



Department
for Work &
Pensions

Estimating the Early Labour Market Impacts of Universal Credit

Updated Analysis

December 2015

DWP ad hoc research report no. 28

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

© Crown copyright 2015.

You may re-use this information (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence. To view this licence, visit <http://www.nationalarchives.gov.uk/doc/open-government-licence/> or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gsi.gov.uk.

This document/publication is also available on our website at:
<https://www.gov.uk/government/collections/research-reports>

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

First published 2015.

ISBN 978-1-78425-687-6

Executive Summary

This is a follow up report on *Estimating the early labour market impacts of Universal Credit – Early Analysis* published in February 2015. It evaluates the short-term impact of Universal Credit on labour market outcomes. Furthermore it describes improvements made to the method and data through extensive sensitivity analysis. It concentrates on the latest data for over 8,000 single unemployed claimants who made a new claim between July 2013 and September 2014 in all of the first ten Pathfinder offices. However, it also considers short-term outcomes for claims made up to February 2015 and longer term outcomes (up to 11 months) for a smaller sample of earlier claims made in the original 4 Pathfinder sites.

The latest estimates suggest that, compared with similar Jobseeker's Allowance claimants making equivalent claims during the same period (July 2013 – September 2015), Universal Credit Pathfinder claimants were 8 percentage points more likely to have been in work within the first 270 days after making their claim. Moreover, at every 30 day interval after making their claim, from 30 days to 270 days, Universal Credit claimants were between 3 and 6 percentage points more likely to be in work than the matched Jobseeker's Allowance comparison group.

On average Universal Credit claimants are estimated to work about 12 days more than the matched comparison group during the first 270 days after they made their claim. We also find evidence of a positive impact on earnings. However, the variability in earnings makes these estimates more uncertain. Furthermore, the estimated impacts on days worked and earnings are both less reliable than the impact on employment status due to the data available.

It is important to note that complex analysis of this nature will continually evolve. Consequently, we would expect the precise point estimates to change over time as more data becomes available and the approach is continually refined. We will continue to update the analysis as more data becomes available, and we will extend the analysis to include more claimants (including different types of claimants), in more areas and for longer periods of time.

Contents

	Executive Summary.....	ii
	List of Abbreviations.....	iv
	List of Figures.....	v
	Non-technical Summary.....	1
1	Introduction.....	4
	1.1 Policy background.....	4
	1.2 The Roll-out of Universal Credit.....	5
2	Evaluation Methodology.....	7
	2.1 Time.....	8
	2.2 Geography.....	8
	2.3 Individuals.....	8
	2.4 Data.....	9
3	Summary of Preliminary Results (published in February 2015).....	12
4	Sensitivity Analysis.....	14
5	Revised Estimates.....	17
	5.1 New data.....	17
	5.2 Revised estimates for short-term employment outcomes.....	17
	5.3 Tracking outcomes for longer.....	18
	5.4 Comparison of outcomes for different cohorts, demographic groups and analysis of matching.....	20
	5.5 Impact on earnings.....	25
	5.6 Sensitivity analysis for latest data.....	27
6	Conclusion and Next Steps.....	29
	Annex A: Sensitivity Analysis to Assess the Internal Validity of the February 2015 Impact Estimates.....	31
	Annex B: Independent NIESR Peer Review.....	71
	Annex C: The Gateway Conditions.....	78
	Bibliography.....	82

List of Abbreviations

CC	Claimant Commitment
DWP	Department for Work and Pensions
HMRC	Her Majesty's Revenue and Customs
IFS	Institute for Fiscal Studies
JCP	Jobcentre Plus
JSA	Jobseeker's Allowance
LEL	Lower Earning Limit
NIESR	National Institute of Economic and Social Research
PAYE	Pay As You Go
PSM	Propensity Score Matching
RTI	Real-Time Information
UC	Universal Credit

List of Figures

Name	Description	Page number
Figure 1	Theory of change underpinning improved labour market outcomes.	5
Figure 2	Evaluation dataset flowchart.	11
Figure 3	Snapshots of labour market outcomes in the original 4 Pathfinder Jobcentre Plus offices and comparable JSA Jobcentre Plus offices for new claims made between July 2013 and April 2014.	12
Figure 4	Cumulative labour market outcomes in the original 4 Pathfinder Jobcentre Plus offices and comparable JSA Jobcentre Plus offices for new claims made between July 2013 and April 2014.	13
Figure 5	Comparison of cumulative employment impacts.	18
Figure 6	Impact of UC on snapshot employment rates for 8,300 new UC claims in 10 offices between July 2013 and September 2014.	19
Figure 7	Impact of UC on cumulative employment rates for 8,300 new UC claims in 10 offices between July 2013 and September 2014.	19
Figure 8	Impact of UC on days worked for 8,300 new UC claims in 10 offices between July 2013 and September 2014.	20
Figure 9	Snapshot employment outcomes broken down by cohort and benefit type.	21
Figure 10	Impact on the likelihood of being employed for July 2013 - July 2014 cohort in original 4 offices.	22
Figure 11	Kaplan-Meier Survival Estimates for July 2013 – September 2014 cohort for 10 offices.	23
Figure 12	Comparison of snapshot employment rates between UC and JSA new claims made in different six-month cohorts.	24
Figure 13	Snapshot employment rates by gender	24
Figure 14	Snapshot employment rates by age	25
Figure 15	Summary of the short term employment outcomes for different cohorts, type of statistics used and time frames.	26
Figure 16	Estimated effect of UC on average earnings for different cohorts.	27
Figure 17	Snapshot employment impact for Original and Second Best Comparator Offices.	28

Non-technical Summary

This ad hoc publication presents our latest and most reliable estimates of the impact that Universal Credit (UC) is having on the labour market outcomes of new claimants. It finds that UC claimants are much more likely to have worked after making a new claim than a matched comparison group of similar people making equivalent claims to Jobseeker's Allowance (JSA) during the same period in similar areas.

UC is designed to increase movement into work and progression in work through:

- Improved financial incentives
- Greater simplicity, flexibility and transparency
- Bringing more people into conditionality regimes

We expect people to respond to improved incentives by more of those out of work moving into work, and those already working increasing their earnings. We are investing around £1.7 billion in the development and delivery of UC. Over the longer term this will deliver wider economic benefits estimated at £6.7 billion (gross benefit) per year. This includes significant value from more people being in work, which will reduce welfare benefit expenditure and increase tax receipts. (DWP, 2014a, p.31)

To evaluate the effectiveness of UC we have examined the labour market outcomes of new claimants.

The Department for Work and Pensions (DWP) is committed to a robust evaluation of UC. Building on a feasibility study by the Institute for Fiscal Studies (IFS), and in consultation with the UC Expert Advisory Group, we developed a method that matches new UC claimants with new JSA claimants and compares their employment and earnings. The method ensures that, as far as possible given existing data, we only compare similar people making similar claims at the same time in similar areas so that we can isolate the impact of UC.

The first preliminary estimates were published in February 2015. These suggested that UC claimants were 5 percentage points more likely to have worked within 120 days of making their claim and had earned more than matched JSA claimants. Due to the data available at the time this analysis was limited to only looking at single unemployed claimants who were eligible to make a new claim to UC in the first four Pathfinder offices between July 2013 and April 2014. This report was independently peer reviewed by the National Institute of Economic and Social Research (NIESR). (DWP, 2015)

Since producing the preliminary estimates, we have:

1. Undertaken a significant programme of sensitivity analysis to address issues highlighted by NIESR's peer review and to test more thoroughly the validity of our method and the reliability of the results. This work has given us increased confidence in both the method and results. It has also led to a small number of refinements in the approach, which ensure that we make the best use of available data and that the results are as robust as possible.
2. Updated the analysis using the latest data on benefit claims, employment and earnings. With a rising caseload and a longer data series we can now measure the impact of UC for more claimants, in more offices over a longer period.

The latest data enables us to estimate the impact of UC on outcomes up to 270 days after the claim was made for over 8,000 single unemployed claimants who claimed UC between July 2013 and September 2014 in the first ten Pathfinder offices. We can also look at the shorter-term impact on more recent claims (those made up to February 2015) and longer-term outcomes for claims made up to July 2014 in the original 4 offices (up to 330 days after the claim was made).

The updated analysis confirms that **UC claimants are more likely to move into employment** than similar JSA claimants. Additional sensitivity analysis suggests that the evaluation method and results are robust and reliable. In particular, in the first ten offices between July 2013 and September 2014 we find:

- New UC claimants were about **8 percentage points** more likely to have been employed within 270 days of making their claim than matched new JSA claimants in similar areas (71% versus 63%). 270 days after making a claim the proportion of UC claimants employed was **3 percentage points** higher than the proportion of matched JSA claimants in work at the same point in time. These positive results hold for claims made at different six month intervals, for different age groups, and for both men and women.
- UC claimants had worked an estimated **12 days more** than their matched JSA counterparts.
- Positive and statistically significant impacts on earnings. These impacts are small relative to the estimated impacts on employment.

These results collectively suggest that UC is encouraging new claimants, and making it more worthwhile and easier for them, to do small amounts of work.

It is more difficult at this stage to estimate the impact of UC on earnings because:

- a) earnings vary a lot, which makes it difficult to detect differences due to UC; and
- b) for these early claimants, we do not have data on, and therefore cannot control for, past earnings.

We will be able to produce more robust estimates of the impact of UC on earnings when we look at new claimants later in roll-out for whom we will have some data on their past earnings.

NIESR have conducted a further independent peer review of this latest report. This concluded *“This updated analysis provides a comprehensive set of sensitivity checks and makes revisions to the methodology in light of these checks. It also uses more developed and updated data and looks at longer term outcomes. Much as the initial assessment, the updated assessment appears to be carried out very carefully and competently.”*

These are encouraging results. They show the positive effect that we would expect in terms of UC moving more people into work. As the caseload of UC claimants continues to grow it will be possible to extend the analysis to look at a wider range of claimants. We will continue developing this analysis, and in 2016 it should be possible to use this approach to estimate the impacts for: claimants with housing costs; all Jobcentre Plus offices in the North West, lone parents; and couples with children. We will also be able to start exploring whether UC is helping people to progress in the labour market for those claims made earlier in the roll-out for whom we will have data on their longer-term outcomes.

It is important to note that complex analysis of this nature will continually evolve. Consequently, we would expect the precise point estimates to change over time as more data becomes available and the approach is continually refined.

1 Introduction

- I. UC is a major reform of the benefit system designed to reduce poverty by making work pay. This report summarises the latest work undertaken by the Department to evaluate the labour market impacts of the early phases of UC. It extends and updates analysis published in February 2015 (DWP). It uses an improved evaluation design and more recent data to estimate the impact of UC for more claimants, in more offices, over a longer period.
- II. This analysis provides the most reliable evidence to date about the extent to which the investment in UC is delivering its ultimate objective to help move more people into work, and earn more. It suggests that new UC claimants are much more likely to have worked after making a new claim than a matched comparison group of similar people making equivalent claims to JSA during same period in similar areas.

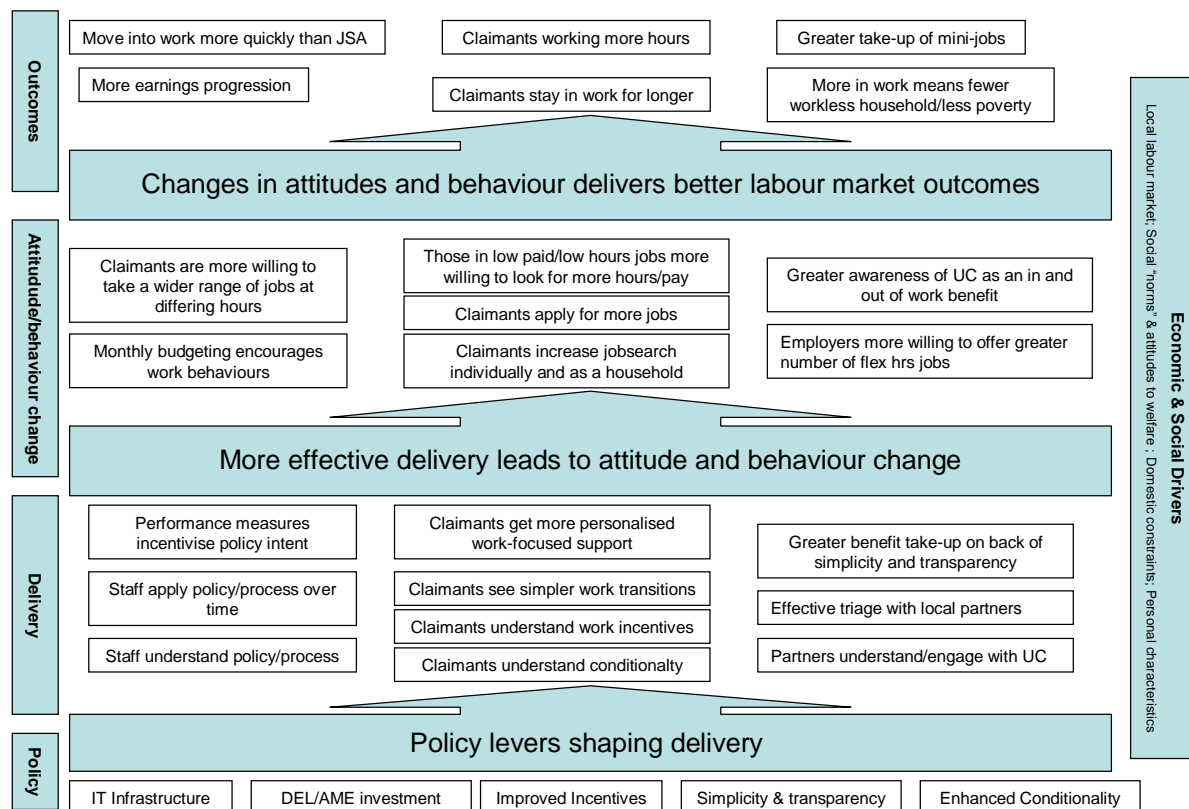
1.1 Policy background

- III. UC is a major reform that simplifies the benefits system by replacing a range of working-age benefits¹ with a single monthly household payment and increases personal responsibility.
- IV. The evaluation is designed using a theory of change approach. Theory of change starts with the assumption that a policy operates in a changing economic and social context, and that the people involved in delivering and experiencing the policy are subject to variable choices and a variable capacity to act. It is the combination of the behavioural context and the policy levers which drive outcomes. Applying theory of change to UC evaluation involves unpacking the underpinning theories behind the policy. *Figure 1* sets out a stylised theory of change showing how UC policy levers are expected to deliver improved labour market outcomes.
- V. The UC Impact Assessment published in December 2012 concluded that up to 300,000 more people are likely to be in work as a result of UC, through improved financial incentives, increased simplicity of the system and increased conditionality (DWP, 2012, p.3).

¹ Income related Jobseeker's Allowance; Income related Employment and Support Allowance; Income Support; Child Tax Credits; Working Tax Credits; Housing Benefit.

Figure 1

Theory of change underpinning improved labour market outcomes



1.2 The roll-out of Universal Credit

- VI. The Department is taking an incremental approach to ensure that UC is introduced safely and securely. Whether an individual can claim UC depends on where they live and their personal circumstances. UC was initially introduced between April and July 2013 in four Pathfinder sites in the North West. These sites were: Ashton-under-Lyne, Wigan, Warrington and Oldham. Progressive roll out of UC began in October 2013 and by spring 2014 ten offices were delivering UC.
- VII. For the new claims included in this analysis, there were a set of gateway conditions so that only certain types of new claims were eligible for UC in those areas where UC had been introduced. In Pathfinder areas these gateway conditions meant that new claims had to be single, not home-owners, without any children and unemployed (i.e. would otherwise have been making a new claim to Jobseekers Allowance (JSA)). Full details of the gateway conditions are in Annex C.

- VIII. From July 2014 the gateway expanded to include couples, and at the same time the singles gateway was extended by dropping some of the original eligibility criteria. From summer 2014 UC started to roll out across the North West of England and eligibility was extended further to include families with children from November 2014. National expansion (for single unemployed claims only) began in February 2015 and will complete by April 2016.
- IX. All the analysis in this report relates to new claims to UC made that met the original eligibility criteria in the Pathfinder sites (either the first four or first ten offices).
- X. The remainder of this report is structured as follows:
- Section 2 – outlines the methodology;
 - Section 3 – summarises the results of earlier analysis;
 - Section 4 – reports the extensive sensitivity analysis carried out on earlier estimates and outlines improvements to the evaluation design;
 - Section 5 – presents the latest estimates using revised methodology and the latest data;
 - Section 6 – conclusions and next steps;
 - Annexes:
 - Annex A: Technical annex on area level differences and comparator offices;
 - Annex B: Independent peer review provided by NIESR;
 - Annex C: List of Gateway Conditions;
 - Bibliography.

2 Evaluation Methodology

This section provides a brief overview of the methodology. More details can be found in earlier publications.

The Department has invited the highest degree of scrutiny and independent peer review to ensure this analysis uses the most appropriate evaluation methods, and to give confidence in the validity and objectivity of the results:

- Evaluation plans were updated and published in July 2014 (DWP, 2014b) following a peer review by the IFS (2014).
- This approach was then applied to produce preliminary estimates, which were peer reviewed by NIESR and published in February 2015. (DWP, 2015).
- NIESR have independently peer reviewed the latest method and estimates (see Annex B).
- This work continues to benefit from the on-going advice and challenge of the UC Expert Advisory Group.

To find out what (if any) difference UC makes to the labour market outcomes of new claimants we have to know what they would have achieved had they claimed the equivalent legacy benefit (in this case, JSA) instead of UC. This is their counterfactual outcome. The counterfactual outcome must be estimated because only actual outcomes under UC can be observed. The counterfactual is estimated using the outcomes of a comparison group of similar people making an equivalent new claim, at the same time, in similar areas, but who claimed JSA because they were in an area that had not yet introduced UC.

For the estimated impact to be reliable the UC ‘treatment group’ (those who actually claim UC) and the JSA ‘comparison group’ (those who would have been eligible for UC had it been available in their area) have to be the same in all relevant respects. The only exception is that one group claims UC rather than JSA. A ‘relevant’ difference is one that affects expected outcomes under the legacy JSA system. The labour market outcomes we expect new JSA claimants to achieve will vary:

- over time,
- between different areas and local labour markets; and
- between different individuals.

The roll-out of UC depends on all of these factors. Consequently, differences in time and/or place and/or individual could bias the impact estimates. Our method focuses on minimising these differences by only selecting new JSA claims made at the same time in similar areas by similar people.

2.1 Time

New UC claims are only compared with equivalent new JSA claims made during the same period. The matching of new UC and JSA claims described below also ensures that the distribution of claims over time is the same between the treated and comparison groups.

2.2 Geography

The comparison group is only selected from areas that are similar to the Pathfinder sites. The idea is that a given type of individual making a new JSA claim at a given time would have the same expected outcomes whether they made their claim in the comparator office or the Pathfinder office.

Comparator offices are offices that have the most similar past employment outcomes for new claims that, as far as we can tell from the administrative data, satisfy the Pathfinder eligibility criteria for UC. We also take into account the size and composition of the new Pathfinder eligible claims each office receives. We focus on Pathfinder eligible new claims because this is the group we are estimating the impact of UC on. It is a relatively small, narrowly defined, and non-representative sub-set of JSA claims. Their labour market outcomes may consequently be very different from those of JSA claimants more generally. Moreover, some offices may do better or worse for this sub-group than they do for JSA claimants as a whole.

The ten most comparable offices are identified for each Pathfinder office in turn. Ten comparator offices are used to help:

- a) Mitigate the risk that the comparator offices might become dissimilar over time;
- b) Increase the pool of individuals available to find matches for the new UC claimants; and
- c) Reflect that no two offices are exactly alike and so any single comparator office would be imperfect.

2.3 Individuals

The approach to ensuring that individuals in the comparison group are the same in all relevant respects as those in the treatment group involves two stages:

1. The first stage restricts the comparison group of new JSA claims to only include those who would have been eligible for UC if they had lived in a UC area. These claimants are ‘apparently eligible’;
2. Then the new UC claimants are matched with these ‘apparently eligible’ new JSA claimants in comparator offices, so that the two groups being compared are the same in terms of all other relevant observables.

