Department for Work & Pensions

GOVERNMENT SOCIAL RESEARCH
Social Science in Government

Employment & Support Allowance: Evaluation of pilots to support Work-Related Activity Group customers with an 18 to 24 month re-referral period

Process and Impact Assessment

January 2019

Research Report 965

A report of research carried out by the Department for Work and Pensions.

This document/publication is also available on our website at: https://www.gov.uk/government/organisations/department-for-work-pensions/about/research#research-publications

If you would like to know more about DWP research, please email: Socialresearch@dwp.gsi.gov.uk

First published January 2019.

ISBN 978-1-5286-0845-9

Value of this research

This is a report of findings from a process and impact evaluation to produce new evidence to inform policy about what works to help Employment and Support Allowance (ESA) and Universal Credit claimants with complex health conditions to improve their prognoses, enabling them to move closer to work and to fulfil their potential.

Trustworthiness

This report is part of the DWP research report series and as such adheres to the Government Social Research publication protocol and the Government Social Research Code for Products. The report has been assured by professionally badged Government Social Researchers in DWP and its production has been supported by other professionally badged analysts from the relevant government analytical services including the Government Statistical Service.

Quality

The impact evaluation was conducted internally by DWP professionally badged analysts using management information and DWP administrative data. The methods and analysis described in the report have been internally peer reviewed with reference to The Magenta Book on evaluation and assured by other analysts in the department.

Executive summary

This report presents findings from a process and impact evaluation of pilots to support ESA (Employment and Support Allowance) WRAG (Work-Related Activity Group) customers with an 18 to 24 month re-referral or 'prognosis' period. The pilots aimed to test the effectiveness of enhanced support to these customers, delivered over the course of two years. Three distinct models were piloted:

- Work Programme Provider (WPP)
- Healthcare Provider (HCP)
- Jobcentre Plus Work Coaches (JCP)

The aim of the pilots was (i) to establish whether enhanced support for this client group had a beneficial effect, and (ii) to identify who is best placed to provide that support.

In each pilot, the outcomes for participants - measured in terms of reduction in days on benefits and employment outcomes - were compared with non-participants who received the standard JCP support offer.

The report provides analysis of (i) how the pilot was implemented, and (ii) the pilot outcomes.

Analysis of the implementation found that deficiencies in administrative data may have led to some people who should have been recruited to the pilot, not participating.

Analysis of outcomes found minimal employment impacts across the three pilots, but some impact on reduction in days on benefit, in particular the JCP and Work Programme pilots.

Contents

Acknowledgements

The Department is grateful to Peter Dawson, Martin Moran and Samuel Smithers of the Department for Work and Pensions (DWP) for producing this analysis and report.

Glossary of terms

Employment and Support Allowance (ESA) – a type of unemployment benefit offering financial support to people who are out of work due to long-term illness or disability.

ESA Work-Related Activity Group (WRAG) – people claiming ESA are placed into two groups depending on the extent to which their illness or disability affects their ability to work. The work-related activity group are required to have regular interviews with an adviser and undertake work-related activities.

ESA WRAG 18-24 month Prognosis Group – once a claimant has been found to be eligible for ESA they will be allocated a prognosis or re-referral date where their entitlement to benefit will be reconsidered. The prognosis group was used to determine whether claimants were eligible for the Work Programme. This pilot extended mandatory referrals to the Work Programme for participants in the 18-24 month Prognosis Group.

Jobcentre Plus Work Coach – front-line DWP staff based in job centres who support claimants into work by challenging, motivating, providing personalised advice, and using knowledge of local labour markets.

Jobseeker's Allowance (JSA) – a type of unemployment benefit paid to people who are out of work and actively seeking work.

Universal Credit (UC) – Universal Credit is a payment to help with living costs for people on low income or out of work. It will replace six existing means-tested benefits and tax credits: Child Tax Credit; Housing Benefit; Income Support; income-based Jobseeker's Allowance (JSA); income-related Employment and Support Allowance (ESA), and; Working Tax Credit.

Work Capability Assessment (WCA) – a requirement of every ESA claim which measures the extent to which illness or disability affects one's ability to work.

Work Choice – a specialist disability employment programme delivered by a range of provider organisations, offering work entry support and up to two years in-work support for people with disabilities.

Work Programme (WP) – an employment support programme delivered by a range of providers with the aim of helping long-term unemployed JSA and ESA claimants find employment.

Abbreviations

DWP	Department for Work and Pensions
ESA	Employment and Support Allowance
HCP	Health Care Professional
HMRC	Her Majesty's Revenue and Customs
IB	Incapacity Benefit
IBR	IB Reassessment
ITT	Intention to Treat
JCP	Jobcentre Plus
JSA	Jobseekers Allowance
JSAPS	Jobseekers Allowance Processing System
LMS	Labour Market System
NINO	National Insurance Number
PRAP	Provider Referral and Payment System
RCT	Randomised Controlled Trial
UC	Universal Credit
WCA	Work Capability Assessment
WP	Work Programme
WPP	Work Programme Provider
WRAG	Work-Related Activity Group

1 Summary

About the pilots

1.1 This is a quantitative process and impact evaluation of pilots providing two years of enhanced support to ESA (Employment and Support Allowance) WRAG (Work-Related Activity Group) customers with an 18 to 24 month re-referral or 'prognosis' period. The pilot has three variants, each with a different provider:

- Work Programme Provider (WP)
- Healthcare Provider (HCP)
- Jobcentre Plus Work Coaches (JCP)

1.2 The aim of the pilots was (i) to establish whether enhanced support for this client group had a beneficial effect, and (ii) to identify who is best placed to provide that support.

Randomisation approach

1.3 The pilots, which were each being tested in a different JCP Group, used a random allocation approach.

1.4 The random allocation error rates varied from three per cent to six per cent depending upon the model and (treatment or control) group but most client characteristics remained reasonably well-balanced. That said, exemption rates were higher than expected (roughly 10 per cent to 30 per cent) and were inconsistently applied.

Analysis of the pilot implementation

1.5 Analysis of the pilot implementation, from the WCA outcome to the recruitment interview and subsequent support, revealed widespread issues. Taking the DWP administrative data at face value, one would conclude that:

- not everybody who should have been recruited was recruited
- some recruits were not in the scope of the pilot
- a lot of recruits did not then undergo the pilot intervention.

1.6 It is likely that these conclusions arise from deficiencies in the administrative data more than they do from poor pilot implementation. However, we are confident that not all members of the treatment group underwent the pilot intervention.

Pilot outcomes

1.7 Despite showing an initial (over the first six months of the pilot) positive impact, the HCP model did not show a statistically significant reduction on benefit dependency when comparing the treatment group with the control.

1.8 The WP model showed a steady increase in the reduction on benefit dependency amongst the treatment group for the first two years of treatment; participants in the WP treatment group were off benefit for an additional 14.6 days over three years compared to the control group. After two years, however, this impact began to steadily decrease.

1.9 The JCP model showed a steady increase in the reduction on benefit dependency amongst the treatment group for the first two years of treatment. After two years this impact has been maintained fluctuating around 2ppts. Compared with the control group, participants in the JCP treatment group were off benefit for an additional 19.3 days over three years.

1.10 There were minimal employment impacts shown across all three pilots. However we should reiterate that these impacts will be underestimating of employment and will not show any self-employment outcomes.

2 Introduction

2.1 These pilots provided enhanced support to ESA customers in the Work-Related Activity Group customers with an 18 to 24 month re-referral or 'prognosis' period. The pilots tested three different models for the delivery of the enhanced support, namely:

- **Healthcare Provider (HCP)-led** – mandatory engagement with HCPs

  As a minimum, this support comprised an initial meeting of not less than 60 minutes duration, followed by at least two further mandatory face-to-face interviews (of at least 45 minutes duration) within the first 26 weeks of the point of referral onto the pilot, and a further two face-to-face interviews after 12 and 18 months (both also of at least 45 minutes duration). Additional HCP interviews were mandatory but additional support was voluntary and was at both the Provider's and the Claimant's discretion. The payment comprised only an attachment fee and not an outcome-related component.

- **Work Programme (WP)** – black box support determined by the WP provider

  Whilst referral to the WP provider was mandatory, the support offered was wholly at the discretion of the WP provider. The WP provider was incentivised by the standard payment model (applicable to the Payment Group 7 group[1]); the pilots do not involve a change to the WP payment model. Participants referred to the WP provider would receive two years of support.

- **Jobcentre Plus (JCP)-led** – enhanced JCP support

  The JCP-led model comprises extra time spent with a JCP Work Coach (530 minutes per year) coupled with improved Work Coach training. It is anticipated that this will lead to increased engagement with employment-related services, leading to improved employment outcomes amongst pilot participants.

2.2 For all three variants, the pilots allow for certain exemptions and easements (e.g. for carers and lone parents with children under school age). All three pilot variants began on 25 November 2013, and nominally continued with enrolment until 29 August 2014. Each participant actively engaged with the pilot for up to 104 weeks, which meant that the pilot support extended to the end of August 2016.

2.3 The pilots are operated in three different JCP Groups. These are the Groups and their participating Districts (as defined at the time of pilot recruitment):

Central England Group: *The Healthcare Provider Model*

Black Country

Derbyshire

Leicestershire and Northamptonshire

Staffordshire and Shropshire

Lincolnshire, Nottinghamshire and Rutland

[1] There are nine payment groups, each reflecting the different levels of support that particular populations of DWP claimant will, on average, need in order to enter employment.

North East Group: ***The Work Programme Model***

Durham and Tees Valley

North East Yorkshire and the Humber

Northumberland, Tyne and Wear

Southern England Group: ***The Jobcentre Plus Enhanced Model***

Greater Wessex

Thames Valley

Devon, Cornwall and Somerset

Surrey and Sussex

Gloucestershire and West of England

2.4 The three different models all employed a Randomised Controlled Trial (RCT) approach whereby half[2] (i.e the 'treatment' group[3]) the potentially eligible participants were directed onto the pilot whilst the other half comprised a 'control' group who experience the 'business as usual' service from JCP.

2.5 The process by which potentially eligible claimants were randomly allocated will be described later when we discuss the quality of the implementation of the random allocation. For now, we simply note that the reason for using an RCT approach was to reduce the chances of systematic differences existing between the treatment and the control groups, and therefore reduce the likelihood that other factors may have contributed to any observed difference in outcomes.

2.6 Here, we refer to differences *within* the different pilot models and not *between* them. Because we randomly allocated within each pilot variant, we can in principle attribute differences (between the treatment and control group) *within* that variant to the pilot impact. However, because we are *not randomly allocating between* the different pilot variants, the different impacts of each pilot model are potentially attributable to factors other than their relative efficacy.

2.7 During the planning phase of the pilots, the number of people assumed to be enrolled onto the pilot was inherently uncertain, partly because we cannot predict future benefit inflows with accuracy, and partly because of limitations in the data available to inform the pilot design. Originally, DWP analysts anticipated participant volumes in the region of:

- 2,000 to 2,500 participants in the HCP-led pilot
- 2,000 to 2,500 participants in the WP-led pilot
- 3,000 to 3,500 participants in the JCP-led pilot

and, for all three variants, an equivalent number of ESA claimants to be in the control group. The extent to which these volumes were realised is discussed later in this report.

[2] In Central England this was changed to a 60:40 split mid-way through the pilot recruitment period in order to increase the number of referrals to the Health Care Professional provider (Ingeus).

[3] The word 'treatment' is a standard term and simply refers to the fact that this group encounters the new intervention. It does not imply that claimants will undergo any medical intervention.

2.8 This report summarises what we know about the implementation of the pilots, and for the JCP model provides an overview of the support provided to pilot participants insofar as we can tell from the available data. We cannot provide a description of the support received by claimants engaged with the HCP and WP models as we cannot access private sector data. However, we can and do identify those pilot recruits who were at least referred to the provider and did begin their provision. This is an important consideration when assessing the extent to which the pilot model was delivered.

2.9 Following this analysis of the pilot processes, we compare the characteristics of the treatment group to that of the control. The findings of this analysis underpin assumptions within the impact assessment.

3 Data Sources

This section of the report briefly summarises the data that is available to the analysis described in this report. We exploit different sources of information for different purposes and each has its own characteristics as follows:

Pilot Marker data: when recruiting claimants into a group, a pilot marker is set on the 'Labour Market System' (LMS). This system is used to administer benefit conditionality and claimant support, and houses a range of markers used to identify people who have been recruited onto these pilots, and their allocated group. This data is very timely and provides us with a complete list of pilot recruits, their location and their exemption status.

Other LMS data: via separate data feeds we also have data on pilot participants' personal characteristics, and on the support they received, both during the claim that led to their participation and prior - and in due course, future - claims. We draw a distinction between this data and the pilot marker data because the data is not specific to the pilot and will, in some cases, have been collected over a much longer period of time.

DWP Benefits Administrative data: this data is sourced from several benefit systems and provides information on all past and present DWP benefit claims (over the past 16 or so years). We only use information for Primary DWP benefits (i.e. income replacement benefits [4]). This information is both more unstable and more out of date than the LMS data owing to various factors such as: the time it takes to collate and process the data, and; backdated claims and retrospective changes (due, for example, to appeals and changes of circumstance). Nominally, the data is only complete up to April 2017.

In our experience, however, data is reasonably complete to one or two months beyond the nominal date so we use information up to and including June 2017 with the caveat that the leading edge of this data will to some extent be subject to incompleteness and retrospective revisions. In principle, owing to the randomised nature of these trials the treatment groups are as likely as their control groups to be affected by these shortcomings, so we can still make comparisons on the basis of the data that we have.

Employment data: we use HMRC's P45 data to identify employment start and end dates. This data is subject to considerable time lags and generally only approaches completeness[5] some six months after the event. As well as being a relatively out of date source, there are known shortcomings in the HMRC data. For example, many employment start and end dates are set to the beginning and end of the tax year and do not reflect the true employment period. Furthermore, many employment records do not have end dates, possibly because the jobs are ongoing or that they

[4] Jobseeker's Allowance, Employment Support Allowance, Income Support, Incapacity Benefit, Carers Allowance and Bereavement Benefit.

[5] By 'completeness' we do not mean that we will have information on every employment spell. Rather, we mean that we will have received all the information that we will ever receive, which will still fall short of all actual employment spells. The size of this shortfall is unknown but is likely to be in the region of 20 per cent to 30 per cent.

have ended without our knowledge. There is also a known and significant under-representation of self-employment as well as jobs where the earnings are lower than the tax threshold (and therefore do not need to be declared to HMRC).

Work Capability Assessment (WCA) data: this information source describes the decisions made by ATOS, the company contracted to carry out the assessments of ESA and ex-IB claimants' capacity for work. The data also includes information on subsequent JCP decisions about the conditionality group that a person should be placed within, and the outcome of any appeals made in response to that decision. This information source is potentially subject to a significant number of retrospective revisions due to the appeals system, which will lead to many original decisions being overturned. This source theoretically allows us to identify people in scope of the pilots, but we also examine WCA decisions following pilot participation as an outcome in its own right on the assumption that changes in self efficacy can manifest in the assessment process. We are using the July 2017 data release.

4 Pilot Implementation

4.1 This section of the report summarises the data that defines which individuals participated in the pilots and when. We characterise the marker data that identifies pilot participants and examine if the random allocation has been properly implemented.

4.2 In all three pilot models, eligible participants are randomly allocated to either a treatment or a control group. The allocation is done on the basis of the last three digits of the National Insurance Number (NINO) on the assumption that this number bears no relationship with any participant characteristic or pilot outcome. If this assumption is true then the NINO provides us with a readily available pseudo-random number generator and a convenient check of the random allocation process[6]. Once a pilot recruit is allocated to a group, an LMS marker should be set in order to record which group each person has been allocated to. The marker also records exemptions but not the exemption reason[7].

4.3 For the majority of the recruitment period, the random allocation was done on a 50/50 basis, save random fluctuations around this figure, i.e. half of all eligible participants put in the treatment group and half the control. An exception was the Central England Group whereby from the 6th May 2014 onwards the split was altered to 60/40 in favour of the treatment group. This change was made because of low referral numbers to the HCP and a request by DWP Commercials to increase those referrals.

4.4 We now examine those markers with a view to identifying a group of treatment and control individuals and the point in time when they were recruited onto the pilots.

4.5 The marker data used by this analysis comprises 21,032 records across all three pilot variants. Two records have no NINO and are therefore of no use to this study. 18,687 individuals are represented in the data because 2,094 people had two or more records within the data. These 2,094 had between them 4,437 records. The duplicates were due to:

- people moving offices: 515 (25 per cent)
- markers being set to treatment or control and exempt on the same day: 386 (18 per cent)
- markers being set to treatment or control and exempt on a different day: 937 (45 per cent)
- markers changing from exempt to treatment or control: 215 (10 per cent)
- markers being changed from treatment to control or vice versa: 287 (14 per cent).

[6] Other methods of random assignment would require us to know what group a person was put in at the point of allocation; collecting this information would be administratively burdensome.

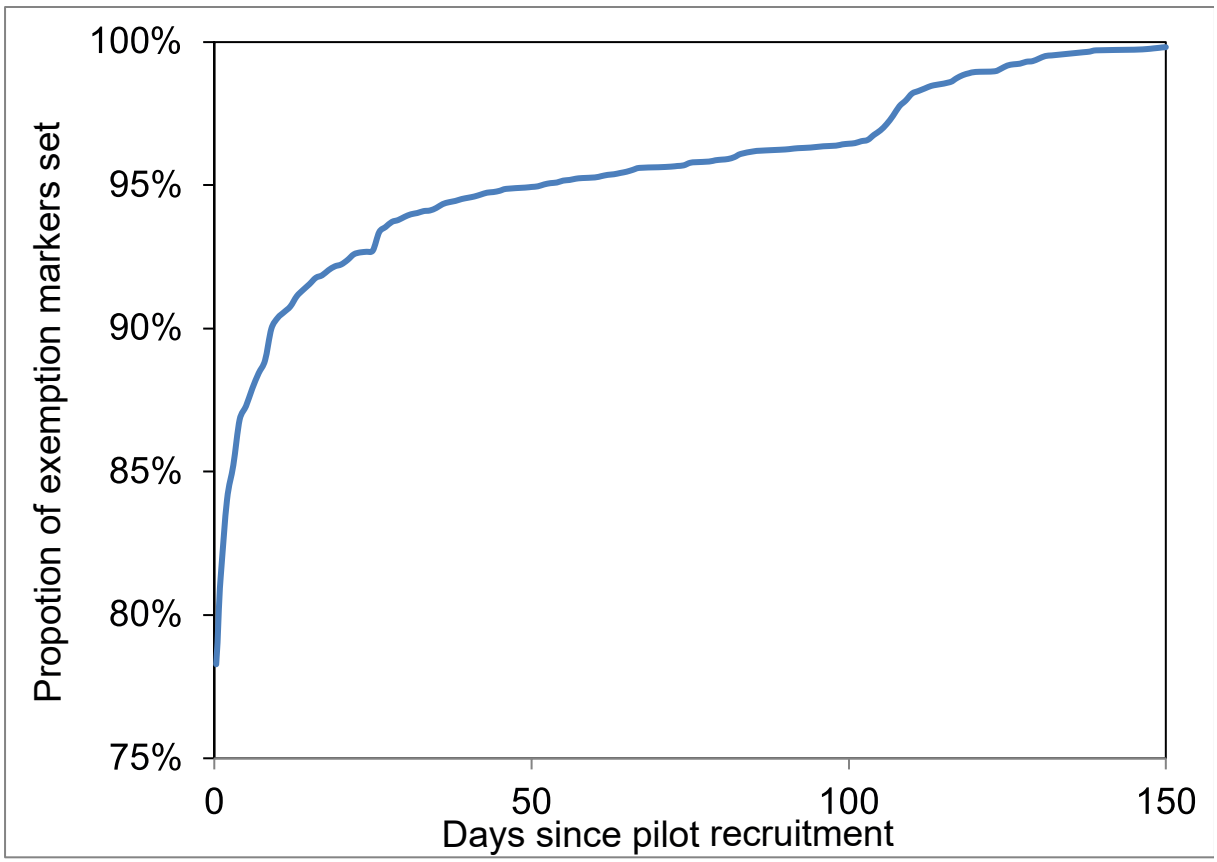
[7] To include exemption reasons on the marker values was deemed to unduly complicate the use of the markers. Further, very few exemptions were expected when the pilot was in its development stage.

4.6 Where a person moves offices, there may have been ample opportunity to receive support as part of the pilot so we retain these cases in our analysis. Because they reflect behaviours we would expect under national roll-out of the pilot model, including them ensures similarity between our estimates and what we might expect under national roll-out.

4.7 Where a person is made exempt we classify that person as wholly exempt regardless of which markers preceded the exemption status. However, where an exemption marker was set prior to a treatment or control marker, we assume that the exemption is erroneous and allocate that person according to the latest information. In other cases, where there was some movement between the treatment and the control group, we allocate that case according to the most recent marker.

4.8 Most (78.4 per cent) exemption markers were set on the day the claimant was recruited to the pilot, and, as chart 4.1 shows, 96.6 per cent within 100 days of pilot recruitment. We assume that the remaining 3.4 per cent of cases had ample time to experience the pilots' effects, and we reset their marker to the value that was set before the exemption was made.

Figure 4.1 The cumulative proportion of exemption markers set since pilot start date



Source: Labour Market System July 2017

4.9 Recruitment to the pilot began in October 2013 with the bulk of recruitment being carried out over the following two years. Indeed, by 1st January 2015 the initial marker for 96.8 per cent of participants had been set. Because there

are doubts as to whether or not those recruited later in the trial period (in 2015 onwards) would have undergone the same process as those recruited previously, we drop the remaining 3.2 per cent of participants from any further analysis, leaving us with 18,086 participants.

4.10 After applying the rules above we find that a significant number (2,785) of people have a single marker value of exempt. In other words, we do not know which group they were nominally allocated to and cannot identify this from their NINO (because it is encrypted). This need not matter if these cases are left out of the analysis. However, for basic checks, such as whether the exemption criterion was applied equally to the treatment and control it is helpful to know the (treatment or control) group that a person was or should have been allocated to.

4.11 A separate source of information on the LMS pilot markers which holds unencrypted NINOs (because it was used for operational checks on the live LMS) and can therefore be used to identify the nominal allocation for these 2,785 cases.

Table 4.1 Actual Compared with Nominal Random Allocation

Actual Allocation	Nominal allocation on the basis of NINO					
	HCP		WP		JCP	
	Treatment	Control	Treatment	Control	Treatment	Control
Treatment	2,506	112	1,840	76	2,425	54
Control	186	2,238	109	1,984	151	2,653
Exempt	503	353	278	230	1,253	1,116
Error rate	5.8%	4.1%	4.9%	3.3%	3.9%	1.4%

Source: Labour Market System: July 2017

4.12 Random allocation was carried out on the basis of the last three digits of the NINO, with those ending in 000 to 499 being allocated to the treatment group and the remainder to the control. An exception was the HCP model. Owing to low numbers of referrals in the first part of the recruitment period, commercial elements of DWP requested that referrals to the HCP be increased in order to meet contractual obligations. Consequently, from 6 May 2014 onwards, people potentially within scope of the HCP model and with a 000 to 599 NINO were allocated to the treatment group.

4.13 In order to identify the nominal allocation, we need to identify which JCP Group or pilot model each person resided within. More specifically, because not every District in the participating Groups took part in the pilot, we need to identify their District. The data available to this study was limited to the JCP office but using that information we were able to identify the District (and therefore the model) for all but 19 of the 18,086 participant records under analysis. These were all located in South Yorkshire and we believe that these markers were set incorrectly in the early days of the pilot.

