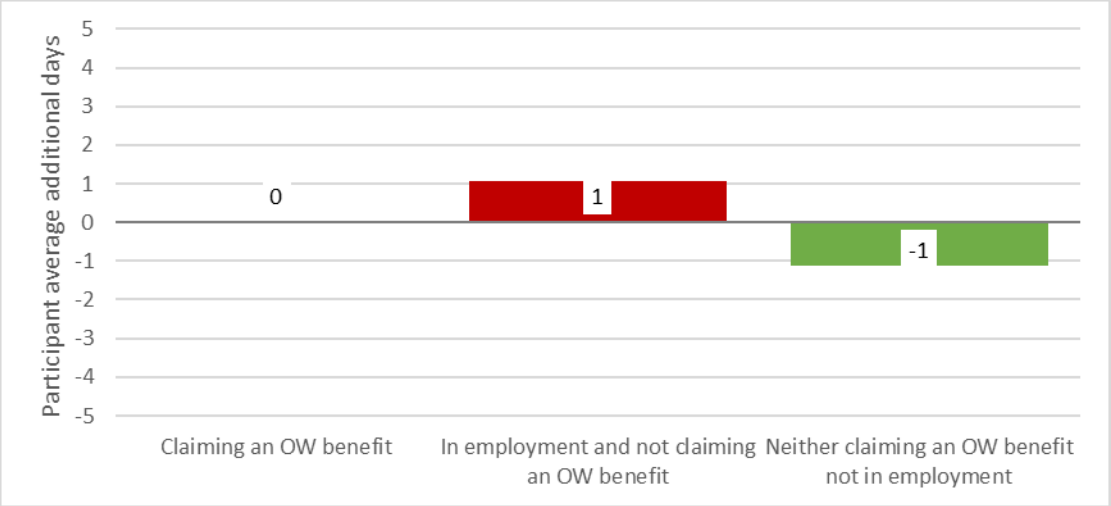


Figure G.12 Additional days in the three primary outcome states over 104 weeks for the ITT cohorts of January 2016 to March 2016 and April 2016 to June 2016



Across the outcome states there is some seasonality and this is seen most clearly in Figure G.10 for the in employment and not claiming benefit state. Those eligible in January to March 2016 were more likely to be in employment 3 months or 15 months later than those referred in April to June 2016. This reflects the fact that the labour market is more buoyant in the period of approximately October to December, when seasonal jobs, particularly in the retail sector, are more prevalent. But taken over the follow-up period as a whole, the differences are minor.

Although it can't be guaranteed that any labour market changes for cohorts from 2016 similarly affect the cohorts from 2017, this is another part of the evidence to support that the difference in outcomes observed between consecutive participant and non-participant groups is mainly attributable to participation in the Work Programme and not from other external changes.

G.3.3. Time effects from ITT groups from during the Work Programme (April-June 2016 and July-September 2016)

To further explore these time effects, cohorts for the same quarters as in G.3.1 except in 2016 were drawn, i.e. individuals eligible between April 2016 and June 2016 and between July 2016 and September 2016. Referrals to the programme were active at the time of these cohorts and they both experienced similar levels of programme participation, 64% and 59% respectively, so if there are no material changes in delivery and the time difference between the two does not have an effect on outcomes, the difference in outcomes is expected to be nil. As was observed in previous graphs, the presence of seasonality might be expected, but the overall impact would still be zero.

Figure G.13 Differences on the likelihood of being receipt of an OW benefit for the ITT cohorts of April 2016 to June 2016 and July 2016 to September 2016

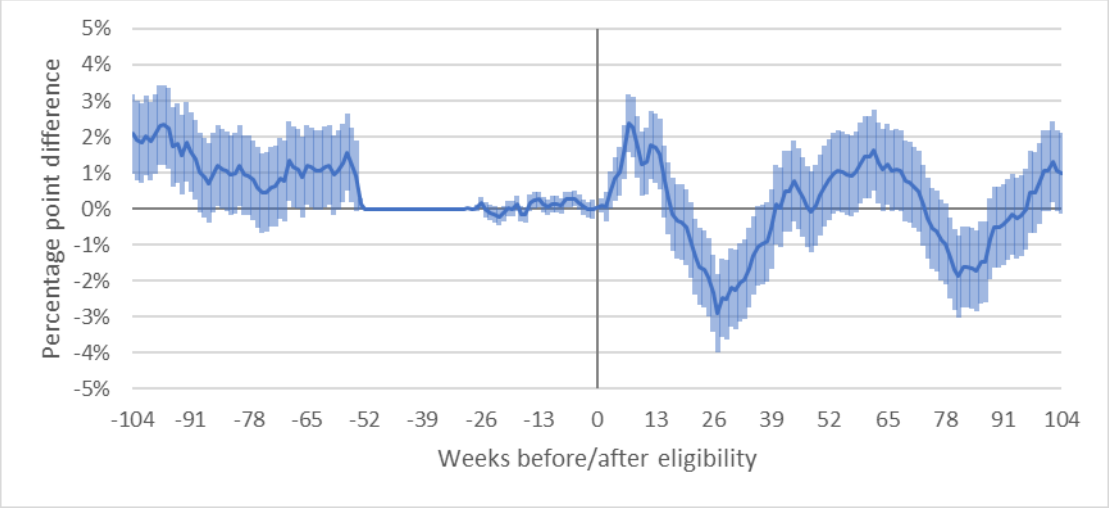


Figure G.14 Differences on the likelihood of being in employment and not claiming OW benefit for the ITT cohorts of April 2016 to June 2016 and July 2016 to September 2016

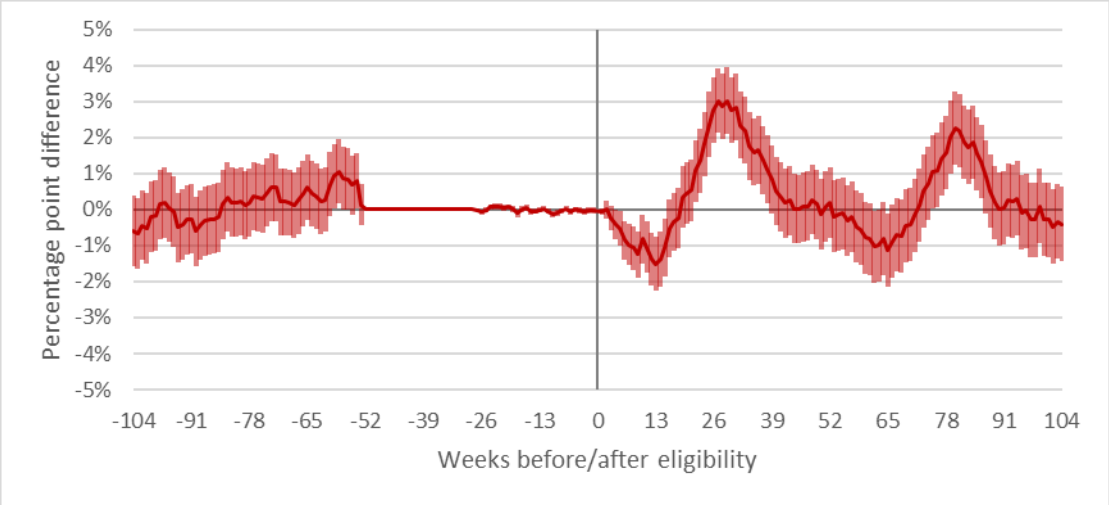


Figure G.15 Differences on the likelihood of neither claiming OW benefit nor in employment for the ITT cohorts of April 2016 to June 2016 and July 2016 to September 2016

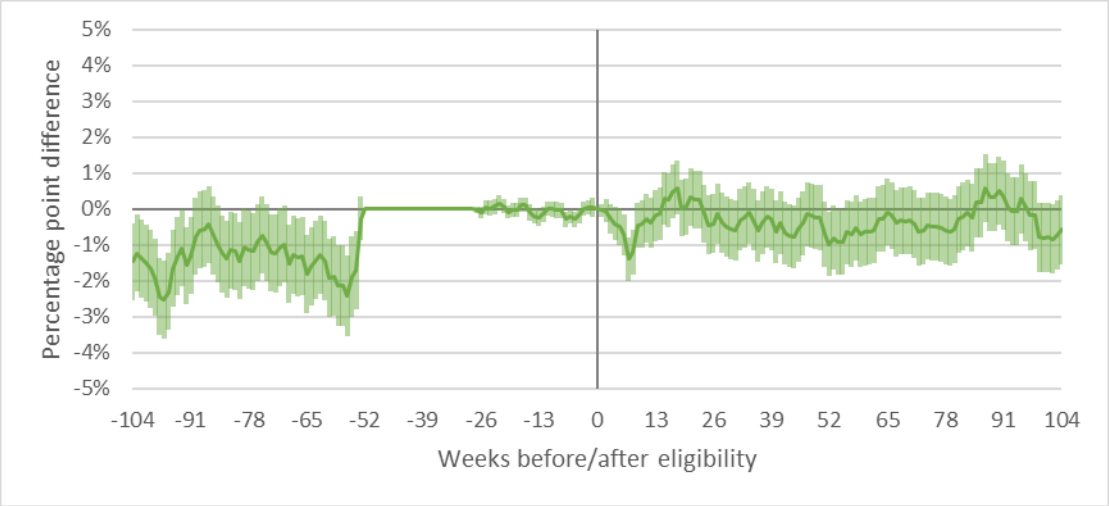
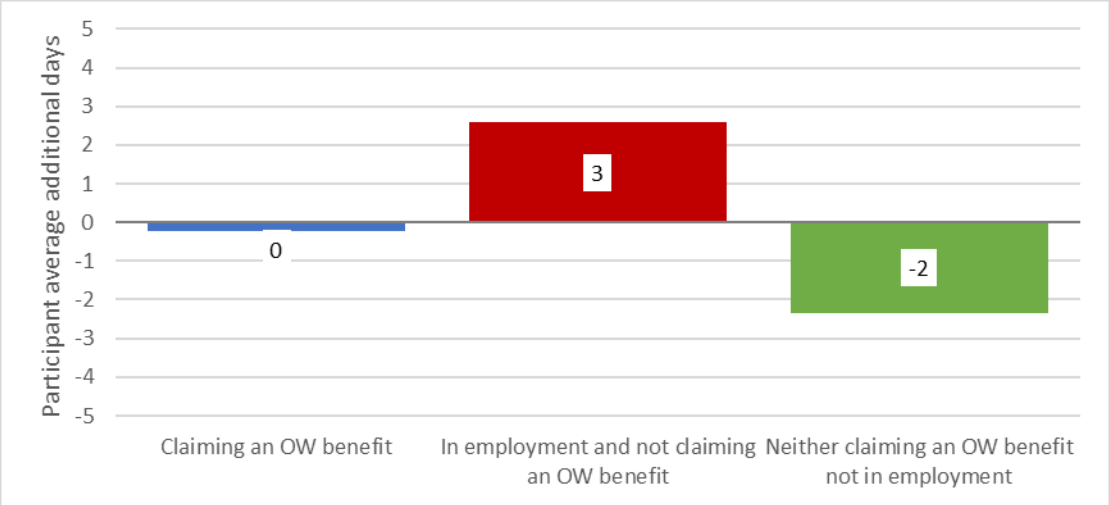


Figure G.16 Additional days in the three primary outcome states over 104 weeks for the ITT cohorts of April 2016 to June 2016 and July 2016 to September 2016



From Figure G.16, it appears that the impacts between these two cohorts from 2016 are much smaller than those of the two equivalent cohorts in 2017, as shown in Figure G.8, and in the case of employment swung in the opposite direction. This suggests there is uncertainty in how changes between cohorts are affecting the impacts, but that these differences are relatively small in magnitude.