We use Propensity Score Matching (PSM) because it:

- a) Overcomes the problem of dimensionality whereby it becomes difficult to identify matches when matching on lots of factors;
- b) Emphasises the importance of, and makes it easy to restrict the analysis to, estimating the impact only on the UC claimants for whom we can identify suitable matches in the non-treated sample; and
- c) Is semi-parametric in that it does not make any restrictive assumptions about how outcomes are determined, i.e. about how the observables affect impacts.

More details of the matching method are in the February report.

2.4 Data

To derive robust impact estimates we need to maximise the sample of UC claimants and the size of the potential comparison group. The bigger the potential comparison group the more chance we have of finding good matches for more of the UC treatment group.

It is prohibitively expensive to conduct a claimant survey of sufficient scale. Therefore, we assembled a wide-ranging evaluation database with data extracted from several DWP administrative systems to get comparable information for all JSA and UC claimants. This data is encrypted for data security and to prevent the identification of individuals, and then made available to analysts with the relevant security permissions via secure data servers. Different data items have then been linked together via encrypted National Insurance numbers, which provide a unique identifier across most DWP claimant data (see *Figure 2*).

Since UC is changing the information collected for benefit claimants and how it is collected the differences in data between the treated and comparison groups is a fourth potential source of bias (in addition to the time, individual and geographical differences discussed above). This is addressed by transforming the data, obtained from the separate UC and legacy systems, and combining it into a single, consistent file.

Whilst this evaluation dataset contains a lot of important information, especially about the past labour market and benefit history of new claimants, it does not contain complete information for all possible variables of interest. For example, not all information on UC eligibility criteria is recorded. Nevertheless, there is evidence from past work that the sort of information held in our administrative data (especially the labour market and benefit histories of new claimants) is sufficient to reliably estimate the employment impact of labour market interventions (Thomas, 2008).

This evaluation dataset is linked to data from the Her Majesty's Revenue and Customs (HMRC) Real-Time Information (RTI) system (more details can be found in the previous version of this report) (DWP, 2015, p.48). RTI requires

employers to submit a range of detailed data to HMRC on or before each payday for each employee.

The RTI contains details of all payments made to employees. It has better coverage of employment than previous systems. People earning below the Lower Earnings Limit² (LEL) are included providing at least one person being paid under the same Pay As You Earn (PAYE) scheme has earned above the LEL in any period in that tax year. RTI can ultimately be used, not just to identify whether people were in work (though that is our starting point), but also to establish how much they were earning, and whether they are increasing their earnings. The RTI does not include information on hours worked, wage rates or earnings from self-employment. RTI data is shared securely between HMRC and DWP and only encrypted non-disclosive data is made available to analysts.

There are a couple of limitations to the current analysis due to the nature of the data available. Firstly, we do not have information about claimants' past earnings or the past earnings in different areas for the claims we focus on here. This reflects that the HMRC's RTI data is relatively new. Consequently, we cannot match claimants or areas on their past earnings. Therefore, estimated impacts on earnings are less robust than the estimated impact on employment status. When we have data on new claims made later in the roll-out of UC we will be able to better control for past earnings at both an area and individual level to produce more reliable estimates of the impact of UC on earnings (DWP, 2015, p.97). In the meantime the estimated impact on earnings should be treated as provisional.

Secondly, the impact estimates represent the impact of UC on employment covered by the RTI. If UC has a differential impact on employment that RTI does not capture this would not be picked up. For example, it is not possible tell from this impact evaluation whether UC might lead to more (or less) self-employment. Similarly, if UC makes employers more likely to report small amounts of earnings then the impact estimates would be biased upwards. These issues are explored through other strands of the evaluation.

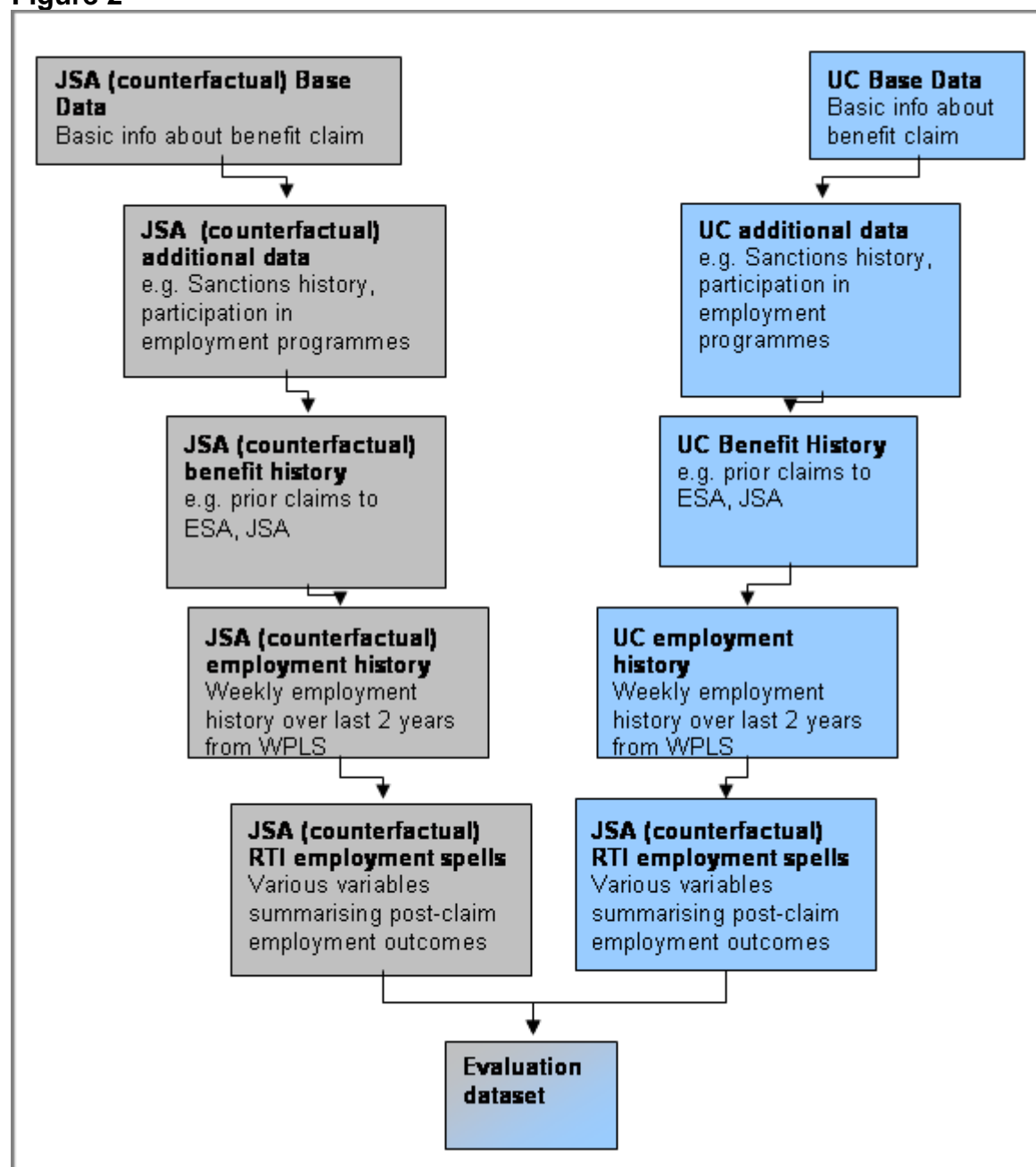
Another possible limitation of this type of analysis not due to data considerations is that entry or anticipation effects could compromise the reliability of the impact estimates. As discussed in "DWP, 2014b" entry and anticipation effects are unlikely to bias our current estimates because during the early phase of UC that this report focuses on:

- a) No-one making new claims to UC during this time would have had the opportunity to enter UC earlier via another benefit type or eligibility route because we focus exclusively on single new claims;
- b) The eligibility and entitlement criteria under both benefit systems will be similar for the types of new claims replaced by UC during Pathfinder;

² Threshold below which no National Insurance contributions are paid.

- c) Many people will not be aware of UC or familiar enough with it for it to produce entry effects from behavioural changes;
- d) And we can select comparison samples far enough away from where UC is introduced.

Figure 2



3 Summary of Preliminary Results (published in February 2015)

Using the methodology outlined above, *Figures 3 and 4* show the outcomes for new UC claims made between July 2013 and April 2014 in the first four Pathfinder offices, compared with matched Pathfinder eligible new JSA claims made during the same period in the non-UC comparator offices. *Figure 3* shows the estimated impact on employment status at various snapshot points in time after the claim was made (30, 60, 90 and 120 days). This shows our estimate of the proportion of claimants employed on a particular day. *Figure 4* shows the cumulative employment impact at 30 day intervals. For example, it shows the impact on the likelihood of having worked at any point within the first 60 days of making a claim or within the first 90 or 120 days. These preliminary results suggested that new UC claimants were more likely than a matched comparison group of similar JSA claimants to be in work at different points in time after the start of their claim. In particular, UC claimants were:

- 5 percentage points more likely to have some work during the first 120 days after they made their claim; and
- 3-4 percentage points more likely to be employed at 30, 60, 90, and/or 120 days after the start of their claim. (DWP, 2014b)

Figure 3

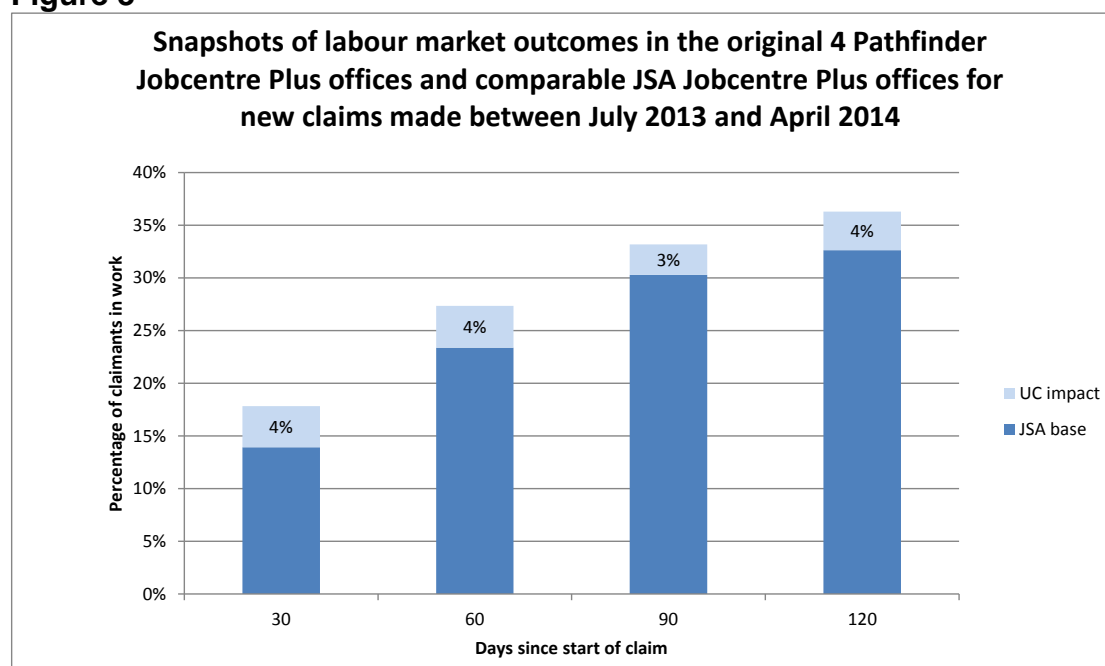
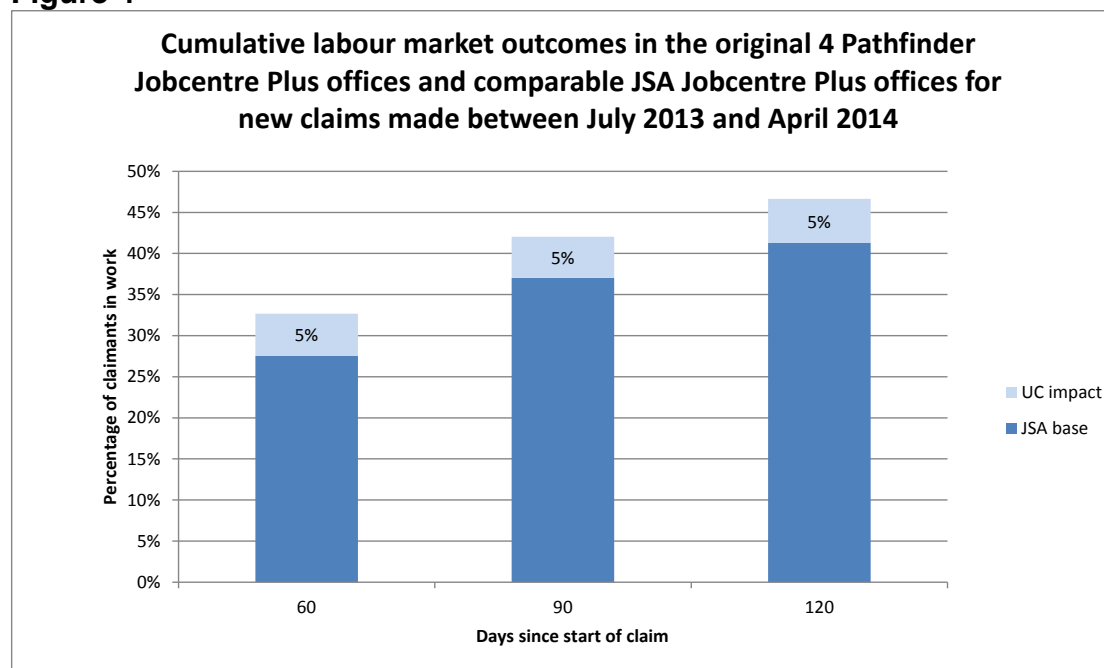


Figure 4



RTI data (supplied by HMRC) was used to derive additional outcome measures to explore in more depth the impact that UC was having on people's labour market outcomes. Using the same matching method we estimated the impact of UC on days employed and earnings since the start of claim. This analysis suggested that during the first four months or 120 days after making a new claim, UC claimants, on average, spent an extra 4 days in work and earned around £50 more in gross earnings than matched new JSA claims. It is not possible to estimate the impact of UC on hours worked or hourly wage rates as this information is not contained in the RTI.

It is important to recognise that these estimates are less robust than the estimated impacts on employment status described above. As noted in Section 2, the evaluation approach controls for past employment histories, eligibility etc. but with existing data it cannot control for past earnings. Consequently, the estimated impacts on earnings are reliable only in so far as past earnings and past employment are correlated. The estimated impact on days worked is less reliable because there are several assumptions required to derive employment start and end dates from the earnings information provided by RTI (see "DWP, 2014b" for details). We estimate days in work crudely by taking the difference between the estimated start and end dates. Given the number of assumptions required and the lack of information about hours worked the estimates of how many days people have worked is only a rough approximation.

4 Sensitivity Analysis

The preliminary impact evaluation published in February 2015 included a range of sensitivity analysis. It found that the estimates were robust to a range of different methods and approaches. Since February 2015 additional data sources have been used to conduct an extensive programme of further sensitivity analysis. The focus of this work has been to explore how well the data and method used controls for potential non-UC differences between the individuals and areas in the treatment and comparison groups.

In other words the analysis has focused on the scope and options for improving the internal validity of the preliminary estimates by improving the method and/or data. This work was conducted using an updated evaluation dataset that included more recent HMRC RTI data and DWP administrative data. These data sources are subject to lags and so tend to become more complete and accurate over time. That is, we might expect the exact magnitude of some impact estimates to change over time simply because some of the data will become more complete. These changes should be small because the analysis does focus on data that already appears to be relatively stable over time, i.e. relatively complete. Some sensitivity analysis has also been done using historical data. For the purposes of this sensitivity analysis we focused on the same new claims that the February report considered, i.e. those made between July 2013 and April 2014 in the original four offices.

The sensitivity analysis carried out since the preliminary estimates were published. The results and implications of this work are summarised below. The analysis has addressed the issues highlighted by NIESR's review of the preliminary analysis. It has also investigated a number of additional issues. Annex A provides more detail.

The original method collectively compared new UC claims in all 4 Pathfinder offices with equivalent new JSA claims in the 40 non-UC offices, which had the most similar historical JSA off-flow rates for Pathfinder eligible claims. The area-level sensitivity analysis focused on the following three issues:

- a. Areas that were similar to the Pathfinder areas in the past may become dissimilar in the future if there are differential trends between them. To investigate this we compared the outcomes in Pathfinder and non-UC comparator areas of new claims unaffected by the introduction of UC. The outcomes of ineligible claimants were marginally better in the Pathfinder areas (see *Figure Av* in Annex A). This is consistent with the existence of differential macro trends favouring the Pathfinder areas. However, *Figure Av* also shows that these ineligible claims had better outcomes in the UC areas *before* UC was introduced. Therefore, these results appear to reflect that we identify the non-UC comparator offices in terms of their past performance for Pathfinder eligible claims and these offices may not be similar to

the UC offices for the ineligible claimants. We will continue to test for differential macro-trends when we look at different areas and time periods.

- b. Offices with similar off-flow rates from JSA may not have similar employment outcomes. We identified comparator offices using new data on historical employment outcomes. This work led to different and more appropriate comparator offices being identified. However, this had little effect on the overall results.
- c. Combining all offices means that some UC claims could be compared with new JSA claims from different types of area. To address this we estimated impacts separately by office. The average impact was largely unaffected although the estimates for individual offices did change.

The original method assumed that someone is employed continuously if there is a short break in their earnings of less than four weeks. However if UC claimants are more likely (less likely) to have short periods of employment this may over (under) estimate their outcomes relative to JSA claimants. When we define a claimant as being in work in a given week only if there is evidence of earnings for that particular week the impact estimates on employment status are marginally lower.

The February analysis only included new UC claims that received a first payment. The comparison group of Pathfinder eligible new JSA claims included all new claims except those with a duration of less than 3 days. We explored the extent to which the definition of a new claim might lead to differences in the types of claims included in the UC and JSA samples. For example, the February analysis might have excluded claims from the UC sample that were included in the JSA comparison group because it can take several weeks to receive a first payment under UC. The results were insensitive to how we defined new UC and JSA claims (see Annex A).

NIESR's review highlighted that our inability to identify Pathfinder eligibility accurately was a cause for concern and worthy of further investigation. Because of this the JSA comparison group will include people who are not the same as the UC claimants. We have investigated the extent to which matching on lots of relevant observables might address this issue. In particular, we have estimated impacts by matching various different sub-groups that can be identified in the data and by artificially creating new groups. The details and results are in Annex A. The results suggest that our matching approach is likely to be effective in controlling for relevant differences between the UC treatment and JSA comparison groups. Nevertheless, we will continue to conduct sensitivity analysis to explore the plausibility of the conditional independence assumption for each phase for the impact evaluation and we will continue to seek ways to improve how we identify UC eligibility. We will also assess the scope for, and value of, using other data sources to further investigate the likely importance and prevalence of factors we cannot identify in the evaluation datasets and the extent to which they might be captured indirectly by the data we do have.

Initially employment outcomes were measured from the date the claim was paid for JSA claims and the date the claim was submitted for UC claims. We have explored the comparability of the different dates available in the UC and JSA data and concluded that these dates we used in February represented the first contact date in the vast majority of cases for both groups. Consequently, these dates are the most appropriate to use. We will continue to explore the scope for conducting sensitivity analysis around the small number of claims where these dates may not represent the first contact date.

The preliminary analysis matched on 300+ observables including weekly employment status during 2 years prior to claim. Since then the accuracy of employment histories has been improved using more frequent and complete information about benefit status from DWP administrative sources. This led to a slight increase in the impact estimate because the unmatched JSA sample has slightly better employment histories than the UC treatment group.