4.14 Table 4.1 shows how the 18,067 cases are distributed across the pilot models and the nominal and actual groups. It shows that treatment group errors are greater than control errors, and that JCP staff were more likely to ‘move’ someone from the treatment group to the control group than they were the

opposite. This is a common phenomenon which we assume occurs because the people doing the random allocation do not think that the trial is appropriate for some individuals so they do not put them in the correct group. We do not dwell here on the other reasons for such errors. Rather, we categorise people according to the actual group that they were allocated to and where we do not know that group, we allocated them according to their nominal group.

Table 4.2 Recruitment Volumes and Exemption Rates

	HCP		WP		JCP	
	Recruits	Exemption Rate	Recruits	Exemption Rate	Recruits	Exemption Rate
Treatment	3,195	15.7%	2,227	12.5%	3,829	32.7%
Control	2,703	13.1%	2,290	10.0%	3,823	29.2%

Source: Labour Market System: July 2017

4.15 Table 4.2 summarises the data in table 3.1 to show the total number of people with an LMS pilot marker within the pilot Districts and their exemption rates. The JCP model has noticeably more exemptions than the other two models. Furthermore, the difference in allocation error rates between the treatment and the control groups has led to a disproportionate number of people in the control. Owing to the change in the allocation method in May 2014, this is less evident for the HCP model though nonetheless still true.

4.16 We note that the pilot volumes compare reasonably well with the expectations outlined in paragraph 2.7. However, as we show, the actual numbers who meaningfully participated in the pilot appears to be somewhat lower than table 4.2 suggests.

4.17 The LMS markers were not set up to record the exemption reasons on the assumption that they would be very few in number and also to simplify the markers as much as possible. However, we can explore possible exemption reasons by checking the status of the pilot participants within DWP's administrative data. The exemption criteria specified in the pilot guidance are as follows:

- lone parents with a child under one;
- full-time carers;
- claimants who are pregnant and within three months of their expected date of delivery;
- claimants in residential rehabilitation for substance misuse;
- claimants who are currently in hospital;
- claimants with a confirmed job start within 90 days;
- claimants currently attending Work Choice; and
- claimants currently attending residential training for adults.

4.18 As well as these exemption criteria, Contributions-Based ESA cases and people who are on the Work Programme at the time of recruitment are out of scope entirely. For some of these criteria we are not able to quantify the number of people to whom they apply. However, table 4.3 shows the proportion of people with a pilot marker (exempt or otherwise) who meet the exemption

criterion that we can quantify, namely: on the Work Programme at the time of the pilot; claiming Carers Allowance; have a Contributions-Based Claim, and; be a lone parent with a child aged under one and/or in the last three months of pregnancy (though we also allow for a few months after the birth given the impracticality of fulfilling conditionality shortly after giving birth).

Table 4.3 All Participants Fulfilling Exemption Criterion

Exemption Criteria	HCP		WP		JCP	
	Treatment	Control	Treatment	Control	Treatment	Control
On WP	3.3%	3.7%	0.9%	3.1%	7.7%	6.5%
Carer	2.2%	1.8%	1.6%	1.7%	2.3%	2.1%
Conts.-Based	34.8%	34.2%	36.3%	35.1%	36.4%	38.1%
Pregnant	0.2%	0.1%	0.1%	0.3%	0.2%	0.3%
Lone Parent	0.5%	0.5%	0.4%	0.9%	0.4%	0.6%
Any Reason	39.2%	38.2%	38.1%	38.9%	44.2%	44.5%

Source: Labour Market System: July 2017 and DWP administrative data April 2017.

4.19 The administrative data we have used records whether the claimant was receiving Contributions-Based ESA at the start of the claim. We were unable to pinpoint if the claimant was still receiving Contributions-Based ESA at the time they were recruited to the pilot; as such we will have overestimated the proportion of recruits that have a Contributions-Based ESA claim.

4.20 Some people meet more than one of these criteria; table 4.3 provides indication of the proportion of participants that do. It is evident that the JCP model set LMS markers for more Contributions-Based individuals than the other two models, hence the higher overall exemption rate. The JCP model also had more people already on the WP, though this reason only makes a relatively small contribution to the differences in the exemption rates. The lower proportion of WP exemptions (or more precisely, people out of scope) in the WP model is possibly due to a clearer understanding of pilot eligibility in those Districts.

4.21 We note that not all of the people fulfilling some exemption criterion had an exemption marker set. Further, we could not find an exemption reason for everybody who did have an exemption marker set. Table 4.4 shows the reason that we did find for the latter group, those who did have an LMS exemption marker set.

Exemption Criteria	HCP		WP		JCP	
	Treatment	Control	Treatment	Control	Treatment	Control
On WP	30.8%	35.7%	6.8%	7.4%	24.8%	23.2%
Carer	7.6%	5.4%	8.3%	7.0%	6.1%	5.2%
Conts.-Based	40.6%	42.8%	50.0%	59.1%	55.3%	60.5%
Pregnant	0.4%	0.0%	0.0%	0.0%	0.5%	0.6%
Lone Parent	0.8%	0.3%	1.1%	1.7%	0.6%	0.7%
Any Reason	71.6%	75.4%	60.8%	69.1%	78.2%	81.1%
No Reason	28.4%	24.6%	39.2%	30.9%	21.8%	18.9%

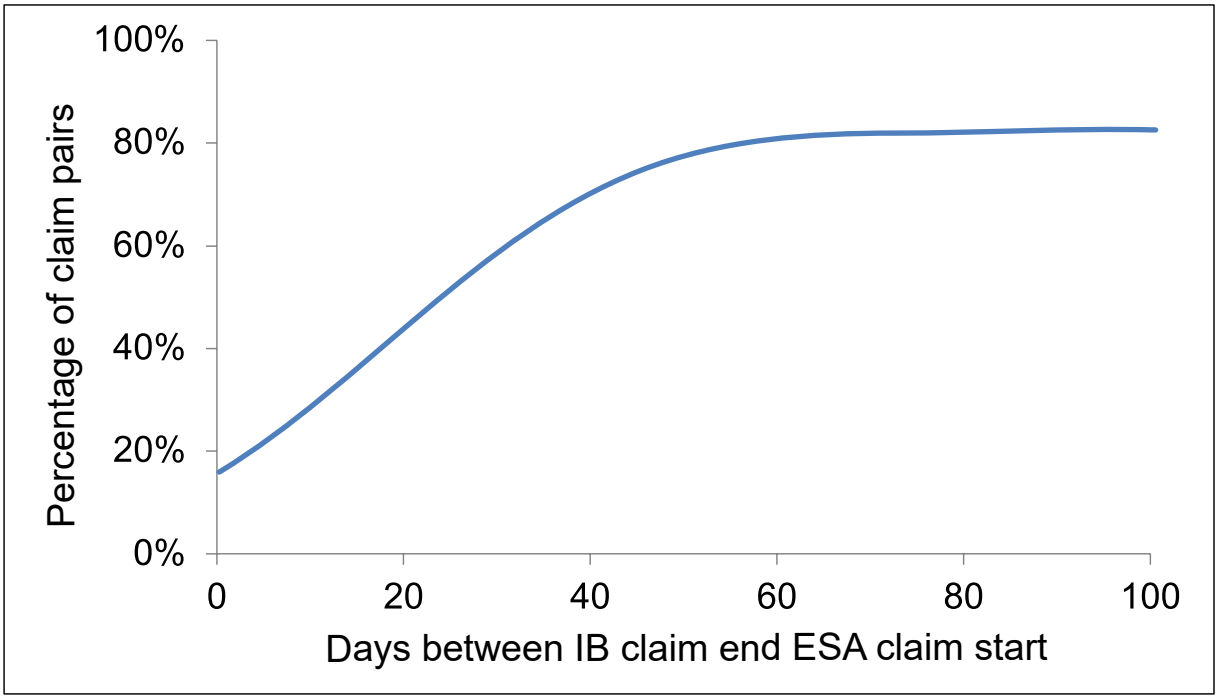
Source: Labour Market System: July 2017 and DWP administrative data April 2017.

4.22 Table 4.4 shows that roughly one-fifth to one-third of people in any particular group do not fit one of the exemption criterion explored. A simple explanation is that these people met one of the criteria that we had no data for . Alternatively, the shortfall may partly be due to inaccuracies in the timings of events. For example, a marker might have been set before or after a particular exemption status applied. Low-matching rates between different data sources will also be a factor. Regardless, the purpose of the exemption analysis was to account for the greater exemption rate within the JCP model and the observation of greater numbers of Contributions-Based cases largely answers this question, so we move on to consider the extent to which people with LMS pilot markers actively participated in the pilots.

4.23 In order to measure participation in the pilot processes, we need to place the people with LMS markers within the wider ESA claim processes. A starting point for this is to identify the ESA claim that led to pilot participation and therein place relevant events such as the WCA and (for the HCP and WP models) provider referrals within the context of that claim.

4.24 We were able to identify an ESA claim for 98.2 per cent of people with an LMS pilot marker (exempt or otherwise) that was live on the day when the marker was first set. For a further 1.2 per cent of people we could identify a claim before and/or after the pilot marker was first set. For around 17 per cent of these claims an ESA claim was started within 60 days of the pilot marker being set. We will retain these claims and shift the pilot start date to the date the claim started. For 0.6 per cent of cases we could find no ESA claim at all. It is not clear why these inconsistencies exist. Given that a person must have had a WCA to qualify for the trial and given the waiting time to the WCA, one might imagine there to be an unambiguous association between the LMS marker and the underlying ESA claim. Claims can sometimes be backdated or temporary NINOs used (though the latter might only apply to the LMS and not the benefits data). Equally, LMS markers might be set later than guidance suggested or in advance of some anticipated meeting but critically, when the ESA claim was not active. Also, NINOs (the common variable that we have used for the match) may have typographical errors in them. There are many possible reasons for the imperfect match but none that we can directly confirm. After excluding claimants that we could not match to an ESA claim, we are left with a dataset of 17,782 pilot participants.

Figure 4.2 Time between qualifying ESA claim and prior IB claim

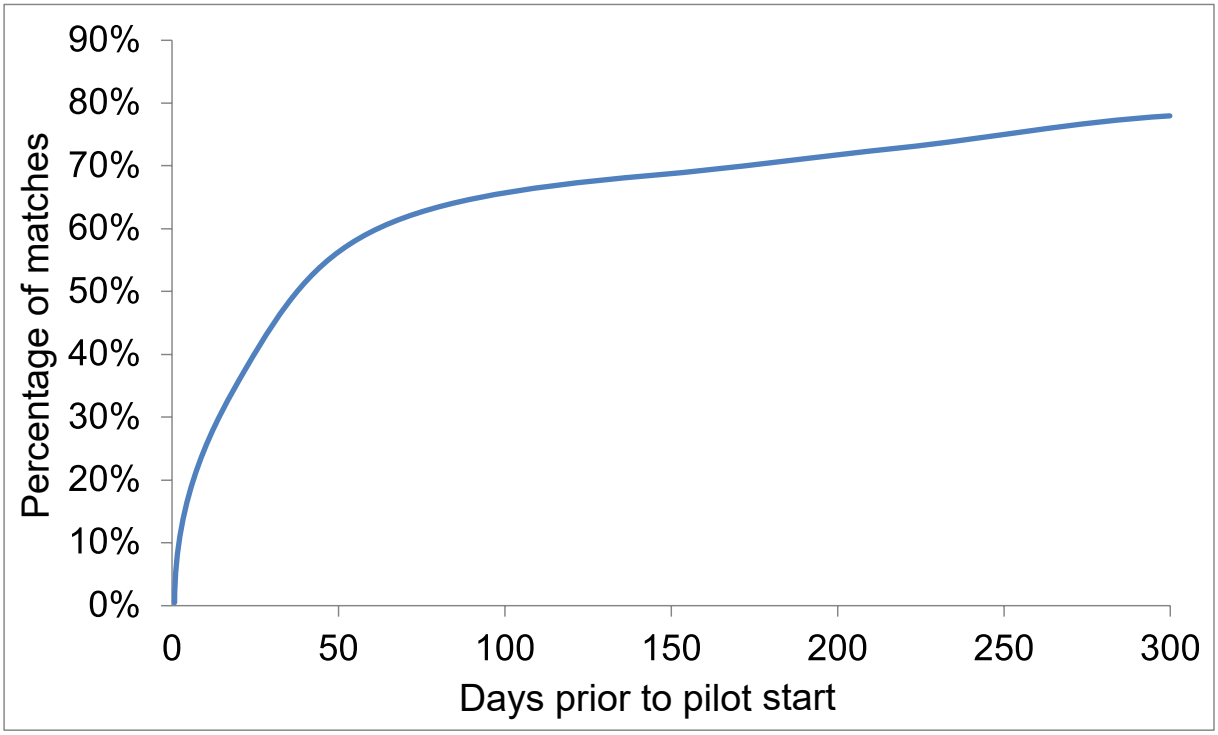


Source: DWP benefits administrative data April 2017

4.25 We note in passing that in scope of the trial were people who were previously on IB but who were part of a managed re-assessment process that led a number of prior IB claimants migrating to ESA. If, through this re-assessment exercise, an ex-IB claimant was placed in the WRAG and they had an 18 to 24 month prognosis then they too would be required to join the trial unless they satisfied one or more exemption criterion. These people can be expected to have a significantly longer history of being reliant upon DWP benefits compared with new/repeat ESA claimants and therefore it is important to distinguish between them.

4.26 We looked for IB claims prior to the ESA claims that we identify above. We found a prior IB claim for 75.8 per cent of all people with an LMS pilot marker. However, when we examine the time between the end of the IB claim and the beginning of the ESA claim we find that they are not all adjoining (see chart 4.2). Indeed, there are a cluster of IB claims that lie within around 60 days of the ESA claim and then a long and fairly linear tail of cases of longer duration. We are aware of an issue whereby ESA claims are not fully 'built' on JSAPS leading to apparent but not real gaps between JSA and ESA claims. However, we are not aware of this affecting adjoining IB and ESA claims as well. Regardless, in light of the patterns in the data it seems prudent to associate prior IB claims with the ESA claim where they lie less than 60 days apart, but not to consider these as part of the migration exercise where the duration is longer. This means that we label 61.2 per cent of cases as 'IBR' (IB reassessment) and the remainder as 'new/repeat' ESA claims.

Figure 4.3 Time between WCA and setting of the pilot marker



Source: DWP benefits administrative data April 2017 and DWP WCA data June 2017

4.27 The next step in tracing the events leading up to participation in the trials is to identify the WCA prior to the date when the LMS marker was set. To identify when this occurred we used the data that feeds into DWP's official WCA statistics. Figure 4.3 shows the time period between the date when the final WCA decision was made[8] and the date when the LMS marker was first set. We could find a prior WCA for 98.3 per cent of all people with an LMS Pilot marker. We do not know the reason for the 1.7 per cent shortfall, but as with some of our previous data merges, non-matches due to incorrect NINOs and other administrative errors will be commonplace.

4.28 Thirty-nine per cent of all the matched WCAs seem to have taken place before the ESA claim that led to pilot participation. This is not implausible because many of the IB re-assessments will take place in the IB claim prior to the ESA one. Indeed 52 per cent do (where we can identify the WCA). Five percent of the nonIBR cases have an assessment which pre-dates the ESA claim. This is more difficult to explain. At face value one might assume that the prior WCA is associated with a prior ESA claim. If this were the case, however, then the WCA would not be expected to lead to participation in the pilot. A purely speculative answer is that the mismatch could be a consequence of the WCA being 'fast-tracked' but somehow not quite aligning to the ESA claim. Regardless, the purpose of this analysis is to measure the timeliness of the WCA with respect to the LMS pilot marker, which is taken to signify the date when each person was recruited to the pilot. We can see that the time to the WCA extends over a considerable period prior to the date when the LMS

[8] There could be several decisions: that recommended by the ATOS assessor; the JCP's decision; the reconciliation between these two, and, finally; a decision following an appeal.

marker was first set. In over 43 per cent of measurable cases, the duration is five weeks or more. In principle a person should have a 'new joiner interview' shortly after the assessment and then be recruited to the pilot.

4.29 Appeals are a significant contribution to this time gap. Where the WCA decision does not appear to have been appealed, the average time between the first WCA decision date and the first setting of the LMS pilot marker is 115 days, whereas for those cases that do appeal the average time difference is 371 days. However, the long duration for the non-appealers shows that the appeals process is not the only cause of delays between the WCA and the pilot marker being set.

4.30 Given the catchment of the pilot one would imagine the outcome of the WCA is a given. When we look at the WCA decision recorded in the data, however, this is in some cases at odds with the pilot intent. Restricting ourselves to the non-exempt participants, we find that 88.1 per cent were put in the WRAG and - in 5.6 per cent of cases - the WCA record indicates that the person was either put in the Support Group or found Fit For Work. In a further 6.3 per cent of cases we either could not find a WCA record, or that record did not explicitly indicate that person to be in the WRAG (e.g. the outcome might be that the 'claim closed during assessment' or the record might be 'clerical'). However, even where the decision was, or could potentially be, to put the person in the WRAG, in 33.4 per cent of cases the prognosis was not recorded as 18 to 24 months. This percentage reduces to 22.6 per cent if we assume that unknown prognoses are all 18 to 24 months.

4.31 In short, the information we have on the WCA prior to pilot participation cannot be used to confirm that the right people have been included in the pilot and at the right time. Not only is there a large amount of variation between the date of the WCA and the date of recruitment (as inferred from the LMS pilot marker), the outcome of the WCA is also often missing or at odds with the nominal qualifying criterion of the pilot.

4.32 Another important question on the subject of the recruitment process is whether everybody who should have been recruited to the pilot actually was recruited to the pilot (be they exempt or otherwise, or in the treatment or control group). We examined every WCA record that indicated a person living in one of the pilot areas had been put into the WRAG and had a prognosis of 18 to 24 months. We then restricted this dataset to those people who had attended a 'new joiner interview' or an 'ESA WP referral' interview during the pilot recruitment period (November 2013 to September 2014). There was little correlation between the timing of the WCA decision and the date of the meeting, so we imposed an artificial limit of three months separation between the two events to give us a degree of confidence that the meeting was associated with the WCA.

4.33 Restricting ourselves to this rather limited number of WCAs, just 53.5 per cent had a pilot marker set. This was true regardless of whether or not we included people who appealed their WCA decision. Further, we could only account for 27.5 per cent of all markers that were set.

4.34 We defined our target WCAs quite tightly. It is possible that meetings attended outside of the pilot period led to pilot participation. Likewise, we have already seen that the conditionality group and the prognosis recorded in the WCA data differs markedly to the group and prognosis implied by the pilot marker data. Therefore, given that pilot markers are not wholly consistent with the WCA data, we should not be too surprised if the WCA data is not wholly consistent with the pilot marker data. These inconsistencies lead us to the conclusion that a check of the implementation of the pilot is not feasible, unless one implicitly trusts the data available to this study, in which case one could only conclude that there were major deficiencies in the way that the pilot was implemented.

4.35 The inconsistencies above may in part be due to this analysis being based upon the wrong WCA. That is, the WCA we are associating with pilot participation is not the one that led to the pilot participation. This means that for whatever reason we have been unable to identify the correct WCA in many cases. Regardless, given the lack of a definitive source of information on WCA outcomes and their prognosis periods, we leave this discussion to consider the extent to which each person meaningfully participated in the pilot.

5 Pilot Participation

5.1 We now compare the marker data described in the previous section with corroborative sources that describe some of the activities that were undertaken following recruitment to the pilot.

5.2 Following a WCA, claimants in the WRAG are required to attend a 'New Joiner Interview' (NJI) whereupon those people in scope of the pilot should be allocated to either the 'treatment' group or the 'control' group. An inspection of the timing of NJIs with respect to the first setting of the LMS pilot marker does not suggest that LMS markers were systematically set after the NJI. Therefore, when examining participation in the pilots, we restrict ourselves to the two-year period starting from the date when the LMS marker was first set.

5.3 Immediately evident from the administrative data on attended meetings is the fact that some people have no evidence of ever attending a work related meeting. There are over 400 'types' of meetings recorded on the LMS, over 80 of which can be associated with (exempt and non-exempt) pilot participants. The purpose of many meetings can be ambiguous (e.g. 'ad hoc interviews') but we have taken a liberal interpretation of the meetings' purpose and retain in our analysis any that appear to have a work focus. In doing so, we have retained 99 per cent of all the meetings that were attended in the two-year pilot period (and held during the ESA claim associated with the pilot).

5.4 Tables 5.1a to 5.1c summarise, by pilot, group and exemption status, the proportion of people with an LMS marker that have any evidence of having attended a single work-focused meeting. Also shown are the average number of meetings attended during the two-year period and the cumulative duration of those meetings. When measuring the cumulative duration of the meetings we take on trust that the nominal duration of the meeting was the actual duration. We also assume that 'new joiner interviews' are 40 minutes, as per the guidance. Other meetings whose duration is not known are assumed to be of the same duration of the average of the known durations (excluding NJIs given that the content of these is quite specific).

5.5 Various things are evident from the tables. Firstly, not all of the non-exempt cases appear to have attended a meeting. For the HCP and WP models this calls into question whether and how they were referred to the provider. For the JCP model it could be said that the non-exempt non-attendees have not participated in the pilot at all.

Table 5.1a JCP meeting attendance for the HCP model

	Not Exempt		Exempt	
	Treatment	**Control**	**Treatment**	**Control**
Attended	95%	97%	65%	63%
Average Number Attended	3.1	5.3	3.3	3.2
Average Cumulative Duration (minutes)	83	139	87	83

Source: Labour Market System: July 2017

Table 5.1b JCP meeting attendance for the WP model

	Not Exempt		Exempt	
	Treatment	**Control**	**Treatment**	**Control**
Attended	96%	96%	87%	92%
Average Number Attended	2.1	5.0	3.8	3.9
Average Cumulative Duration (minutes)	70	141	106	109

Source: Labour Market System: July 2017

Table 5.1c JCP meeting attendance for the JCP model

	Not Exempt		Exempt	
	Treatment	**Control**	**Treatment**	**Control**
Attended	97%	96%	82%	81%
Average Number Attended	13.2	5.2	4.1	4.1
Average Cumulative Duration (minutes)	451	151	120	119

Source: Labour Market System: July 2017

5.6 A lot of the exempt groups have not attended any meetings. We do not know the reasons for mixed attendance of meetings. Perhaps some exemption reasons rule out meetings at all, whereas some just rule out the more intensive requirements of the pilot model, hence why there is some evidence of ongoing work-focused interviews (i.e. on average the exempt have attended between 3 and 4 meetings). Regardless, exempt cases are not of major concern if we are to trust the LMS pilot marker.