Overall, although it can't be guaranteed that any labour market changes occurring after the closure of referrals similarly affect cohorts whilst referrals were active, this is another part of the evidence to support that the difference in outcomes observed between consecutive cohorts is mainly attributable to participation in the Work Programme and not from other external changes.

G.3.4. Time effects between cohorts

Although the main ITT cohorts are drawn from consecutive time periods, January 2017 to March 2017 and April 2017 to June 2017 respectively, the total time period still covers a range of six months. There is a possibility that there may have been some changes in labour market conditions over this length of time. To test to see whether any material time effects can be observed, the ITT cohorts have been restricted to two consecutive months: 6,363 individuals from March 2017 and 6,173 individuals from April 2017. However, the drawback of these sub-cohorts, is that the level of participation is lower at around 26% for the March 2017 ITT sub-group, compared to the whole January 2017 to March 2017 ITT group which had a 38% programme participation rate. Therefore, one would expect the impacts to be smaller, even if there are no time effects.

Figure G.17 Impact on the likelihood of being receipt of an OW benefit for the March 2017 ITT cohort compared to the April 2017 ITT cohort

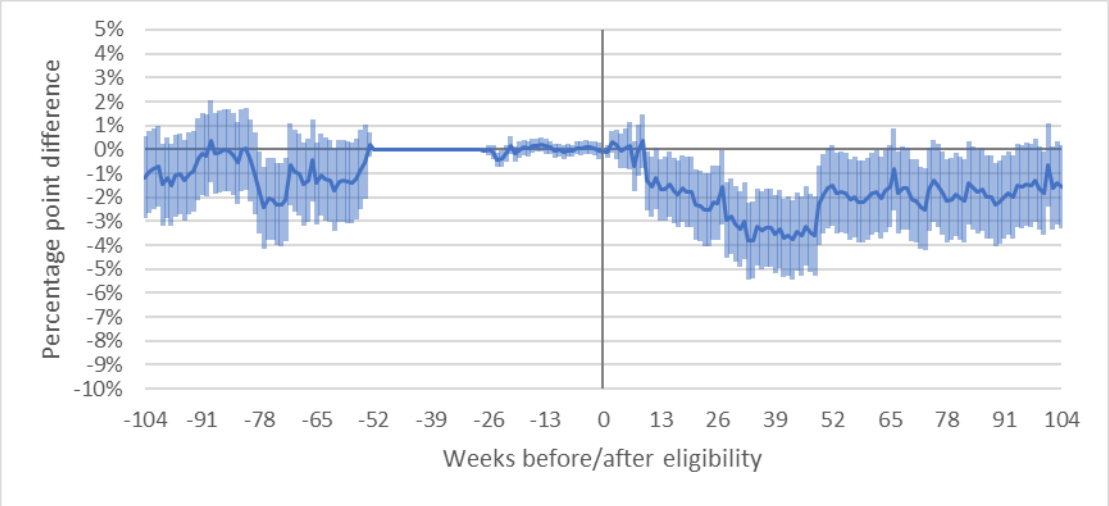


Figure G.18 Impact on the likelihood of being in employment and not claiming OW benefit for the March 2017 ITT cohort compared to the April 2017 ITT cohort

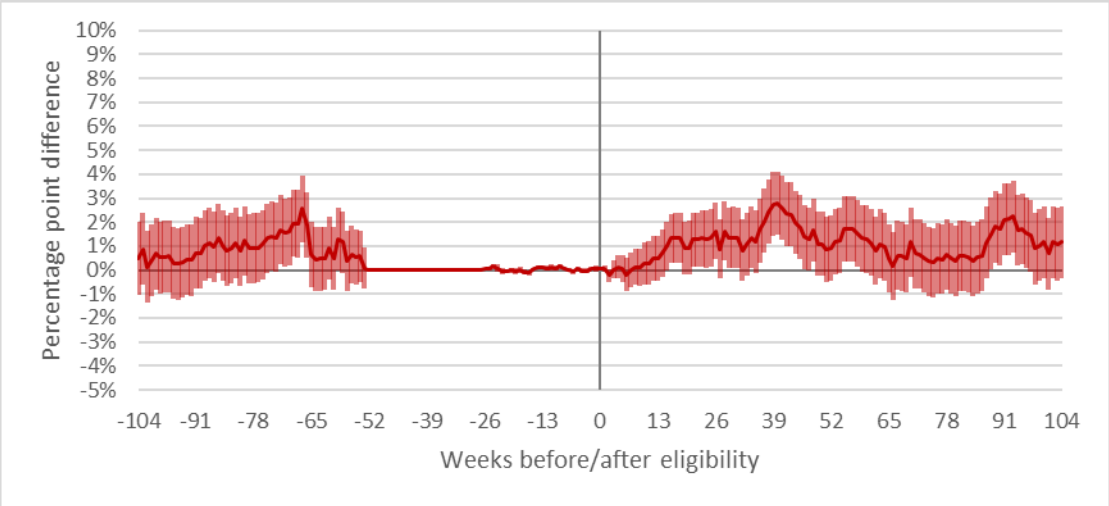


Figure G.19 Impact on the likelihood of neither claiming OW benefit nor in employment for the March 2017 ITT cohort compared to the April 2017 ITT cohort

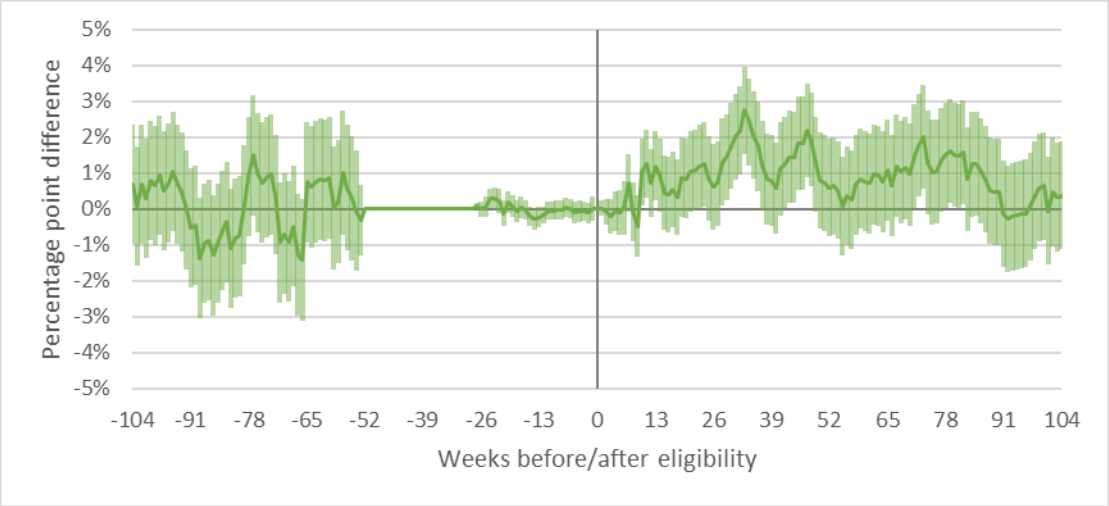


Figure G.20 compares the calculated additional days in each of the primary outcome states for the comparison of March 2017 and April 2017 ITT sub-cohorts, to the whole ITT cohorts, as shown in section 4.3.

Figure G.20 Additional days in the three primary outcome states over 104 weeks for the March 2017 ITT cohort compared to the April 2017 ITT cohort and the wider ITT cohorts

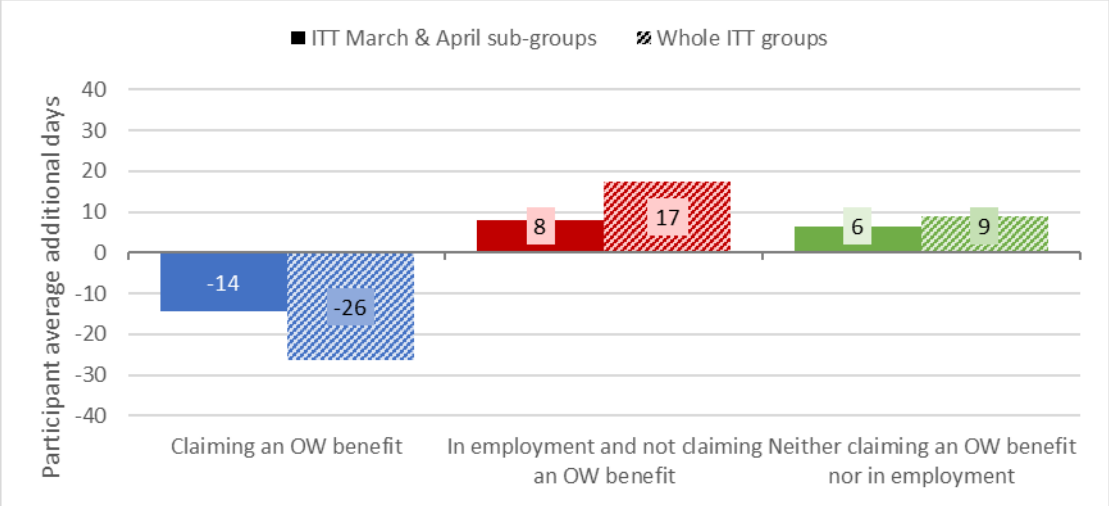


Figure G.20 shows that the impacts are smaller to those seen for the wider cohort – this is likely to be due to the lower programme participation rate, but it could also be influenced by any time effects present.

G.3.5. ITT comparison between October-December 2016 and April-June 2016

The following charts show results for a ITT cohort defined as individuals meeting the eligibility criteria between October 2016 and December 2016 compared to the previously defined ITT comparison group, i.e. individuals eligible for the programme between April 2017 and June 2017. This timeframe allows at least 90 days to elapse between individuals becoming eligible and the programme closing to referrals in

March 2017. This new ITT treatment group has a higher proportion (55%) of individuals referred.