Overall the results of this extensive sensitivity analysis suggest that the impact estimates published in February are robust. That is, all of the sensitivity analysis suggests that UC does lead to claimants being much more likely to have some work after making a claim than they would have had if they had remained under the legacy system. It also led to the following changes in the methodology to increase the internal validity of estimates:

- i) New comparator areas are used with historically similar employment outcomes for apparently eligible new claims rather than those with similar JSA off-flow rates. The identification of comparator offices still takes into account the size and composition of their apparently eligible new claims.
- ii) Estimate impacts separately by office to better control for area-level differences between UC and JSA claimants.
- iii) Made some modest improvements to how we identify Pathfinder eligibility.
- iv) Improved consistency between the new claims included in the treatment and comparison groups by including all new UC claims that sign a Claimant Commitment (i.e. not just those who receive a payment) and all new JSA claims that either sign a CC or attend their initial jobseeker interview.
- v) Focus on outcome measures that are less susceptible to the assumptions and rules we have to make to estimate employment spells from the earnings information contained in the RTI.

Our original estimate was that new UC claimants were 5 percentage points more likely to work within 120 days of making a new claim than a matched comparison group of new JSA claimants. After making these changes to the method, and using the updated evaluation data, this impact estimate rises to 8 percentage points.

5 Revised Estimates

5.1 New data

The revised analysis uses new HMRC RTI data covering cohorts starting from July 2013 up to February 2015. As before, this data is combined with various DWP data sources to produce a single consistent evaluation dataset. This dataset allows us to estimate the impact of UC on more claimants over a longer period and to extend the analysis to measure the impact on claims made in all ten Pathfinder offices (previously the analysis was restricted to the original four offices).

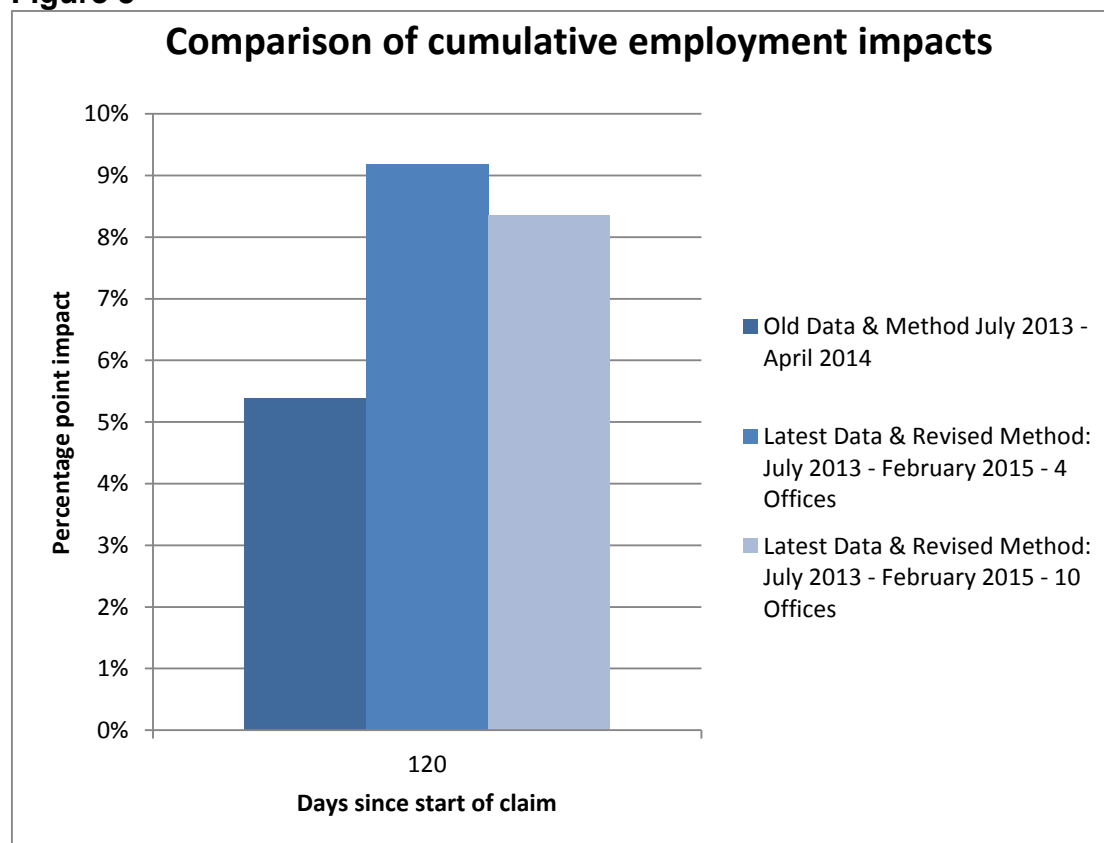
Ideally we want to measure outcomes for as many claimants as possible, for as long as possible. However, in deciding which cohort of new claims to focus on there is a trade-off between sample size and the length of time we can estimate impacts over. In this report, the main focus is on new claims made between July 2013 and September 2014 in all ten Pathfinder offices. For these claims employment outcomes can be tracked for 270 days (nine months). However, we also look at the:

- a) short-term impact (up to 120 days) on all claims made up to February 2015 in the first ten offices; and
- b) longer-term outcomes for claims made in the original 4 offices between July 2013 and July 2014 (up to 11 months after the start of the claim).

All of the analysis relates only to the types of claims that were first eligible for UC, namely new claims by single unemployed people with no children, no housing costs and who met a range of other eligibility criteria (see Annex C).

5.2 Revised estimates for short-term employment outcomes

Figure 5 shows the percentage point increase in the chances of being employed within 120 days of making a claim. It compares the results reported in February 2015 for around 6,000 new claims made in the original four offices between July 2013 and April 2014 with the results using the refined method and more recent data in the first four offices and also in the first ten offices.

Figure 5

These results confirm that using a refined method and more up-to-date data on more claims, UC has a positive and statistically significant impact on the probability of working after making a new claim. Our latest results suggest the size of the effect is larger than initially estimated – at around 8 percentage points. This compares with 5 percentage points reported previously. This change is partly due to changes in the method and partly due to the data being updated. However, the precise impact estimate is expected to change as the analysis continues to evolve. It is also important to note that the impact is on the probability of being employed at any time during the period. It does not reflect how much time people were employed either in terms of days worked or the duration of employment.

5.3 Tracking outcomes for longer

A key development since the preliminary estimates were published is that the more recent RTI data allows us to track employment outcomes for longer.

Figure 6 shows the percentage of JSA claimants that we estimate are employed at different points in time after making their claim, and the positive differential for UC claimants. These estimates relate to over 8,000 new claims made in the first ten offices between July 2013 and September 2014. At every point in time UC claims are more likely to be in work than matched new JSA claims made in similar areas during the same period. For example, 51% of UC claimants are in work nine months after their claim started compared with 48% of matched JSA claimants. It is important to note that those employed 270

days after making their claim may be different to those who were employed at earlier points in time. The larger cumulative estimates relative to the snapshots suggests that the increase in employment chances with UC may be due to UC making it more worthwhile and easier for people to accept short-term temporary work.

Figure 6

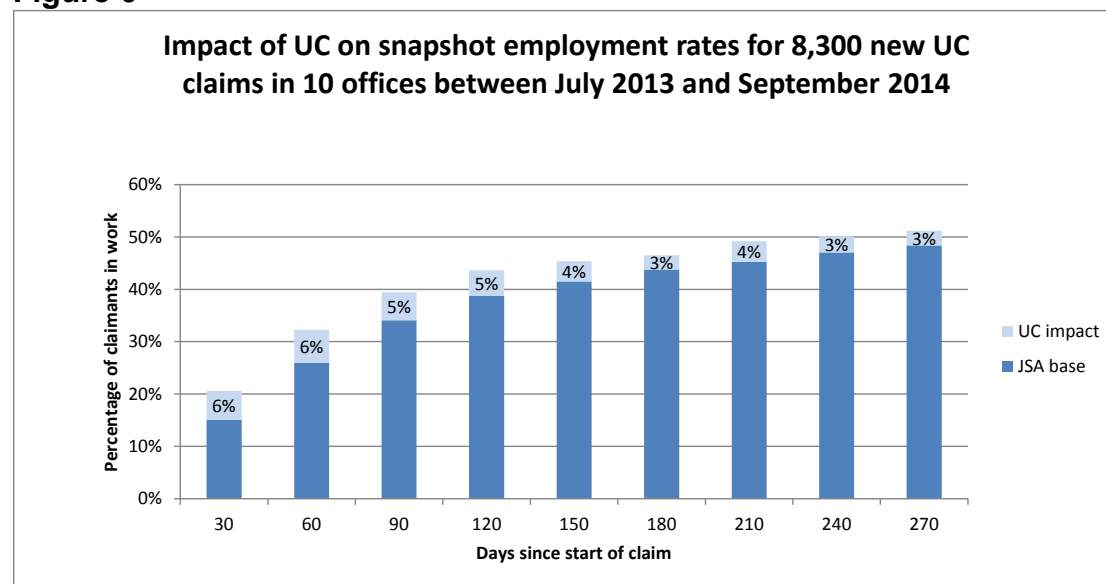


Figure 7

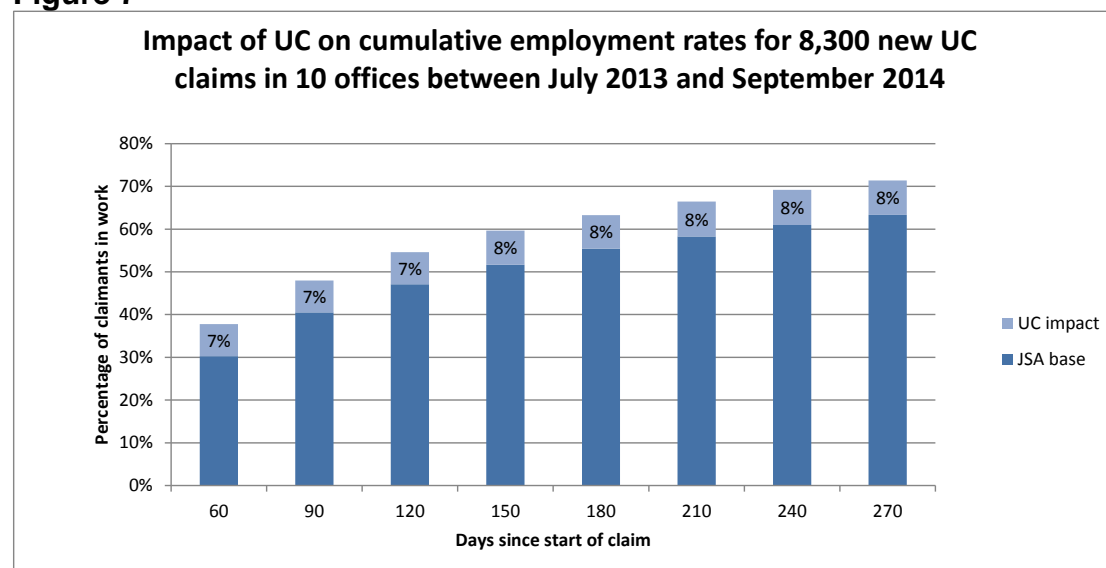


Figure 7 shows the impact of UC on the chances of having any paid work within different periods of time after making a new claim. For example, it shows that UC claimants are 8 percentage points more likely to have worked within 270 days of making their claim than matched JSA claimants (71% compared with 63%).

Figure 8

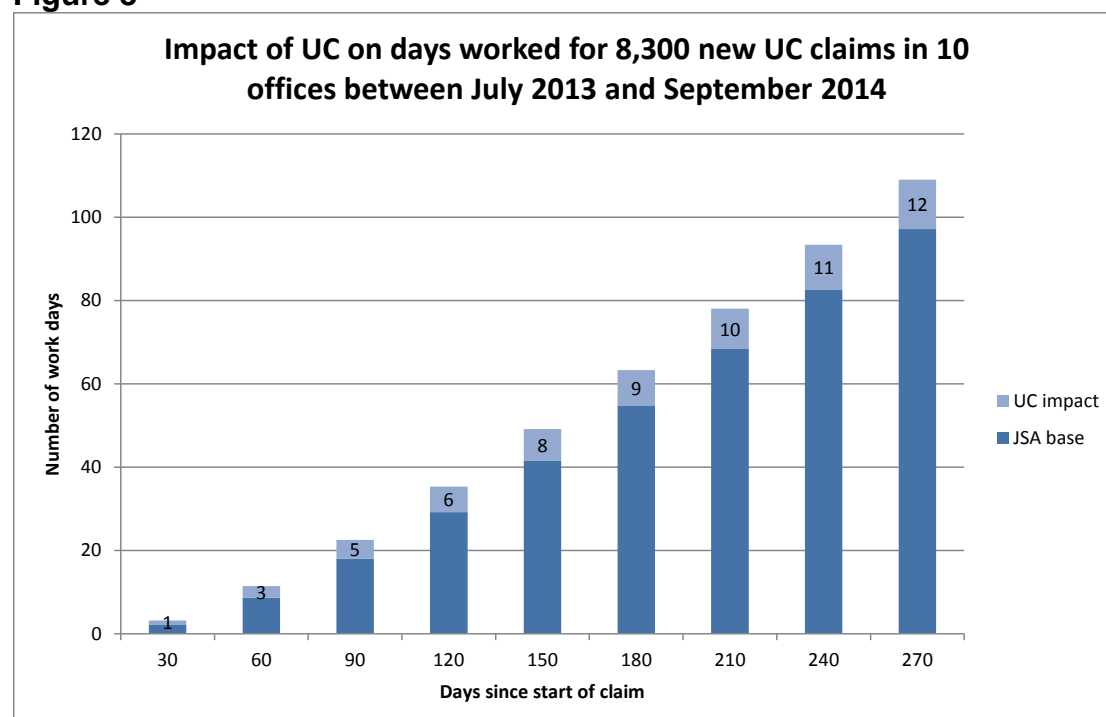


Figure 8 shows the impact on days worked since the beginning of the claim. On average UC claimants work about 12 days more, in the first nine months after making a new claim than matched JSA claimants (109 days compared with 97 days). This outcome measure, for both groups, is only a rough approximation because of the assumptions needed to derive employment start and end dates from the detailed earnings information available in the RTI.³

5.4 Comparison of outcomes for different cohorts, demographic groups and analysis of matching

Figure 9 shows for the snapshot estimates that the positive UC differential holds for different cohorts. Also, for the smaller sample of new claims made between July 2013 and July 2014 in the first four Pathfinder sites we can see that the positive impact is still sustained at 330 days (eleven months after claim start).

³ For example, this method may show that a period of employment started on the 1st of March and ended on the 31st of March, but it does not necessarily follow that the individual was actually working every single day for a continuous 31 days.

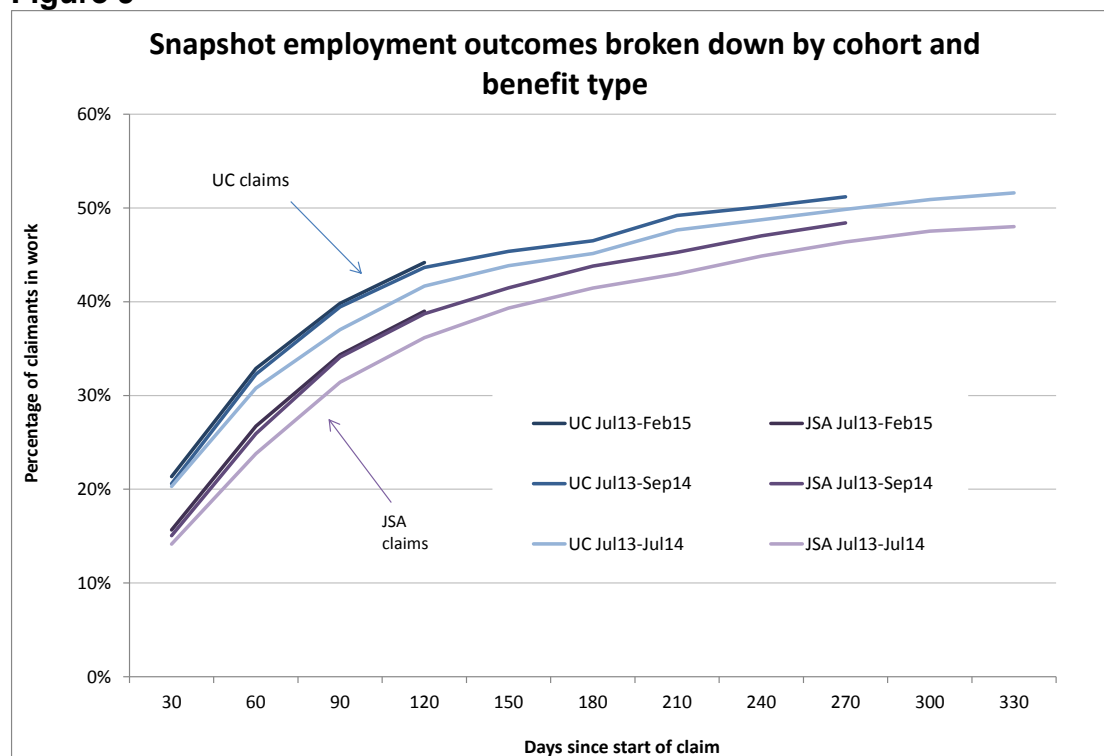
Figure 9

Figure 10 shows the estimated impact of UC on weekly employment status. The results of the matching showed that there were no statistically significant differences in the employment status of the UC treatment group and the JSA comparison group in any individual week prior to the start of their claim. However, whilst none of the individual differences were statistically significant the UC group consistently had a slightly higher probability of being employed prior to making their new claim. This difference between the UC and JSA groups whilst very small looks reasonably stable over time. It suggests that the matching may not have completely controlled for all the factors that affect the chances of being employed between the treatment and comparison groups.

Since this small pre-claim difference is stable over time it suggests that any factors that we might not be capturing in the matching model are not changing differentially over time between the two groups. Therefore, we can deduct the average pre-treatment difference from the estimated differences both pre- and post-treatment to eliminate these time constant unobservable differences.

Figure 10 defines an individual as being employed in a given week if they have evidence of earnings in that particular week. It shows that for new claims made in the original 4 Pathfinder offices between July 2013 and July 2014 there was no difference in the probability of being employed prior to making their claim (after differencing out the small pre-treatment difference). After UC there is a statistically significant difference between the UC and JSA groups. This impact is significant throughout the whole tracking period.