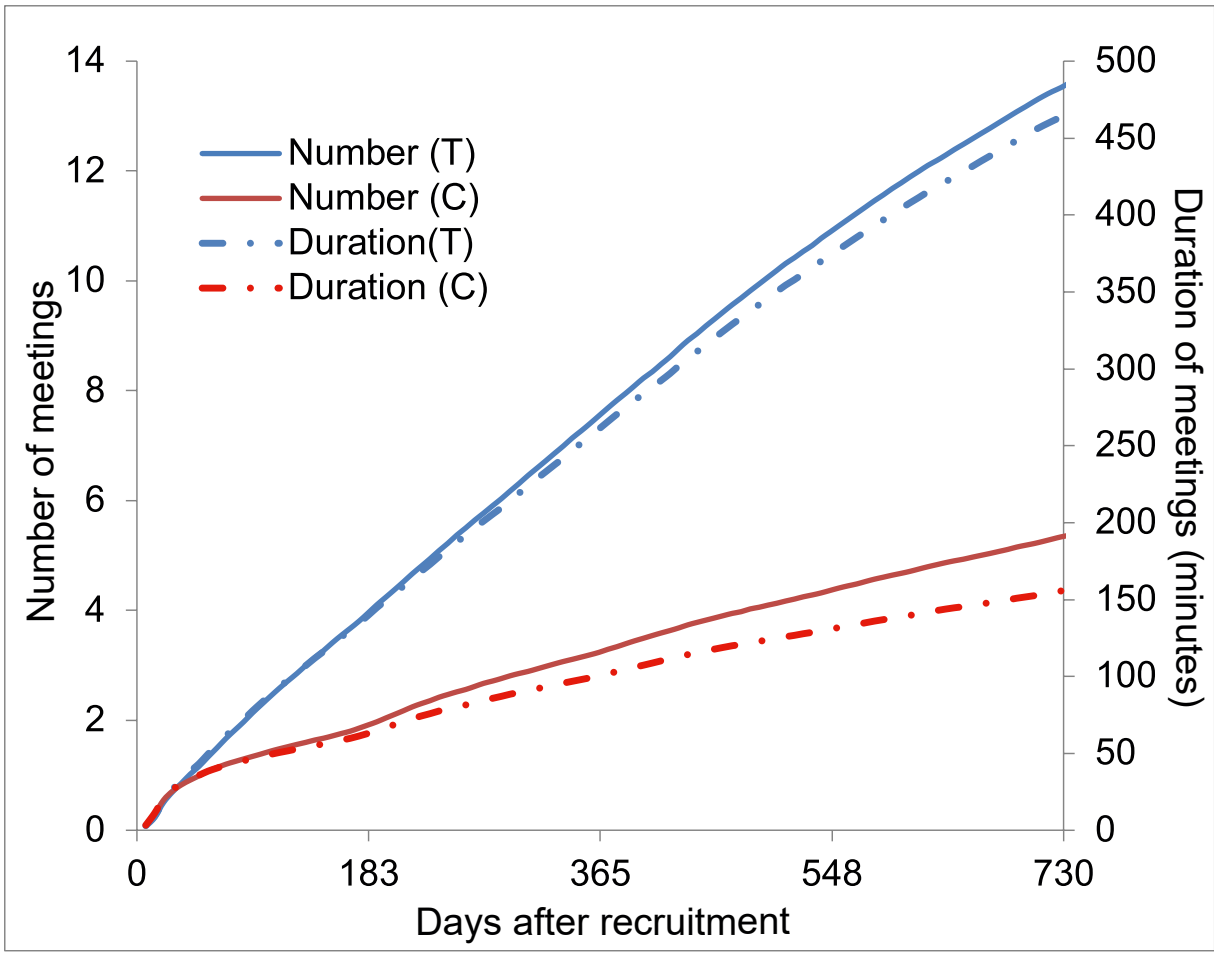
5.7 The non-exempt cases broadly follow our expectations in that for the HCP and WP models the treatment group attend fewer meetings than the controls (because the former have been handed over to a provider), though why the treatment group have attended, on average, more than one work-focussed

meeting is not so clear. For the JCP model the converse is true whereby the treatment group have attended substantially more meetings than the control, as expected. Also, the control groups are supported fairly consistently from model-to-model with the average number of meetings attended being approximately five.

5.8 We did check the extent to which people who appealed their WCA Decision were more or less likely to attend a meeting than those who did not appeal. We note that the data source we have used only tells us about completed appeals and not ongoing appeals, but that caveat aside, what we find is that those who appeal are if anything more likely to attend a work-focused meeting. It may be that the attendance was motivated by a desire to express a view on the WCA outcome and explore options. Regardless, what we can conclude is that appeals do not appear to be the reason why some people do not attend any work-focused meetings.

5.9 Figure 5.1 illustrates the cumulative number of meetings and minutes spent in meetings for the JCP model. The chart shows the cumulative number and cumulative duration of work-focused meetings attended by the non-exempt treatment and control groups.

Figure 5.1 Cumulative Number and Duration of Work Focussed Meetings Attended by non-exempt treatment (T) and control (C) Participants in the JCP Model



Source: DWP benefits administrative data April 2017 and DWP WCA data June 2017

5.10 It does appear likely that the amount of Work Coach support delivered to the treatment group is roughly half that allowed under the pilot model. The chart does not describe the totality of support. Where a person is referred to training or a partner organisation, that activity will not be captured by this data. Further, we are not assured about the accuracy of the meetings data, but - given that the control group does seem to receive the nominal amount of support (two years of 88 minutes per year, typically meeting once every six months) - this lends credibility to the data for the treatment group.

5.11 We do not show similar charts for the HCP and WP models because multiple JCP Work Coach meetings are not an integral part of those models, and tables 5.1a to 5.1c indicate consistent treatment across all three (non-exempt) control groups. We do note that 61 per cent of the HCP non-exempt treatment group have attended more than one meeting, whereas the equivalent figure for the WP model is 50 per cent which suggests that JCP Work Coaches may have been 'supplementing' the HCP support to a greater degree than they did for the WP support.

5.12 The discussion so far has established engagement in the pilot process amongst the JCP non-exempt treatment group. We now go on to consider the HCP and WP non-exempt treatment group. We have various sources of information that indicate whether a person was considered for referral to one of the pilot providers, whether they were actually referred and whether they actively took part in that provision, namely:

- LMS Pilot Opportunities
- LMS WP Opportunities
- WP Evaluation data
- PRaP Payments.

5.13 The LMS Pilot Opportunities are not of great significance. The LMS pilot Opportunities are supposed to be set for all non-exempt treatment group participants, but not doing so has no material effect upon participation. Indeed, in most cases the JCP opportunity does not indicate that the support has started, but that may be because the Opportunity is simply being used as a marker and little more.

5.14 The WP Opportunities signify an intention to refer to a WP provider and as such they provide a more concrete indication of how an individual was dealt with. The WP Evaluation data tracks what happened subsequently. That is, the date a person started on the WP and any outcomes subsequently achieved. Finally, the PRaP data records actual payments to the provider which is perhaps the most definitive confirmation of participation. PRaP covers both the WP and HCP model, whereas the LMS Opportunities do not capture the HCP referrals (though the HCP pilot Opportunity does in principle signify a referral). LMS Opportunities and PRaP are linked and PRaP feeds into the WP evaluation database, so these three sources should be very consistent.

5.15 Table 5.2 summarises the extent to which there is evidence that the people in the non-exempt treatment group in the HCP model actually participated in the contract provision. Only 69.7 per cent of non-exempt treatment cases had a PRaP payment made to Ingeus (the HCP provider) on their behalf. A further 3.5 per cent had a record on PRaP that suggested the referral process

had begun, but the participant had not formally started the Programme, nor had a payment been made. Of the 30.3 per cent of non-starters, 3.9 per cent had been referred to the WP in the past or were on the WP at the time of recruitment to the pilot, which may account for the absence of a start on the HCP Programme. However, in 22.5 per cent of cases we could find no reason for the absence of a start on the Programme.

Table 5.2 Evidence of provision attendance in the HCP model

	Non-Exempt		Exempt	
	Treatment	Control	Treatment	Control
Started HCP provision	69.7%	5.2%	2.6%	1.2%
Previous WP referral	2.0%	4.4%	20.2%	24.1%
Referred but not started	3.5%	1.1%	0.8%	0.3%
On WP after pilot start	0.3%	1.2%	1.0%	1.2%
No evidence	24.5%	88.0%	75.4%	73.2%

Source: Labour Market System: July 2017 and DWP Provider Referral and Payment System: August 2017.

5.16 A small number of exempt and/or control participants actually had evidence of a PRaP payment being made (to Ingeus) on their behalf. Also, a number had been or were still on the WP at the time of recruitment onto the pilot which partly explains their exemption (see the earlier section on an analysis of exemptions).

5.17 Turning to the WP model (table 5.3) we find a similar situation whereby 22.7 per cent of the nonexempt treatment cases do not appear to have been referred to a WP provider. In 2 per cent of cases this will be because they had already been referred to the WP. In a further 2 per cent of cases we have evidence of a referral after the pilot claim.

5.18 There is also some evidence of control cases and exempt cases (within both the treatment and control groups) starting a WP spell despite the status of the pilot marker; 3.7 per cent, 3.0 per cent and 4.0 per cent respectively. A significant number of exempt cases had started a WP spell prior to pilot recruitment but this is expected given that this group are out of scope of the pilot.

Table 5.3 Evidence of provision attendance in the WP model

	Non-Exempt		Exempt	
	Treatment	Control	Treatment	Control
Started WP provision	75.8%	3.7%	3.0%	4.0%
Previous WP referral	1.8%	2.8%	19.7%	15.0%
Referred but not started	1.5%	1.4%	2.6%	1.3%
Referred after pilot claim	1.9%	3.5%	3.3%	2.7%
No evidence	19.6%	88.7%	72.1%	76.5%

Source: Labour Market System July 2017 and DWP Provider Referral and Payment System: August 2017.

5.19 Table 5.4 below is based solely on the PRaP data and therefore does not include any reference to LMS Opportunities that indicate WP participations. Neither, of course, does the table refer to any participants of the JCP model. This table provides a summary of the payments made to the providers for the HCP and WP models.

Table 5.4 Payments to WP and HCP providers

	HCP	WP
Participants	**2654**	**1922**
Pilot Provider Payments	1849	1455
	69.7%	***75.7%***
Ingeus UK Ltd	1849	0
Avanta Enterprise Ltd	0	1080
G4S Regional Management	0	180
Newcastle College Group	0	195

Source: DWP Provider Referral and Payment System: August 2017

5.20 Finally, table 5.5 summarises evidence of WP participation in the JCP model. Consistent with the previous discussion, a sizeable number of exempt cases were or had been on the WP in the past. A non-negligible number (almost 10 per cent) of non-exempt cases also had some present or past attachment to the WP. For 83 per cent of the treatment cases and 79 per cent of the controls their WP spell was complete by the time they were recruited onto the pilot. This is in line with the pilot design, but it does raise the issue of whether these claimants behaved differently during the pilot period because of that prior participation in the WP. For the remaining 17 per cent and 21 per cent respectively, who were on the WP at the time of recruitment onto the pilot, technically these were out of scope of the pilot and should not have appeared in the pilot marker data (or at any rate be marked as exempt).

Table 5.5 Evidence of work programme provision in the JCP model

	Non-Exempt		Exempt	
	Treatment	Control	Treatment	Control
Started WP provision	2.1%	2.9%	7.5%	8.0%
Previous WP referral	9.1%	9.8%	30.2%	26.9%
Referred to WP after pilot claim	1.3%	0.8%	1.1%	0.6%

Source: Labour Market System July 2017

5.21 We conclude this discussion by noting that one per cent of the control group were referred to the WP after recruitment to the pilot whereas guidance was that all control clients should receive standard JCP Work Coach support (of 88 minutes per year) in order to ensure a consistent baseline across all three pilot models.

5.22 There are various reasons why a person may be notionally recruited onto the pilot but not be referred to a provider. For example, they may never have attended a meeting, they may have left benefits shortly after being recruited or they may appeal their WCA decision. Table 5.6 summarises the extent to which these possibilities may have occurred, allowing three months for the ESA claim

to end, though we acknowledge that this period of time is arbitrary. The JCP model is not included in the table because external (to DWP) provision is not an intrinsic component of that model.

5.23 Table 5.6 demonstrates that people who do not have a meeting record are far less likely to be referred to a provider. However, appealers are marginally more likely to be referred to a provider. The slightly higher rate may be because the more intensive regime created a greater incentive to appeal the WCA decision, but this is pure speculation and the effect is not a dramatic one, if indeed there is an effect. People who close their ESA claim shortly after recruitment to the pilot are significantly less likely to be referred to a provider, but this by no means explains the majority of the shortfall in referrals.

5.24 We complete this section with a summary of the 'Opportunities' that non-exempt pilot participants have been referred to in the JCP model. There are many such Opportunities, so, for the purpose of monitoring the pilot activities, we have grouped them in categories. These categories have been constructed and agreed with managers of the JCP model and therefore reflect Operational views of the purpose of those referrals.

5.25 That said, the majority of attended Opportunities are concerned with skills, health checks and skills provision (see Table 5.7 for figures), so we have grouped all other categories into a single 'Other' type.

Table 5.6 Link between provision and other factors

	HCP		WP	
Factor	Proportion of total	Proportion referred	Proportion of Total	Proportion referred
Attended Meetings	95.3%	74.9%	95.8%	79.4%
No Meetings	4.7%	37.9%	4.2%	58.0%
Appealed WCA decision	11.7%	77.2%	12.2%	80.3%
No appeal of WCA decision	88.3%	72.7%	87.8%	78.3%
Claim closed after 3 months	2.3%	40.0%	1.8%	62.9%
Claim open after 3 months	97.7%	74.0%	98.2%	78.8%

Source: Labour Market System: April 2017 and DWP WCA data June 2017

5.26 The 'Other' category consists of: local initiatives; specialist support for, for example, substance abusers, and; various other forms of provision. There is no correct level of use of Opportunities. Rather, it is whatever is appropriate to the claimants' circumstances and the resources available to the Work Coach. Accordingly, we simply present the data in table 5.7 'as is' and note that there is a clear contrast between the number of Opportunities attended by the treatment group and the control which is entirely consistent with the increased focus on the former.

5.27 There have been some referrals to Opportunities over and above those presented in table 5.7. However, these do not yet appear to have been attended and therefore they are of less interest.

Table 5.7 Attendance of "Opportunities" within the JCP Model

	Treatment	Control
Work- Related	9.9%	3.7%
Job- Related	1.4%	0.6%
Skills- Related	24.9%	12.9%
Other	11.5%	11.8%

Source: Labour Market System July 2017

5.28 This section has discussed at length the relationship between the LMS Pilot markers; the qualifying criterion for the pilot, and the events that we would expect to follow recruitment (or not to follow - in the case of control and exempt participants), and; the role that WCA appeals may have played in anomalous cases. The intention behind this analysis was to quantify the extent to which the pilot design had been adhered to and therefore provide an informative context for the impact analyses which will follow shortly. What we have demonstrated is that there are some minor and some not-so-minor inconsistencies between different data sources, and these prevent us from confirming that the right people have been recruited to the right pilot group at the right time. The analysis also suggests that not everybody who was recruited to the pilot actually underwent the pilot intervention.

5.29 As well as checking the integrity of trial implementation, the analysis presented in this section was intended to establish whether there would be merit in dropping certain observations to produce a 'cleaner' sample of trial participants. The data does not allow us to make a clear judgement over what is a 'good' and 'bad' observation, so we make no decision at this stage. However, we return to this issue later on in this report when deciding which observations to include in our impact assessment.

5.30 Before moving on to the analysis of the pilots' impact, we briefly overview 'Decision Maker Activity' (DMA) during the pilot. DMA refers to the process by which doubts about whether or not a person has complied with their benefit conditionality are raised with a 'Decision Maker'. The Decision Maker then decides whether or not a sanction should be applied to that person's benefit. Conditionality for ESA claimants is less demanding than for JSA claimants, and they are only sanctioned for failing to attend a mandatory meeting, or failing to participate in mandatory work-related activity.

5.31 Table 5.8 shows, for each pilot, the proportion of participants who, to date, have had at least one doubt raised about them. Also shown is the average number of doubts per person (excluding those with no doubts).

Table 5.8 ESA Doubts Raised During the Pilots

	Non-Exempt		Exempt	
	Treatment	Control	Treatment	Control
HCP Model				
Per cent with doubts	17.3	4.2%	8.1%	8.5%
Avge no. of doubts	2.1	1.6	1.6	1.9
WP Model				
Per cent with doubts	25.8%	3.6%	3.7%	2.7%
Avge no. of doubts	2.2	2.1	1.9	3.2
JCP Model				
Per cent with doubts	8.7%	3.0%	4.2%	8.3%
Avge no. of doubts	2.0	1.3	1.6	2.2

Source: The Appeals Service Decision Maker and Appeals System, May 2017

5.32 Table 5.9 shows the proportion of participants who actually incurred a sanction over the same time period, and the average number of sanctions incurred. Table 4.10 shows the reason for those sanctions, i.e. the proportion of all sanctions that were for failing to attend a mandatory meeting, and the proportion that were for failing to participate in mandatory work-related activity.

Table 5.9 ESA Sanctions Incurred During the Pilots

	Non-Exempt		Exempt	
	Treatment	Control	Treatment	Control
HCP Model				
Per cent with sanctions	8.1%	2.2%	3.2%	2.6%
Avge no. of sanctions	1.8	1.3	1.4	1.9
WP Model				
Per cent with sanctions	8.0%	1.7%	1.9%	2.7%
Avge no. of sanctions	1.7	1.7	1.0	1.7
JCP Model				
Per cent with sanctions	3.2%	1.7%	2.2%	2.9%
Avge no. of sanctions	1.8	1.1	1.3	1.5

Source: The Appeals Service Decision Maker and Appeals System, May 2017

5.33 Tables 5.8 and 5.9 show several things:

- amongst the non-exempt, treatment claimants are more likely to have a doubt raised against them and to incur a sanction
- however, the difference is less marked for the JCP model when compared with the HCP and WP models
- exempt claimants also appear to be subject to doubts and sanctions but at generally lower levels, the exception being with JCP where the exempt are subject to more doubts and sanctions

- despite having the same conditionality, the controls exhibit some variation in the levels of doubts and sanctions. This may reflect differences in local practices that may also be contributing to the differences between the treatment groups
- for the non-exempt treatment and control groups in the HCP and JCP models, most sanctions are for failing to attending a mandatory meeting. For every other group and/or model most sanctions are for not participating in work related activity.

5.34 To summarise then, the approach to the enforcement of ESA conditionality for the treatment group appears to have been markedly different in the JCP model compared with the other two models. This is perhaps due to contractors following sanctions guidance more rigidly, perhaps due to a perceived lack of autonomy and discretion. Whereas JCP Work Coaches may have enforced conditionality more leniently, with a view to establishing a better long-term relationship with their clients. Alternatively, it may simply be that the JCP treatment group were more compliant with the regime. Regardless, we speculate no further but mention that qualitative interviews with both pilot staff and pilot participants - described in the research report accompanying this report - may provide some insight into this difference in the levels of doubts and sanctions.

Table 5.10 Reasons for the ESA Sanctions

	Non-Exempt		**Exempt**	
	Treatment	Control	Treatment	Control
HCP Model				
Failed to attend	94.4%	62.1%	26.1%	23.5%
Failed to participate	5.6%	37.9%	73.9%	76.5%
WP Model				
Failed to attend	7.6%	35.1%	20.0%	20.0%
Failed to participate	92.4%	64.9%	80.0%	80.0%
JCP Model				
Failed to attend	77.4%	83.3%	8.7%	15.2%
Failed to participate	22.6%	16.7%	91.3%	84.8%

Source: The Appeals Service Decision Maker and Appeals System, May 2017

6 Characteristics of Pilot Participants

6.1 Before considering outcomes and the impact that the pilot intervention had upon those outcomes it is important to be assured that the treatment group and the control group were similar in their personal and labour market characteristics. This section of the report explores that subject.

6.2 We collated a range of data from the various DWP administrative systems. Annex A provides a set of tables that show breakdowns such as: gender; age; ethnicity; partner and parental responsibilities; International Classification of Disease codes, and; geographical distribution of the treatment and control participants for each of the three pilot models. We also checked a range of other indicators, such as: 'sought occupations'; refugee status, and; so called 'harder to help' indicators such as substance abusers, low qualifications and ex-offender status. The quality of this wider data is known to be questionable and in some cases is poorly-populated. For this reason, we do not include that data in this report but we do note that what checks were done did not suggest a sizeable selection bias on the basis of those characteristics.

6.3 Regarding the data that we do include in the Annex, it is important to consider how we define our treatment and control groups. The previous section highlighted various imperfections in the pilot implementation: random allocation errors, inconsistencies in the application of the exemption criterion, and people who do not appear to have undergone the pilot intervention.

6.4 In the annex we restrict ourselves to the non-exempt claimants but we include all people with a pilot marker regardless of their subsequent work-related activities. We also allocate people to the (treatment or control) group they should be in rather than the group they were in. We do this to hypothetically reduce the amount of 'selection bias' in the data. In other words, we use an 'Intention to Treat' approach. This simply means that we analyse people on the basis of what should have happened to them, and not what did happen to them. This is a common approach and is employed because to do otherwise would potentially introduce differences between the treatment and control group that are not due to the pilot intervention.

6.5 We note though, that there are other ways in which selection bias may enter our pilot sample. For example, the difference in the sizes of the treatment and control groups is too large to be due to chance alone. This suggests that some people in scope of the pilot may not have been allocated to either of the treatment or control groups or the exemption criteria may have been applied differently for one group compared with the other. Owing to the quality of the WCA data we cannot pinpoint who should have been included in the trial and given the incomplete data on the exemption criterion, neither can we independently identify who should and should not have been included in the trial on that basis. Indeed, we have already established that some active participants do seem to fulfil one or more of the exemption criteria.

6.6 Regardless, the data presented in Annex A is too extensive to present or discuss in detail here. For now, we simply note that the differences between the treatment and control groups are usually less than 1ppt, and statistical tests

suggest that the (nominal) random allocation process has produced groups that are reasonably well balanced. There are some relatively large differences. The JCP model, for example, has significant differences in the proportion of people who prefer not to record their ethnicity, as does some of the data not shown in Annex A such as some of the disabilities recorded on the LMS or the prevalence of substance abusers. As this data is offered voluntarily by the claimant these differences may simply reflect the closer working relationship that Work Coaches have with the treatment group. That may well be true about the differences in disabilities and 'harder to help' indicators, however, it doesn't explain the lower level of recorded ethnicities amongst the treatment group. Regardless, given the number of characteristics that we have measured, we would expect some seemingly large differences to occur through chance alone. Aside from a small number of discrepancies that may be due to recording differences and not true differences, no single measure gives us particular cause for concern.

6.7 We comment generally that treatment (and control) participants are moderately more likely to be male, mostly white and over 40. Roughly half suffer mental health difficulties and a not insignificant number (six per cent to nine per cent) have been screened as having a 'basic skills' need. The majority (~80 per cent) do not appear to have a dependent child but, given that a similar number do not claim the adult dependent allowance, this may reflect under-reporting. That is, given the age of many participants, we might expect more of them to be partnered and have children although it is plausible that the children of those that are parents are no longer dependent upon them.