Figure G.21 Impact on the likelihood of being receipt of an OW benefit for the earlier ITT treatment group compared to the ITT comparison group

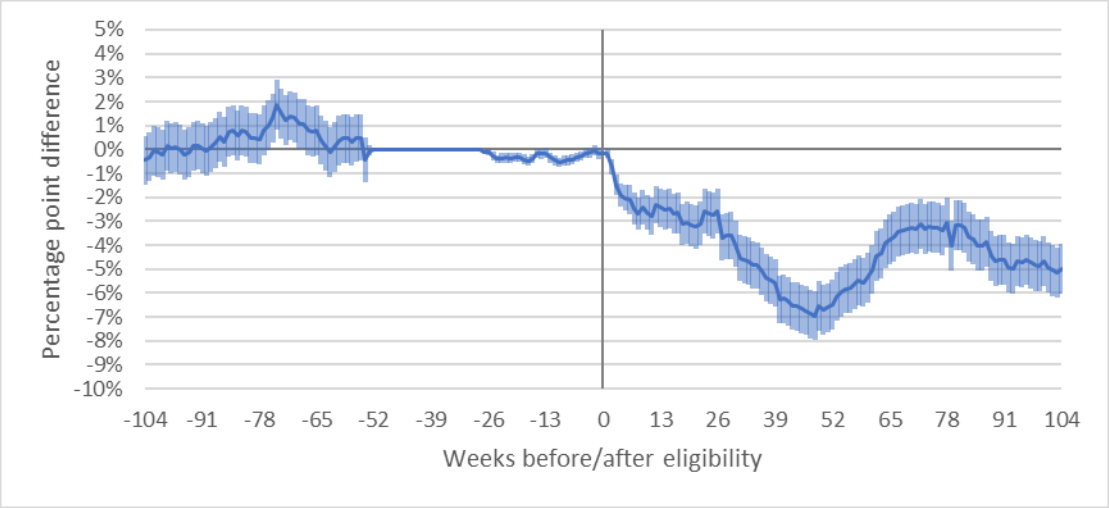


Figure G.22 Impact on the likelihood of being in employment and not claiming OW benefit for the earlier ITT treatment group compared to the ITT comparison group

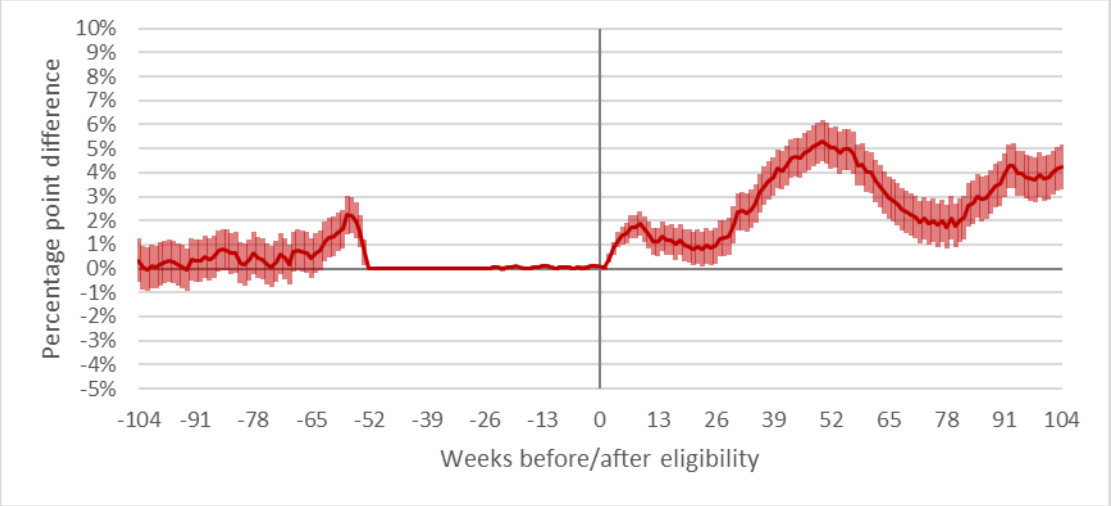
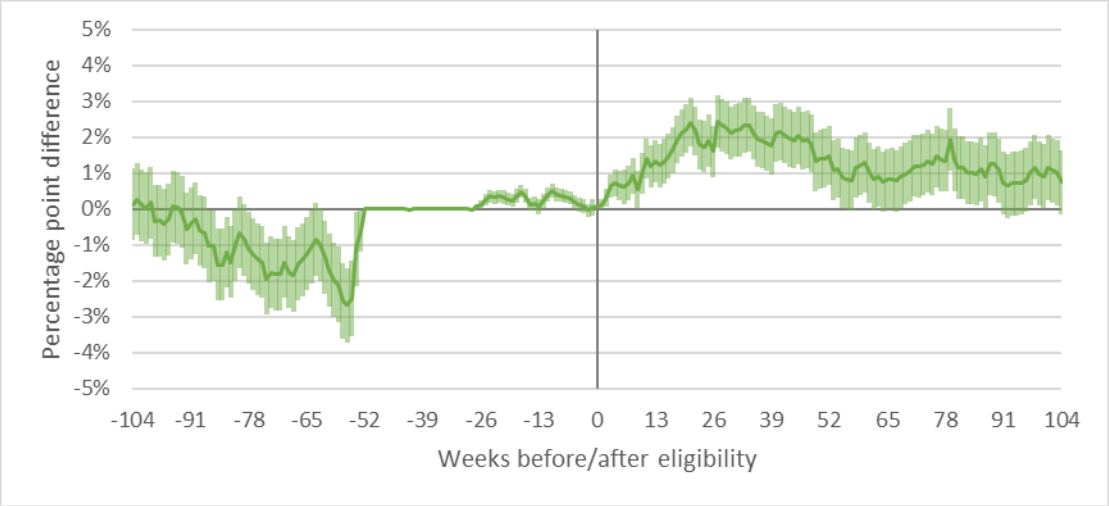


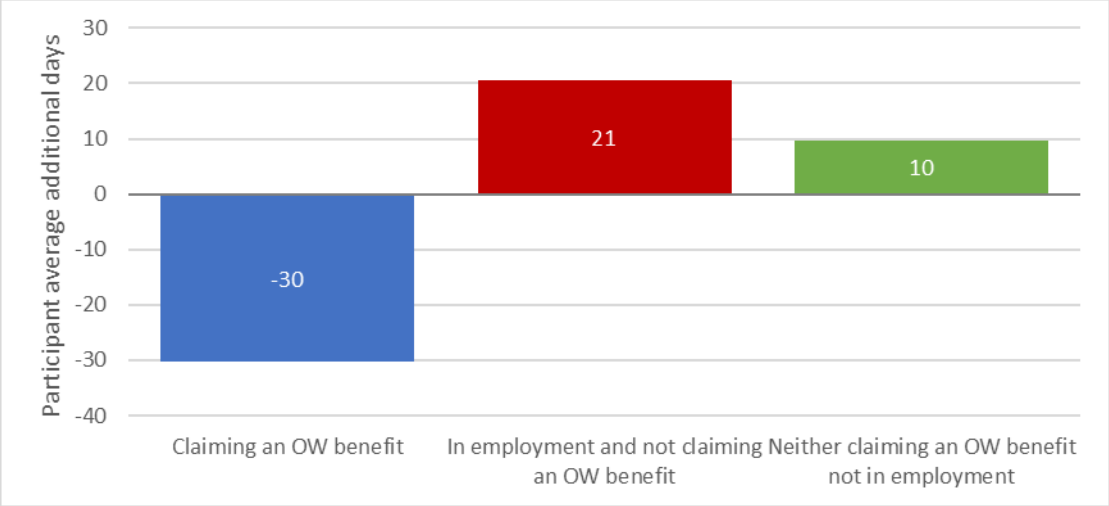
Figure G.23 Impact on the likelihood of neither claiming OW benefit nor in employment for the earlier ITT treatment group compared to the ITT comparison group



This matching of this additional ITT treatment group for October 2016 to December 2016 to the ITT comparison group shows similar levels of impact as seen for the ITT treatment group for January 2017 to March 2017. The observed impact seen in Figure G.21 - D11 above show a slightly higher impact than seen in the analysis using the later ITT treatment group in section 4.3. This is most likely due to the higher proportion of Work Programme participants in the earlier ITT group, but could also be driven by the larger time differences between the treatment and comparison groups.

Figure G.24 below shows the observed additional days over 104 weeks for the three mutually exclusive outcome states.

Figure G.24 Additional days in the three primary outcome states over 104 weeks for the earlier ITT treatment group compared to the ITT comparison group



G.3.6. Cohorts expanded to 6-month windows

The main analysis considers two cohorts on an Intention to Treat (ITT) basis, with each cohort spanning 3 months. These groups could be expanded further, for example spanning 6 months each, with the acknowledgment that as the cohorts encompass a much larger time period they could therefore be more susceptible to external time effects.

The following charts show the calculated impact of these expanded cohorts, i.e. individuals eligible for the programme between October 2016 and March 2017 as the treatment group and between April 2017 and September 2017 as the comparison group. In this case the treatment group has a Work Programme participation rate of 46% compared to the main analysis where the participation rate was 38%.