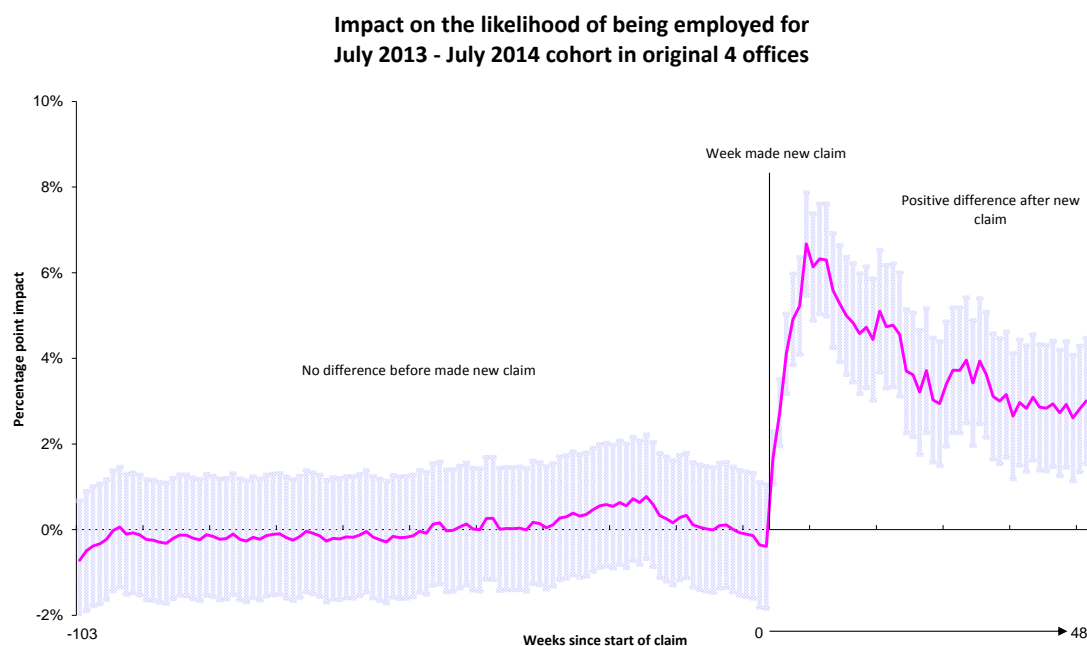
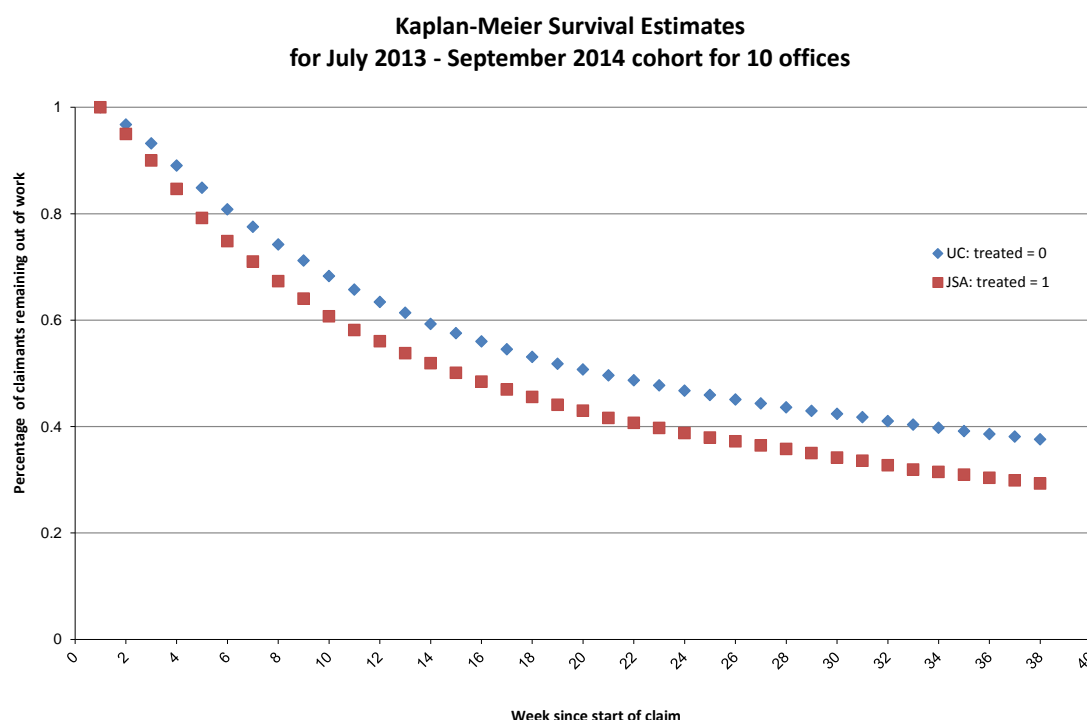
Figure 10

Figure 10 focuses on earlier claims to provide a longer follow-up period. Similar results are found when we look at the weekly employment status of new claims made in the first ten offices between July 2013 and September 2014. As expected, the estimated impact falls over time. This reflects that the proportion of JSA claimants in work, at any given snapshot, rises over time anyway (see *Figure 7*). Consequently, the people who are still not in work several months after claiming are more likely to be further away from the labour market and require additional support to move into work.

Figure 11 plots the Kaplan-Meier survival graphs for the matched UC and JSA samples from the July 2013 – September 2014 new claims. The graphs plot the fraction of each group that have not started employment as time elapses (equivalently 1 minus the fraction shown gives the proportion who have become employed since making their initial claim). The chart shows that UC claimants move into work more quickly than matched JSA claimants in similar areas made during the same period. For example, by week 25, 62% of UC claimants had worked compared with 46% of JSA claimants. The difference between the survival curves is statistically significant. Further work will extend this analysis to consider the duration of these first employment spells.

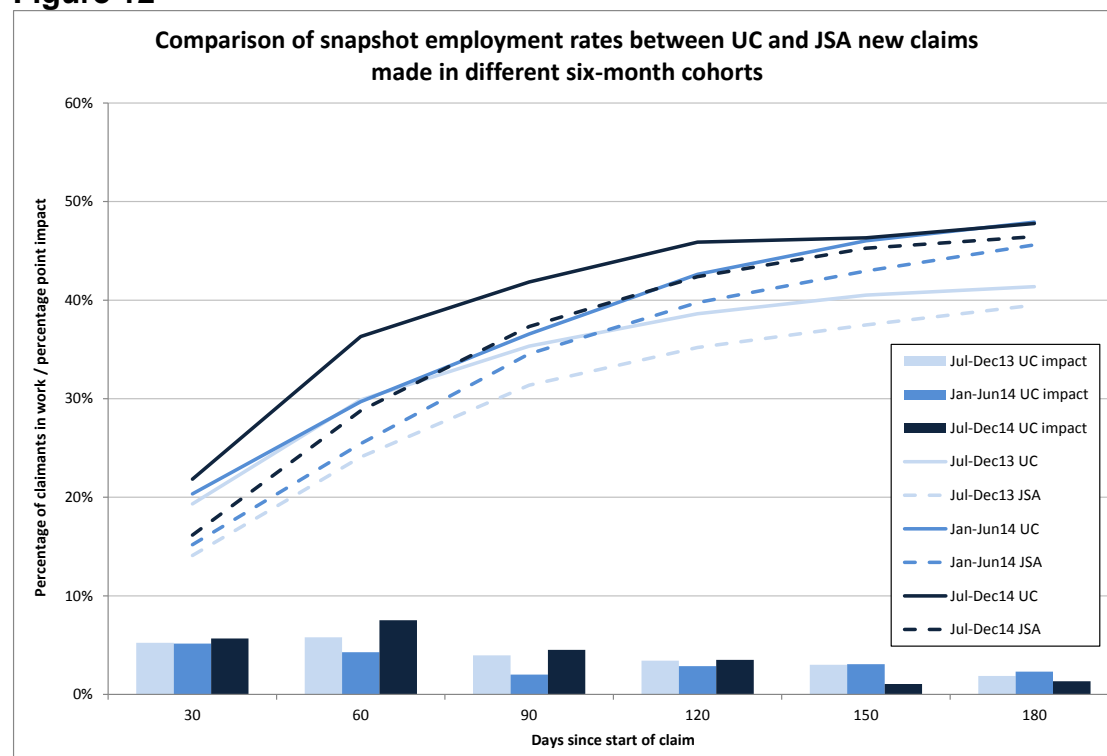
Figure 11



Given the available sample size and reasonably narrow characteristics of the treatment group, the opportunities for sub-group analysis are limited. Nevertheless, it is important to understand whether there are any differences in the size of the UC impact over time and for different groups. With the data available it is possible to look at results for claims made in different 6 month cohorts, by gender, and by age.

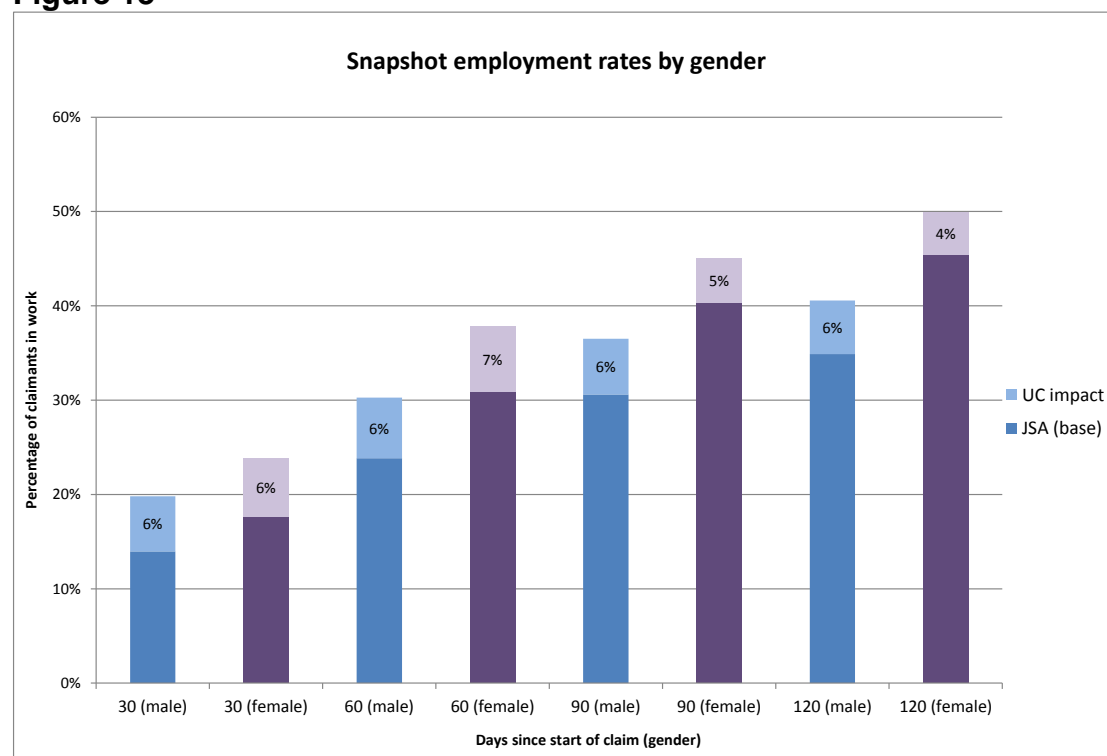
Figure 12 compares employment rates for new UC and JSA claims made in the first four Pathfinder offices in three different six month cohorts: July to December 2013, January to June 2014, and July to December 2014. This shows that whilst the precise employment rates are different for claims taken at different points in time (as we would expect given seasonality and wider labour market trends), the percentage point impact is not significantly different between cohorts. This is encouraging as it shows that the positive effect of UC on employment is so far sustained over time. As more data becomes available we will be able to monitor this effect for longer.

Figure 12



It is also possible to look at the impact by gender. *Figure 13* compares employment rates for a sample of 6,400 male and 3,800 female UC claims made in the ten Pathfinder offices between July 2013 and February 2015.

Figure 13



This shows that at each 30 day interval there is no significant difference in the impact of UC compared to JSA between the male and female sample.

Similarly, *Figure 14* shows there is no significant difference when looking at the impact of UC separately by age group (looking at nearly 7,000 UC claims made by under 25 year olds and over 3,500 UC claims made by over 25 year olds in the ten Pathfinder offices between July 2013 to February 2015).

Figure 14

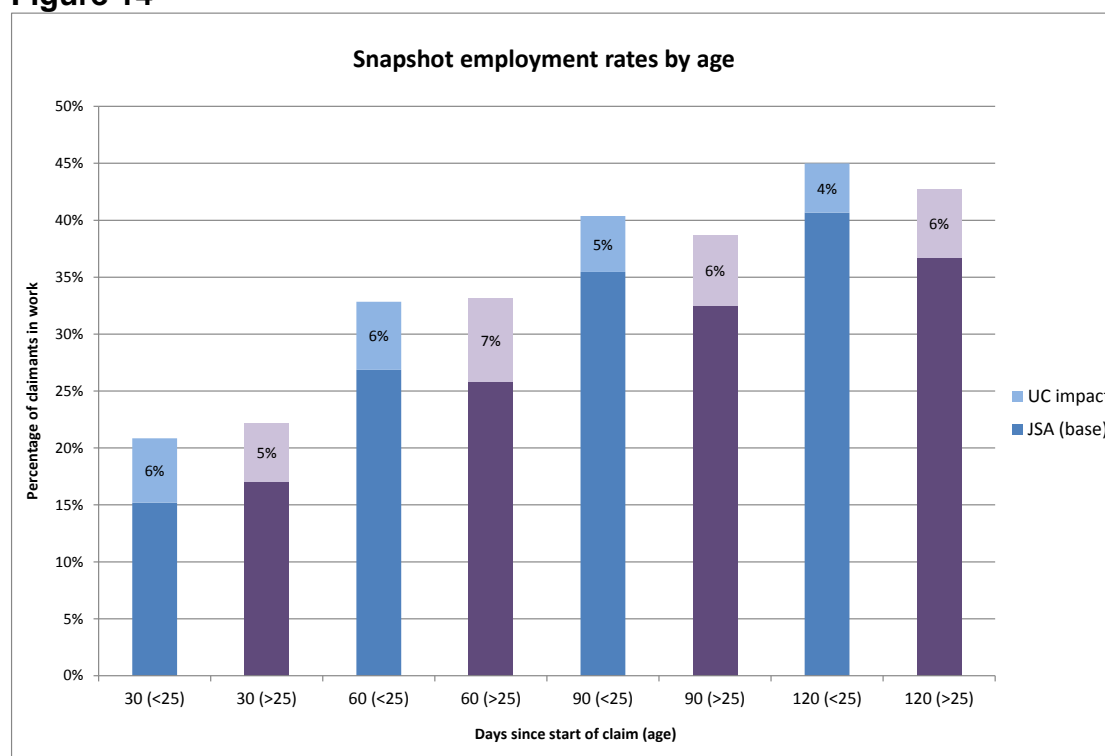


Figure 15 summarises the short term employment outcomes for different cohorts, type of statistics used and time frames.

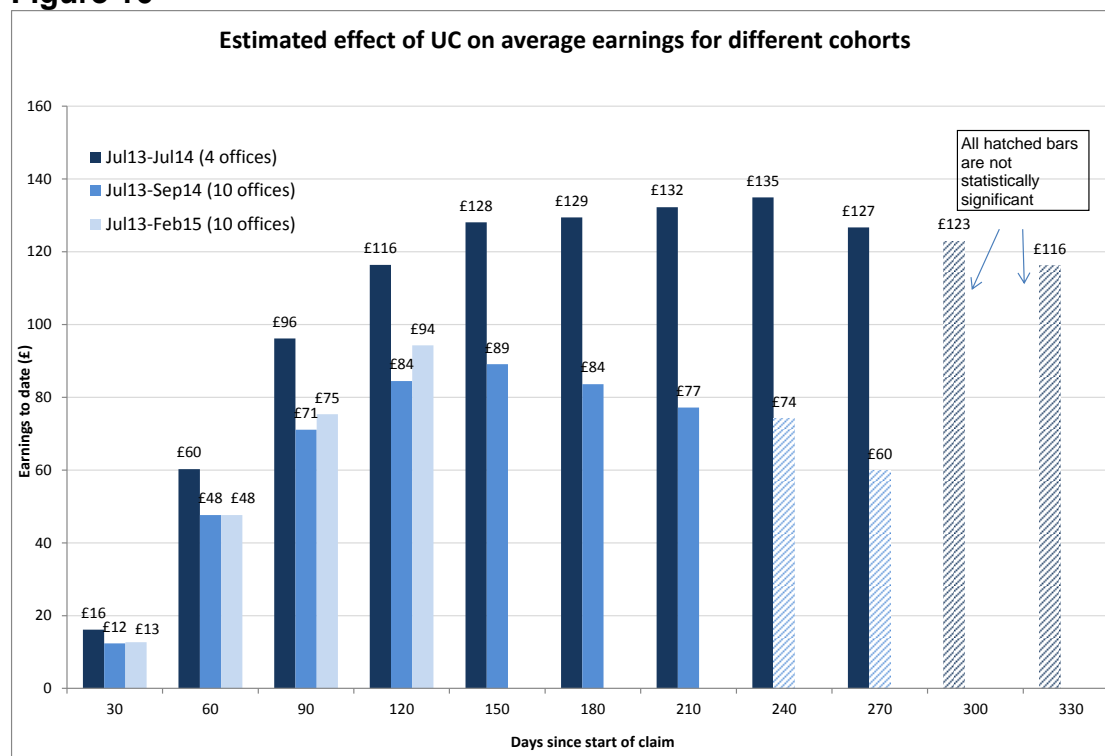
5.5 Impacts on Earnings

As with earlier analysis, we have used the RTI data to derive additional outcome measures to explore in more depth the impact that UC is having on labour market outcomes. Using the same matching method we can estimate the impact of UC on earnings since the new claim was made. However, for the reasons highlighted in Section 2 these estimates are more uncertain (subject to very wide confidence intervals due to the variability in earnings) and are likely to be less reliable because of the matching method. Consequently, these estimates are less reliable than the estimated impacts on employment status.

Figure 15

	Cohort							
	July 2013 - April 2014 (Old Data & Method)		July 2013 - July 2013		July 2013 - September 2014		July 2013 - February 2015	
Number of offices	4	10	4	10	4	10	4	10
Snapshot employment impact at (%):								
30 days	4%	-	6%	-	6%	6%	6%	6%
60 days	4%	-	7%	-	7%	6%	7%	6%
90 days	3%	-	6%	-	5%	5%	6%	5%
120 days	4%	-	6%	-	5%	5%	6%	5%
150 days	-	-	5%	-	4%	4%	-	-
180 days	-	-	4%	-	3%	3%	-	-
210 days	-	-	5%	-	4%	4%	-	-
240 days	-	-	4%	-	3%	3%	-	-
270 days	-	-	3%	-	3%	3%	-	-
300 days	-	-	3%	-	-	-	-	-
330 days	-	-	4%	-	-	-	-	-
Cumulative employment impact within first (%):								
60 days	5%	-	-	-	8%	7%	9%	8%
90 days	5%	-	-	-	8%	7%	9%	8%
120 days	5%	-	9%	-	8%	7%	9%	8%
150 days	-	-	-	-	8%	8%	-	-
180 days	-	-	-	-	8%	8%	-	-
210 days	-	-	-	-	9%	8%	-	-
240 days	-	-	-	-	9%	8%	-	-
270 days	-	-	-	-	9%	8%	-	-
300 days	-	-	-	-	-	-	-	-
330 days	-	-	-	-	-	-	-	-

Figure 16 shows the estimated effect of UC on average earnings, for different cohorts of new claims made: in the first four offices between July 2013 and July 2014; in all ten Pathfinder offices between July 2013 and September 14; and all ten Pathfinder offices between July 2013 and February 2015. It suggests that there is a modest positive impact of UC on earnings. However, this effect is not always statistically significant and there are wide confidence intervals around these estimates. For example, for our central estimate new claims earn around £80 more than matched JSA claims during the first six months, but the approximate confidence interval is from £20 to £150. Given this variability, the estimated impact on earnings must be treated with caution. In future, it will be possible to match claims and areas on the basis of earnings using historical RTI, which should produce more robust estimates.