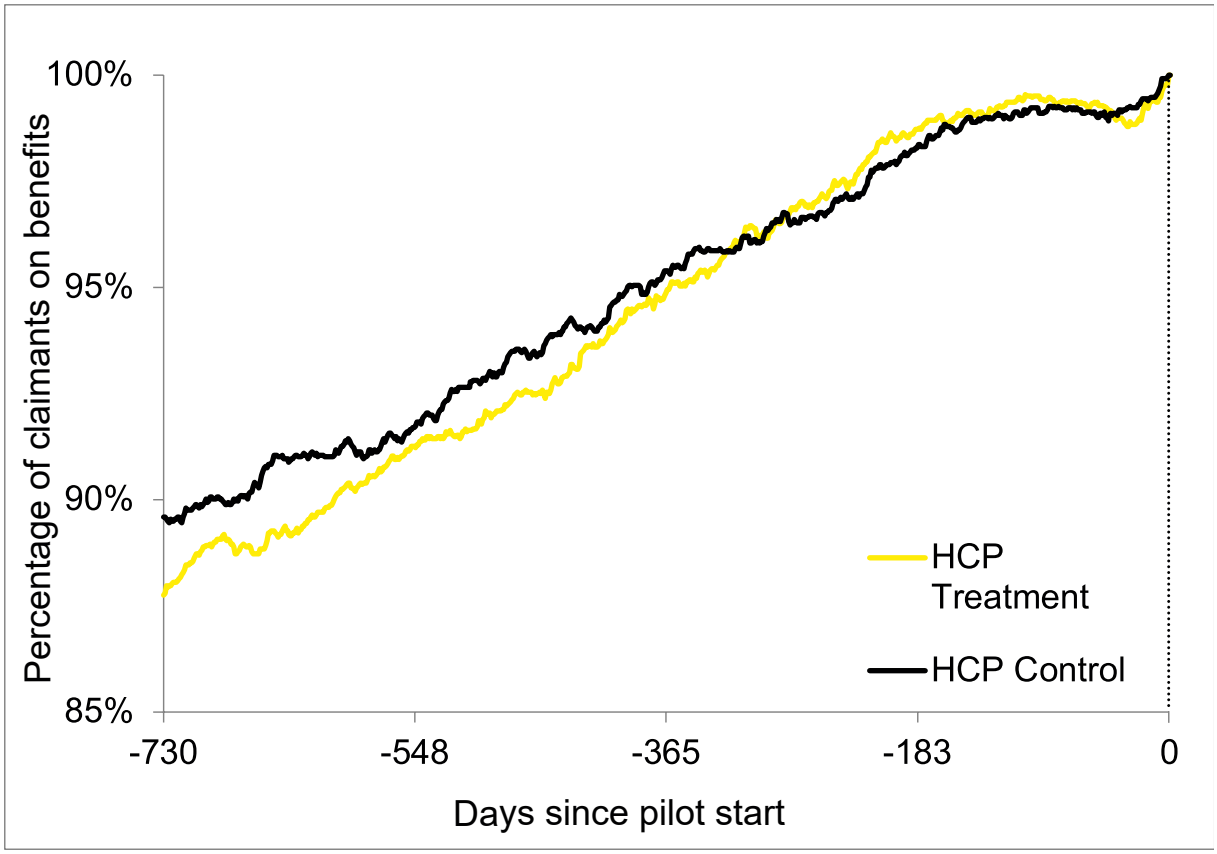
6.8 Whilst we do not show this data, we note that there is very little evidence of significant numbers of alcohol and substance abusers amongst the pilot participants. Similarly, it appears there are very few people in the other 'harder to help' categories such as ex-offenders or refugees.

6.9 In conclusion, the data so far paints a picture of an older, lower-paid, and perhaps less well-educated population (based on sought occupations, skills and qualifications) though this data is highly incomplete and non-validated. To explore labour market status further, we show in figures 6.1a to 6.1c the benefit status of the treatment (and control) participants in the period of time prior to recruitment to the pilots. We include this data because a person's reliance upon DWP benefits in the past is a very good (but not perfect) indicator of people's reliance upon benefit in the future. Differences in prior benefit status can therefore be expected to result in differences in benefit status after the pilot irrespective of any pilot impact.

6.10 For the HCP model (figure 6.1a), the control group were marginally more dependent upon benefits one to two years prior to the pilot but more recently there is closer agreement, with the control group now being marginally less reliant upon benefits. This change could be due to random variation, but it is nonetheless a real difference that could play a minor role in future differences in labour market outcomes.

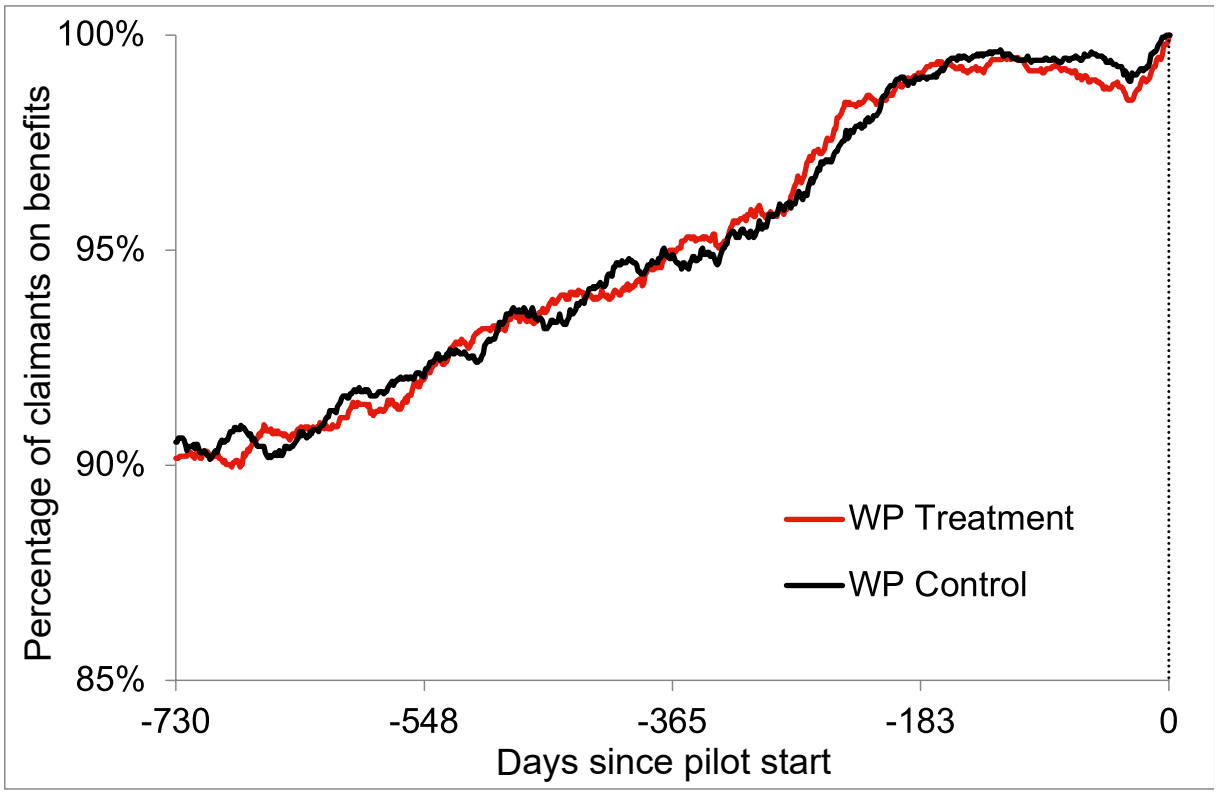
6.11 The WP model (figure 6.1b) shows a stronger agreement in past reliance upon DWP benefits. The JCP model (figure 6.1c) has the greatest agreement on historical benefit dependency.

Figure 6.1a Benefit caseload for HCP model prior to pilot participation



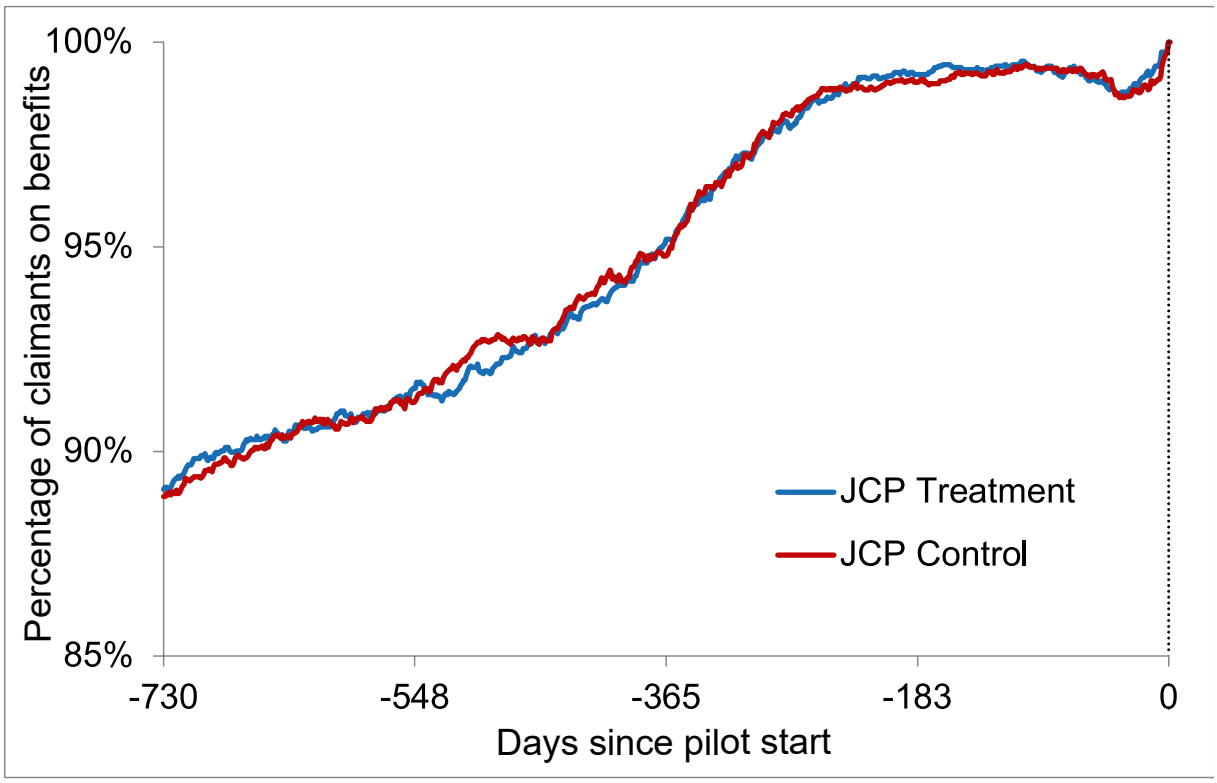
Source: DWP benefits administrative data: April 2017

Figure 6.1b Benefit caseload for WP model prior to pilot participation



Source: DWP benefits administrative data: April 2017

Figure 6.1c Benefit caseload for JCP model prior to pilot participation

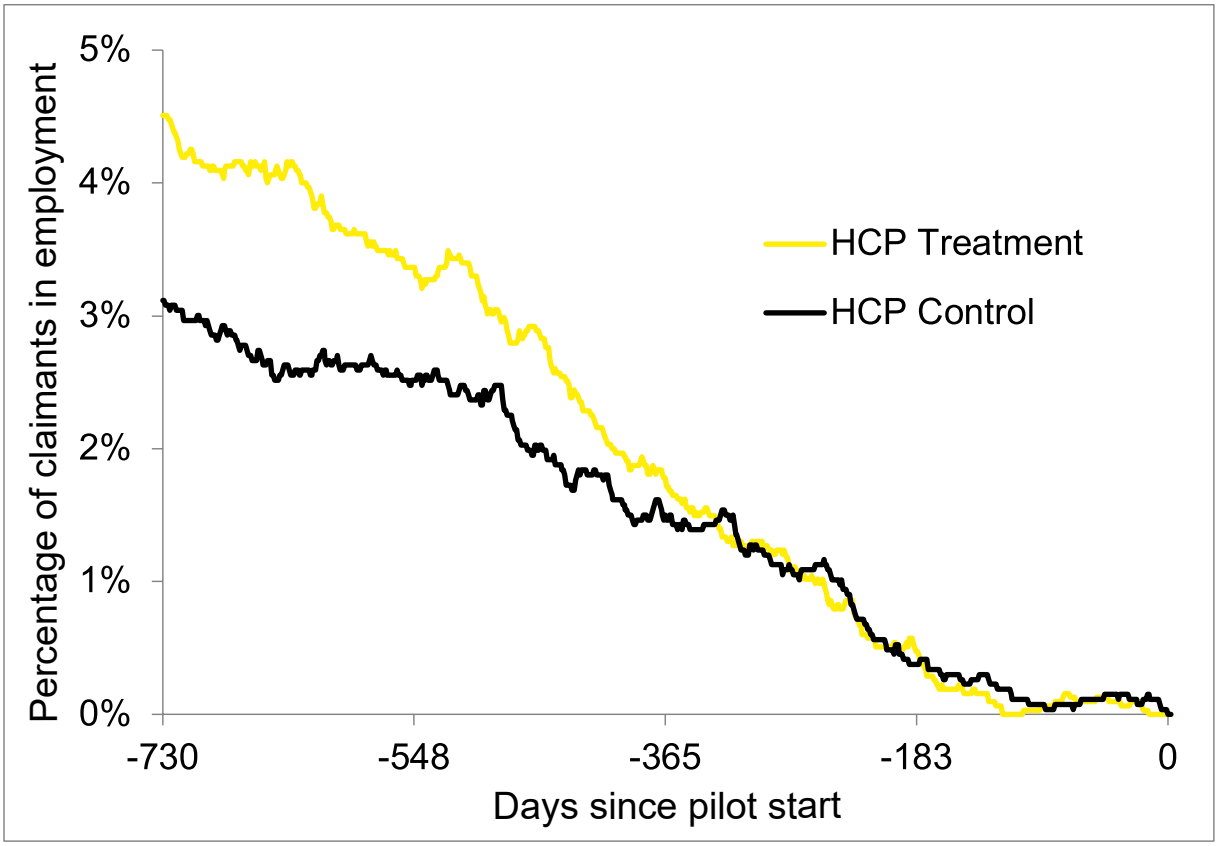


Source: DWP benefits administrative data: April 2017

6.12 Figures 6.2a to 6.2c show pre-pilot employment rates, as inferred from HMRC's P45 data, which we know is deficient in multiple respects. The charts corroborate the employment data in that HCP participants were more likely to be in employment (i.e. less likely to be on benefits) one to two years prior to pilot participation. We see a generally better agreement for the WP model here yet we also see that the control group were less likely to be in employment one to two years prior to the pilot. Chart 6.2b does not contradict chart 5.1b because not being on DWP benefits does not mean that a person is necessarily in employment. There are many other statuses that a person could find themselves in or it may simply be their jobs were less likely to be recorded in HMRC's P45 data.

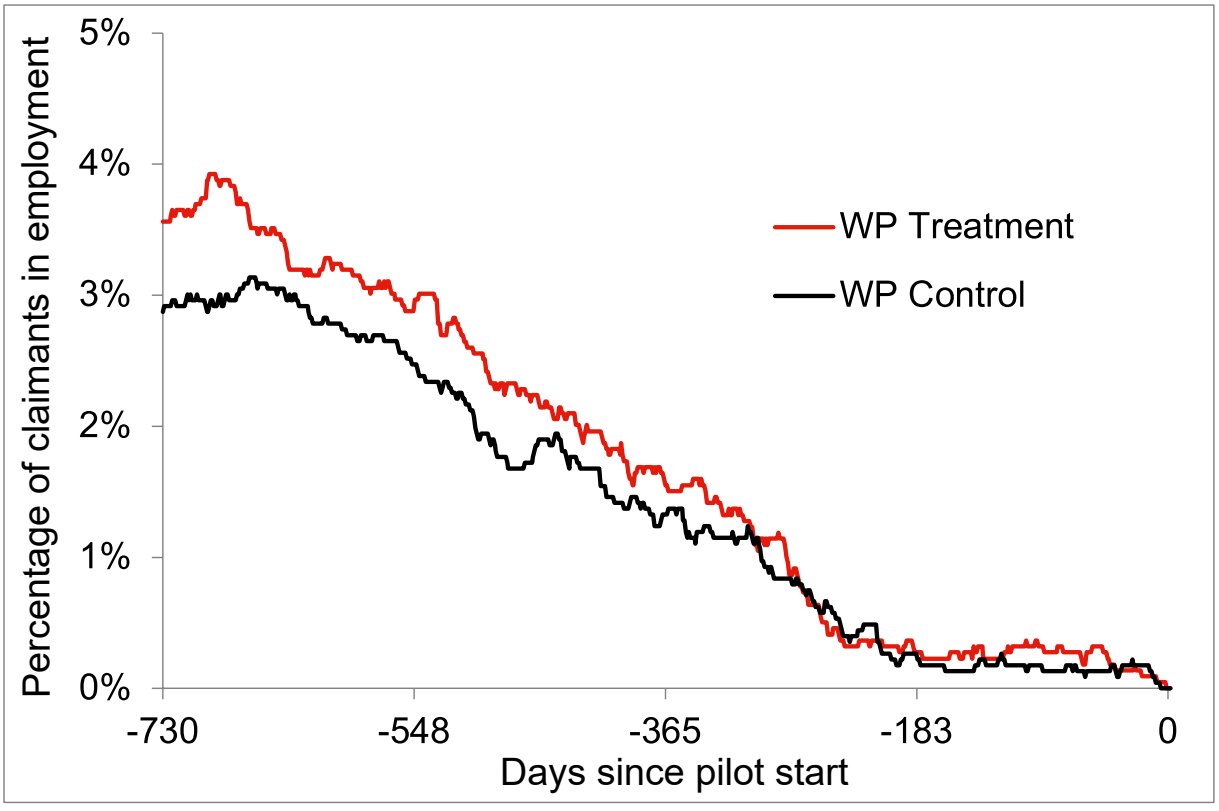
6.13 On the JCP model (figure 6.2c) the treatment group are certainly, on average, less likely to be in employment prior to pilot participation. However, the difference is less marked than the difference in benefits (chart 5.1c). As with the prior charts though, this may simply be due to shortfalls in the HMRC data rather than real differences in labour market attachment.

Figure 6.2a Employment for the HCP model



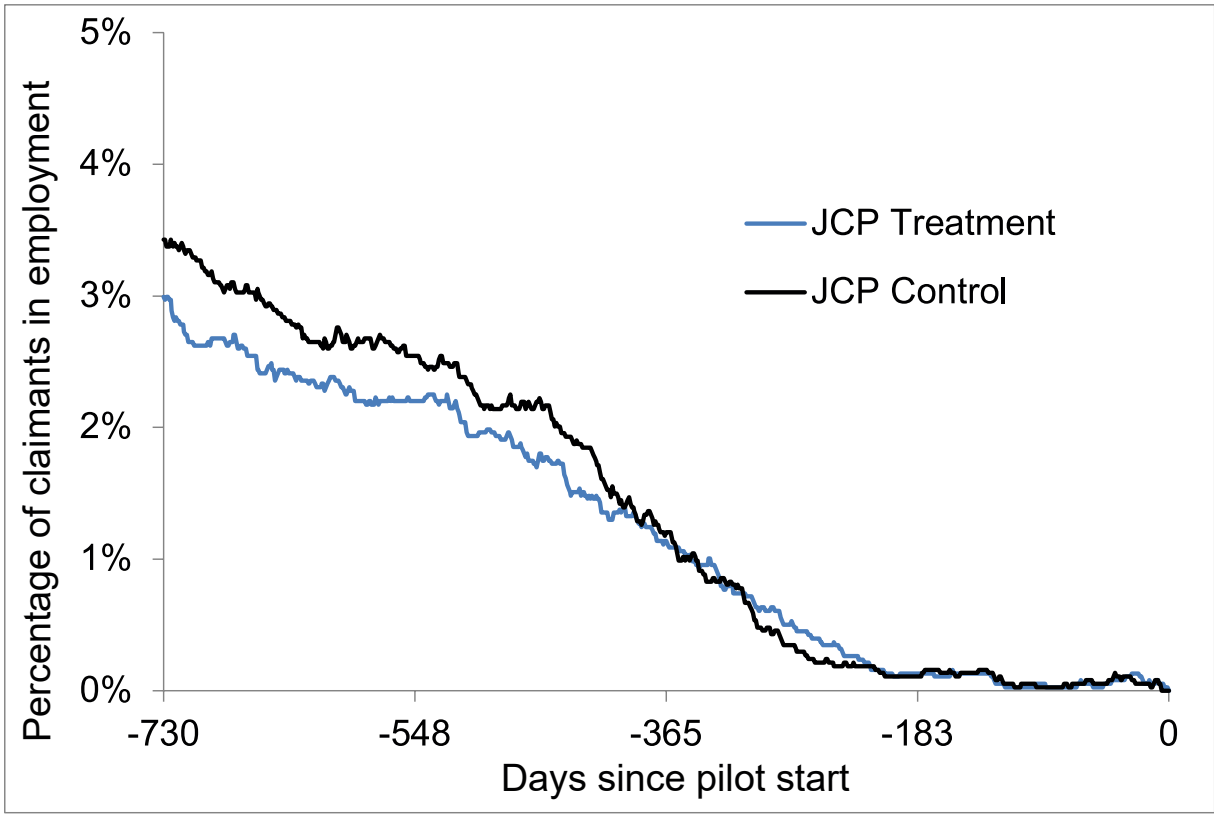
Source: HMRC P45 data: August 2017

Figure 6.2b Employment for the WP model



Source: HMRC P45 data: August 2017

Figure 6.2c Employment for the JCP model



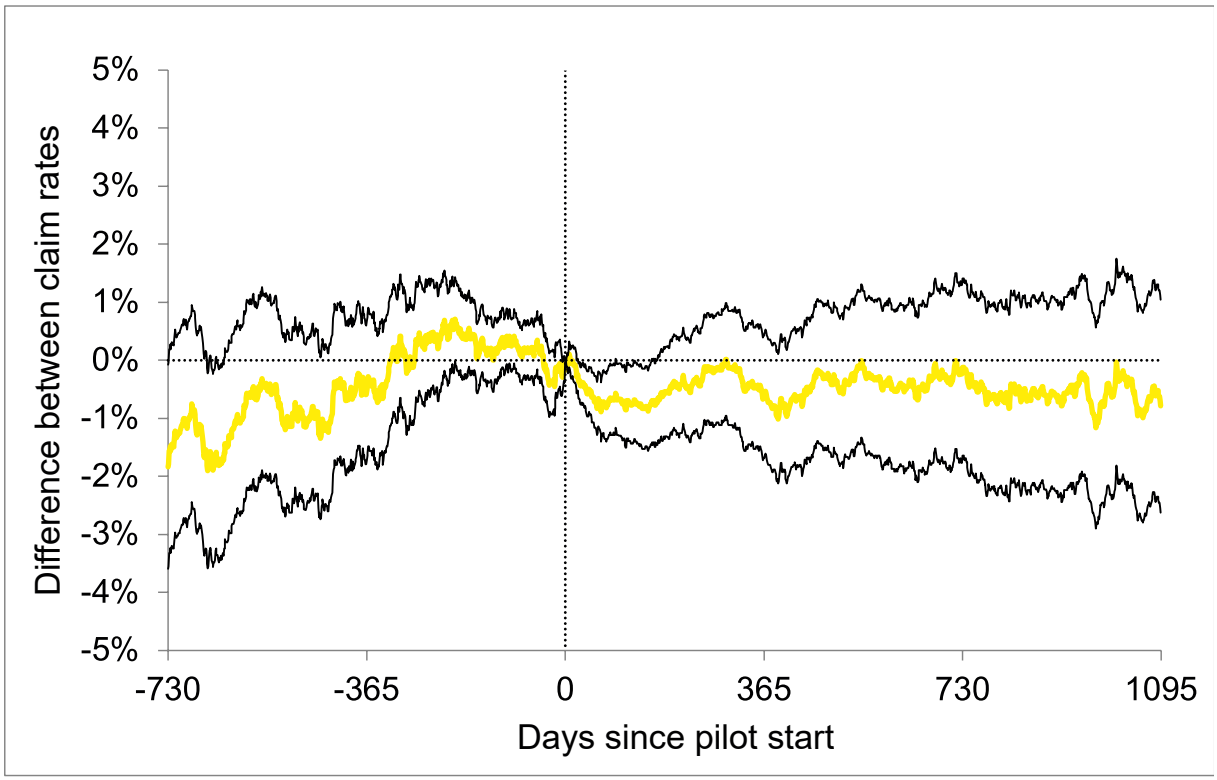
Source: HMRC P45 data: August 2017

7 Post-Pilot Outcomes

7.1 This section of the report summarises measurements made up to three years after the start of pilot participation, and the confidence that we can have in those measurements. Figures 7.1a to 7.1c show the differences in benefit receipt before and after recruitment to the pilots for each of the three models. Figures 7.2a to 7.2c show the equivalent employment data based upon HMRC's P45 data. The differences are generally small and when considering the margins of error (which are the upper and lower black lines and which indicate the range of values that we can be 95 per cent sure that the true impact lies within) the differences in the post-recruitment period are often not significantly different to zero. But we can have some confidence in the differences albeit somewhat less than 95 per cent.