Figure G.25 Impact on the likelihood of being receipt of an OW benefit for the 6-month ITT treatment group compared to the 6-month ITT comparison group

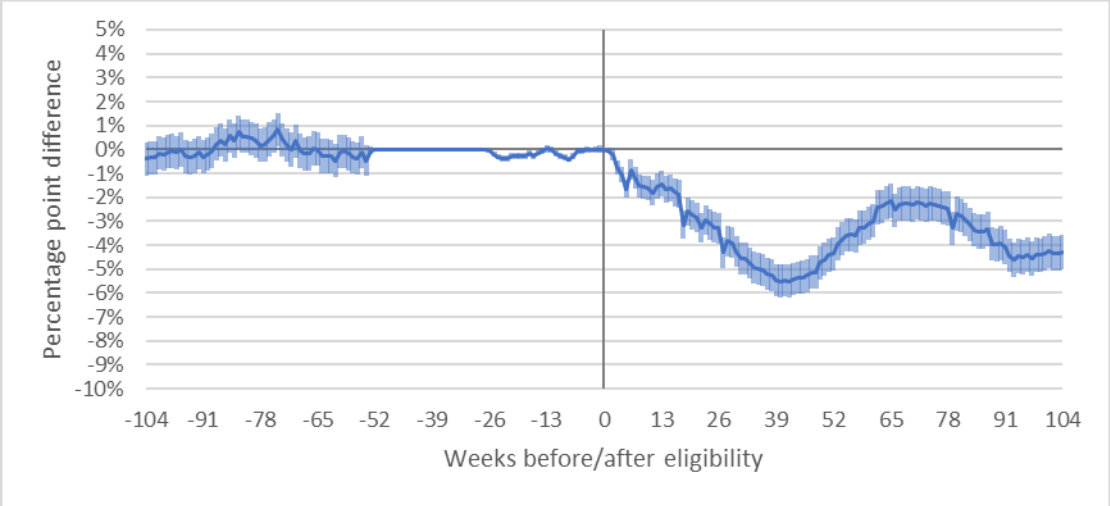


Figure G.26 Impact on the likelihood of being in employment and not claiming OW benefit for the 6-month ITT treatment group compared to the 6-month ITT comparison group

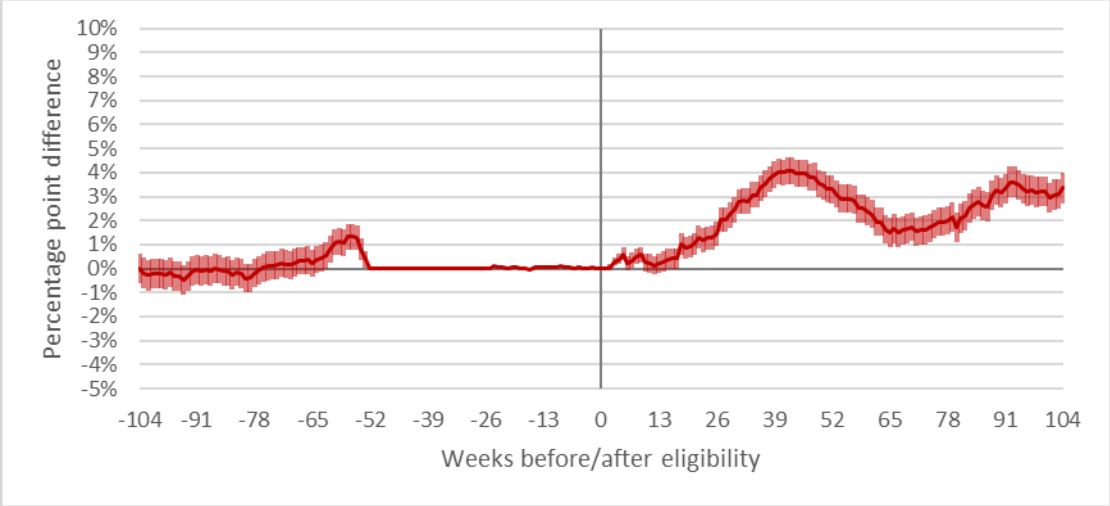
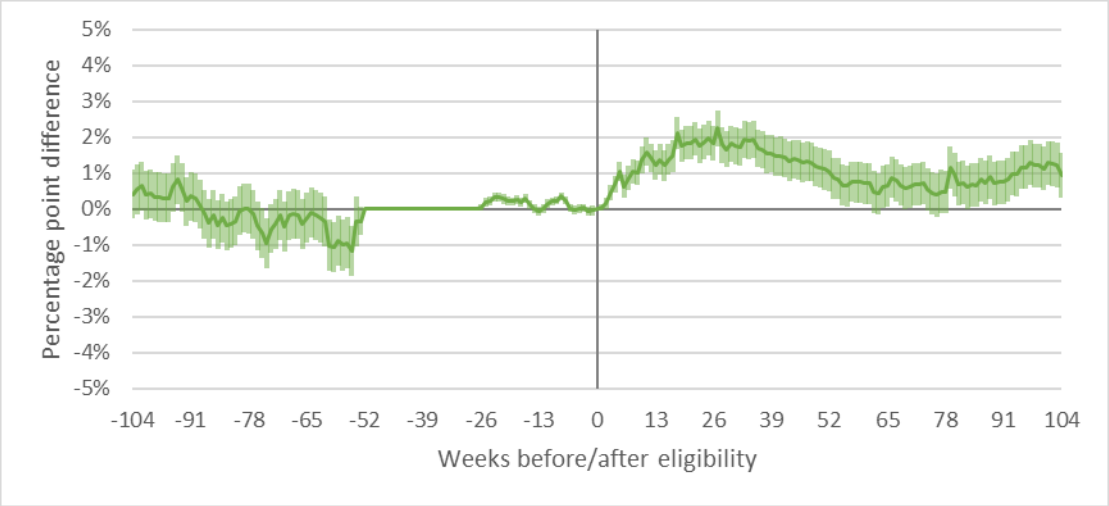


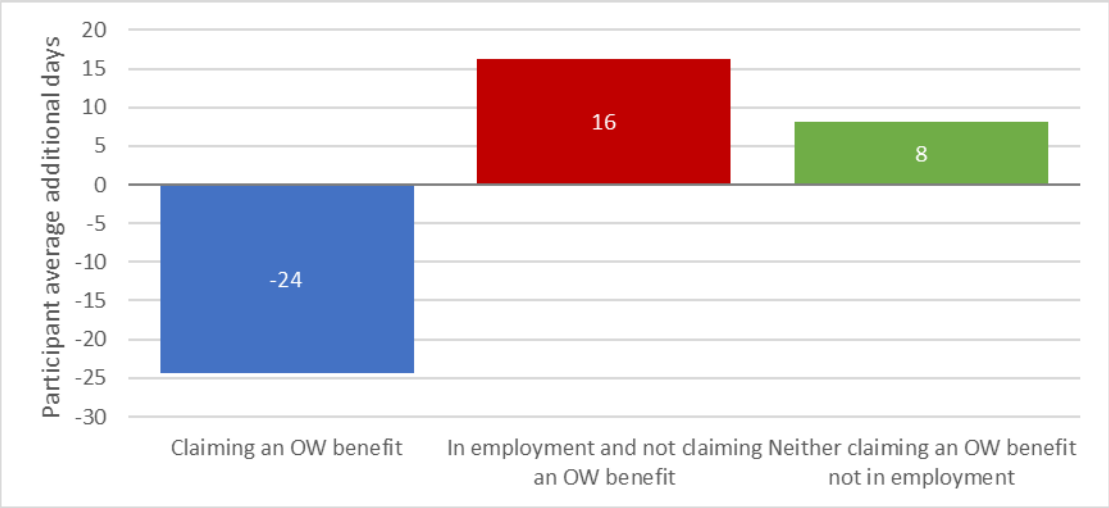
Figure G.27 Impact on the likelihood of neither claiming OW benefit nor in employment for the 6-month ITT treatment group compared to the 6-month ITT comparison group



These show similar results to the main ITT groups explored in section 4.3, however it should be noted that the employment impact, seen in Figure G.26, shows some difference between the 6-month cohorts at around 52- to 65-weeks prior to programme eligibility. This difference is not present for the main analysis, seen in Figure 4.6.

Figure G.28 below shows the observed additional days over 104 weeks for the three mutually exclusive outcome states.

Figure G.28 Additional days in the three primary outcome states over 104 weeks for the 6-month ITT treatment group compared to the 6-month ITT comparison group



G.3.7. Summary of time effect sensitivities

Table G.1 below summaries the additional days in each of the primary outcome states for the sensitivity group testing.

Table G.1 Summary of additional days in each of the primary outcome states for ITT sensitivity groups

Group	Days in receipt of OW benefit	Days in employment and not in receipt of OW benefit	Days neither in receipt of OW benefit nor in employment
ITT treatment group January 2017 – March 2017 matched to ITT comparison group April 2017 – June 2017	-26	17	9
ITT group April 2017 – June 2017 matched to ITT group July 2017 to September 2017	9	-6	-2
ITT group January 2016 – March 2016 matched to ITT group April 2016 – June 2016	0	1	-1
ITT group April 2016 – June 2016 matched to ITT group July 2016 to September 2016	0	3	-2
March 2017 ITT cohort compared to the April 2017 ITT cohort	-14	8	6
ITT group October 2016 – December 2016 matched to ITT group April 2017 – June 2017	-30	21	10
ITT treatment group October 2016 – March 2017 matched to ITT comparison group April 2017 – September 2017	-24	16	8

Appendix H. Analysis of Participant and Non-participant cohorts

This section the outcome states for the comparison between participants referred to Work Programme Payment Group 2 between January 2017 and March 2017 and compares them to the selected group of non-participants with pseudo-referral dates between April 2017 and June 2017. Appendix E describes in further detail the selection of these two groups.

H.1. Primary outcome states

Figure H.1, Figure H.3, and Figure H.5 show the proportion of each cohort at any point in time two years before and after the programme for the three primary outcome states: OW benefits, employment and not on OW benefits and neither OW benefit nor employment.

By subtracting the proportion of non-participants in a given state from the proportion of participants in the same state for each week the estimated percentage-point impact can be found, as shown in Figure H.2, Figure H.4 and Figure H.6 for OW benefits, employment and neither OW benefit nor employment, respectively.

Figure H.1 Proportion of participant and non-participant cohorts (post-matching) in receipt of an OW benefit

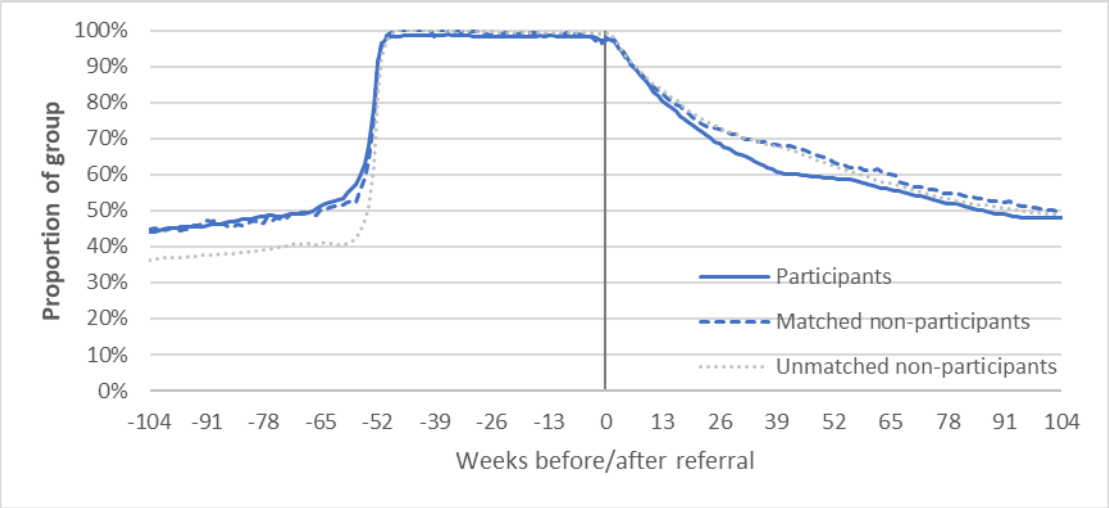


Figure H.2 Impact on the likelihood of being receipt of an OW benefit for the participant cohort compared to the non-participant cohort