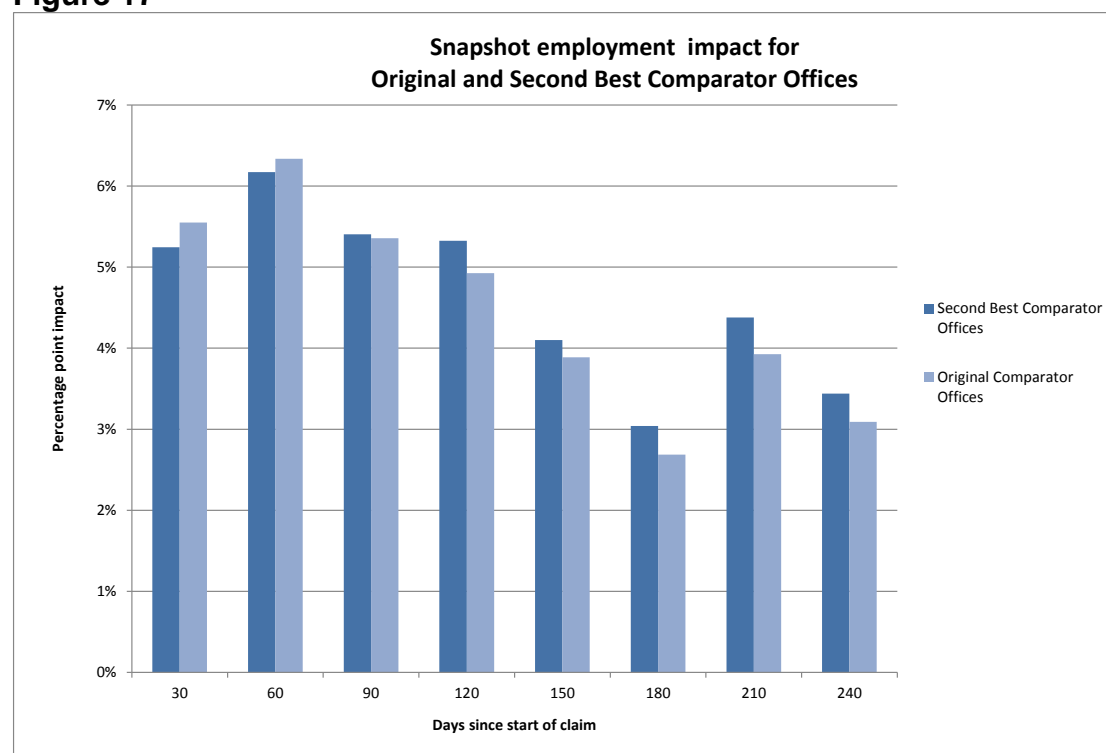
Figure 16

Whilst the estimated impacts on earnings and days worked are less reliable they do suggest that the relatively big and sustained impact of UC on the chances of being employed may well reflect that UC encourages claimants, and makes it more worthwhile and easier for them, to do small amounts of paid work. That is, the relatively modest impact on earnings suggests that whilst UC claimants are more likely to have some work it appears that this additional work probably involves relatively few hours at relatively low wages. Moreover, the small snapshot impact estimates suggest that UC might be making it more worthwhile and easier for claimants to take-up short-term work. However, further analysis is required before we can be confident that we understand the impact of UC on earnings and the duration of employment.

5.6 Sensitivity Analysis for latest data

Most of the sensitivity analysis undertaken since February has focused on the validity of the preliminary estimates for the new claims made in the original 4 Pathfinder offices between July 2013 and April 2014. However, we have replicated key aspects of the sensitivity analysis using the latest data and looking at more claims in more offices. For example, we re-estimated impacts using the next best ten comparator offices for each UC office. *Figure 17* shows that this has no significant impact on the estimates.

Figure 17



We have also explored whether there is any evidence of differential trends by estimating the impact on the JSA residual group⁴ and ineligible claims over similar time periods before and after the introduction of UC. This found no evidence of differential trends. In particular, the outcomes of other groups did not improve in the UC offices relative to the non-UC comparator offices after UC was introduced. Annex A provides details.

We also carried out detailed sensitivity analysis to explore the potential implications of only being able to partially identify Pathfinder eligibility for JSA claimants. Again, Annex A provides details. The results from estimating impacts on different sub-groups, we can identify in the data, suggest our central estimates are robust. For example, when we estimate the impact on all apparently eligible claims in the UC offices (including those that continue to claim JSA – the JSA residual) the matching diagnostics improve and the estimated impact remains significant and positive. As expected the actual impact estimate is smaller. This reflects that only some of the apparently eligible group in the UC offices in this analysis actually receive the treatment, i.e. claim UC. However, the size of this estimated impact is consistent with the central estimate when we take into account the proportion of apparently eligible claims in the UC offices that actually claim UC.

⁴ i.e. group that falls into Apparently Pathfinder Eligible (APE) group where new claimants that still claimed JSA after UC had been introduced for Pathfinder eligible claimants.

6 Conclusions and Next Steps

This report represents an important development in building robust evidence on the impact of UC on labour market outcomes. It updates and extends the preliminary impact report published in February 2015. It uses a refined method, more up-to-date data and looks at the impact on more claims in more areas over a longer period.

The results are encouraging and in line with the positive impact that we would expect to see given the changes in financial incentives, increased simplicity, and additional conditionality. The central estimate is that UC claimants are 8 percentage points more likely to have worked within the first 270 days of making their claim than a matched comparison group of JSA claimants who made equivalent claims during the same period in similar areas. The chances of being employed at particular points in time tend to be smaller which might indicate that some of the increase in the likelihood of becoming employed is due to UC leading to an increase in short-term work. These positive results hold for claims made at different six month intervals, for different age groups, and for both men and women. An extensive range of sensitivity analysis suggests that the evaluation method and results are robust and as reliable as they can be given the data available.

On average UC claimants are estimated to work about 12 days more than the matched comparison group during the first 270 days after they made their claim. We also find evidence of a positive impact on earnings. However, the variability in earnings makes these estimates more uncertain. Furthermore, the estimated impacts on days worked and earnings are both less reliable than the impact on employment status due to the data available.

Overall, these results suggest that UC is encouraging claimants, and making it more worthwhile and easier for them, to do small amounts of work.

Whilst these results are encouraging, they are still early estimates. It must also be remembered that with the available data it is only possible to estimate employment effects for a relatively narrow group of single unemployed claimants, who are not representative of the UC population as a whole⁵.

There is much more to do as UC continues to rollout with national expansion for single unemployed claims due to complete by April 2016, a growing caseload of family claims and the start of national rollout of the full digital service to all types of legacy claims from May 2016.

⁵ Though it is important to note that for other groups the expected impact of UC will be greater because of higher relative effects on effective marginal deduction rates and participation tax rates (e.g. for lone parents working less than 16 hours per week); or more significant changes in the conditionality regime (e.g. for ESA-type claims before the Work Capability Assessment, partners, and in-work claimants).

We will continue to update the UC evaluation dataset as more data becomes available, and will extend the analysis to include more claimants, in more offices, new groups of claimants, and for longer periods of time. We will also do more to improve estimate and understand the impact of UC on earnings.

As with any complex analysis of this nature, the precise point estimates should be expected to change over time – the key finding is that UC is having the intended effect of increasing movement into work.

Annex A: Sensitivity Analysis to Assess the Internal Validity of the February 2015 Impact Estimates

This annex summarises additional sensitivity analysis conducted since February to further test the internal validity of the preliminary impact estimates published in February 2015. Whilst we conducted a range of sensitivity analysis as part of the February analysis we have since been able to exploit more data sources and consider more aspects of the methodology and data. Much of this work has focused on issues raised in NIESR's review of the February analysis. It is in three main parts covering:

- i) Controlling for geographical differences and identifying comparator offices;
- ii) Controlling for individual level differences and assessing Pathfinder eligibility; and
- iii) Assessing the accuracy and consistency of the evaluation data.

Part 1 Geography – Controlling for Area Level Differences and Identifying and Using Comparator Offices

The outcomes we expect new JSA claimants to achieve vary depending on where they are because of differences between: local labour markets; labour market policies; how well policies are delivered, etc. We know that the likelihood of claiming UC rather than JSA depends on geography because of the geographical roll-out. This makes geography a potential source of bias, i.e. because it affects both expected outcomes and the likelihood of treatment.

DWP (2015) describes in detail the approach used to identify comparator areas that are similar as possible to the Pathfinder offices. In summary, we identify those offices that have historically similar outcomes for apparently Pathfinder eligible new claims to JSA taking into account the size and composition of their monthly on-flows.

Identifying comparator offices is difficult. No two offices have the same outcomes over time or the same volume and mix of claimants, i.e. no offices are identical. The outcomes within offices vary hugely over time. This reflects a range of factors including changing economic conditions, changes in the size and composition of claimants, etc. The best we can do is to identify non-

UC comparator offices that are most similar to the Pathfinder offices. Even amongst the most alike offices outcomes can diverge significantly during some periods. It is partly because of this volatility that we identify and use the ten most comparable offices for each Pathfinder office. Using ten comparator offices for each Pathfinder office also helps:

- a) Mitigate the risk that the comparator offices might become dissimilar over time. That is, whilst one or two offices might start to diverge from their respective Pathfinder office it is much less likely that all ten will diverge and in the same way at the same time.
- b) Increase the pool of individuals that we can use to find matches for the new UC claimants.

In our original sensitivity analysis we re-estimated impacts using alternative comparator offices. This helps guard against changes in our chosen comparator offices leading to bias. However, it does not mitigate the risk that non-UC changes in one of the four Pathfinder offices might bias the results. This risk is relatively high given that until recently we have been constrained to looking only at the original 4 offices. This reflects that for the February 2015 report we only had long enough to track the outcomes of enough UC claimants in the original four offices.

Despite these challenges we found that the UC and comparator areas were similar historically and that a given type of apparently Pathfinder eligible JSA claimant would have had the same expected outcome irrespective of whether they had claimed in a given UC office or its respective comparator offices. This is a direct result of how we identify comparator offices. We can check that it is true using the historical Atomic Data Store (ADS) data that we use to identify comparator offices. For example, we find that there is no difference between the outcomes of apparently eligible JSA claims made in Ashton and those made by matched claimants in Ashton's ten comparator offices between April 2011 and April 2013. The estimated difference in the off-flow rate between claimants in Ashton and its comparator offices was just 0.1 percentage points and is statistically insignificant with a t-statistic of only 0.2. Moreover, matching makes little difference, which reflects that the new claims in these areas look very similar.

The sensitivity analysis since February has focused on the following additional questions to further explore how effective the method is at controlling for potentially relevant differences between UC and non-UC areas:

- i) Do offices with historically similar off-flow rates also have the most similar past employment outcomes for new claims that are apparently eligible for Pathfinder (APE)?

- ii) How well does the way we use comparator offices control for area level differences?
- iii) Do historically similar offices remain similar after UC is introduced in the Pathfinder offices? In other words, is there any evidence of differential area trends that we might be conflating with the impact of UC?

Comparator offices based on employment outcomes rather than off-flow rates

The preliminary evaluation identified comparator offices based on their JSA off-flow rates for apparently Pathfinder eligible new claims during the two years before UC started in April 2013. The analysis took into account the size and composition of the new claims offices received in terms of age, gender and JSA claim history. This approach reflected the data available at the time and an expectation that an office's off-flow rate would be highly correlated with its employment outcomes, and therefore would be a suitable proxy. The JSA off-flow rate was measured after 60 days.

The preliminary impact estimates focus on employment outcomes within 120 days of claiming and controlled for a wider range of information about the type of claim. So, the implicit assumption was that offices with similar off-flow rates from JSA within 60 days for apparently Pathfinder eligible new JSA claims given age, gender, JSA history and on-flow size would also have similar off-flow rates to *employment* within 120 days for the same types of claims given a wider range of information about the claim and employment history of the claimants.

To test this assumption we constructed an individual-level historical dataset that:

- a. Has outcome measures more similar to those we obtain from the RTI but derived from WPLS;
- b. Includes the same detailed information about past benefit and employment history, past participation in DWP programmes, sanctioning, etc. as we use in the main evaluation; and
- c. Uses an improved approach to identify Pathfinder eligibility.

We aggregate this data to office-level. We then repeated the original comparator office analysis focusing on past employment outcomes as measured using the WPLS and taking into account the wider range of information about the claimants. Due to the volume of claims the data was only initially obtained for new claims made between March 2012 and April 2013.

In doing the analysis at office-level it is not possible to include all the same explanatory variables in the Probit model we use to estimate the probability of treatment as we do at an individual-level. Instead we use a more parsimonious model. For example, rather than having dummy variables for each week in the two years prior to the new claim to indicate what proportion of the cohort was in WPLS employment during that week we just take an average percentage of time spent employed by the monthly on-flow cohort of new claims during the two years before their claim.

So, this updated comparator office analysis uses outcomes and contextual data that are much more similar to the data that we use in the main evaluation. However, we aggregate and simplify the data to use it at an office-level. Estimating similar models using the individual-level data and office-level dummies produces the same results, but is much more computationally intensive.

Column 1 *Table A1* lists the ten “best” comparator offices for each of the original four Pathfinder offices using the original historical data, which focused on JSA off-flow rates for the apparently Pathfinder eligible (APE) group. Column 2 lists the ten “best” comparator offices using the richer historical dataset and focusing on employment outcomes. The new historical data leads to very different comparator offices. For Ashton, Oldham and Warrington none of the ten “best” offices using the employment outcome and latest historical data were originally included in the ten “best” comparator offices. Only Wigan has (three) comparator offices that were in the top ten based on both the original data and the new historical data.

This lack of overlap is surprising. We had expected the JSA off-flow rate at 60 days to be highly correlated with the WPLS employment off-flow rate within 120 days. Consequently, we had expected more overlap between the comparator offices in each case. However, we know that, as well as focusing on different outcomes (albeit ones we expect to be highly correlated); there are other differences between the two historical datasets because:

- a) They cover different (albeit overlapping) periods; and
- b) The new data better identifies Pathfinder eligible new claims.

Column 3 in *Table A1* shows the “best” comparator offices when we use the new richer historical data, but use the off-flow rate within the first 120 days from JSA rather than employment outcomes and confine the analysis to March 2012-April 2013. It is surprising that when everything else is the same except for the outcome measure we still get very different comparator offices.

Amongst Ashton, Oldham and Warrington there is only one office that is in the top ten based on both their historical performance on employment outcomes and their past performance on the JSA off-flow rate.

Table A1

Original (Dec 12)	Employment Off-Flow June 15	JSA Off-Flow June 15
Ashton Under Lyne JC	Ashton Under Lyne JC	Ashton Under Lyne JC
Barrow JC	Blackburn Orchard House JC	Carlisle JC
Bury JC	Kentish Town JC	Dalkeith JC
Cheetham JC	Kirkby St Chads JC	Leigh JC
Dalston JC	Lewisham JC	Longsight JC
Edmonton JC	Palmers Green JC	Newton Aycliffe JC
Longsight JC	Rochdale JC	Peckham JC
Longton JC	Rusholme JC	Pontypridd JC
Newcastle City JC	Stourbridge JC	Rochdale JC
Seacroft JC	Wavertree JC	Shettleston JC
Southport JC	Wolverhampton Molineux House JC	Southport JC
Oldham JC	Oldham JC	Oldham JC
Beeston JC	Dagenham JC	Cardiff Charles Street JC
Cardiff Charles Street JC	Erdington JC	Cardiff West JC
Kirkby St Chads JC	Hoxton JC	Crossgate House Doncaster JC
Nottingham Station Street JC	Huyton JC	Dundee Wellgate JC
Peckham JC	Leeds Park Place JC	Leeds Southern House JC
Poplar JC	Lewisham JC	Nottingham Station Street JC
Rochdale JC	Middlesbrough Central JC	Peckham JC
Tottenham JC	Rochdale JC	Peterborough JC
Walsall Bayard House JC	Walthamstow Forest Road JC	Rotherham JC
Worsley JC	West Bromwich JC	Stockport JC

Original (Dec 12)	Employment Off-Flow June 15	JSA Off-Flow June 15
Warrington Nolan House JC	Warrington Nolan House JC	Warrington Nolan House JC
Bedford JC	Carlisle JC	Hanley JC
Bexleyheath JC	Chelmsford JC	Kennington Park JC
Edinburgh High Riggs JC	Chichester JC	Lincoln Orchard Street JC
Leicester New Walk JC	Chorley JC	Loughton JC
Mansfield Hillhouse JC	Hessle JC	Mansfield Hillhouse JC
Redbridge JC	High Wycombe JC	Newcastle City JC
Scunthorpe JC	Nuneaton JC	Norwich JC
Southend JC	Port Talbot JC	Peterborough JC
Stockport JC	Redhill JC	Poplar JC
Streatham JC	Torquay JC	Stockport JC
Wigan JC	Wigan JC	Wigan JC
Barnsley JC	Barnsley JC	Hull Britannia House JC
Batley JC	Chatham JC	Hull Market Place JC
Crossgate House Doncaster JC	Grays JC	Huyton JC
Dundee Wellgate JC	Harrow Kings House JC	Leicester Wellington Street JC
Leicester Wellington Street JC	Ipswich JC	Mansfield Hillhouse JC
Lincoln Orchard Street JC	Leicester Charles Street JC	Peterborough JC
Newport Charles Street JC	Leicester Wellington Street JC	Sheffield Cavendish Court JC
Rotherham JC	Romford JC	St Helens JC
Sheffield Cavendish Court JC	Sheffield Cavendish Court JC	Tottenham JC
Southend JC	Wakefield JC	Wakefield JC

These results reflect that these two different outcomes are not as highly correlated as we expected. Further exploration shows that the association between employment and benefit off-flow outcomes differs between offices and varies within offices over time. For example, there are even some offices for which there is a negative association between off-flow rates and employment outcomes. This is likely to reflect the small number of

observations per office (13) and the small number of apparently eligible new JSA claims in small offices, which makes their outcomes volatile. This office level variation in the association between off-flows and employment outcomes is why we end up with different comparator offices when we look at employment outcomes rather than benefit off-flows.

Since we are evaluating the impact of UC on employment we want comparator offices with similar employment outcomes. There remains a risk that offices with historically similar employment outcomes as measured by WPLS may not have had historically similar employment outcomes had it been possible to measure them using RTI. This risk reflects the difference in coverage between RTI and WPLS. However, in the absence of historical RTI data the best measure of employment outcomes available is based on the WPLS. When we come to estimate the impact of UC on new claims made later in the roll-out we will be able to start using RTI data to identify comparator areas and in the individual-level matching.

Figure (Ai) shows that the original comparator offices for Warrington have similar historical outcomes in terms of their JSA off-flow rates. However, *Figure (Aii)* shows that Warrington has better employment outcomes than these comparator offices. Therefore, even in the absence of UC we might have expected Warrington to have better employment outcomes post-UC, i.e. because Warrington was doing better before UC.