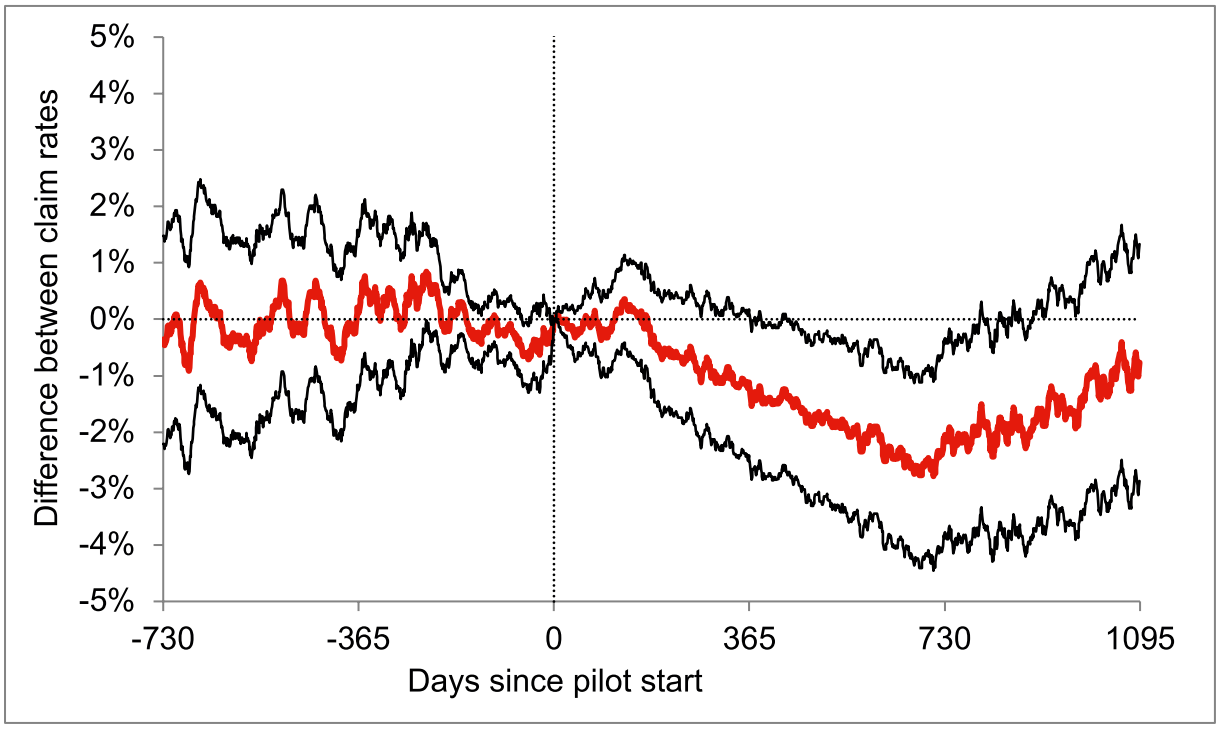
7.2 The HCP model does not exhibit a clear impact upon benefit outcomes to within a 95 per cent degree of confidence. One could argue, as we have done above, that the impact is almost measurable with a 95 per cent degree of confidence, and, for some short periods of time, the impact does pass this threshold. However, the difference in benefit receipt prior to pilot participation calls into question the cause of the post-pilot differences (we explore this issue in more detail later).

Figure 7.1a Benefit impact for the HCP model. A positive difference indicates the treatment group have a higher benefit rate than the control group



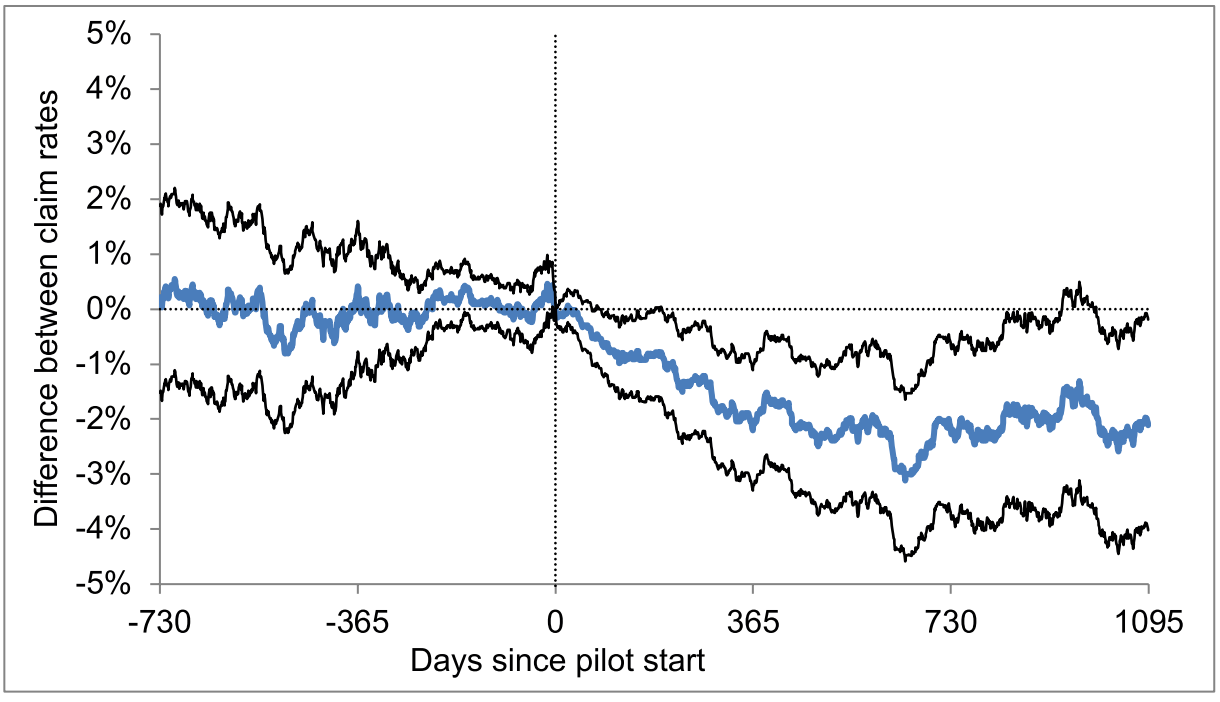
Source: DWP benefits administrative data: April 2017

Figure 7.1b Benefit impact for the WP model. A positive difference indicates the treatment group have a higher benefit rate than the control group



Source: DWP benefits administrative data: April 2017

Figure 7.1c Benefit impact for the JCP model. A positive difference indicates the treatment group have a higher benefit rate than the control group



Source: DWP benefits administrative data: April 2017

Table 7.1 Accumulative additional days off benefit

Days off benefit			
Measured after	HCP	WP	JCP
6 months	1.1	0.1	1.1
12 months	1.7	1.6	3.8
18 months	2.6	4.5	7.5
24 months	3.3	8.7	11.8
30 months	4.3	12.4	15.7
36 months	5.3	14.6	19.3

7.3 The percentage point impact from figures 6.1a-6.1c can be converted to a days off benefit measure at any given point in time by summing the daily impacts to a given point. Days off benefit for each of the three models after 6, 12, 18, 24, 30, and 36 months are given in table 6.1.

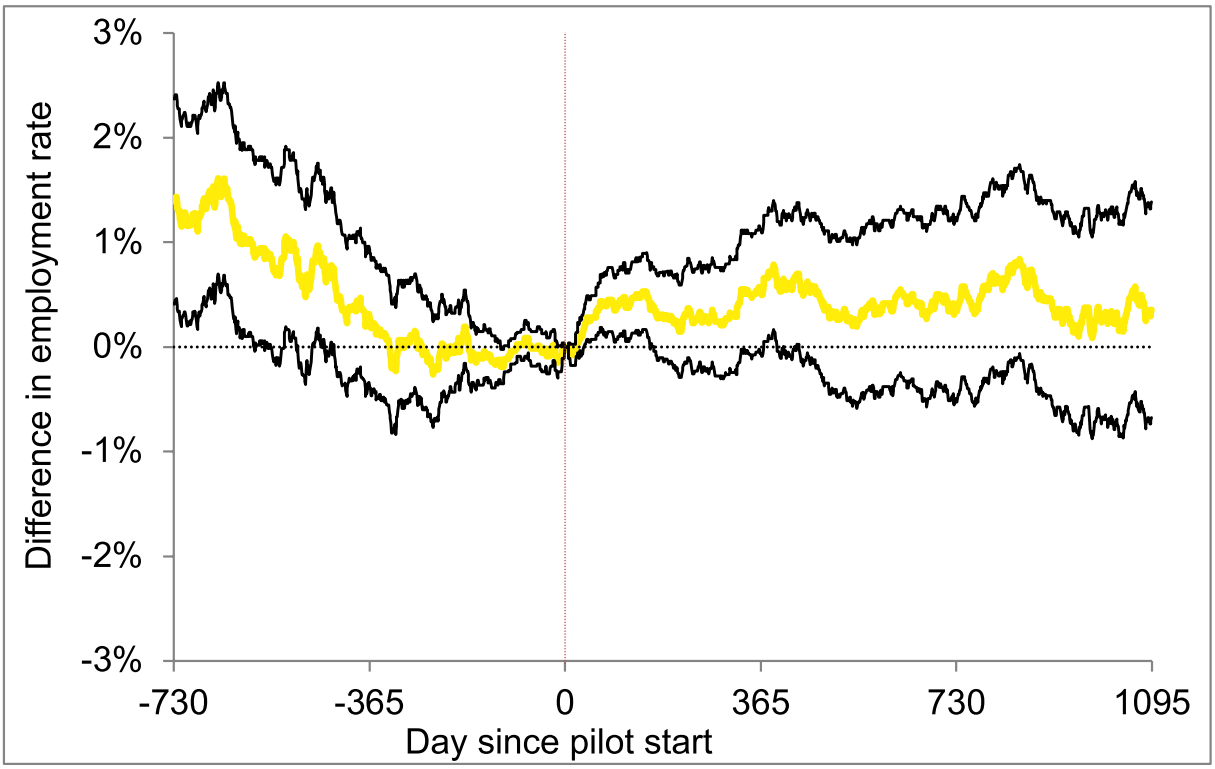
7.4 The WP model shows an unambiguous impact from roughly 12 months post-pilot onwards. The pre-pilot period is sufficiently well aligned that we do not call into question the post-pilot difference. The impact seems to be sustained to the two-year point after which the impact lessens for the final year, with the difference after three years not being statistically significant at the 95 per cent level. This reduction in impact may be due to support for claimants on the Work Programme ceasing after two years.

7.5 The JCP model begins to show a significant impact almost straight away. This impact steadily increases to be as large as 3ppt after 21 months. This impact seems to lessen at the beginning of the third year but appears sustained around 2.5ppt throughout the third year. The alignment of benefit histories between the groups prior to the pilot assures us that the impact is real, and not caused by some unexplained bias in the allocation method.

7.6 As discussed in the previous section there are some issues with the P45 employment data. However, these issues withstanding, we see that the employment impacts (figures 6.2a-c) for each of the three models are reflective of the benefit impacts.

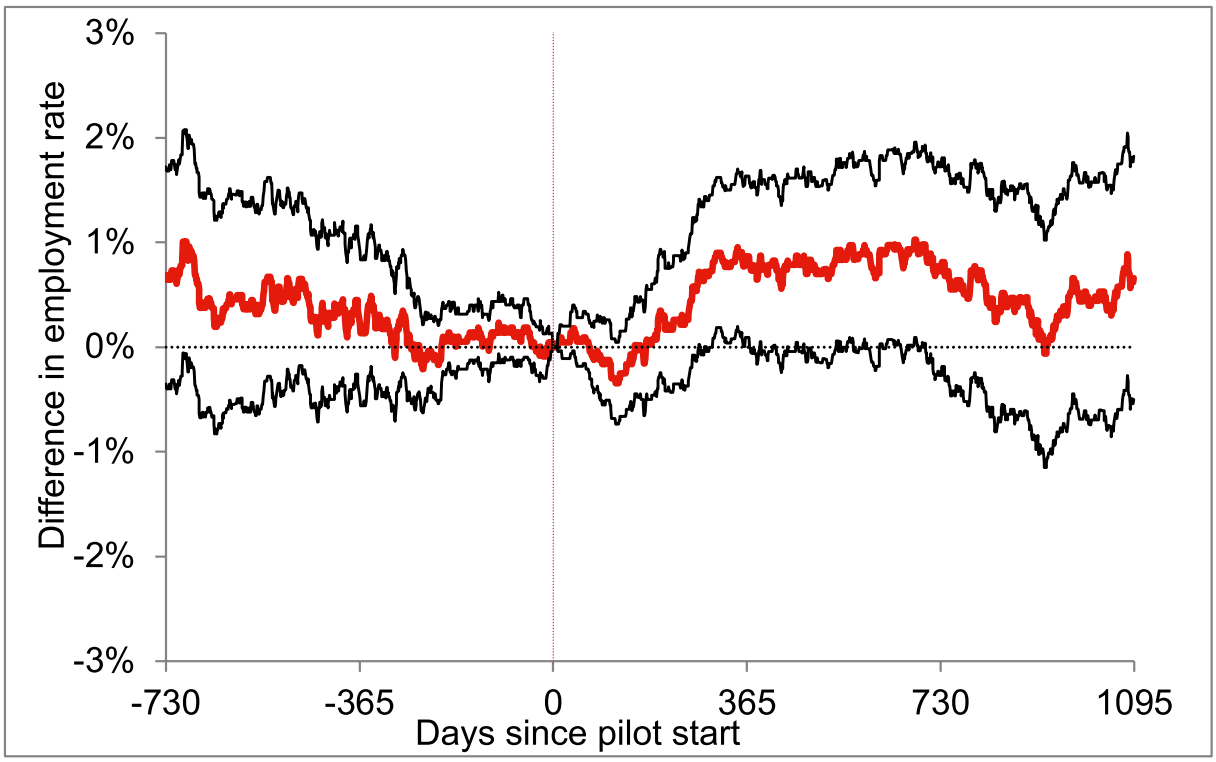
7.7 The differences in employment outcomes for each of the three models never rise above 1ppt and the margins of error indicate that this difference is not statistically significant for a sustained time period. The difference is, however, always positive, indicating that the pilot models can improve employment outcomes compared with business as usual as undertaken by the control groups.

Figure 7.2a Employment impact of the HCP model. A positive difference indicates the treatment group have a higher employment rate than the control group



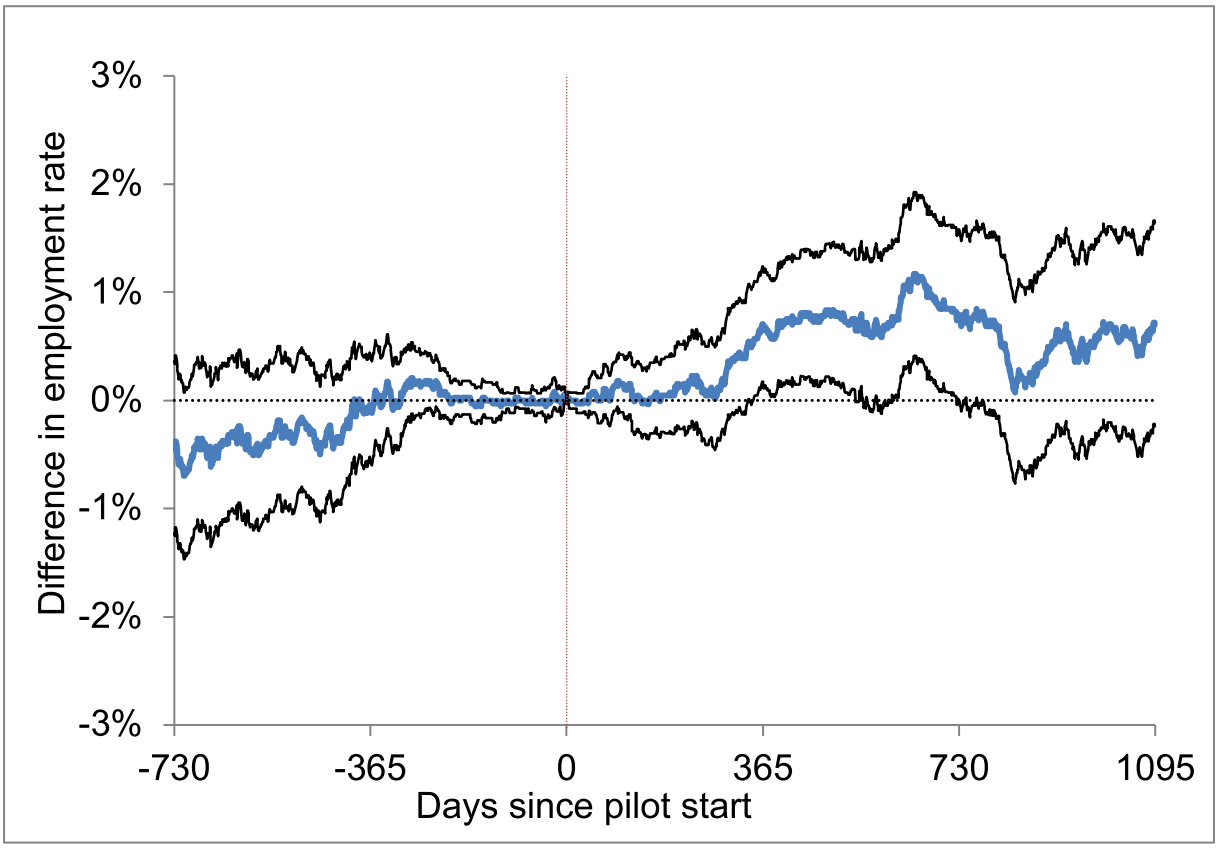
Source: HMRC P45 data: August 2017

Figure 7.2b Employment impact of the WP model. A positive difference indicates the treatment group have a higher employment rate than the control group



Source: HMRC P45 data: August 2017

Figure 7.2c Employment impact of the JCP model. A positive difference indicates the treatment group have a higher employment rate than the control group



Source: HMRC P45 data: August 2017

7.8 Figures 7.1a-c and figures 7.2a-c show the difference in benefit receipt and employment outcomes respectively. The actual rates of benefit receipt and employment figures are given in Annex B.

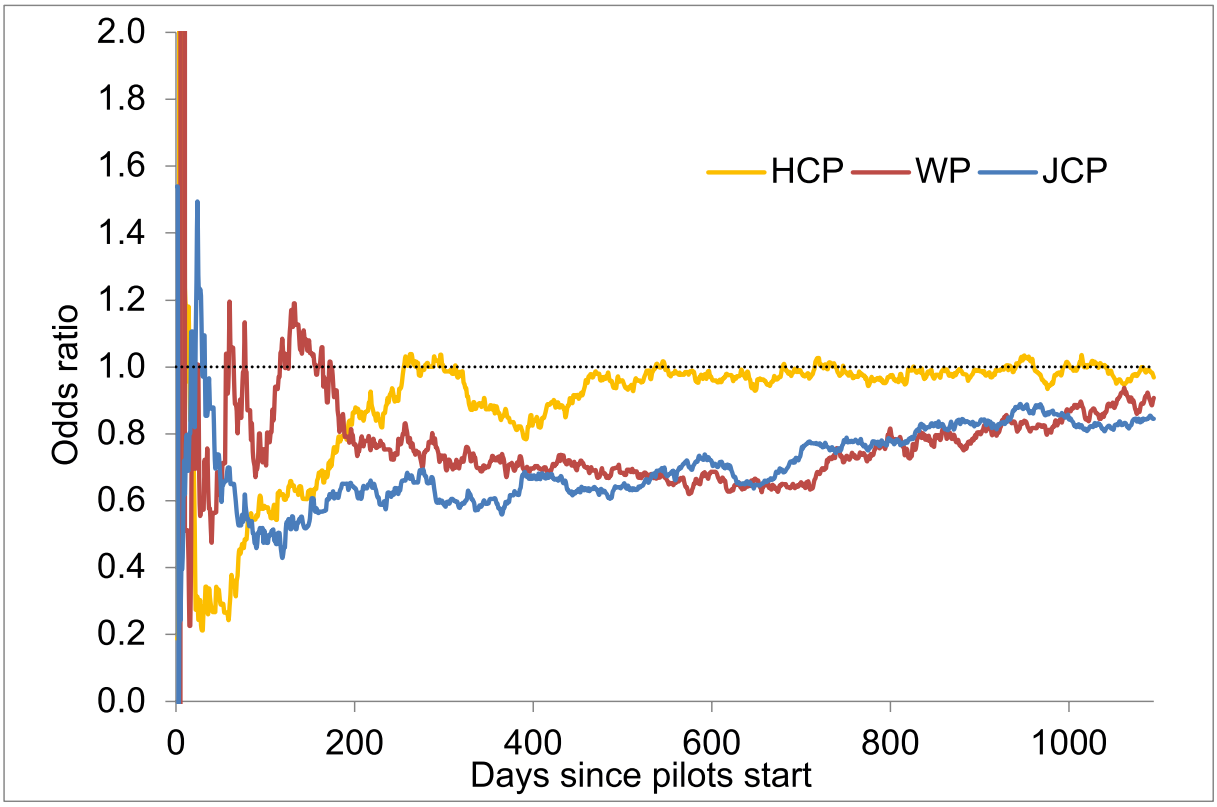
7.9 We have used 'Logistic Regression' to measure impact, in order to enable a more unbiased comparison of impacts, and corroborate the 'face value' interpretation of the outcome data. Logistic Regression allows us to model the impact of a particular intervention whilst controlling for other 'confounding factors', i.e. those which have an association (causal or otherwise) with the outcome of interest and are more prevalent in one of the groups than another. In other words, we expect the existing differences between groups to have some bearing on benefit and employment outcomes. If an RCT produced unbiased treatment and control groups then there would be no confounding factors; both groups would equally exhibit all measurable characteristics.

7.10 Figure 7.3 summarises, for the three different models, the differences in benefit receipt due to pilot participation after controlling for various factors, such as benefit histories, and most of the characteristics given in Annex A (except where they affect very few people). We cannot accurately control for a characteristic where very few people demonstrate that characteristic because there is statistical uncertainty regarding the influence of that characteristic. Benefit histories are aggregated into fortnightly blocks with a variable for each block to record whether or not the person was on benefits at any point during that fortnight.