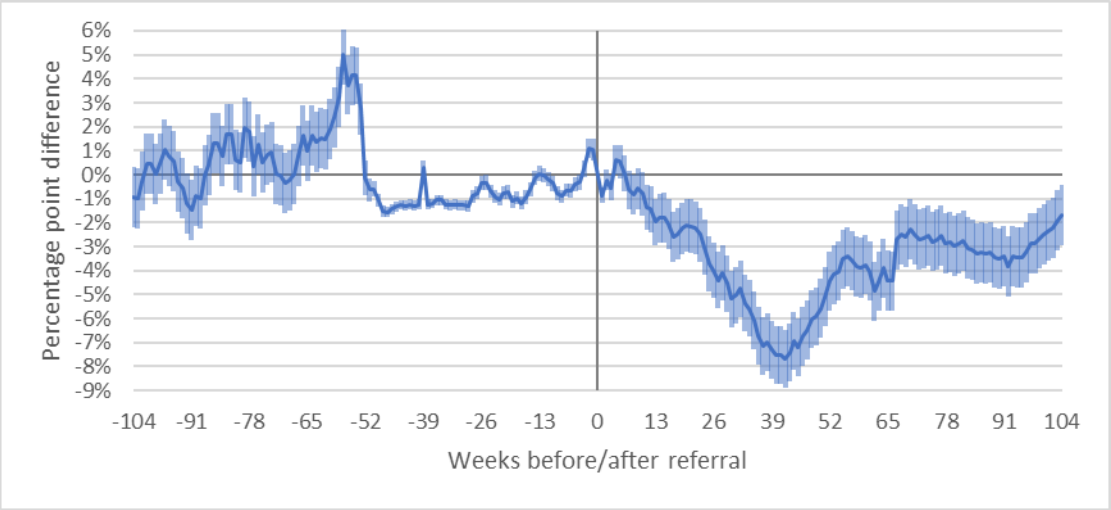


Figure H.3 Proportion of participant and non-participant cohorts (post-matching) in employment and not claiming OW benefit

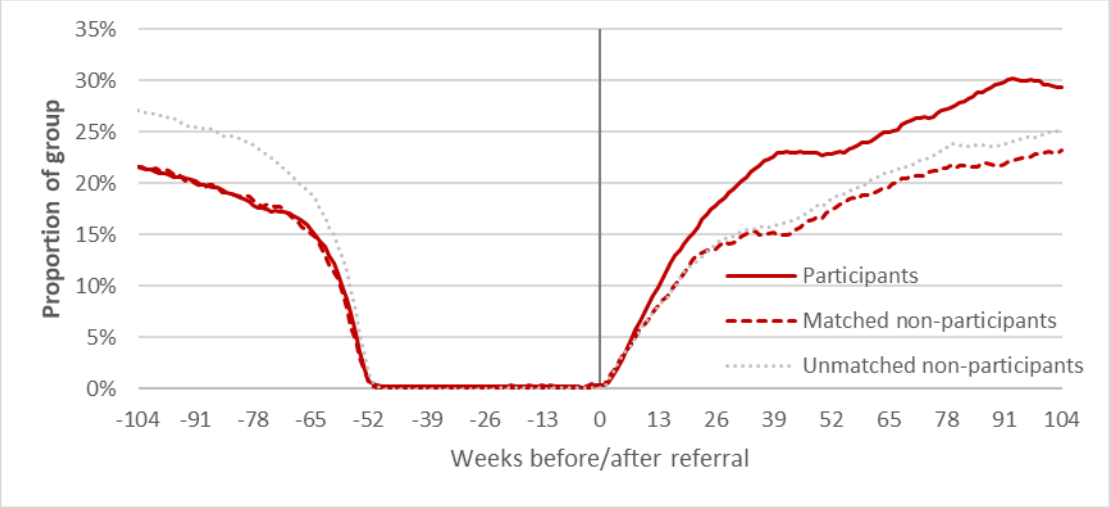


Figure H.4 Impact on the likelihood of being in employment and not claiming OW benefit for the participant cohort compared to the non-participant cohort

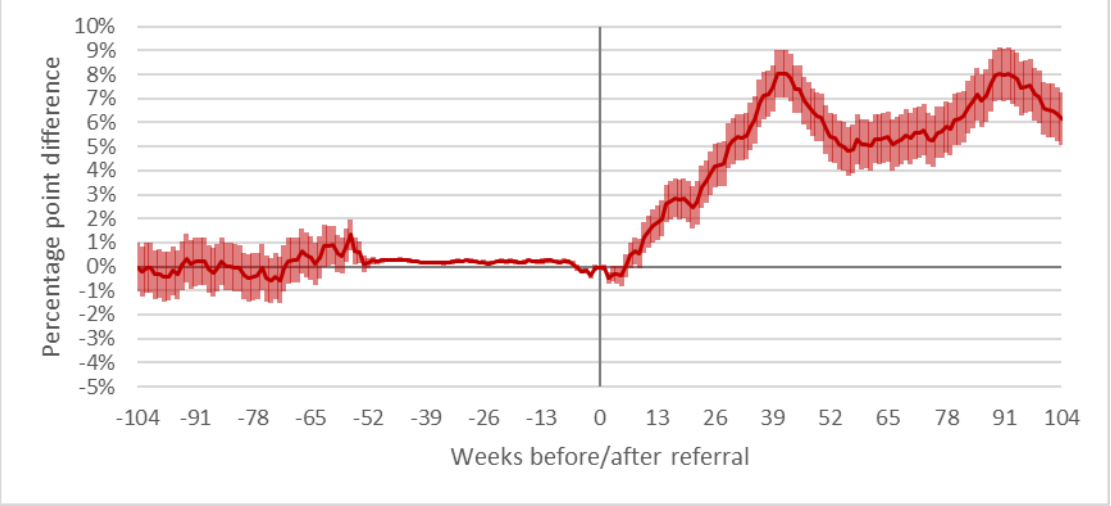


Figure H.5 Proportion of participant and non-participant cohorts (post-matching) neither claiming OW benefit nor in employment

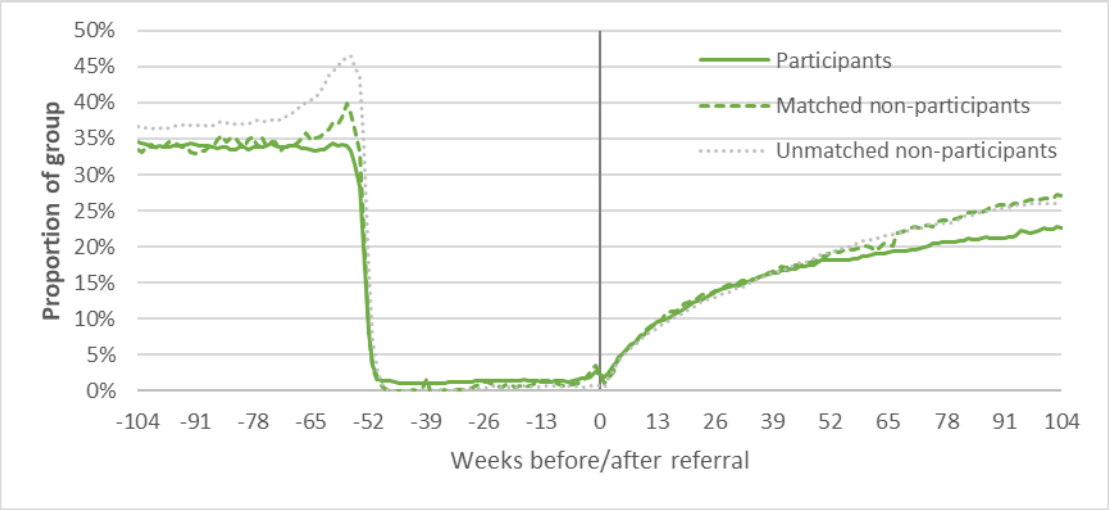


Figure H.6 Impact on the likelihood of neither claiming OW benefit nor in employment for the participant cohort compared to the non-participant cohort

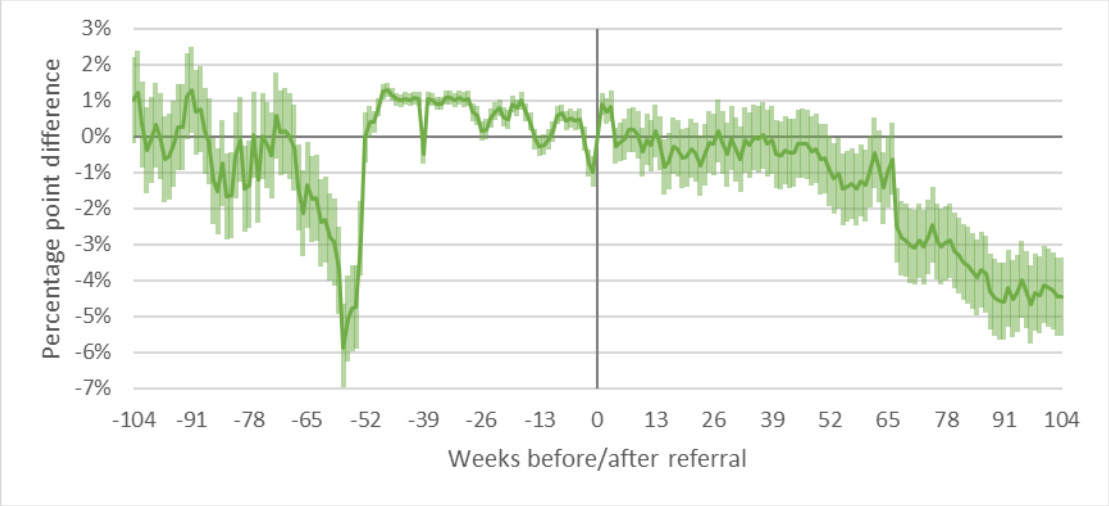


Figure H.1 to Figure H.6, show that prior to the referral date, although matching has made a large difference across outcome states and it is fairly evenly matched for employment outcomes, many differences remain for OW benefit outcomes and neither on benefit nor in employment outcomes, particularly around the period around 52 to 65 weeks prior to the referral date, or equivalent remain. These differences suggest that there is some bias remaining and possibly indicates that there is also unobserved bias that has not been captured. It is clear that matching also makes a larger difference to the pre-referral outcomes than the post-referral outcomes – but this has been observed in some other evaluations which use matching.