Figure (Ai)

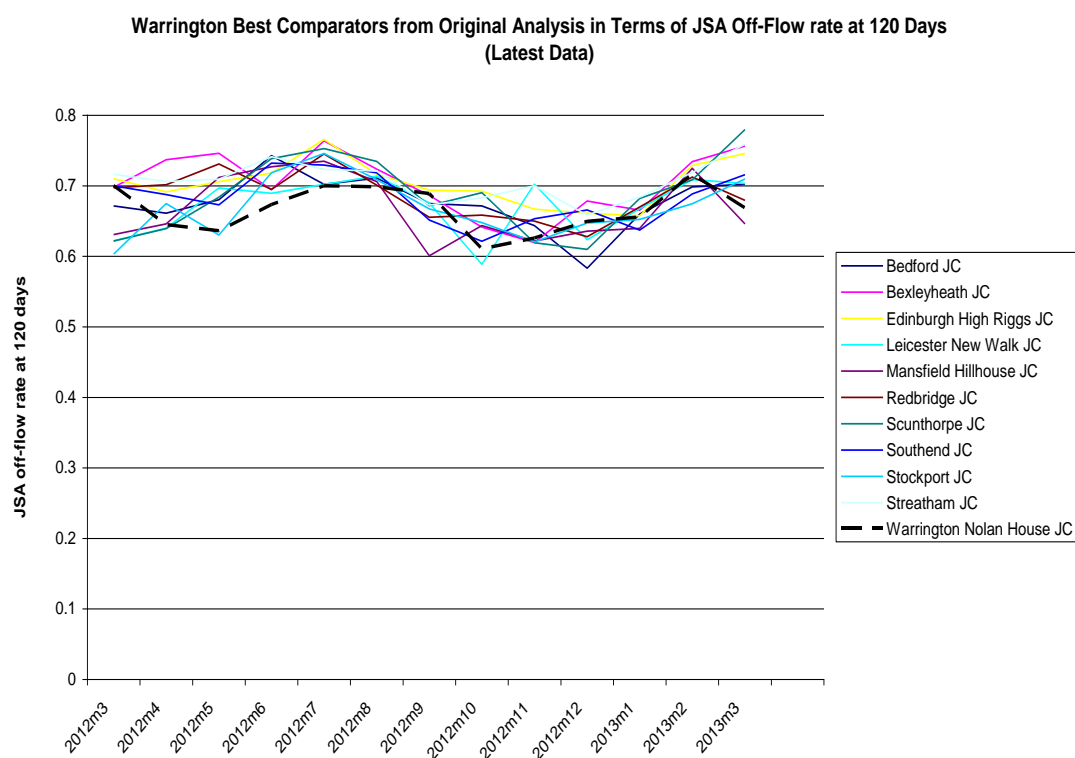


Figure (Aii)

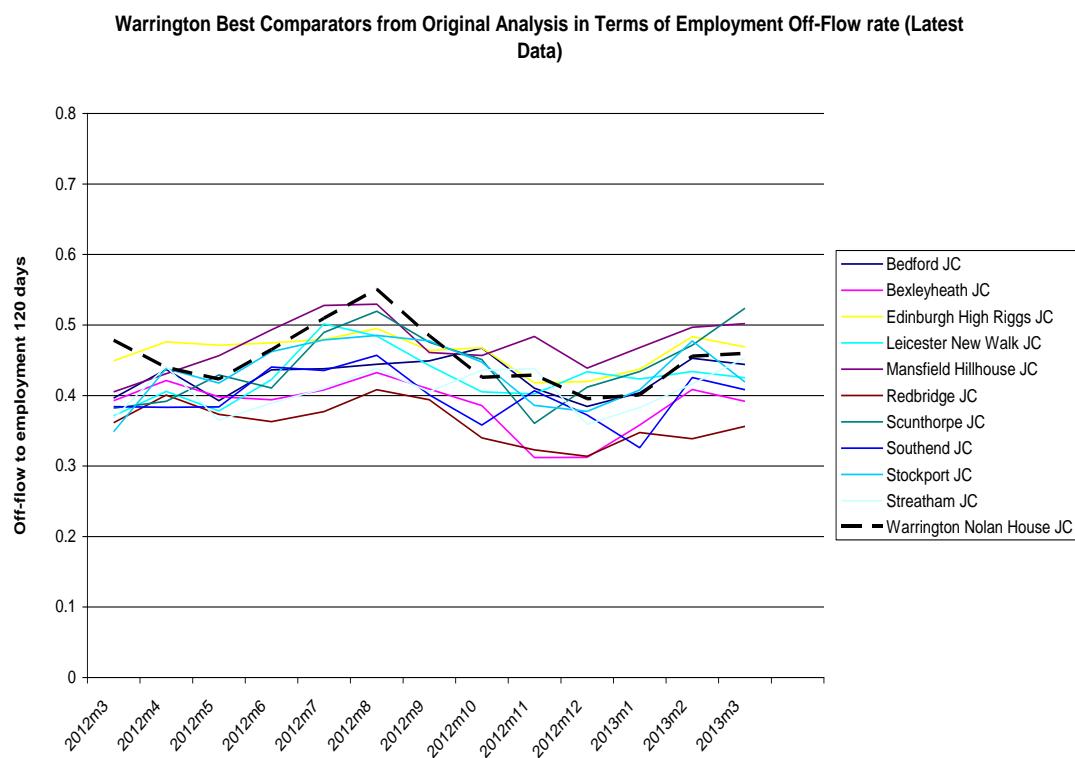
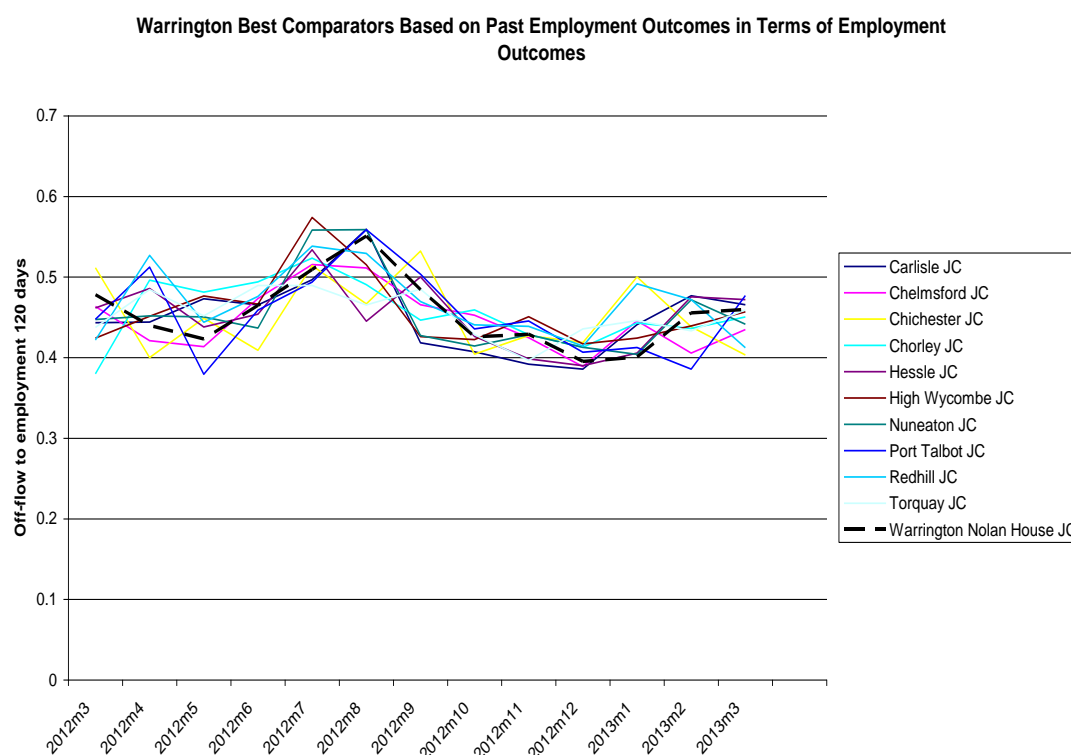


Figure (Aiii) shows Warrington relative to its “best” comparators when we identify comparator offices using past *employment outcomes* and look at their past *employment outcomes*. As one would expect, these offices are much more similar in terms of their past employment performance than the comparator offices we identified based on their past JSA off-flow rates. The impact estimate for Warrington new claims between July 2013 and October 2014 falls from nearly 12pp using the original comparator offices to 6pp using the comparator offices based on past employment outcomes.

Similar patterns emerge for the other four Pathfinder offices. The new comparator offices inevitably are better aligned to their respective UC office in terms of their past employment outcomes, but are not as well aligned as the original comparator offices in terms of their JSA off-flow rates. Overall, changing the comparator offices to those identified using past employment outcomes does not alter the overall results markedly. Whilst it does affect the estimates for individual offices these changes balance out. So the overall estimated impact remains about the same as before.

Figure (Aiii)



Individual Office Analysis

In the preliminary evaluation we looked in totality at all new UC claims in the four Pathfinder offices and all new apparently eligible JSA claims in all the non-UC comparator offices together. This approach ensures that the share of new claims is balanced between each Pathfinder office and its respective comparator offices. For example, if 20% of all new UC claims are made in Wigan then 20% of all apparently eligible new JSA claims will be made in the non-UC comparator offices for Wigan. However, it does not necessarily mean that all the new UC claims in Wigan are matched against new JSA claims in offices that are historically similar to Wigan.

We can better ensure that we are only comparing new UC claims with new JSA claims in similar areas by estimating the impacts separately for each UC office and only including in the comparison group those new JSA claims made during the same period by similar people in the comparator offices for that particular UC office.

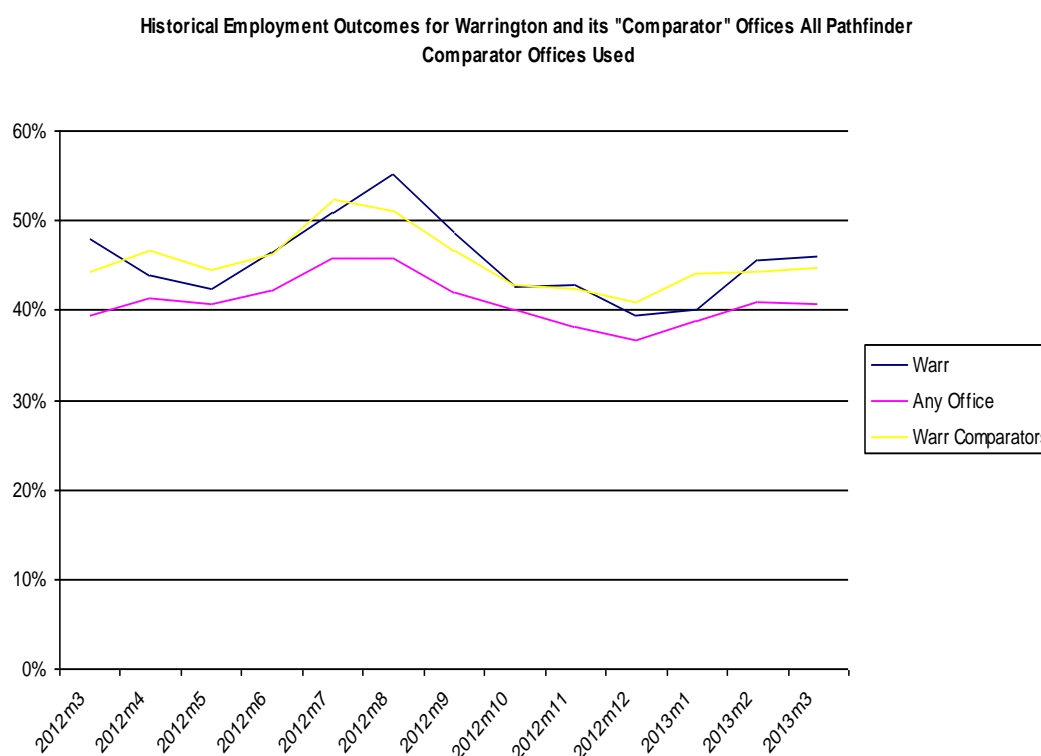
Earlier we mentioned that a non-UC change in a given Pathfinder office might bias our central estimate. Moreover, this risk was relatively high for the Pathfinder evaluation given the small number of Pathfinder offices. The fact that there are positive and statistically significant impacts in all four Pathfinder offices suggests that the impacts are more likely to be due to UC than something else.

It is worth noting that the matching is less effective at an individual office level. This is not surprising because the sample sizes are much smaller. The samples appear to be matched more closely and for a greater proportion of UC claimants in Ashton, Warrington and, to a slightly lesser extent, Oldham. The results imply that in the main evaluation some of the UC claims in Wigan are matched with new claims outside Wigan's own comparator offices. This is possible because the original matching method only ensured that the whole samples are balanced, i.e. that there is the same proportion of new JSA claims in the Wigan comparator offices as there are new UC claims in Wigan.

To further explore the implications of constraining the matched comparison group to the comparator offices specific to a particular UC office we estimated the impact separately for Ashton and Warrington with and without the comparator offices for the other UC offices. For Ashton this made little difference. However, for Warrington the impact estimate was much bigger (11pp versus 6pp) when we allowed matched JSA claims to be taken from

offices that were not the best comparator offices for Warrington. We looked at which offices the matched JSA claims were taken from in the less constrained approach and averaged their pre-UC employment outcomes. *Figure (Aiv)* shows that the non-UC offices where the matched JSA claims were taken tended to have worse employment outcomes for the same claimant group before UC compared with Warrington. This suggests that these claims were probably in different local circumstances and would have done worse than the UC claims even in the absence of UC (i.e. because of other differences between their areas). Again, this suggests that estimating results separately by office is likely to be more robust.

Figure (Aiv)



Offices where the impact estimate changes most using the new comparator offices tend to be those where the impact becomes smaller. Visual inspection of their historical outcomes leads to similar conclusions to those reached for Warrington above. That is the original comparator offices tended to have similar off-flow rates but worse employment outcomes historically and were consequently likely to lead to overestimates of the impact of UC.

Arguably it is reassuring that the impact estimates based on using offices with historically similar employment outcomes leads to impacts that are more similar across offices. That is, with the same treatment at the same time for the same group of claimants we might expect similar impacts across areas. This is not necessarily the case. However, it is possible that the impact of UC could vary between areas and this is something we will continue to explore as UC rolls out.

When we estimate results separately by office we have to make do with a smaller sample size. It also increases the processing time. We also have found that it can be more difficult to find good comparator offices for some of the smaller offices. This can sometimes reflect that the UC office is relatively unusual and so there are fewer offices that come anywhere close to them in

terms of past outcomes. It also reflects the greater volatility in outcomes that can arise when offices are dealing with relatively few claims each month. A small number of different types of claimants can lead to a relatively big change in the office-level outcome. As we start to evaluate the impact of UC on new claims made during the North West expansion this issue should become less prevalent, i.e. because we will have more typical offices delivering UC for which we can find very similar non-UC offices to draw the comparison group from.

The constraint on the sample size means common support becomes a bigger issue, particularly for the smaller more recent UC offices. This is despite the fact that the ratio of apparently Pathfinder eligible new JSA claims in comparator offices to the number of new UC claims is at least as great as it is in other offices. The problem arises because of the large number of explanatory variables. It is much more likely with fewer UC claims to find that some combination of values for the explanatory variables completely determines outcomes for some observations. That is, there will be combinations of the explanatory variables that only occur for the new UC claims or for the new JSA claims. This issue is more likely for values of the explanatory variables that are less common. For example, past ESA claims are much less prevalent than past JSA claims or past employments. Consequently, many of the ESA dummies can drop out of the Probit treatment model due to collinearity.

We estimated a more parsimonious treatment model to allow for this. For example, we replaced the ESA dummies with a summary variable indicating the proportion of weeks during the two years prior to the claim that an individual claimed ESA. This leads to a better specified treatment model. However, this approach led to an increase in the estimated impact of UC and deterioration in the quality of matching in some of the small offices.

The primary purpose of the Probit treatment model is to balance the observables between the treatment and comparison groups. Arguably it does not matter whether the model is well-specified or whether some observations are completely determined. What matters is whether the resulting matched samples are well balanced. Consequently, it seems prudent to retain the full participation model to get better quality matching and therefore more reliable results.

Differential Macro-Trends

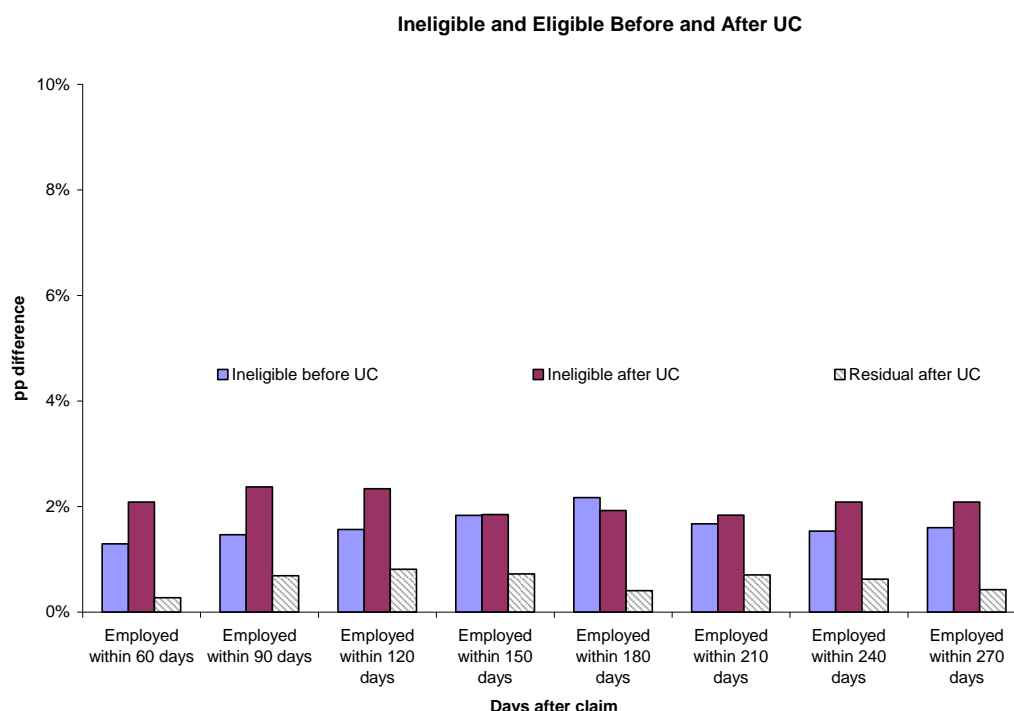
If there were other things happening at the same time as the UC roll-out in the UC (or non-UC offices) that might influence the outcomes we would expect new claimants to achieve then our estimated impact of UC would be conflated with these other changes. For example, a new factory might open in a Pathfinder office after UC was introduced. This could generate a lot more job opportunities for new claimants. We would expect that if this kind of exogenous development did happen then it should also mean that non-UC claimants in the Pathfinder office would have better expected outcomes than similar claimants in the comparator offices where there has not been a new factory.

To test whether there were any differential macro-trends after UC was introduced in the Pathfinder offices we estimated the impact of UC on new claims that were ineligible for UC. These ineligible claimants should not have been affected (at least not directly) by UC being introduced in the Pathfinder offices.

One ineligible group is the JSA residual group in the UC offices. These are the Apparently Pathfinder Eligible (APE) new claims that still claimed JSA after UC had been introduced for Pathfinder eligible claims. The results suggested that there was no difference at all in the likelihood of being employed at some point within 120 days of making a claim for the JSA residual group compared with the apparently eligible new JSA claims in the comparator offices. This suggests that there were not other things happening in the Pathfinder offices apart from UC that contributed to UC claimants' better outcomes assuming the matching approach controls for all relevant factors.

We extended this analysis to look at other groups that were unaffected by the introduction of UC in the Pathfinder offices. Initially we conducted this analysis using the original comparator offices (those based on historical off-flow rates) and using our original approach whereby all offices were included in a single matching exercise. Since then we have repeated the analysis using the latest comparator offices and stratifying the matching by office. *Figure Av* shows the latest results. There appears to be a small statistically significant positive impact on ineligible claimants in the UC offices. This could suggest that there were other things happening in the UC areas that coincided with the introduction of UC, which might have contributed to the better outcomes achieved by UC claimants.

Figure Av



However, it is important to remember that the comparator offices are those with similar historical employment outcomes for Apparently Pathfinder Eligible new JSA claims. It may be that these comparator offices had better outcomes than the UC offices for the apparently ineligible new claims before UC. We know from earlier work that if we look at all new JSA claims we identify different comparator offices compared with when we only focus on apparently eligible new claims. *Figure Av* also shows that there was a small positive impact on the ineligible claimants prior to UC as well. This suggests that the impact on ineligible claims post UC just reflects that these particular comparator offices achieve worse outcomes for these types of claims and that this was true before as well as after UC. Consequently, it does not suggest that there were differential area-level trends that might have contributed to the positive impact found for UC claims.

Conclusions and implications from geographical sensitivity analysis

The sensitivity analysis led to changes in our evaluation approach in the main report. In particular, we now:

- a) Aggregate the estimated impacts of individual UC offices to better control for area-level differences; and
- b) Use comparator offices with historically similar employment outcomes.

We have not found evidence that the estimated impact of UC might be due to differential macro-trends between the UC and non-UC comparator offices that coincided with the roll out of UC. For example, UC has had a positive impact in all four Pathfinder offices. This makes it less likely that the differences we are detecting are due to other things happening. This is also supported by the fact that the estimated impact of UC looks to be exactly the same when we look at more offices over longer time scales. We also find no evidence of differential trends in the UC areas. Ineligible claims appear to have similar relative outcomes pre and post UC in the UC and comparator offices.

Part 2 - Controlling for Individual Level Differences and Identifying Pathfinder Eligibility

People making new claims are individuals – they are different. New claimants who have spent more time in work and less time on benefits are more likely to get a job more quickly than people who have spent less time working and more time on benefits. Similarly, people's outcomes will vary depending on age, gender, marital status, education, ethnicity, number and age of dependent children etc.

We want the individuals in the comparison group to be the same - in terms of everything that might affect their outcomes – as those in the treatment group. Our approach to achieving this involves two steps. First we seek to subset the comparison group of new JSA claimants. So that it only includes those claimants who would have been eligible for UC if they had lived in a UC area. Using the Jobseekers Allowance Payment System (JSAPS) and Housing Benefit data (SHBE) we assess new claims to JSA against the UC eligibility criteria shown in Column 1 of *Table 2*.