7.11 Another consideration is that Logistic Regression tells us the impact upon the 'odds' of having an outcome. The method doesn't in the first instance tell us how much more likely a person is to have an outcome depending upon some factor, but it provides us with an odds ratio (OR), which is simply the ratio of the odds of having an outcome should you exhibit some characteristic, and the odds of having an outcome should you not have that characteristic. For example, if the odds of males having an outcome were 4 to 1 and females 5 to 1, then the OR would be (4/1)/(5/1)=0.8. Note that this is not the same as saying that men are 20 per cent less like to get a job because the actual probabilities are 80 per cent (4 to 1) and 83.3 per cent (5 to 1). The difference in probability is actually 3.3ppts or 4 per cent of the baseline, which is somewhat smaller than the 0.8 that the OR measures.

7.12 Figure 7.3 shows the OR (of participants compared to non-participants) of being on a primary DWP benefit at various points in time after pilot participation. The HCP model initially seems to show a significant positive impact (an OR below 1 signifies a lower likelihood of being on benefits). That impact then reduces to the point where, by nine months or so, there is no discernible impact upon benefit receipt.

Figure 7.3 Odds ratio for each of the pilot models



Source: DWP benefits administrative data: April 2017

7.13 The WP model fluctuates between showing a positive and negative impact within the first six months then settles down to show a positive impact for the remainder of the first two years. At the start of the third year the impact begins to lessen and is negligible at the end of year three. This behaviour is similar to that which we observed in figure 7.1b, and indicates that the WP model shows the greatest impact within the first two years of the programme.

7.14 Figure 6.3 initially appears to show that JCP pilot participants are more likely to be claiming benefits in the first four weeks of pilot participation. This appears to be reflected in figure 6.1c, which suggests that more pilot participants were on benefits shortly after pilot participation. After this period, we see a positive impact which steadily declines over time before plateauing at around 0.8 for year three.

7.15 Margins of error are not shown in figure 6.3 because to do so would confuse the presentation of the key messages to be taken from that chart. Annex D presents each model's impact in isolation to the others and provides margins of error so that the periods of time when the impacts are statistically significant (with a 95 per cent degree of confidence) is evident.

7.16 After the initial volatility in the short-term impacts, the regression analysis suggests that the HCP model has little influence upon benefit receipt; after six months the OR is an average of 0.96. The WP model has an average impact of 0.79 and the JCP model an average impact of 0.72 after this period.

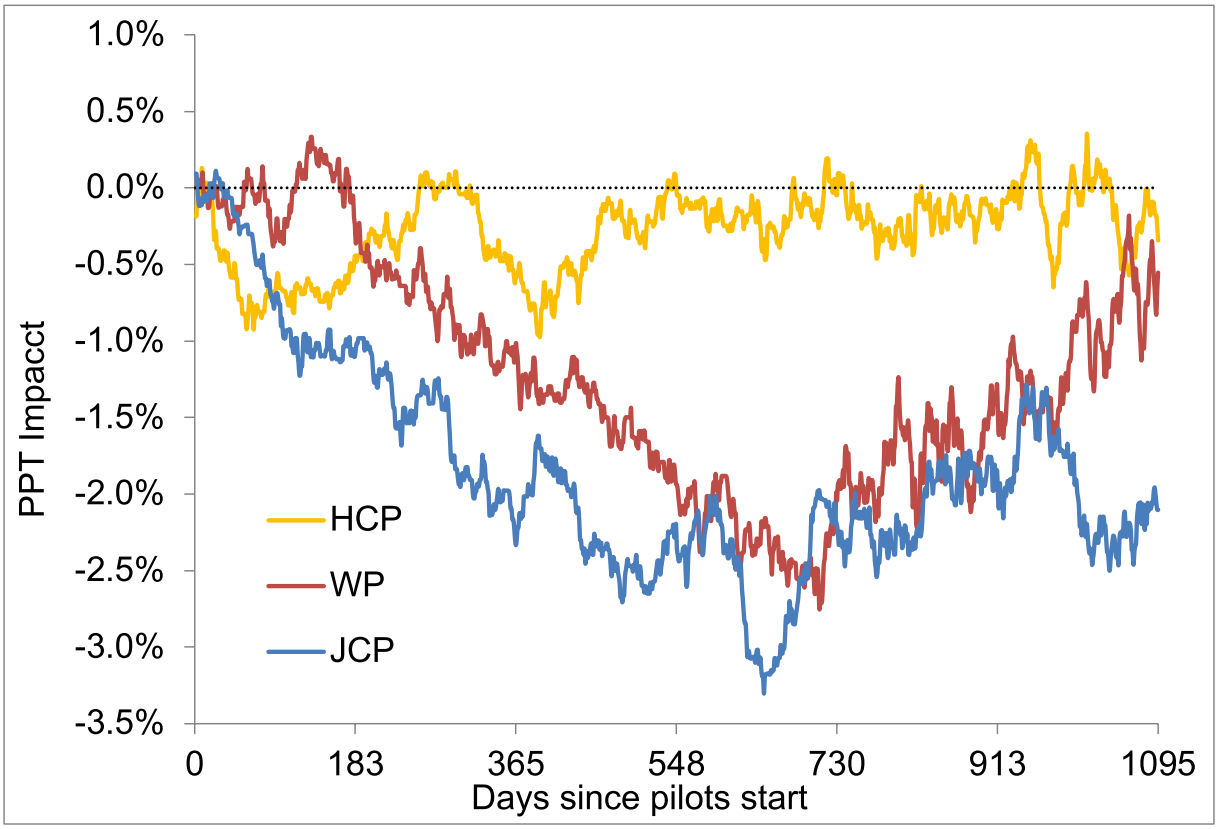
7.17 It is important to bear in mind that the OR exaggerates the relative likelihood of being on benefits, as illustrated by the example of gender in paragraph 7.11. After translating the OR to the absolute risk-reduction of being off benefits due to being in the treatment group (figure 7.4), we find the results taken from the Logistic Regression mirror that of those found from the analysis of the empirical data. More details of the transformation between OR and absolute relative risk are given in Annex D.

7.18 Furthermore, we can again calculate the additional days off benefit the logistics model predicts for the treatment group (Annex D table D.1). After three years, the HCP model has achieved a predicted impact of 3.1 days, the WP model: 13.4 days, and the JCP model 20.1 days.

7.19 In correcting for any biases, the logistic model actually predicts a higher impact for the JCP model and suggests that the HCP and WP models would have slightly smaller impacts than those reported in table 7.1. However, the percentage point impacts from the logistic models only slightly differ from those found above and lie comfortably within the margins of error. As such, we can be assured that any bias is not significantly altering the results and we will report the findings from analysis of the empirical data as our main findings.

7.20 The summary figures shown above include every non-exempt pilot participant. We note that one consequence is that the true impact will have been diluted by the inclusion in the treatment group of people who did not receive the pilot intervention but we have no way of isolating an equivalent control group for the subgroup who did receive the intervention. Further, owing to the ITT (Intention to treat) approach that we have taken, some people in the control group did receive the intervention, further reducing the contrast between the treatment and the control group's exposure to the pilot intervention.

Figure 7.4 Percentage point impacts from the logistic regression model

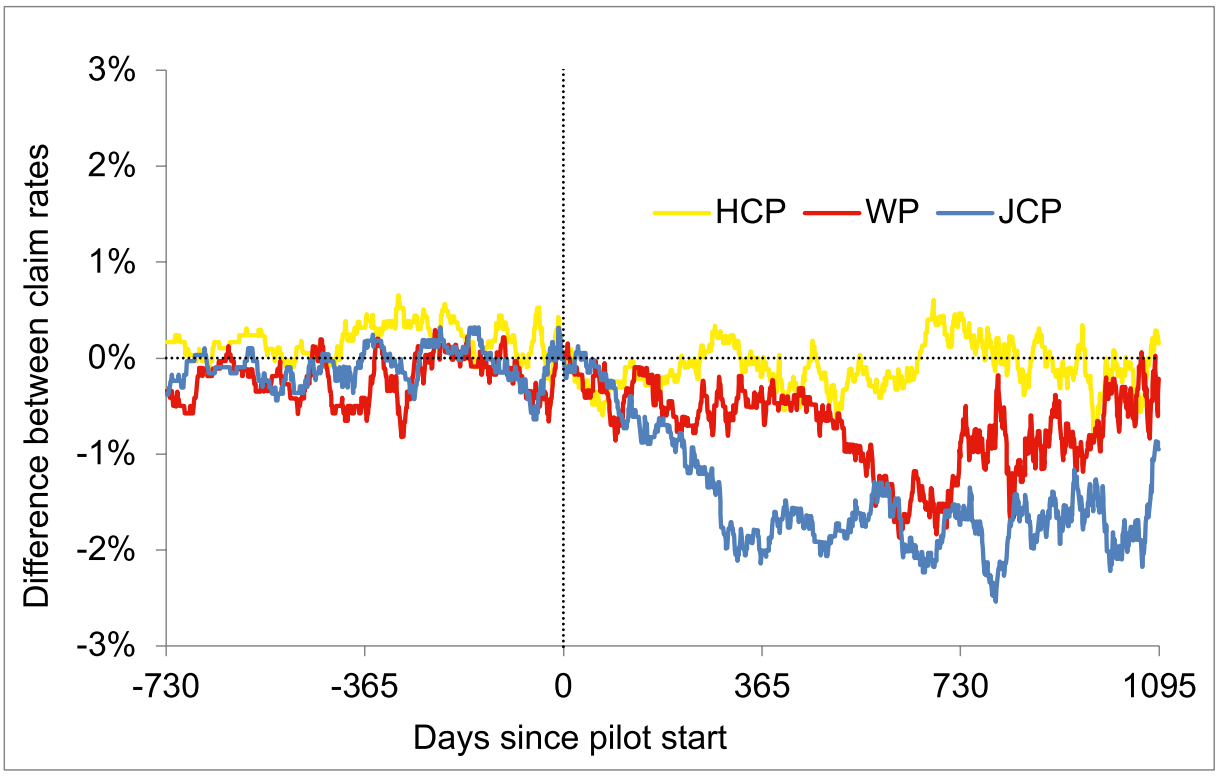


Source: DWP benefits administrative data: April 2017

7.21 These high-level findings obscure observations for subgroups. For example, if we look at the subgroup who we believe to be IB migration cases (figure 7.5), the HCP model shows a consistently negative impact upon benefits. The JCP and WP model show consistently positive impacts, but the latter reduces after a time whereas the JCP model does not show a diminishing impact.

7.22 In addition to the caveat above, we should note that the good agreement between the treatment and control groups' benefit histories is not a good indicator of the absence of bias. Nearly all participants will have spent nearly all of the prior two years on DWP benefits, which gives little scope for differences in those histories to emerge. Nonetheless, there are few significant differences in the other characteristics that we have examined and certainly none that would account for the differences in performance that we observe. Indeed, after three years, the HCP model shows only 1 additional day off benefit, the WP models shows, 8 additional days, while the JCP model shows 16 additional days off benefit.

Figure 7.5 IB Reassessment Benefit impacts



Source: DWP administrative data: April 2017

8 Conclusion

8.1 In conclusion, recruitment to the ESA 18 to 24 month prognosis pilots has been subject to some errors, with random allocation error rates varying from two per cent to six per cent depending upon the pilot model and the pilot (treatment or control) group. We have not been able to confirm that everybody in scope of the pilots has been allocated to a pilot group because of the poor quality of WCA data. There may have been inconsistencies in the application of exemption criterion insofar that, where we could confirm exemption status, people with that status could be found in both the treatment and the control groups.

8.2 We also found that not everybody with a pilot marker set, and who was not exempted, seemed to undergo the pilot intervention that was being tested in their area. There may be valid reasons for this but we were not able to pinpoint them. Regardless, the consequence was that we carried out an ITT analysis and not one where the treatment group was restricted to active pilot participants. As a result, any impacts we measured are likely to have been diluted by the presence of non-participants in our treatment group and active participants in our control group.

8.3 Our assessment of the net impact upon benefits suggests that the HCP model has an average 0.47ppt impact upon benefit receipt after the first six months, which translates to additional 5.3 days off benefit per treatment group participant after three years.

8.4 The WP model showed a consistent improvement in impacts over the first two years of the program with a maximum impact of 2.78ppts; this impact lessens over the third year to 0.77ppts. Over the three years the WP model yields on average an additional 14.6 days off benefit per treatment group participant.

8.5 The JCP model appears to show a sustained impact over the three averages with an average impact of 2.02ppts from month six which does not show any sign of lessening. Over the three years the JCP model yields on average an additional 19.3 days off benefit per treatment group participant.

8.6 There were minimal employment impacts shown across all three pilots. However, we should reiterate that these impacts will underestimate employment and will not show any self-employment outcomes.

Annex A: Characteristics of the pilot populations

Table A.1 Characteristics of the HCP model participants

Differences which are statistically significant are marked with a double asterisk

Characteristic	Treatment	Control
Active Participants	**2654**	**2324**
Gender		
Male	53.7%	54.7%
Female	46.3%	45.3%
Ethnicity		
White	85.3%	85.7%
Black	1.6%	2.1%
Asian	5.2%	4.7%
Mixed	0.8%	0.9%
Chinese/Other	0.9%	0.6%
Prefer not to say	5.3%	5.5%
Unknown	0.9%	0.6%
Age at Start of Pilot		
16 to 24	9.0%	9.2%
25 to 29	4.9%	5.1%
30 to 39	17.7%	16.6%
40 to 49	28.6%	27.9%
50 to 59	33.5%	33.6%
60 or Over	6.4%	7.5%
Primary Condition		
Certain Infectious and Parasitic Diseases	0.9%	0.9%
Neoplasms	0.6%	0.8%
Diseases of the Blood and Blood forming organs and certain diseases involving the immune mechanism	0.2%	0.0%
Endocrine, Nutritional and Metabolic Diseases	1.8%	1.3%
Mental and Behavioural Disorders	50.6%	52.2%
Diseases of the Nervous System	6.1%	6.5%
Diseases of the Eye and Adnexa	0.7%	0.6%
Diseases of the Ear and Mastoid Process	0.6%	0.5%

Characteristic	Treatment	Control
Diseases of the Circulatory System	2.6%	3.1%
Diseases of the Respiratory System	1.8%	1.4%
Diseases of the Digestive System	1.2%	1.6%
Diseases of the Skin and Subcutaneous System	0.8%	0.4%
Diseases of the Musculoskeletal system and Connective Tissue	18.7%	18.1%
Diseases of the Genitourinary System	0.5%	0.4%
Pregnancy, Childbirth and the Puerperium	0.0%	0.1%
Certain Conditions Originating in the Perinatal Period	0.0%	0.0%
Congenital Malformations, Deformations and Chromosomal Abnormalities**	0.2%	0.5%
Symptoms, Signs and Abnormal Clinical and Laboratory findings, not elsewhere classified	8.0%	8.1%
Injury, Poisoning and certain other consequences of external causes**	3.9%	2.8%
Factors influencing Health Status and Contact with Health Services	0.8%	0.7%
Skills Needs		
Basic Skills Need	8.5%	8.6%
English as a second language**	1.3%	0.6%
Number of Children		
1 Child	9.6%	9.3%
2 Children	6.4%	6.8%
3 Children	3.1%	2.6%
4 or More Children	2.2%	1.7%
Age of Youngest Child		
0 to 2	3.1%	2.4%
3 or 4	2.0%	1.5%
5 to 10	6.7%	5.9%
11 to 15	5.2%	6.5%
16 or Over	4.9%	4.8%
Unknown	0.9%	0.5%
In Receipt of Partner Allowance	21.0%	20.7%

Source: DWP benefits administrative data: April 2017 and Labour Market System: July 2017

Table A.2 Characteristics of the WP model participants

Differences which are statistically significant are marked with a double asterisk

Characteristic	Treatment	Control
Active Participants	1922	2038
Gender		
Male	51.5%	52.6%
Female	48.5%	47.4%
Ethnicity		
White	94.8%	94.3%
Black	0.1%	0.1%
Asian	0.7%	1.0%
Mixed	0.4%	0.3%
Chinese/Other	0.3%	0.4%
Prefer not to say	3.3%	3.4%
Unknown	0.4%	0.4%
Age at Start of Pilot		
16 to 24	7.0%	6.9%
25 to 29	4.6%	4.6%
30 to 39	15.0%	16.4%
40 to 49	28.5%	29.6%
50 to 59	37.7%	36.9%
60 or Over**	7.1%	5.6%
Primary Condition		
Certain Infectious and Parasitic Diseases	0.5%	0.4%
Neoplasms	0.4%	0.8%
Diseases of the Blood and Blood forming organs and certain diseases involving the immune mechanism	0.1%	0.1%
Endocrine, Nutritional and Metabolic Diseases	1.0%	1.1%
Mental and Behavioural Disorders	51.4%	49.7%
Diseases of the Nervous System**	5.7%	7.6%
Diseases of the Eye and Adnexa	0.9%	0.6%
Diseases of the Ear and Mastoid Process	0.4%	0.6%
Diseases of the Circulatory System	2.8%	2.7%
Diseases of the Respiratory System	1.8%	1.8%
Diseases of the Digestive System**	1.7%	0.9%
Diseases of the Skin and Subcutaneous System	0.5%	0.6%
Diseases of the Musculoskeletal system and Connective Tissue	17.8%	16.1%

Characteristic	Treatment	Control
Diseases of the Genitourinary System	0.5%	0.7%
Pregnancy, Childbirth and the Puerperium	0.1%	0.0%
Certain Conditions Originating in the Perinatal Period	0.0%	0.0%
Congenital Malformations, Deformations and Chromosomal Abnormalities	0.3%	0.2%
Symptoms, Signs and Abnormal Clinical and Laboratory findings, not elsewhere classified	9.6%	10.6%
Injury, Poisoning and certain other consequences of external causes	3.9%	3.9%
Factors influencing Health Status and Contact with Health Services	0.9%	1.4%
Skills Needs		
Basic Skills Need	7.5%	8.8%
English as a second language	0.1%	0.1%
Number of Children		
1 Child**	8.0%	10.4%
2 Children**	5.0%	6.7%
3 Children	2.3%	2.8%
4 or More Children	1.9%	1.9%
Age of Youngest Child		
0 to 2	1.9%	2.5%
3 or 4	1.1%	1.5%
5 to 10	5.9%	7.0%
11 to 15**	4.1%	6.0%
16 or Over	4.7%	5.9%
Unknown	0.8%	0.7%
In Receipt of Partner Allowance	19.9%	20.6%

Source: DWP benefits administrative data: April 2017 and Labour Market System: July 2017

Table A.3 Characteristics of the JCP model participants
Differences which are statistically significant are marked with a double asterisk

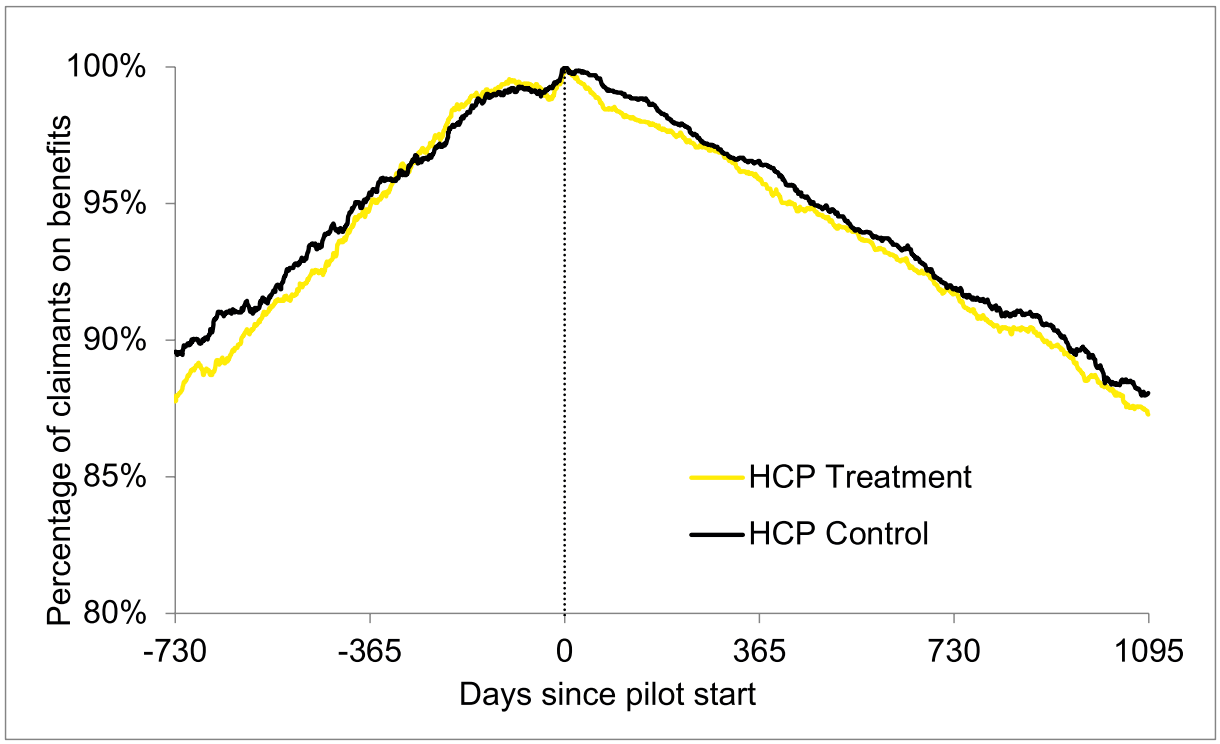
Characteristic	Treatment	Control
Active Participants	2554	2657
Gender		
Male	53.5%	53.2%
Female	46.5%	46.8%
Ethnicity		
White	85.6%	87.4%
Black	1.3%	1.8%
Asian	2.0%	2.5%
Mixed	0.8%	0.9%
Chinese/Other	1.1%	0.8%
Prefer not to say**	9.2%	6.1%
Unknown**	0.1%	0.5%
Age at Start of Pilot		
16 to 24	8.6%	7.6%
25 to 29	4.4%	4.2%
30 to 39	17.3%	16.4%
40 to 49	28.6%	30.3%
50 to 59	34.6%	34.0%
60 or Over	6.5%	7.6%
Primary Condition		
Certain Infectious and Parasitic Diseases	1.1%	1.2%
Neoplasms	0.6%	0.4%
Diseases of the Blood and Blood forming organs and certain diseases involving the immune mechanism	0.1%	0.2%
Endocrine, Nutritional and Metabolic Diseases	2.0%	1.5%
Mental and Behavioural Disorders	51.0%	49.3%
Diseases of the Nervous System	6.0%	6.8%
Diseases of the Eye and Adnexa**	1.7%	0.9%
Diseases of the Ear and Mastoid Process	0.4%	0.5%
Diseases of the Circulatory System	3.0%	3.2%
Diseases of the Respiratory System	2.0%	1.5%
Diseases of the Digestive System	1.5%	1.6%
Diseases of the Skin and Subcutaneous System	0.4%	0.5%
Diseases of the Musculoskeletal system and Connective Tissue	17.0%	17.8%