The results show that following referral, or equivalent, participants spent more time in employment and less time in receipt of OW benefits than the matched group of non-participants. As the participant and non-participant groups are not drawn from the same time period, the peaks in differences observed in employment at around 39- and 91-weeks post-referral are most likely due to seasonal effects as a result of

drawing two cohorts from different times. It is expected that these would average out over the two-year tracking period. Also it should be noted that although the employment impact stays high, there appears to be some decay in impact on OW benefits.

The additional days the participant group is estimated to have spent in each of the defined outcome states can be calculated over the 2-year tracking period.

Figure H.7 Additional days in the three primary outcome states over 104 weeks for the participant cohort compared to the non-participant cohort

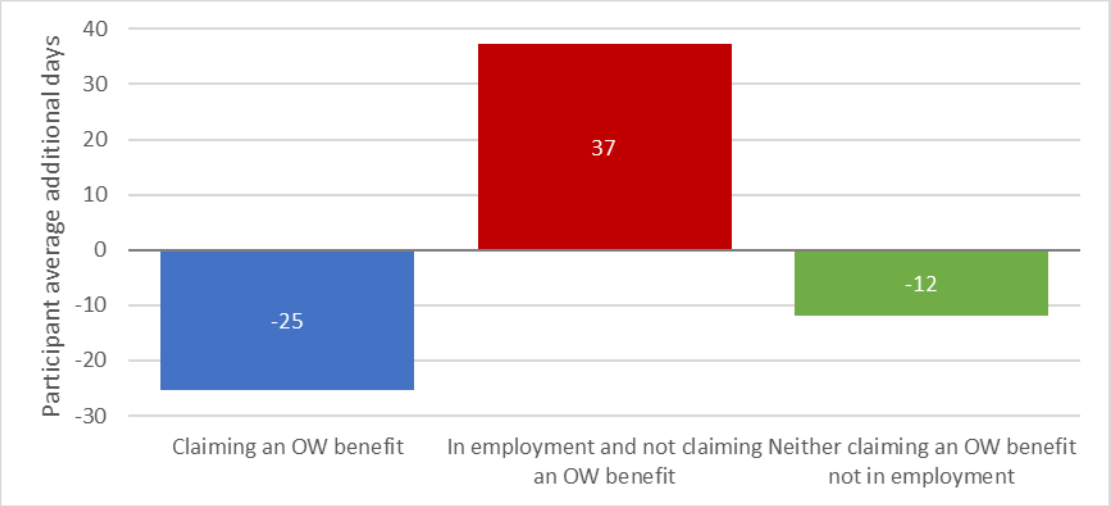


Figure H.7 shows that, compared to the matched non-participant group, over 2 years the participant group spent:

- Less time on OW benefits – 25 fewer days
- More time in employment – 37 more days
- Less time neither on OW benefits nor in employment – 12 fewer days

H.2. Additional outcome states

As defined in section 4.2, four additional outcome states have been monitored. Figure H.8 and Figure H.9 show the difference between the two groups and therefore the estimated impacts.

Figure H.8 Impact on the likelihood of being in receipt of JSA or Universal Credit with Searching for Work conditionality for the participant cohort compared to the non-participant cohort

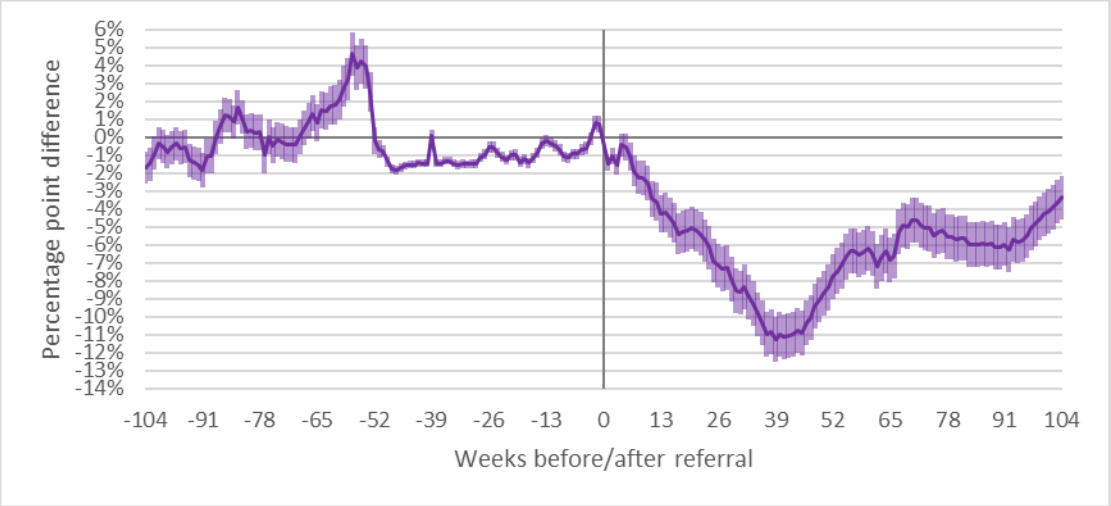
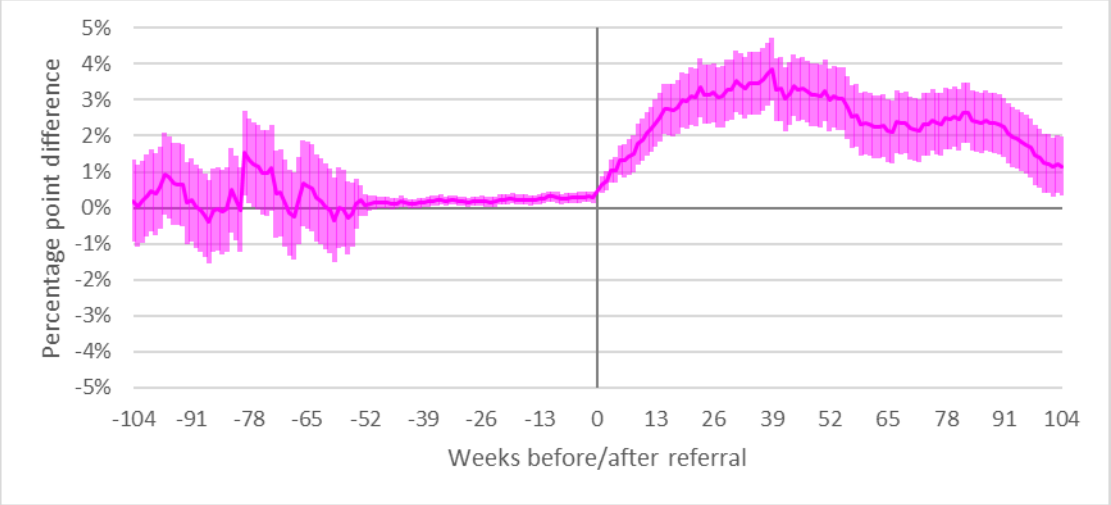


Figure H.9 Impact on the likelihood of being in receipt of the remaining OW benefits for the participant cohort compared to the non-participant cohort



Although Figure H.1 and Figure H.2 showed that participants spent less time in receipt of OW benefits than the non-participant group, Figure H.8 and Figure H.9 show there are differences between the individual benefits. Matching for JSA and UC with Searching for Work conditionality looks to be similar to the matching on OW benefits, but the matching appears to be better for the remaining OW benefits.

In similar way as to the ITT groups, participants claim JSA or UC with Searching for Work conditionality for less time than the non-participant group, which is offset somewhat by the participant group having more time claiming the other OW benefits.

As for the ITT groups, Figure H.10 shows the 'overlap' of claiming an OW benefit and being in employment has been monitored as a separate outcome type.

Figure H.10 Impact on the likelihood of being in receipt of an OW benefit and in employment for the participant cohort compared to the non-participant cohort

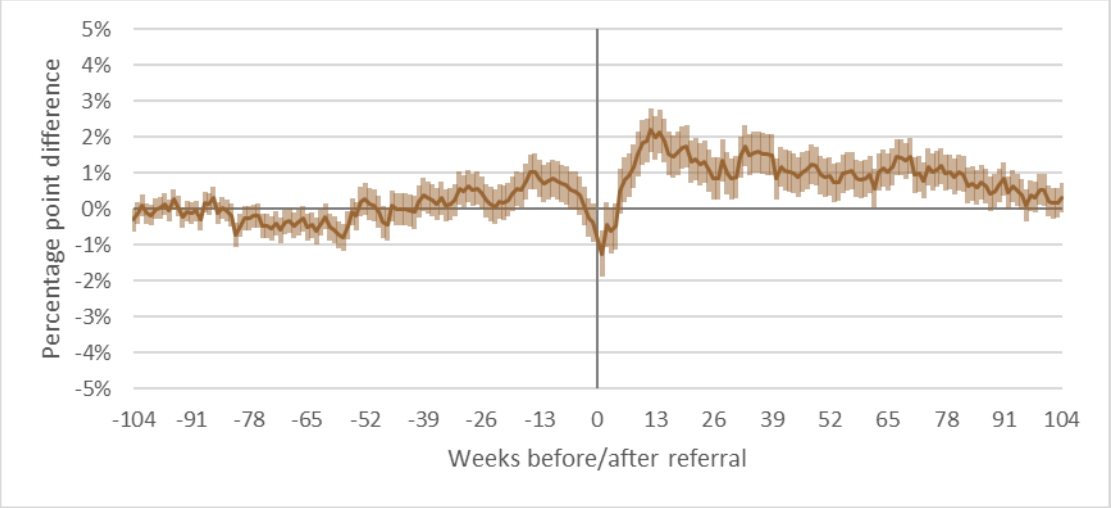


Figure H.10 shows that there is a small significant difference, and therefore impact, on the likelihood of being both on an OW benefit and in employment. Later in the section a sub-group analysis of participants with and without evidence of UC claims prior to programme referral is presented using the time in employment irrespective of receipt of OW benefits outcome state.

The final set of graphs show an additional outcome state considers claims to disability benefits in the form of Disability Living Allowance (DLA) and Personal Independence Payments (PIP). These benefits can be claimed regardless of an individual's employment status, and can therefore coincide with claims to OW benefits or spells of employment.