Column 2 highlights that there are many criteria that we cannot assess eligibility against because we do not capture them in the data. There are also some criteria in Column 1 that we only capture imperfectly. For example, new claimants are ineligible for UC if they own or partly own the home where they live but we only exclude those who receive support for mortgage interest. We call claimants “**apparently Pathfinder eligible**” (APE) if they appear to be eligible based on the administrative data (the criteria in Column 1). We call new claimants “**actually or genuinely Pathfinder eligible**” (GPE) if they meet the entire criteria in *Table 1*. The criteria in grey are those that JSA claimants must satisfy. These *should* be automatically met by both groups.

Table 2 – Eligibility for UC under Pathfinder

Eligibility Criteria for UC under Pathfinder	
Have information about in administrative data	Do not have information about in administrative data
Have a national insurance number	Not pregnant or given birth in last 15 weeks
In correct postcode area	Not represented by a Personal Acting Body or a Corporate Body ⁴
Be a British citizen and have lived in the UK for the last 2 years	Not challenging/appealing a decision on ESA, JSA, IB, SDA, IS, Housing Benefit (HB) and Working Tax Credit (WTC).
Be aged between 18 and 60 and 6 months.	Not waiting for a decision on a claim to Income Support, JSA, ESA, Incapacity Benefit, Severe Disablement Allowance, Working Tax Credit, Child Tax Credit or Housing Benefit
Do not own home or partially own the home they live in (specifically do not receive support for mortgage interest).	Be unemployed or have household earnings below an agreed threshold (to be agreed), and each have individual earnings of less than £330 a month if over 25years and £270 if under 25years.
Not homeless or living in temporary or supported accommodation.	Not responsible for children/qualifying young persons who are fostered, adopted or being looked after.
Do not have capital exceeding the set limits (£6000+)	Not responsible for children/qualifying young persons who are registered blind or have a disability benefit
Not receiving Carers Allowance or Disability Living Allowance	Not carer for someone with a health condition or disability
Not receiving of Housing Benefit ¹	Not a company director or in a limited liability partnership
Not had a previous JSA or ESA claim ending within 2 weeks of this claim beginning ²	Not self-employed/expecting self-employment earnings in the next month
Not required to pay child maintenance for a child	Do not have a household member who is in the armed forces (regular or reserves) and who is away in connection with that role
Not a couple ³	Not in any form of education or training or about to start a course of education or training within the next month.
Have no children	Fit for work
	Have a bank, building society, Post Office account or a current account with a credit union
	Not receiving Income Support, Severe Disablement

¹ This condition was removed from Mid-June 2014 with the introduction of a new singles gateway onto UC.

² This condition was removed from Mid-June 2014 to reflect the introduction of a new singles gateway onto UC.

³ Couples were admitted in pilot areas from June 2014 and in all live offices from end-July 2014

⁴ When another person or organisation acts on someone's behalf for benefit claim purposes.

Apparent and actual eligibility could be different if there is measurement error. For example, we might classify people as apparently eligible when they are actually ineligible because the data:

- i) Does not capture all the eligibility criteria. So some people might satisfy the criteria in Column 1, but not all the criteria in Column 2; or
- ii) Is wrong so we conclude that some people meet all the eligibility criteria in Column 1 when they do not; or
- iii) Captures some criteria imperfectly; or
- iv) Some combination of i)-iii).

If the administrative data is wrong or captures some eligibility criteria imperfectly we might also conclude that some claimants are apparently ineligible when they are actually eligible.

This makes it difficult to identify a comparison group of people who are the same as those making new UC claims in the Pathfinder offices. Ideally we want to identify those new JSA claimants who would actually have claimed UC if they had they been in a Pathfinder office.

Table 3 shows the picture could be complicated further by delivery error. Delivery error means that actual eligibility may not align with the actual outcome. For example, genuinely eligible claimants might end up claiming JSA and/or some ineligible claimants might end up claiming UC because of delivery error.

For simplicity we assume there is no delivery error or that it is negligible. This means that the outcome is always consistent with actual eligibility. So, only genuinely eligible people claim UC and all ineligible people claim JSA. This seems plausible. It implies that staff can assess eligibility accurately and that they do assign people accordingly. This rules out *c*, *d*, *e*, and *f* in *Table 3*⁶.

⁶ Suppose there was delivery error that led to ineligible people claiming UC. Also, suppose there was no measurement error, so apparent eligibility was always the same as actual eligibility. Then by confining the comparison group to apparently/actually eligible new claims we would exclude some people who would have claimed UC due to delivery error.

We also assume that if someone is genuinely eligible they will be apparently eligible. We cannot check this assumption because we do not have the same data on the new UC claimants that we use to assess the eligibility of new JSA claimants. Hypothetically, some actual UC claimants could be genuinely eligible but apparently ineligible because of measurement error. However, if we assume that the data we have on the criteria in Column 1 of *Table 2* is accurate then we also rule out *b* and *d*.

This means the treatment group of new UC claims will only comprise new claimants who are genuinely and apparently eligible – group *a*. However, the comparison group of apparently eligible new JSA claims in non-UC offices will comprise:

- Group *a* = genuinely and apparently eligible new claims. These are likely to be similar to the new UC claimants except for their geography; and
- Group *g* = a residual group who are apparently eligible but actually ineligible because we cannot identify eligibility accurately for these people in the data.

We cannot identify which JSA claimants are in group *a* and which are in group *g*. However, we can distinguish between these groups in the UC offices. We know there were around 5,800 new claims to UC (group *a*) in the original 4 Pathfinder offices between July 2013 and April 2014 and approximately 6,500 apparently eligible new claims to JSA (group *g*) in the same offices during the same period.

We know these two groups are different by virtue of the fact one claims UC and the other claims JSA. In turn, we think this will reflect that they differ in terms of the unobservable eligibility criteria. *Figure (Avi)* shows that these groups are also different in terms of the observables we have about them in addition to the Pathfinder eligibility criteria we use in step 1. We know these observables affect expected outcomes. Consequently, it is unsurprising that there is a big difference in outcomes between the UC and JSA residual groups in the Pathfinder offices. 45% of the JSA residual group were employed at some point within 120 days of their claim compared with 55% of new UC claimants.

We cannot distinguish between APE and GPE new claims in the comparator offices. However, we have no reason to think that the composition of all APE claims would be different in the comparator offices than it is in the UC offices. Therefore, it is likely that around half of all APE claims in the comparator offices would be genuinely eligible claimants who would have claimed UC had they been in a UC office and about half of them to be actually ineligible. In

other words, half of the comparison group would be similar to the UC treated group and half would be significantly different in relevant respects.

Identifying which new JSA claims are apparently eligible is the first step to eliminating relevant individual-level differences between the treatment and comparison groups. The second step matches new UC claimants with apparently eligible new JSA claimants in comparator offices so that the two groups we compare are the same in terms of all other relevant *observables*. These are things we can measure in addition to the eligibility criteria that we match on directly in step 1 to identify the apparently eligible new JSA claims.

The observables we match on include things that we know affect the expected outcomes of new JSA claimants. Critically, we have detailed information about new claimants' employment and benefit history, information about their participation in other DWP programmes, their sanctioning history, recent claim behaviour, age and gender. Various studies suggest that the rich administrative data available and particularly the detailed information we have about people's past labour market and benefit claim history may be sufficient to obtain reliable estimates. This is true even if the detailed labour market and benefit history do not reflect all the usually unobserved factors that might bias the results such as motivation, attitudes to work, etc. (e.g. "Caliendo *et al.*, 2014").

Arguably the decision to claim UC or JSA is not entirely voluntary and will, in many cases, be driven by need. Consequently, unlike many voluntary active labour market policies there is less risk that some of these unobserved factors attributes will bias the impact estimates, because we have less reason to think that they will differ between the treatment and comparison groups.

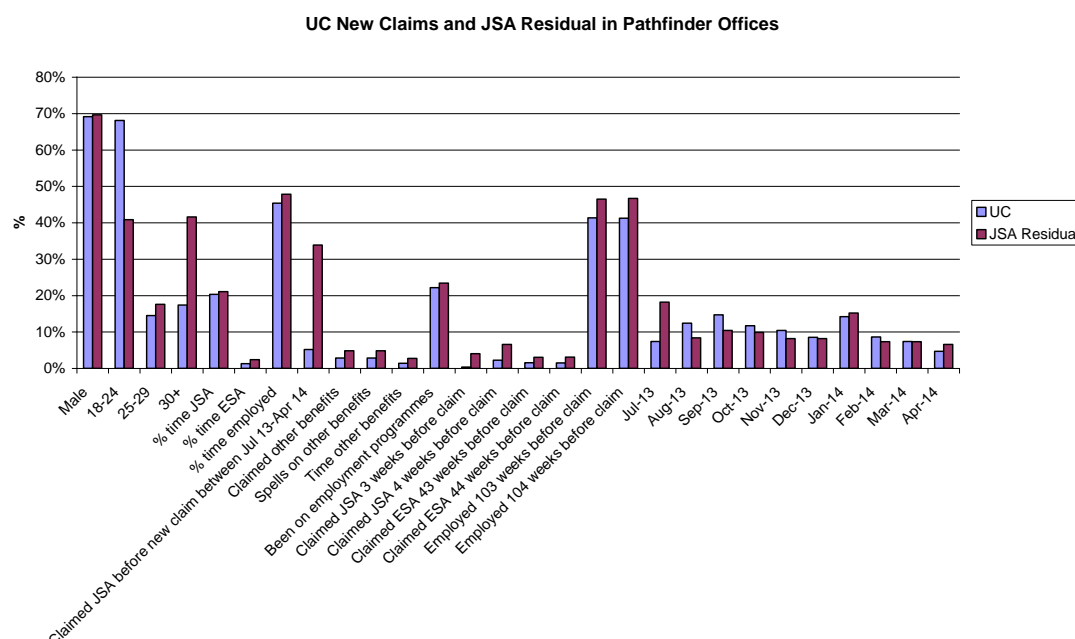
The validity of our approach depends on these two steps satisfying the conditional independence assumption in terms of claimants' *individual* characteristics. That is, conditional on all the individual level observables we match on, either directly in step 1 or through PSM in step 2, the expected outcomes of both groups would be the same under the legacy system. This means that there are no other individual level factors that:

- i) Differ between the matched treated UC claimants and the JSA comparison group; and
- ii) Affect their expected outcomes under JSA.

Table 3 – Identifying the Comparison Group for the UC Treated Group

	Unobserved	Administrative Data	Outcome/D ecision	Possible Reason between Discrepancies Between Genuine/Actual Eligibility, Apparent Eligibility and Outcome	
a)	Genuinely Eligible	Apparently eligible	Claims UC		
b)	Genuinely Eligible	Apparently in-eligible	Claims UC	Measurement error - on some criteria in admin data wrongly classed as ineligible when are eligible	
c)	Genuinely Eligible	Apparently eligible	Claims JSA	Delivery error - should be claiming UC	JSA Residual
d)	Genuinely Eligible	Apparently in-eligible	Claims JSA	Measurement error - on some criteria in admin data wrongly classed as ineligible when are eligible - and delivery error - should be claiming UC	
e)	Genuinely In-eligible	Apparently eligible	Claims UC	Measurement error - on some criteria in admin data wrongly classed as eligible when are in-eligible and/or missing eligibility criteria in admin data and/or imperfectly proxied eligibility criteria in admin data) and delivery error - should be claiming JSA	
f)	Genuinely In-eligible	Apparently in-eligible	Claims UC	Delivery error - should be claiming JSA	
g)	Genuinely In-eligible	Apparently eligible	Claims JSA	Measurement error (on some criteria in admin data wrongly classed as eligible when are ineligible and/or missing eligibility criteria and/or imperfectly proxied eligibility criteria)	JSA Residual
h)	Genuinely In-eligible	Apparently in-eligible	Claims JSA		

Figure Avi



The key question is whether step 2 eliminates any individual level differences that remain after step 1. That is, we know the apparently eligible comparison group is different from the treatment group because we cannot identify all the UC eligibility criteria accurately. Moreover, we think that some of these differences are likely to affect the outcomes we would expect new JSA claimants to achieve. For example, we might expect pregnant claimants and those awaiting a decision on another benefit claim to be less likely to move into work shortly after making a new JSA claim.

Step 2 could eliminate a lot of these remaining unobservable *relevant* differences. However, the Conditional Independence Assumption (CIA) will not be satisfied (in terms of individual level characteristics) if:

- Some unobservable differences remain. This means the unobservables are not perfectly correlated with the observables and so are not balanced between the treated and untreated groups even if the observables are balanced. This seems quite likely; and
- These residual differences have an additional independent affect on expected outcomes over and above all the observables we match on in steps 1 and 2.

We cannot test the CIA directly. However, we can assess its likely plausibility to some extent by seeing how well we can match:

- i) UC claimants with the JSA residual in Pathfinder offices (groups *a* and *g* in UC areas);

- ii) UC claimants with apparently eligible new JSA claims in non-UC offices (group *a* in Pathfinder offices with groups *a* and *g* in non-UC offices);
- iii) The JSA residual group in Pathfinder offices (group *g* in UC offices) with the apparently eligible new JSA claims in non-UC offices (groups *a* and *g* in non-UC offices);
- iv) And UC new claims and the JSA residual in Pathfinder offices with the apparently eligible new JSA claims in non-UC offices (groups *a* and *g* in both UC and non-UC offices).

Figure (Avii) illustrates these different groups and *Table 4* summarises the results. An obvious limitation of this analysis is that we can only assess the quality of matching achieved on the observables. The observable variables we use are the same as those used in the main evaluation.

The unmatched JSA residual group in UC offices is clearly different from those making new UC claims. For example, the likelihood ratio test shows that if we include all the observables in the treatment model it significantly improves the explanatory power. This result is expected since *Figure (Avi)* showed that they were different in terms of some selected observable characteristics.

Figure Avii

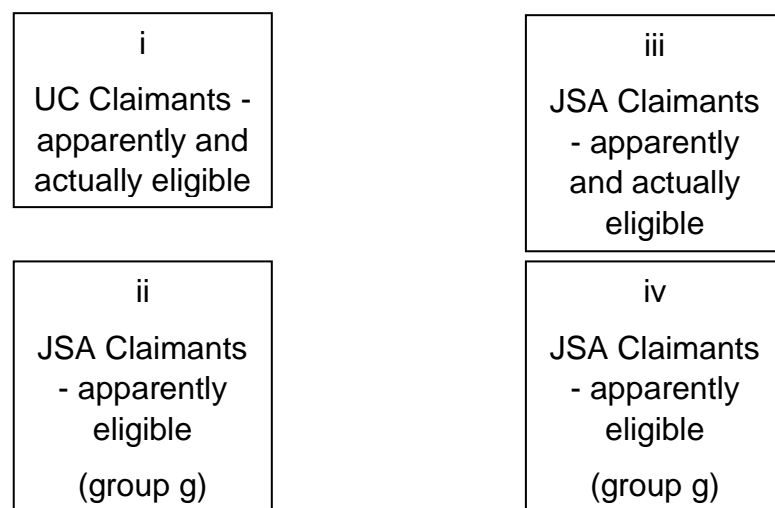
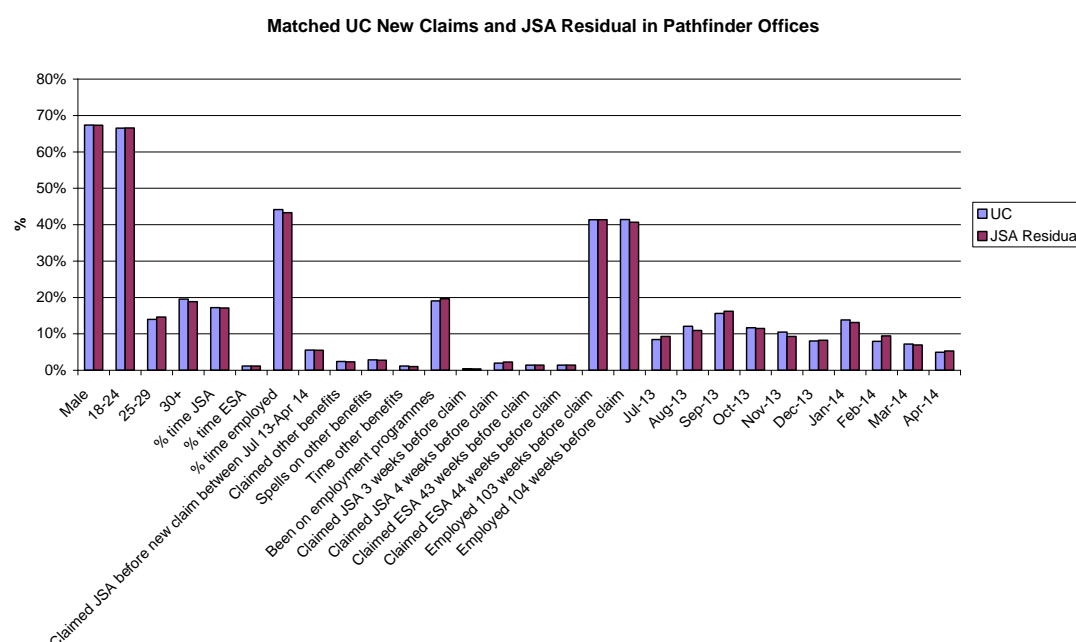


Table 4

	UC vs JSA residual in Pathfinder offices	UC vs Apparently Eligible new JSA comparator offices	JSA residual Pathfinder offices vs Apparently Eligible new JSA comparator offices	UC & JSA residual in Pathfinder offices vs Apparently Eligible new JSA comparator offices
Pseudo R2 before	0.226	0.123	0.052	0.055
Pseudo R2 after	0.031	0.002	0.002	0.001
Impact pp (employed within 120 days)	8.9	5.9	-0.3	2.8
SE	0.02	0.01	0.01	0.01
t value	5.8	8.2	-0.5	5.4
Common Support	3,686	5,641	5,498	11,189
Off Support	2,052	97	89	136
% Off Support	36%	2%	2%	1%
Mean bias before	6	5	7	5
Mean bias after	1	0	1	0
Max bias before	56	58	26	26
Max bias after	6	2	2	2
Likelihood ratio test before p>chi2	0	0	0	0
Likelihood ratio test after p>chi2	0.966	1	1	1

Table 4 shows that we can match the two groups so that the resulting matched treated and untreated groups are balanced in terms of all observables. This suggests that there are new claimants in the JSA residual group who are similar to the new UC claimants in terms of all the observables that we know are important for determining expected outcomes. Figure (Aviii) illustrates this graphically by showing the same characteristics as Figure (Avi) but for the matched samples.

Figure (Aviii)

However, there is a substantial lack of common support. This estimate only relates to two-thirds of those making a new UC claim. We cannot find suitable matches in the JSA residual group for a third of new UC claimants. So as NIESR's review highlighted: *"This illustrates that UC claimants are a distinct group, having (observed, and potentially and of more concern unobserved) characteristics associated with higher employment probabilities. The question*

is whether those factors that can be measured for both UC claimants in the Pathfinder offices and PE JSA claims in the comparator offices are sufficient to capture the impacts of unobservable characteristics of UC claimants on employment outcomes. If not, there is some concern that there may be an upward bias to the central estimates.”

For the two thirds of UC claimants that we can find suitable matches for in terms of all the observables we think there are similar people in the JSA residual group whose outcomes provide a reliable guide to what these UC claimants would have achieved under the JSA regime. We know the matched groups must still differ in terms of at least some of the unobservable eligibility criteria because one claims UC and the other JSA. However, they are balanced in terms of all the observables, which we know affect their expected outcomes. The impact estimate when we use the JSA residual group as the comparison group is bigger than our central estimate. This is consistent with there being some unobserved differences between the groups remaining after matching which contribute to the better outcomes amongst the UC group. However, there are other possible explanations too. For example, it might be that the impact of UC varies between different types of individual and that the impact is bigger for the two-thirds of UC claimants that we can find matches for amongst the JSA residual group.

In the main evaluation the comparison group comprises all apparently eligible new JSA claims in non-UC areas. *Table 4* shows that the explanatory power of the treatment model is lower for this comparison group than when the comparison group comprises APE JSA claims in UC offices. This reflects that these groups are more similar than the UC and JSA residual claimants in UC areas (Column 2 in *Table 4*). *Figure (Aiv)* illustrates this graphically. This is what we expected, because we think this comparison group includes people who are similar to the:

- The JSA residual in the Pathfinder offices (group *g*); and
- UC claimants (group *a*) the genuinely eligible group.