Characteristic	Treatment	Control
Diseases of the Genitourinary System	0.6%	0.7%
Pregnancy, Childbirth and the Puerperium	0.0%	0.0%
Certain Conditions Originating in the Perinatal Period	0.0%	0.0%
Congenital Malformations, Deformations and Chromosomal Abnormalities	0.0%	0.2%
Symptoms, Signs and Abnormal Clinical and Laboratory findings, not elsewhere classified	8.8%	8.4%
Injury, Poisoning and certain other consequences of external causes**	3.2%	4.4%
Factors influencing Health Status and Contact with Health Services	0.6%	0.8%
Skills Needs		
Basic Skills Need	6.3%	6.3%
English as a second language	0.5%	0.6%
Number of Children		
1 Child	9.3%	10.6%
2 Children	5.3%	5.7%
3 Children**	3.0%	1.7%
4 or More Children	1.5%	1.3%
Age of Youngest Child		
0 to 2	1.8%	2.2%
3 or 4	1.4%	1.2%
5 to 10	6.8%	5.7%
11 to 15	4.7%	4.9%
16 or Over	5.2%	5.6%
Unknown	0.5%	0.8%
In Receipt of Partner Allowance	16.9%	16.7%

Source: DWP benefits administrative data: April 2017 and Labour Market System: July 2017

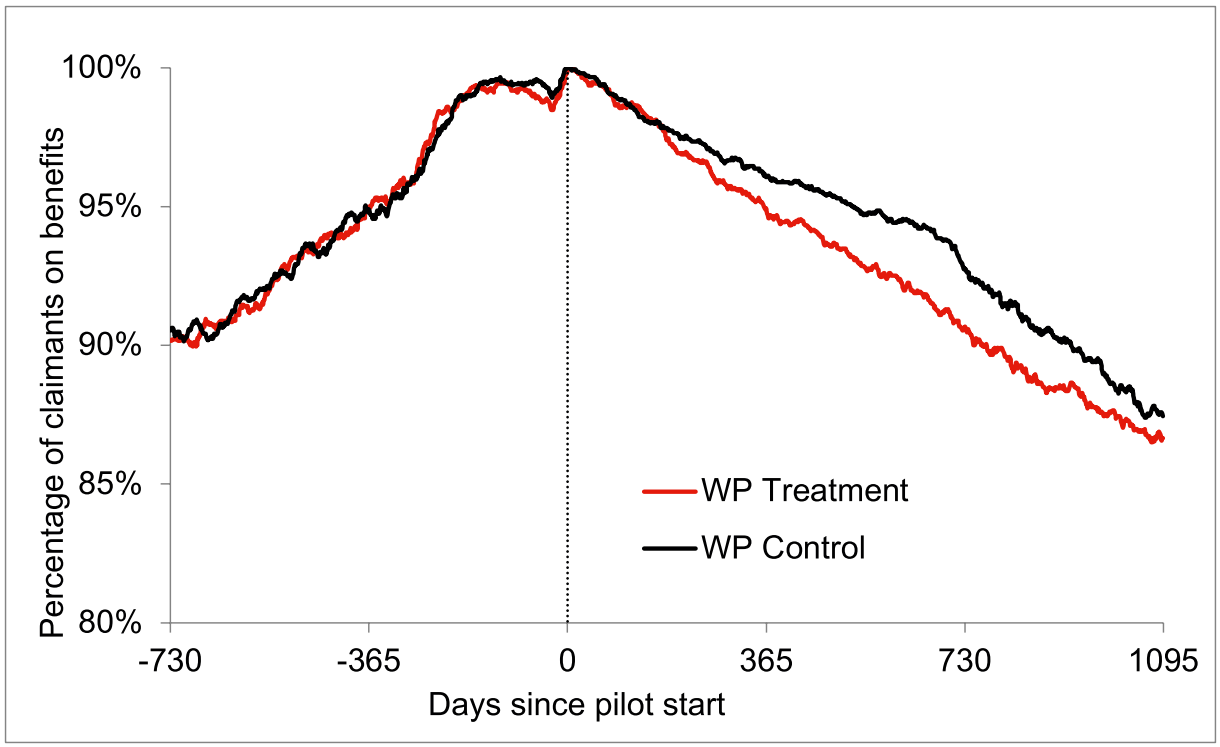
Annex B: Benefit and Employment rates

Figure B.1 Benefit profiles for the HCP model



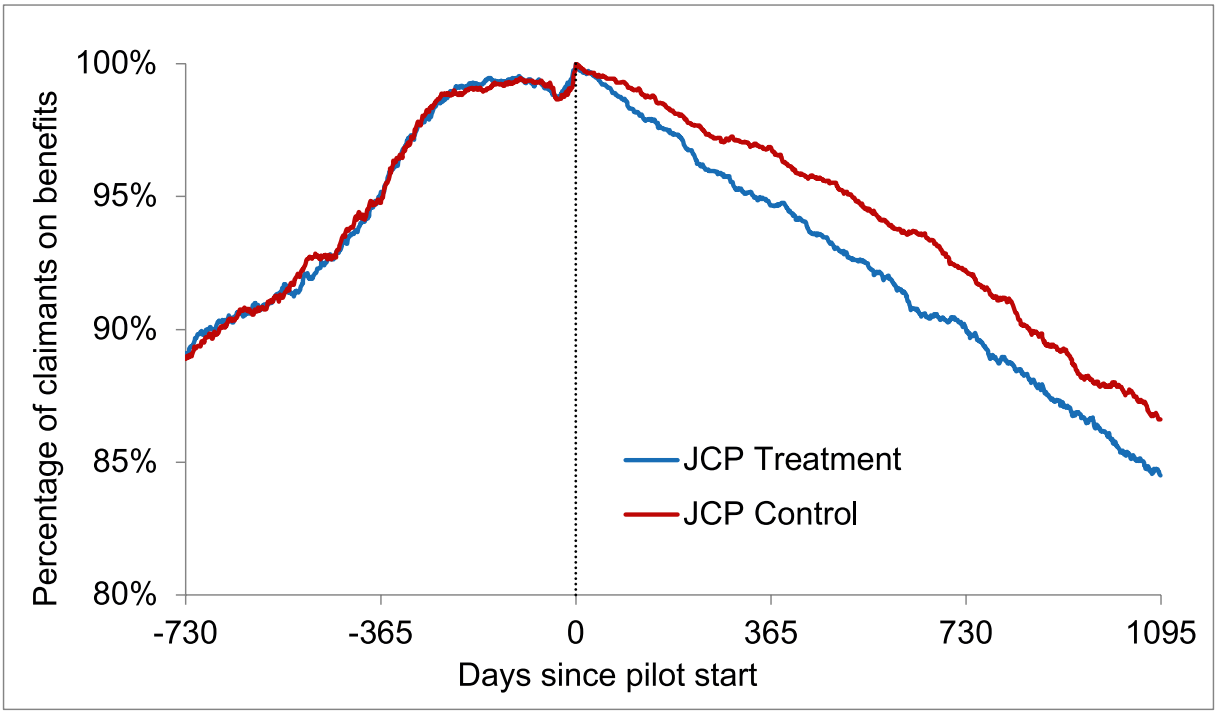
Source: DWP benefits administrative data: April 2017

Figure B.2 Benefit profiles for the WP model



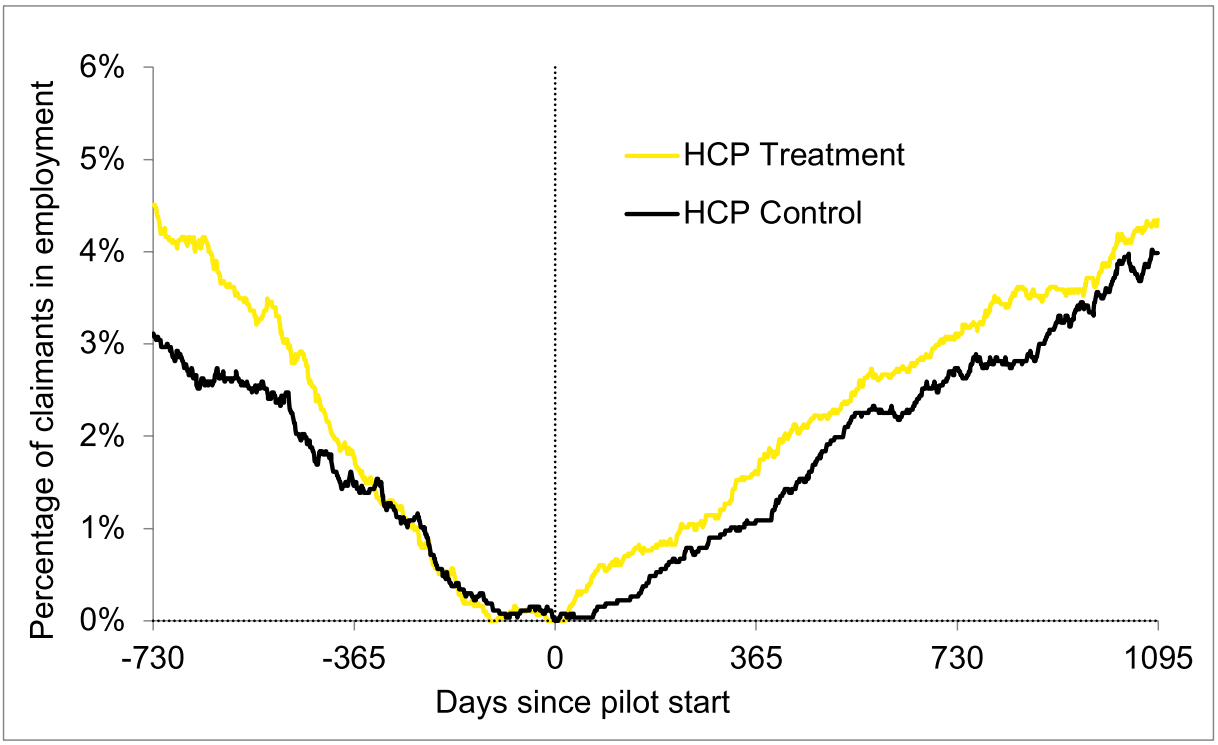
Source: DWP benefits administrative data: April 2017

Figure B.3 Benefit profiles for the JCP model



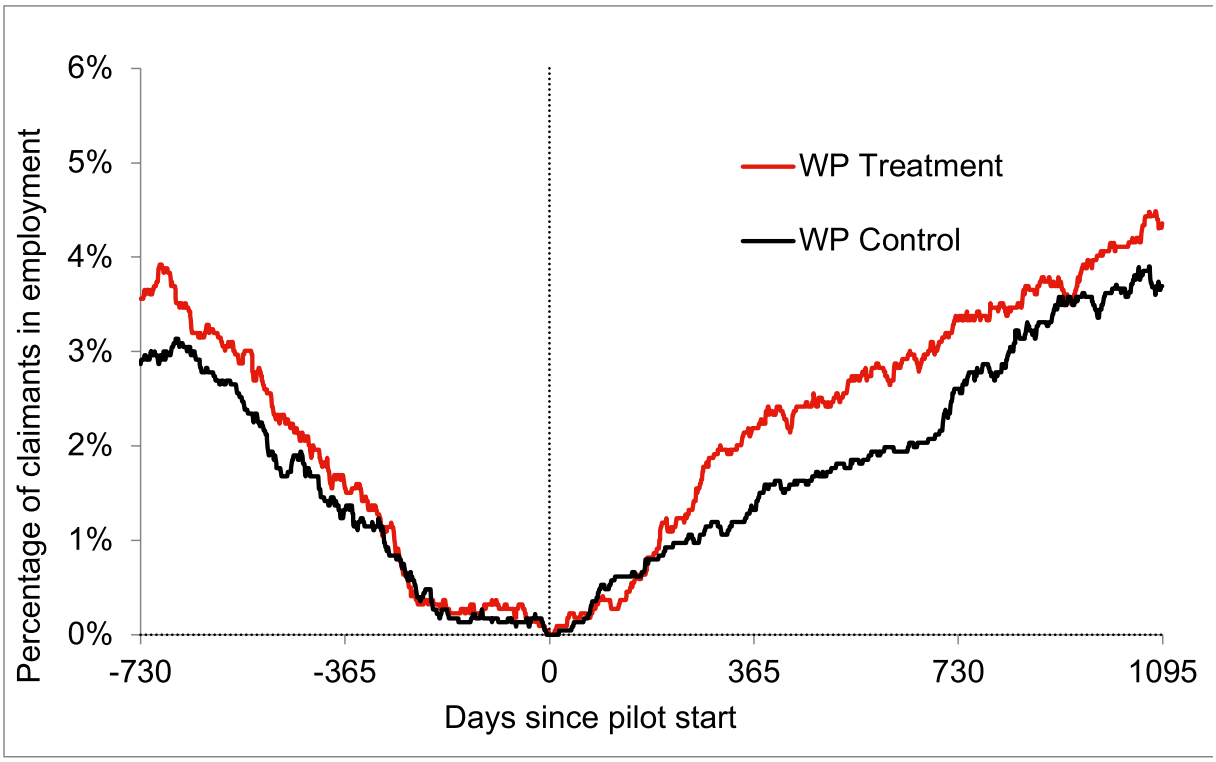
Source: DWP benefits administrative data: April 2017

Figure B.4 Employment profiles for the HCP model



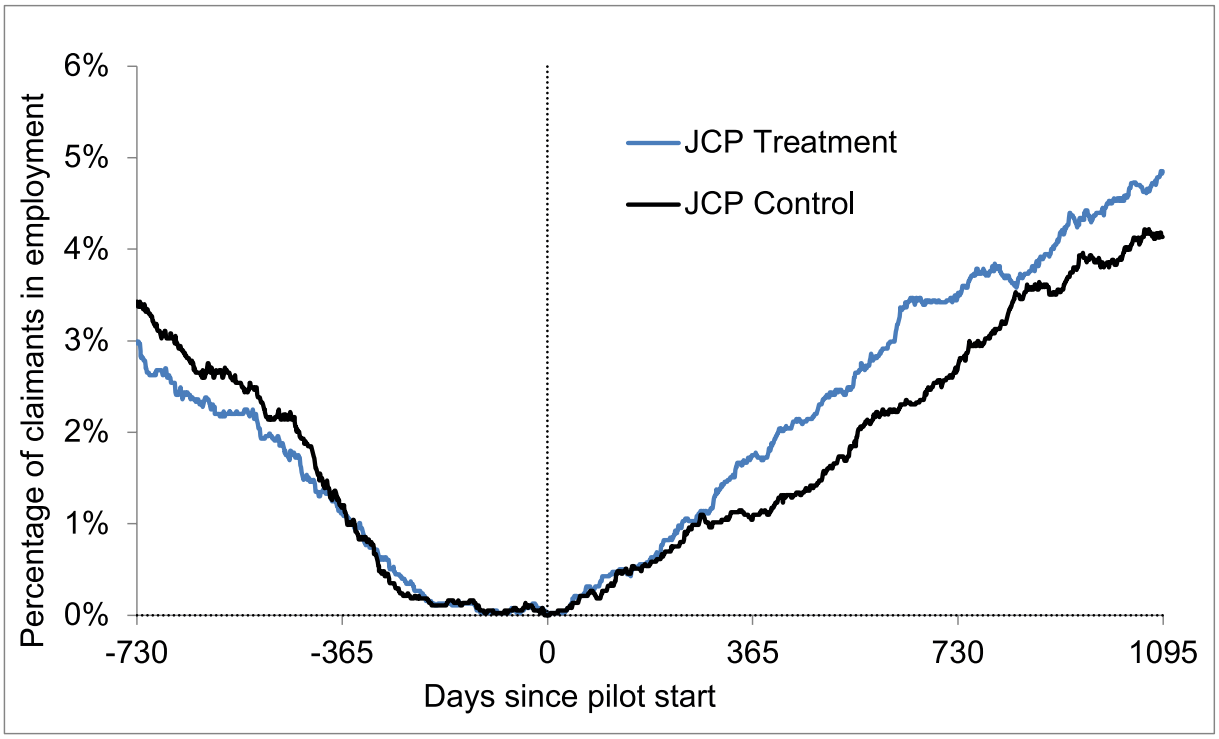
Source: HMRC P45 data: August 2017

Figure B.5 Employment profiles for the WP model



Source: HMRC P45 data: August 2017

Figure B.6 Employment profiles for the WP model



Source: HMRC P45 data: August 2017

Annex C: Accumulative days off benefit and in employment

The following figures show the cumulative days off benefit for each of the three models. The cumulative number of days off benefit after x number of days is the sum of the percentage point difference calculated in figures 5.1a to 5.1c up to day x and is the area bounded between the difference line and the x-axis. If this area is positive, it indicates more days are spent on benefit by the treatment group than the control group, whereas a negative area indicates a positive impact. These findings are summarised at six-month intervals in table 5.1.

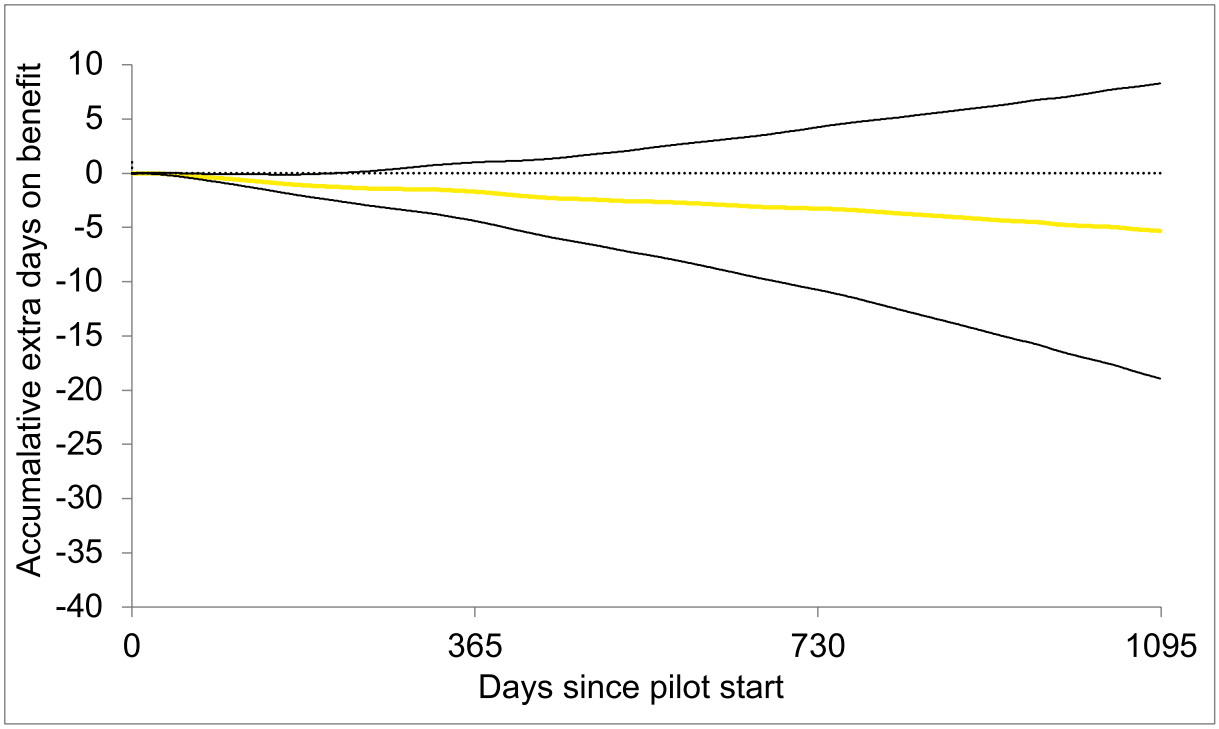
The confidence intervals are calculated by summing the 95 per cent confidence intervals from figures 5.1a to 5.1c. This method has assumed there is complete dependence between one data point and the next, and is probably an overestimate in the range of uncertainty. As such, though the upper bounds of the WP model (figure C.2) touch upon the x-axis indicating the results are not statistically significant at the 95 per cent confidence level, one should not discount that the overall trend shows an impact.

The confidence intervals used in figures 5.1a to 5.1c and in 5.2a to 5.2c use the standard formula for difference in proportion between two populations, that is:

$$CI_{\pm} = p_0 - p_1 \pm 1.96\sqrt{\frac{p_0(1-p_0)}{n} + \frac{p_1(1-p_1)}{m}}$$

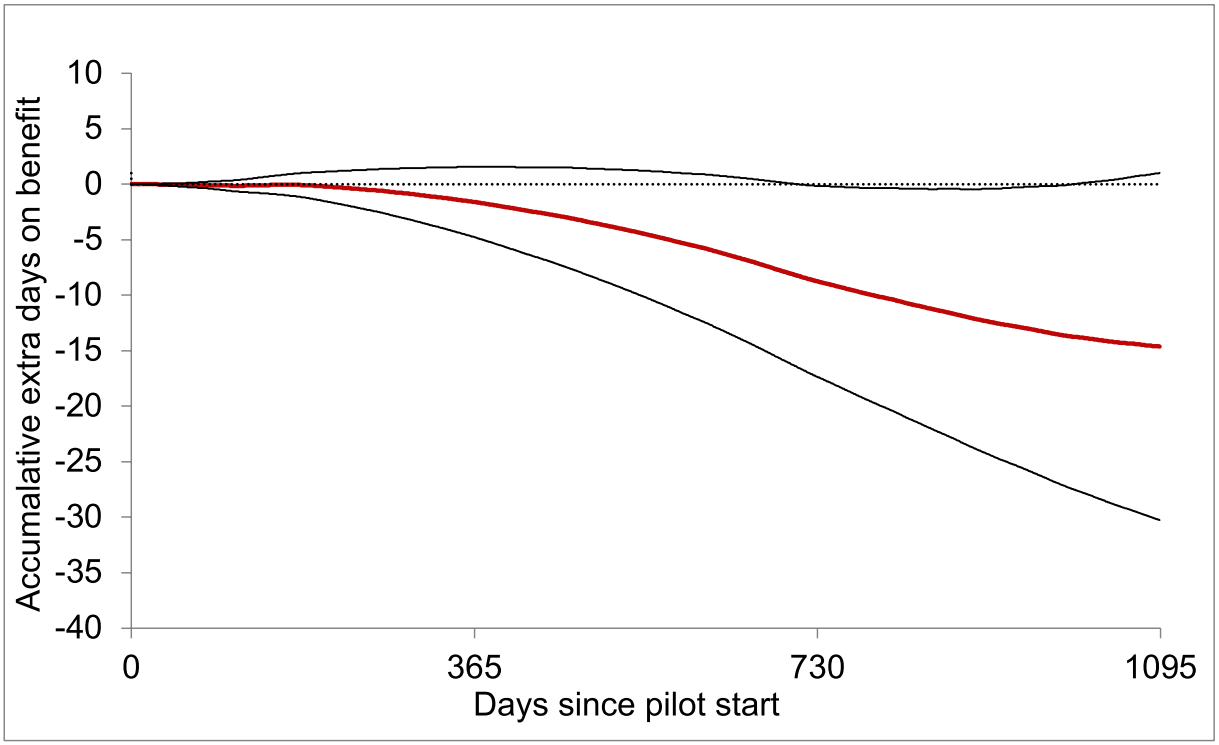
where p_0 and p_1 are the proportions in the control and treatments groups respectively and n and m are the sizes of these groups.