Figure H.11 Proportion of participant and non-participant cohorts (post-matching) in receipt of DLA or PIP

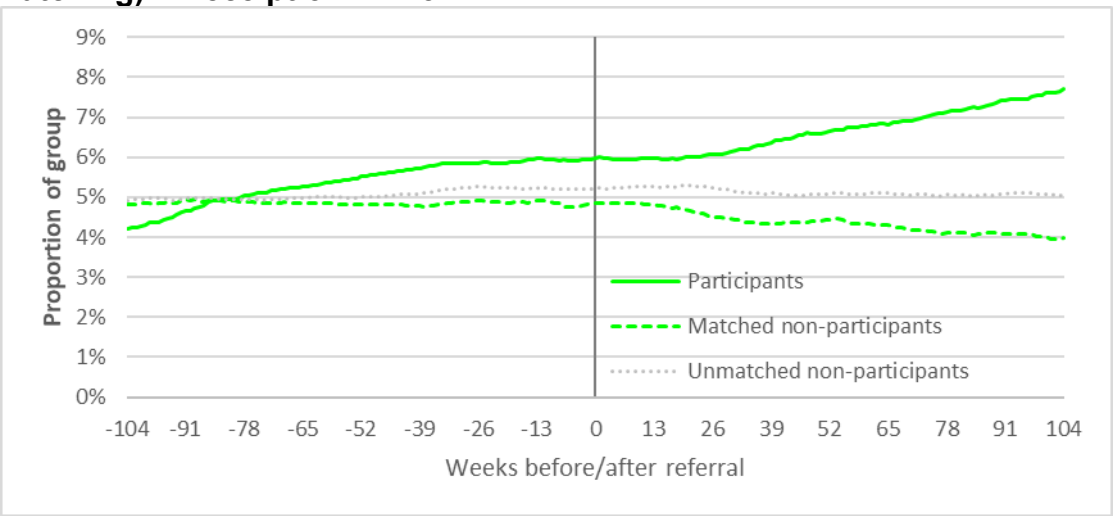


Figure H.12 Impact on the likelihood of being in receipt of DLA or PIP for the participant cohort compared to the non-participant cohort

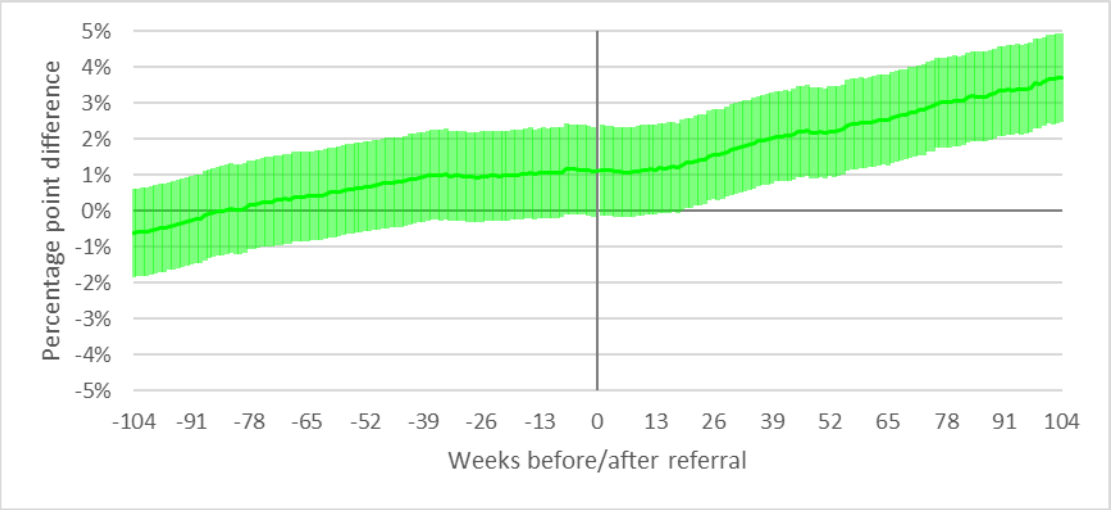


Figure H.11 shows that there is some divergence of programme participants and non-participants in their uptake of DLA or PIP before their referral, or equivalent, to Work Programme Payment Group 2. The proportion of the participant group in receipt of DLA or PIP increased from around 4% to 6%, while the non-participant group stayed at a rate of around 5%. After referral, the participant group increased to over 7%, while the non-participant group fell to around 4% in receipt of DLA or PIP at the end of the two-year tracking period. These differences are seen as the observed impact in Figure H.12.

It is important to recognise that the increase in the levels of DLA and PIP for the participant group is not observed in the impacts for the Intention to Treat (ITT) groups. This suggests that there may be some selection bias within the participant group that the matching is unable to control for. This might lead to other factors, for example, those participating in the programme are more engaged with the benefit system and more effectively signposted to other appropriate support. Linked to this, there is an overlap and association between claiming ESA, part of remaining OW benefits here, and DLA/PIP - therefore given the observed increase in the OW benefits, it is plausible to expect that some of those diverted to ESA might also claim PIP as well. Finally, whilst PIP is intended to replace DLA as a benefit, the two benefits are administered differently which may lead to differences as PIP continues to roll out.

These four additional outcome states can also be measured in terms of the additional days the participant group spends in each state beyond the non-participant group. Unlike the outcomes seen in Figure H.7, the outcomes in Figure H.13 are not mutually exclusive.

Figure H.13 Additional days in the four additional outcome states over 104 weeks for the participant cohort compared to the non-participant cohort

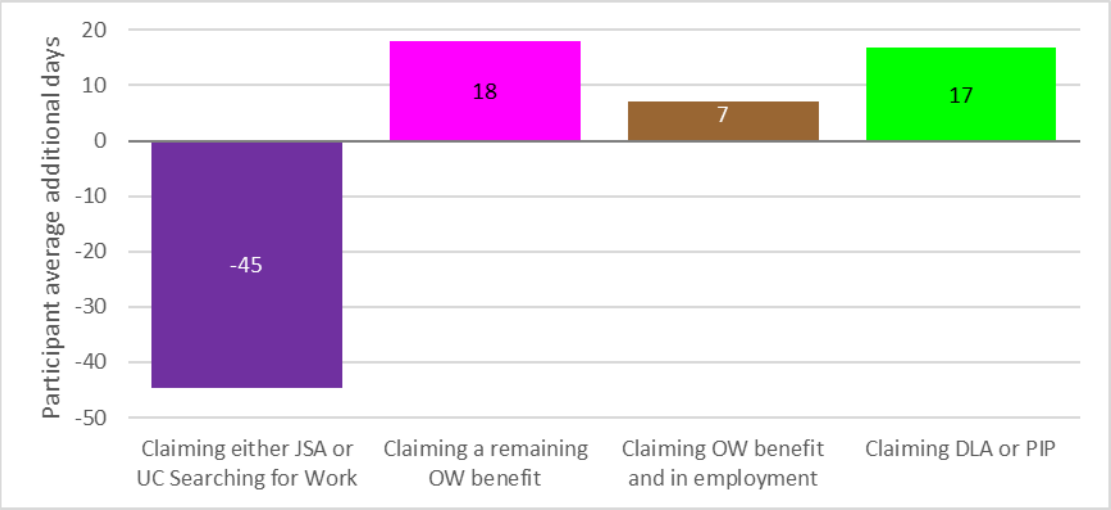
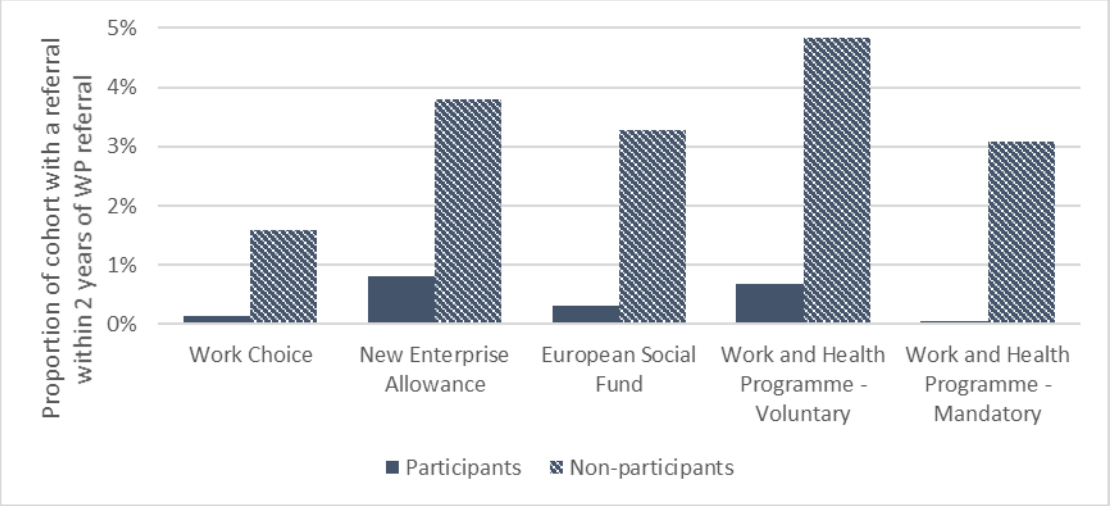


Figure H.13 shows that there was an increase in time spent on other OW benefit types, such as ESA, IS and IB. This increase did not outweigh the decrease in time on JSA or UC with Searching for Work conditionality that was seen. Also around one additional week was spent with an overlap of OW benefits and employment was observed for the participant cohort. There was an increase in time spent claiming DLA or PIP, which can be claimed by individuals in or out of employment or claiming other benefits simultaneously. This increase in time claiming DLA or PIP was similar in size to the observed increase in time claiming other OW benefit types, which are generally related to the claimant having a health condition.

H.3. Engagement with other contracted provision

Figure H.14 shows the proportion of the participant and non-participant cohorts with at least one referral to another contracted programme in the two years following referral, or pseudo-referral, to Work Programme Payment Group 2. These referrals are not mutually exclusive, and it is possible that an individual could have referrals to multiple programmes. Overall just under 15% of the non-participant cohort were referred to other contracted provision, compared to 2% of the participant group.

Figure H.14 Referrals to other contracted provision within 2 years of Work Programme referral for the participant and non-participant groups



As for the ITT groups, if other provision does have an impact, then observed impact estimates here reflect the difference between the impact of the programme and any other provision. Given that the uptake of other provision is higher for non-participants, then assuming that the net impact of other provision is non-negative, this will lead to a small understatement of the true impact.

H.4. Summary of findings from participant and non-participant sub-group analysis

Table H.1 below summarises the findings from the additional analyses deployed to try and quantify sources of bias, namely:

- Sub-groups of March 2017 participants and April 2017 non-participants
- Additional groups of participants from January 2016 to March 2016 and April 2016 to June 2016
- Additional groups of non-participants from April 0217 to June 2017 and July 2017 to September 2017
- Sub-groups of participants and non-participants differentiating between those with and without evidence of a Universal Credit claim prior to programme referral
- Sub-groups of participants and non-participant excluding those with propensity scores of more than 0.9.