Figure C.1 Accumulative days on benefit for the HCP model



Source: DWP benefits administrative data: April 2017

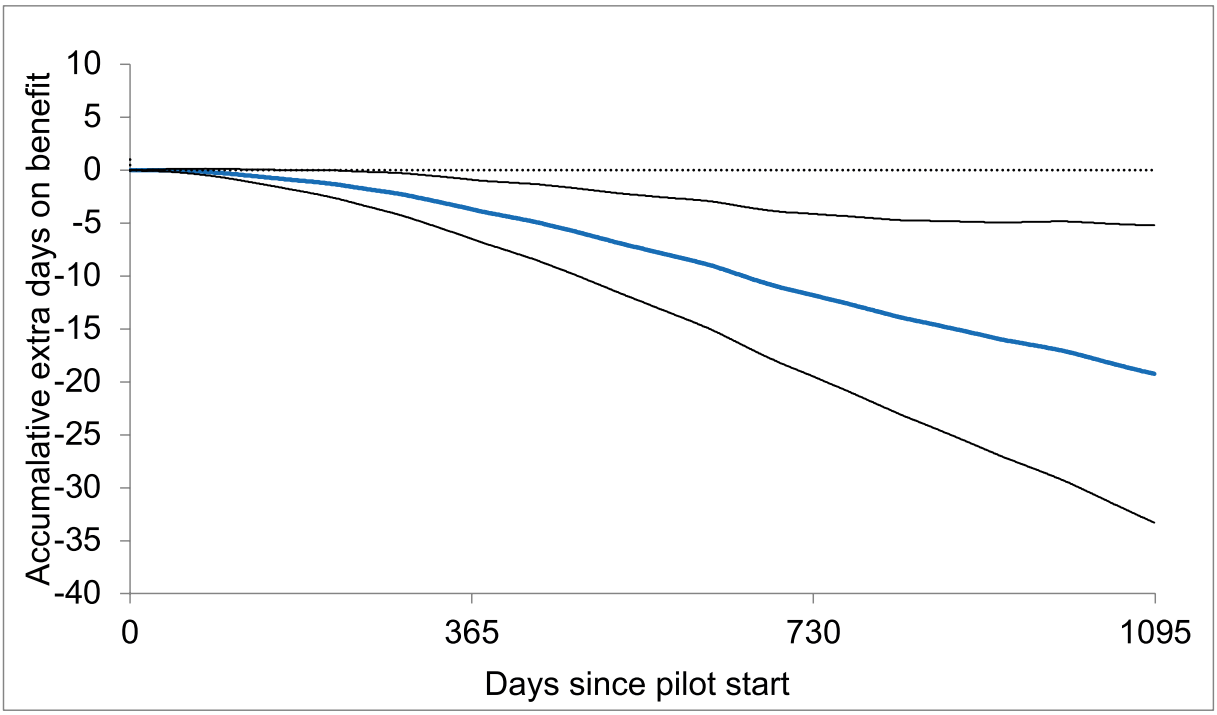
Figure C.2 Accumulative days on benefit for the WP model



Source: DWP bene its administrative data: April 2017

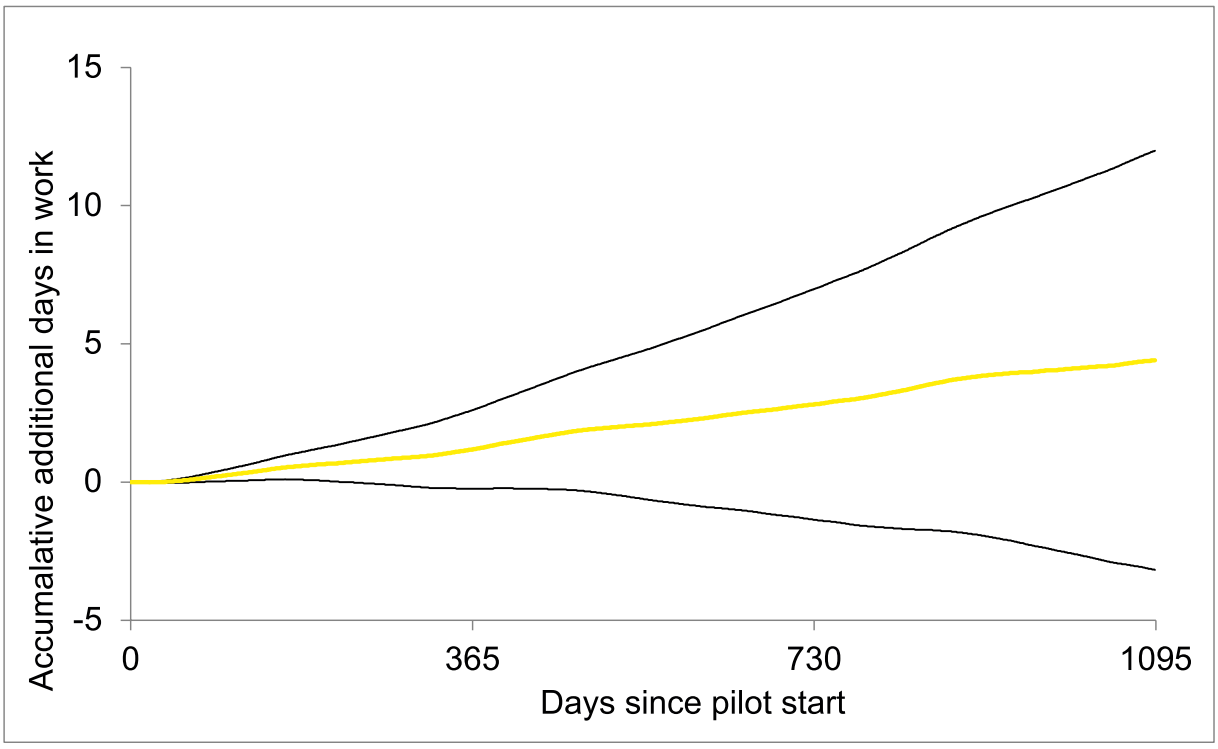
Note: The upper and lower black lines indicate the margin of error for the cumulative extra days on benefit for each model, in other words: the range of values that we can be 95 per cent sure that the accumulative extra days of benefit lie between.

Figure C.3 Accumulative days on benefit for the JCP model



Source: DWP benefits administrative data: April 2017

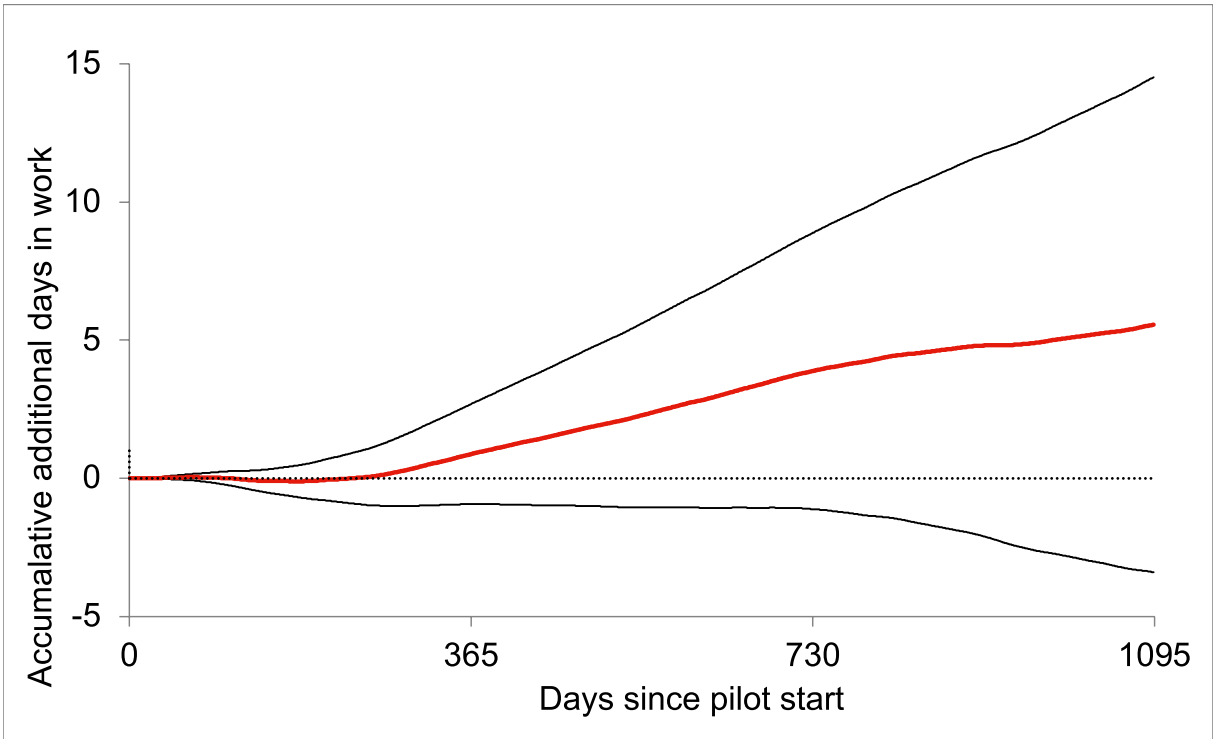
Figure C.4 Accumulative days in employment for the HCP model



Source: HMRC P45 data: August 2017

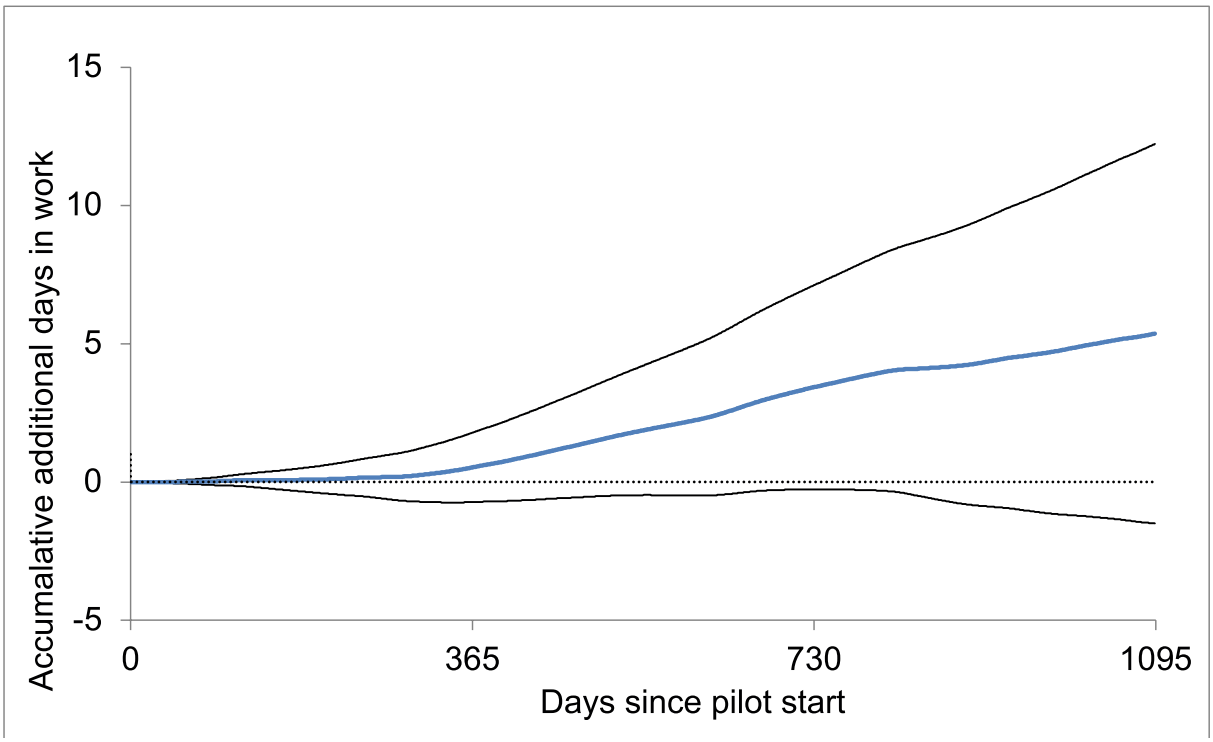
Note: The upper and lower black lines indicate the margin of error for the cumulative extra days on benefit (C.3) or in employment (C.4) for each model.

Figure C.5 Accumulative days in employment for the WP model



Source: HMRC P45 data: August 2017

Figure C.6 Accumulative days in employment for the JCP model



Source: HMRC P45 data: August 2017

Note: The upper and lower black lines again indicate the margin of error for the cumulative extra days in employment for each model.

Annex D: Logistic Regression

The logistic model we have developed uses the benefit status on any given day as the dependant variable. We have included the standard characteristics given in tables A.1 to A.3 as covariates in the model. We have also included benefit history and the date the claimant started the trial as covariates. We have used the Logistic Regression procedure in SAS 9.3 for all our logistics regression modelling. Among the outputs available from this procedure are the odds ratio (OR) and Wald confidence limits at the 95 per cent level on this OR.

The nominal allocation (we are still using the intention to treat methodology) to the treatment or control group is added in as a binary dummy variable with the control coded as zero and the treatment group coded as one. In this way, the OR outputted from this model can be treated as the factor that the odds on remaining on benefits would increase by if the claimant was part of the treatment group. As such, an OR of one indicates there is no benefit to being in the treatment group, and OR greater than one indicates that the being in the treatment group increases the odds of remaining on benefits relative to the control group, and an OR of less than one indicates that being in the treatment group reduces a claimant's odds of being on benefit relative to the control group.

As mentioned above, directly reading the OR can mislead when trying to measure the impact. In order to give a meaningful figure for the impact that allows us to calculate the cumulative days off (or on) benefit, as predicted by the logistic regression model, we must calculate the absolute risk reduction.

This measure can be defined as where *RR is* the relative risk and p0 is the probability of being of a claimant being on benefits if they are part of the control group. If p0 were small then the OR would approximate the relative risk and we could just use that in our calculation of the absolute risk reduction. However, as the rate of remaining on benefits is always above 85 per cent we cannot use this approximation. We relate the relative risk to the OR through the formula:

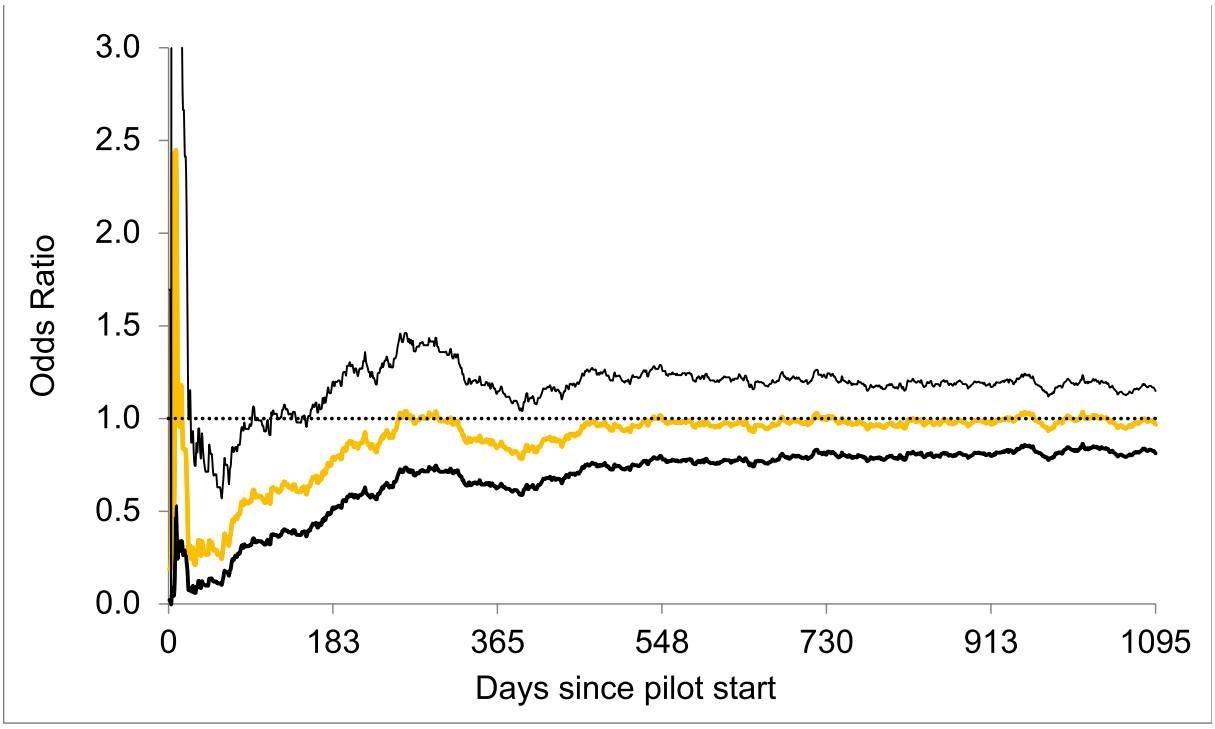
$$RR = \frac{OR}{(1 - p_0) + p_0 \times OR}.$$

Which is easily defined through manipulation the definition of relative risk and the OR:

$$RR = \frac{p_1}{p_0}, \quad OR = \frac{p_1/(1 - p_1)}{p_0/(1 - p_{0)}}.$$

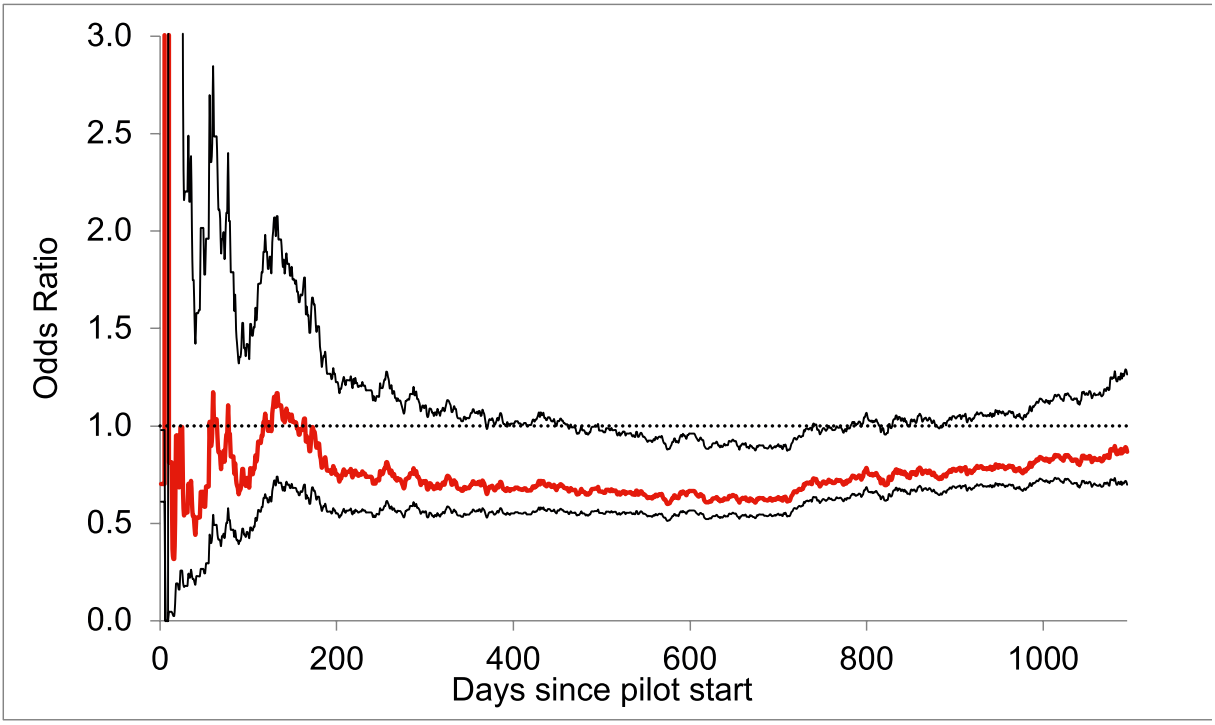
Figures D.1-3 give the ORs for the three years following the pilot start for each of the three models with the associated Wald confidence limits. Table D.1 shows the cumulative days off benefit every six months.

Figure D.1 HCP Model



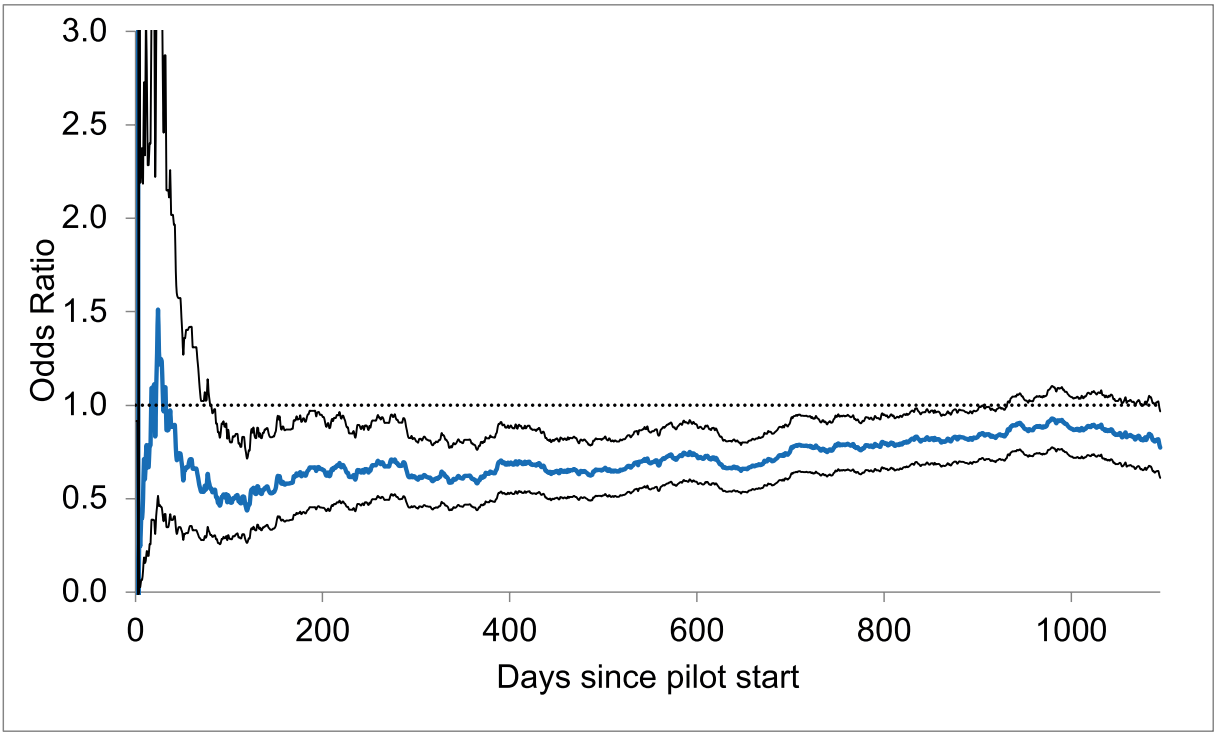
Source: DWP benefits administrative data: April 2017

Figure D.2 WP Model



Source: DWP benefits administrative data: April 2017

Figure D.3 JCP Model



Source: DWP benefits administrative data: April 2017

Table D.1 Accumulative days off benefit as calculated using the logistic regression approach

Days off benefit			
Measured after	HCP	WP	JCP
6 months	1.1	0.1	1.2
12 months	1.5	1.5	4.1
18 months	2.2	4.2	8.1
24 months	2.5	8.3	12.7
30 months	2.8	11.5	16.6
36 months	3.1	13.4	20.1

Source: DWP benefits administrative data: April 2017