Table H.1 Summary of participant additional days in each of the primary outcome states for participant groups

Label	In receipt of OW benefit	In employment and not in receipt of OW benefit	Neither in receipt of OW benefit nor in employment
Participant / non-participant comparison from January 2017 to March 2017 with April 2017 to June 2017	-25	37	-12
March 2017 participants and April 2017 non-participants	-20 (-5)	34 (-3)	-15 (+3)
Time effects from comparing participants from January 2016 to March 2016 with April 2016 to June 2016	-6	2	3
Time effects from comparing non-participant from April 2017 to June 2017 with July 2017 to September 2017	5	-3	-2
Participant / non-participant sub-groups without prior UC claims	-37 (+12)	39 (+2)	-1 (-11)
Participant / non-participant sub-groups with prior UC claims	8 (-33)	29 (-8)	-36 (+24)

Participant / non-participant comparison excluding propensity scores over 0.9	-27 (+2)	34 (-3)	-7 (-5)

Figures in brackets are differences with main participant/non-participant comparison

Appendix I. Additional Cost Benefit Analysis

In addition to cost benefit ratios calculated using the primary outcome state of in employment and not in receipt of an Out of Work benefit, further analysis has been carried out using the additional outcome type of in employment irrespective of Out of Work benefit status.

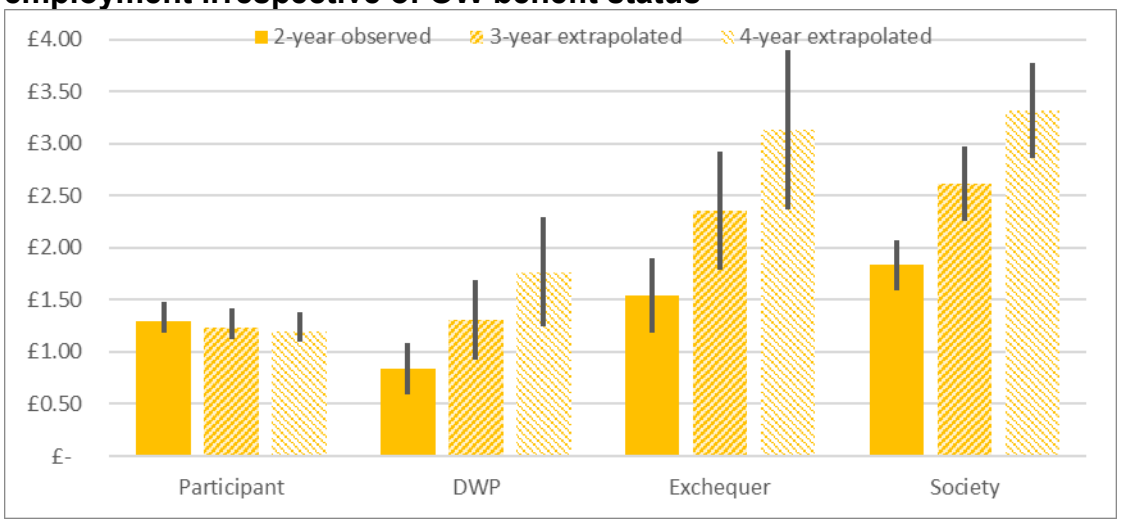
This outcome type may be more appropriate for use with cohorts with a large proportion of individuals in receipt of Universal Credit (UC). This is because a design feature of UC not present in legacy benefits is the ability for individuals to be earning through work and in receipt of benefits. In the case of the cohorts used in this analysis, the proportion of individuals in receipt of UC rather than legacy benefits is around 35%, leaving the remainder in receipt of legacy benefits. Therefore, it is deemed more appropriate to use the primary employment outcome state, as described in section 5.7.

Table I.1 and Figure I.1 mirror Table 5.3 and Figure 5.4 as seen in the main cost benefit analysis of chapter 5.

Table I.1 Cost Benefit ratio estimates from ITT cohorts using in employment irrespective of OW benefit status

		Participant	DWP	Exchequer	Society
Impact of participation	2 year observed	£ 1.30	£ 0.84	£ 1.54	£ 1.83
	3 year extrapolated	£ 1.23	£ 1.31	£ 2.36	£ 2.62
	4 year extrapolated	£ 1.20	£ 1.77	£ 3.14	£ 3.32
Range of impact	2 year observed	£1.18 to £1.48	£0.59 to £1.08	£1.18 to £1.90	£1.59 to £2.08
	3 year extrapolated	£1.13 to £1.41	£0.92 to £1.69	£1.79 to £2.92	£2.26 to £2.97
	4 year extrapolated	£1.09 to £1.38	£1.24 to £2.29	£2.37 to £3.90	£2.86 to £3.77

Figure I.1 Cost benefit ratios for impacts estimated over four years using in employment irrespective of OW benefit status



Glossary

AME	Annually Managed Expenditure
CBA	Cost benefit analysis
CEPMI	Contracted Employment Programme Management Information
CPA	Contract Package Area
DEL	Delegated Expenditure Limit
DLA	Disability Living Allowance
DWP	Department for Work and Pensions
ESA	Employment and Support Allowance
HMRC	HM Revenue and Customs
IB	Incapacity Benefit
Intention to Treat	Cohorts of individuals based on the date they met the programme eligibility criteria, regardless of whether they went on to be referred to the programme
IS	Income Support
ITT	See Intention to Treat
JCP	Jobcentre Plus
JSA	Jobseeker's Allowance
MPL	Minimum Performance Level
NHS	National Health Service
NIL	Non-Intervention Level
Non-participant group	Individuals with a pseudo-referral to Work Programme Payment Group 2 between April 2017 and June 2017
ONS	Office for National Statistics
OW benefit	Out of Work benefit
Participant group	Individuals referred to Work Programme Payment Group 2 between January 2017 and March 2017
PbR	Payment by Results
PIP	Personal Independence Payment
PRaP system	Provider Referrals and Payments system
PSM	Propensity Score Matching
RTI	Real Time Information
SCBA	Social Cost Benefit Analysis
SMD	Standardised Mean Difference
UC	Universal Credit

References

Adams L., Oldfield K., Riley C. and James AS. 'Destinations of Jobseeker's Allowance, Income Support and Employment and Support Allowance Leavers 2011' Department for Work and Pensions research report 791, 2012. (viewed on 1 July 2020)

Beale I., Bloss C. and Thomas A., 'The Longer Term impact of the New Deal for Young People', Department for Work and Pensions Working Paper No.23, 2008.

Caliendo, Mahlstedt and Mitnik, 'Unobservable, but Unimportant? The Influence of Personality Traits (and Other Usually Unobserved Variables) for the Evaluation of Labor Market Policies' IZA discussion paper 8337, 2014. (viewed on 1 July 2020)

Card, D., Kluve J. and Weber A., 'What works? A meta-analysis of recent active labor market program evaluations', *Journal of the European Economic Association* 16(3): 894-931, 2017.

Carter E. and Whitworth A. 'Creaming and Parking in Quasi-Marketised Welfare-to-Work Schemes: Designed Out Of or Designed In to the UK Work Programme?' Journal of Social Policy, 2015 Apr; 44(2): 277–296.

National Audit Office, 'The Work Programme' HC 266, 2014. (viewed on 1 July 2020)

Coca-Perraillon. 'Local and Global Optimal Propensity Score Matching' SAS Global Forum 2007, paper 185-2007. (viewed on 1 July 2020)

Cohen J. Statistical Power Analysis for the Behavioural Sciences, p25, 1977, Available online

Crepon B. and van den Berg GJ. 'Active Labour Market Policies' Working Paper 2016:17

Department for Work and Pensions, 'Alternative Claimant Count statistics: background information and methodology' (viewed on 1 July 2020)

Department for Work and Pensions, 'Universal Credit transition rollout schedule March 2018 to December 2018' (viewed on 1 July 2020)

Department for Work and Pensions, 'Work Programme provider guidance' (viewed on 1 July 2020)

Department for Work and Pensions, 'Work Programme statistics' (viewed on 1 July 2020)

Department for Work and Pensions, 'Work Programme (Universal Credit) provider guidance' (viewed on 1 July 2020)

Dorsett R. and Lucchino P. 'The Work Programme: factors associated with differences in the relative effectiveness of prime providers' Department for Work and Pensions ad hoc research report 26, 2016. (viewed on 1 July 2020)

Foster S. and others. 'Work Programme evaluation: Operation of the commissioning model, finance and programme delivery', Department for Work and Pensions research report 893, 2016

Fujiwara D. 'The DWP Social Cost-Benefit Analysis framework (WP86)' Department for Work and Pensions working paper 86, 2010. (viewed on 1 July 2020)

Greenberg, D., Ashworth, K., Cebulta, A. and Walker, R., 'Do Welfare-to-work programmes work for long'. Journal of Applied Public Economics, Vol 25, Issue 1, p27-53, 2004.

Haigh, R. and Woods, J. 'Work Experience: a quantitative impact assessment' Department for Work and Pensions research report 917, 2016.

Hasluck, C and Green, AE. (2007) What works for whom? A review of evidence and meta-analysis. Department for Work and Pensions research report 407, 2007.

Hillmore, A., Marlow S., Ainsworth P. 'Impacts and Costs and Benefits of the Future Jobs Fund' Department for Work and Pensions Ad-hoc statistical analysis, 2012

HM Revenue and Customs, 'Real Time Information: improving the operation of Pay As You Earn' (viewed on 1 July 2020)

HM Treasury, 'The Green Book: appraisal and evaluation in central government' (viewed on 1 July 2020)

Lechner, M., Miquel, R. and Wunsch, C. 'Long run effects of public sector sponsored training in West Germany', Journal of European Economic Association, Vol 9, Issue 4, 742-784, 2011

Newton, B., Meager, N, Bertram, C., Corden, A and others. 'Work Programme evaluation: Findings from the first phase of qualitative research on programme delivery' Department for Work and Pensions research report 821, 2012. (viewed on 1 July 2020)

Newton, B., Meager, N, and Sainsbury, R. 'Work Programme evaluation: the participant experience report' Department for Work and Pensions research report 892, 2014. (viewed on 1 July 2020)

Office for National Statistics. 'Unemployment rate (aged 16 and over, seasonally adjusted)' (viewed on 1 July 2020)

Prince S., Marlow S., Hillmore A., Pritchard D., and Lally J. 'Early Impacts of Mandatory Work Activity' Department for Work and Pensions Ad-hoc statistical analysis, 2012

Scholz F. and Ingold J. 'Activating the 'ideal jobseeker': Experiences of individuals with mental health conditions on the UK Work Programme', Journal Human Relations, 2020

Ward, R., Woods, J. and Haigh R. 'Sector-based work academies: a quantitative impact assessment' Department for Work and Pensions research report 918, 2016. (viewed on 1 July 2020)