However, these two groups are still significantly different from one another. Again, this is what we expect given that we think the comparison group includes the JSA residual group *g* as well as the genuinely eligible group *a*.

The results of the matching does indeed suggest that there is more overlap between this comparison group and the new UC claimants. Common support is not an issue. We can find suitable matches for 98% of new UC claims based on the observables using a bandwidth of 0.0001. Moreover, the quality of matching is much higher. The Pseudo R^2 for the matched sample falls to 0.002 and the Likelihood Ratio statistic falls to 36 (see *Table 4*).

Of course, we still do not know if the groups are balanced on all relevant individual level factors. In particular, we do not know if they are balanced in terms of the unobserved eligibility criteria that we know cannot be balanced in the raw sample. However, the fact the groups are so well balanced on the rich range of observables that we know are relevant to outcomes and that these observables are likely to be correlated with most *relevant* unobservables suggests that it is likely achieving a good balance on the relevant observables will improve the balance on many of the relevant unobservables. However, how far matching on observables indirectly balances relevant unobservables remains an untestable assumption.

Figure (Aiv)

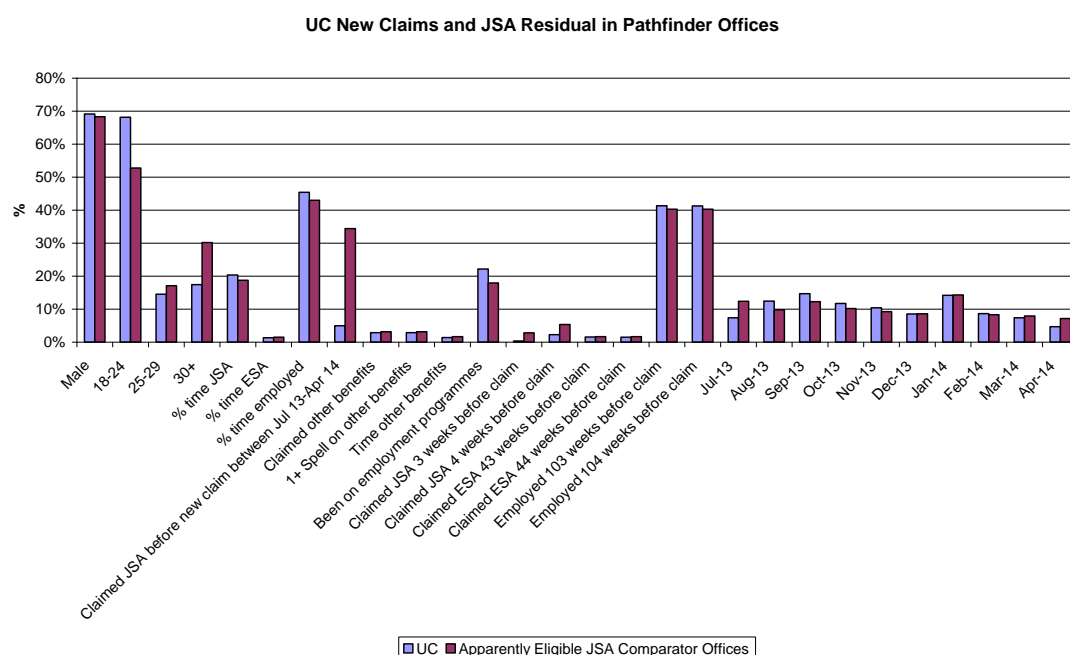


Table 4 shows that we can identify very good and complete matches for the JSA residual group in the Pathfinder offices using the comparison group of apparently eligible new JSA claims in non-UC offices. This supports our earlier assumption that the apparently eligible new JSA claims in non-UC offices is likely to include within it a JSA residual group (group g) that is similar to the one we can separately identify in the UC areas.

Table 4 shows that even prior to matching these groups are more similar than the groups we have compared already. This may be a little surprising since we think the comparison group in the non-UC offices includes GPE new claims as well as those that are only APE. It suggests that, in the raw samples, the apparently eligible new claims in non-UC offices comprise of more people who, based on the observables, are more similar to the JSA residual group in the Pathfinder offices than they are similar to the UC treatment group. However, it could also reflect that the residual type group is

bigger in the UC areas than the genuinely eligible group and so it may also be bigger in the non-UC areas.

After matching there are no significant differences between the JSA residual group and the matched comparison group taken from the apparently eligible new JSA claims in non-UC offices. NIESR's review concluded that: "It is encouraging that both Pathfinder office groups (UC and PE non-UC) can be balanced on the key covariates used when considering comparator offices, and that the difference in outcomes between the non-UC group in the Pathfinder offices and the matched controls is zero (rather than negative, as might be expected given the potential selection bias discussed above)". This *"(...) may suggest that the combination of covariates that are used in the matching process are sufficient to capture the effects on outcomes of the unobservable characteristics that determine selection into UC."*

This analysis has shown that we can identify a good matched comparison group in terms of all the 300+ observable variables we have data on for two-thirds of UC claimants by using the JSA residual claims in UC offices. This suggests that when we draw the comparison group from all APE claims in the comparator areas we will probably compare some new UC claims with new APE JSA claimants who would not have been eligible for, or claimed UC even if they had been in a UC area. In other words, we would compare some new UC claims with that half of the comparison group that we know must be different from the UC claims.

To explore this issue further we took half the JSA residual group in the UC offices and put them in the comparison group along with all the APE claims in the comparator offices. We then estimated the impact on UC only claims. Whilst the matched comparison group was mainly made up from APE claims in the comparator areas this just reflected the bigger number of these claims relative to the JSA residual claims. About two fifths of the JSA residual group is included in the matched comparison group (compared with a third of the APE group from comparator areas).

If we include the other half of the JSA residual group in the treatment group, whilst keeping half in the comparison group, then just over half of the JSA residual group are in the matched comparison group (and just over 40% of the APE claims in the comparator areas are in the matched comparison sample).

These results are intuitive. In the second scenario the treatment group is bigger and more diverse. Therefore, it is unsurprising that more people from both the JSA residual and the APE claims in comparator areas are included in the matched sample. The increase in the number of claims in the matched

comparison sample is bigger for the JSA residual claims. Again this is intuitive. We might expect that the JSA residual claims we put in the comparison group to be more likely to be matched to the JSA residual claims we added to the treatment group.⁷

This analysis shows that, in terms of the observables, the JSA residual claims are more similar to some UC claims than any of the APE claims in the comparator offices. This is true even though the APE claims in the comparator offices will include claims that would have actually claimed UC had they been in a UC office. So, when we estimate the impact of UC on UC claims using a comparison group of all APE claims in comparator offices we will probably compare many new UC claims with JSA claims that are only apparently eligible, but who would not have actually claimed UC had they been in a UC area. We know these claims must be different in unobservable respects even though we can match them very closely on all the observables.

These unobservable differences are only important if they affect expected outcomes under the legacy system. Intuitively we might expect many of the eligibility criteria we do not currently capture to affect expected outcomes, e.g. pregnancy or awaiting the outcome of another benefit claim. However, many of these things will also be highly correlated with the observables, we do include and which we know are important for determining outcomes, such as employment and benefit histories. It is only if the unobservables have an additional or independent impact on expected outcomes over and above all the observables we include that they would bias the results.

There is limited scope to significantly improve how we identify Pathfinder eligibility. Even if we add other criteria for which we might be able to find data we would still categorise many actually ineligible claims as apparently eligible.

We can make the treatment and comparison groups more alike by defining the treatment group as all apparently eligible claims in UC offices. This means the treated group will include both new UC claimants *and* JSA residual claims in UC offices (groups *a* and *g* (and *i* and *ii*) in *Figure (Aix)*). The comparison group remains the same as before. It too comprises all apparently eligible new JSA claims in comparator offices (groups *a* and *g* in *Figure (Aiii)*).

The rationale for this approach is that the Pathfinder eligible group in the UC offices (groups *i* and *ii*) should be very similar to the Pathfinder eligible group

⁷ If we match on the nearest neighbour only we get the same sort of result. More of the JSA residual are included in the matched comparison sample when the treatment group includes half the JSA residual group.

in the comparator offices (groups *iii* and *iv*). We know that *i* and *ii* together are more similar to groups *iii* and *iv* than the actual UC claims by themselves in terms of the observable factors we match on. They are also likely to be more alike in terms of unobserved eligibility criteria. We have no reason to think that these should be any systematic significant differences between *i+ii* and *iii+iv*.

Arguably, defining the treatment group in this way also makes the comparator offices more appropriate. This reflects that we identify these based on their historical outcomes for all apparently Pathfinder eligible new JSA claims (since we cannot confine the analysis to the genuinely pathfinder eligible).

The last column of *Table 3* reports the results from matching all apparently eligible new claims in the Pathfinder offices (i.e. UC claims plus the JSA residual) with all apparently eligible new claims in the non-UC areas. As we would expect these groups are more similar prior to matching and the matching diagnostics are very strong. We can identify very good matches for 99% of the apparently eligible claims in Pathfinder offices from the apparently eligible claims in non-UC offices. Interestingly, even though about half of the Pathfinder group do not receive the treatment (i.e. still claim JSA as the JSA residual) there is still a significant impact on outcomes overall because of the impact on the subset who do receive UC. Moreover, the estimated impact on all APE claims in the UC offices was just under half the estimated impact on actual UC claims. This is what we would expect *a priori* given the share of UC claims amongst all apparently Pathfinder eligible claims and given the estimated impact on the JSA residual group was close to zero and statistically insignificant. This implies that all of the impact on the apparently Pathfinder eligible claims was due to its impact on actual UC claimants.

These results suggest that it would be reasonable to derive an estimate of the impact on actual UC claims (i.e. those who receive the treatment) from the estimated impact on the apparently eligible group by apportioning the estimated impact according to the share of UC claims in the apparently eligible group in the UC offices. The rationale being that:

- a) The estimated impact should be more reliable because there should be no differences (observed or unobserved) between the matched treated and comparison groups; and
- b) Though less reliable the estimated impacts on actual UC claims only and the JSA residual claims only would support the hypothesis that the impact on all APE claims is solely due to UC's impact on those who receive the treatment, i.e. the actual UC claimants.

The results in *Table 4* used our original evaluation methodology. We have since made several changes to our method. For example, we are estimating

impacts separately by each UC office and then aggregating the results and we have identified more appropriate comparator offices based on past employment outcomes rather than JSA off-flow rates.

We have updated the sensitivity analysis using our new methodology. So, for each office we estimate the impact on three treatment groups:

- a) All Pathfinder eligible new claims;
- b) Actual UC claims; and
- c) The JSA residual.

A priori we expect:

- A zero impact on the JSA residual group because they should not receive any treatment if there are no relevant unobservables missing from the matching; and
- The impact on all APE claims to be lower than the impact on new UC claims in proportion to the share of JSA residual claims.

However, if we estimate the impact separately for the different groups a)-c) there will be a lack of consistency in which claims are included in the different analyses, which could make the results difficult to compare. For example, when we estimate the impact of UC on all apparently eligible new UC claims in Warrington between July 13 and April 14 there are 1,224 new UC claims that are in the treatment group for which we can find suitable matches in the JSA comparison group. When we confine the treatment group to actual UC claims only then there are 1,100 new UC claims in the treatment group for which we can find matches.

Moreover, it is not just a case of there being 124 more new UC claims included in the matched treatment group when we estimate the impact on all apparently eligible new claims. There are actually 229 UC claims included in the treatment group when we look at all eligible claims that are not in the treatment group, when we estimate the impact on actual UC claims only. Analogously, there are 105 new UC claims in the treatment group when we estimate the impact on actual UC claims only which are not in the treatment group when we estimate the impact on all apparently eligible new claims in UC offices.

The same applies to the residual group. There are more and different residual claimants included in the treated sample when we estimate the impact on all apparently eligible new claims in UC offices compared with when we estimate the impact on JSA residual claims only.

The nature of the claims included and excluded in each model varies. For example, the UC claims in the treatment group only when we look at all eligible claims spent an average of 52% of the last two years in employment whilst those UC claims who were in the treatment sample only when we look at actual new claims the proportion of time spent in work during the two years prior to claim was 57%.

It is important to remember that we only estimate the impact on claims in the treatment group for whom we can find claims in the comparison group with a similar Propensity Score. This is because we restrict the estimates to those in the treatment group on “common support”. The treated UC claims who are on common support varies depending on whether we estimate the Propensity Score using all APE claims or just actual UC claims. This is because the Propensity Score for a given individual will differ between these two models.

When we estimate the impact on all APE claims the treatment model becomes solely the probability of being in a UC area rather than a comparator area. That is, it just depends on differences between all APE individuals in the UC and all APE claims in the comparator offices. The matched comparison group will be the same as the typical UC/JSA residual claim in the UC offices.

When the treatment group comprises only UC claims, the treatment model also reflects the probability of being a UC claimant rather than a JSA residual claimant. In this case the matched comparison group will be the same on average as the typical UC claim in the UC offices in terms of the observables. We know the UC claims differ from the JSA residual group in both observable and unobservable ways. Consequently, the model that only includes UC claims as the treated group has more variation between the treatment and comparison group, which leads to more explanatory power. This just reflects that the two groups are more different. So matching is more important for achieving a balance between them.

The matching algorithm we use constrains matches to a bandwidth of 0.0001. This means that we can only find a suitable match for a new claim in the treatment group if there is a new claim in the comparison group with a Propensity Score within 0.0001 of the treated claim’s Propensity Score. Since the PS for a given individual will differ between the two models (for both the treated and comparison group claims) then the claims that are on and off support will also differ between the two models. If we do not constrain the matched claimant to have a PS within a certain bandwidth of the treated claimant then we can find matches for virtually all claimants in both models. It

is only a handful of isolated cases with very high propensity scores where this is not the case.

So when we estimate the impact on all APE claims the UC claims for which we can find matches are different from the UC claims for which we can find matches when we estimate the impact on UC claims alone. The same applies to the estimated impact on the JSA residual claims. Therefore, we would not necessarily expect consistency between the estimated impacts for all three different treatment groups because the people we look at differ between the different estimates.

To improve the comparability, between the impact estimates for the three different treatment groups, we can confine the impact estimate for the UC only and JSA residual only treatment groups to include only those claims that we can find matches for when we estimate the impact on all eligible claims. For example, in the case of Warrington, we can take the 1,224 UC claims that we find matches for when we estimate the impact on all eligible claims and then we estimate what impact UC has on these 1,224 UC claims.

This produces more intuitive results. There are no statistically significant impacts on the JSA residual group in any of the four original offices and the impact on UC claims is in proportion to the share of UC claims amongst the APE group in the UC offices. The same is true when we extend the analysis to the first ten UC offices and the time period to new claims between July 2013 and October 2014. It makes no material difference whether the estimates for the UC only and JSA residual only claims are generated by re-matching or by just comparing the outcomes using the weights obtained from matching all APE claims in the UC and non-UC offices.

Conclusions and Implications of APE and GPE Analysis

We know that estimating the impact on UC only claims means that we compare many UC claims with new claims that would not have claimed UC even if they had been in a UC area. We also know that many of these JSA residual claims will be different from the UC claims even though they are more similar, in terms of the observables, than some of the Genuinely Apparently Pathfinder Eligible claims in the comparator offices. Matching on the observables may also balance some relevant unobservables (including the unobserved eligibility criteria). This is likely to be true given how important many of the observables are in determining outcomes and the likely association between them and some of the unobserved eligibility criteria.

However, there must still be other unobserved differences between them since the JSA residual type claim would not have claimed UC. We do not know if these remaining differences are relevant or not. That is, after we control for all the relevant observables (and indirectly control for relevant unobservables that are correlated with the observables) would the JSA residual type claim still have a different expected outcome under the legacy system than the UC claimant because of the remaining unobserved differences between them?

We have seen that estimating the impact on all APE claims in UC offices reduces the observable differences between the treatment and comparison group. It should also reduce unobserved differences (including the unobserved eligibility criteria). This is because roll-out does not depend on unobservable factors. Roll-out could be correlated with unobservables if these unobservable factors vary between the UC and comparator offices. However, we have no reason to think is the case.

So, estimating the impact of UC on all apparently eligible claims in UC offices should generate a more reliable estimate. The matched comparison group will be balanced with the typical APE claim in the UC office rather than the typical UC claimant. The reason for including all APE claims in UC offices in the treatment group is that this treatment group will be more comparable in terms of observables and unobservables to the APE comparison group. It will still involve comparing some new UC claims with JSA residual type claims in the comparator offices. However, we will also compare JSA residual type claims in the UC offices with some genuinely eligible new JSA claims in the comparator offices. So, this approach should achieve a better balance between the treatment and comparison groups overall. We can see that this is true in terms of the observables but it should also be true in terms of the unobservables, including the unobserved eligibility criteria.

The impact on all APE claims will underestimate the impact on UC claimants. Nearly half of the treatment group do not receive the treatment. We can derive an impact on UC claims by apportioning the impact according to the share of UC claims amongst all APE claims in UC offices. This produces an estimate that is virtually the same as that which we obtain when we estimate the impact on UC claims only. This suggests that the main estimates using UC only claims as the treated group are reliable, i.e. because they are entirely consistent with the more reliable estimated impact on APE claims. In future we will continue to estimate impacts on the residual and APE groups as part of the sensitivity analysis to help assess the likely reliability of the main estimates.

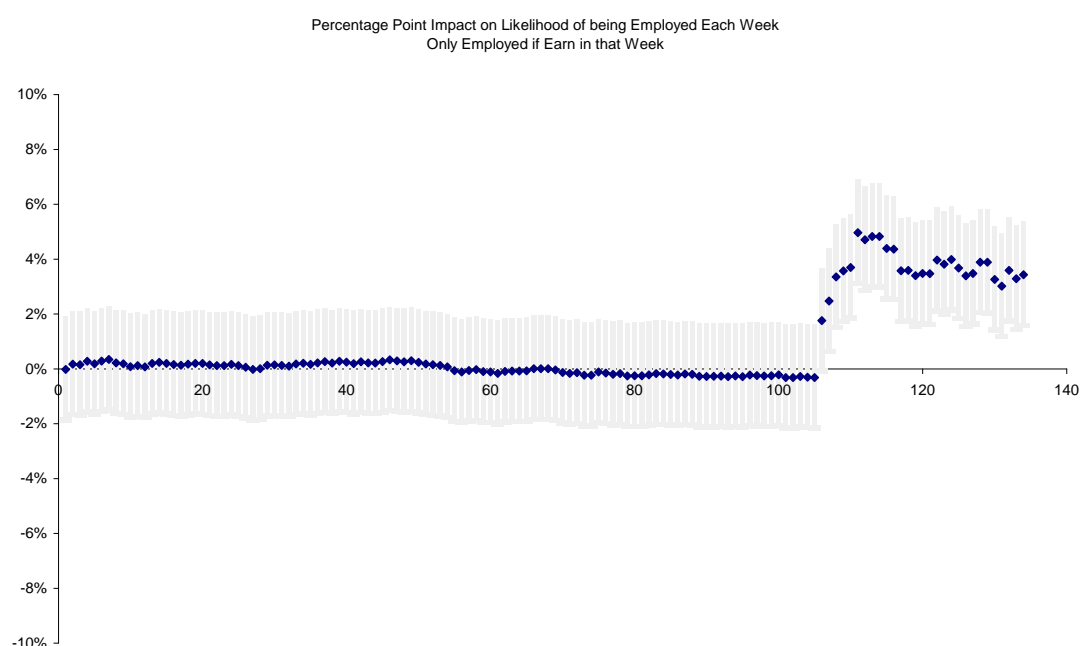
We will also explore the scope for assessing the likely prevalence and importance of the different unobserved eligibility criteria amongst the APE claims and the extent to which they might be correlated with observables. This work would require data sources with information on the same observables as we use in the main evaluation and information on some of the unobservables we do not capture in the administrative data. So, the first stage will be to review potential data sources and the scope and value of using them.

Potential sources are likely to be from surveys and involve smaller samples than the evaluation datasets. This is likely to mean that any conclusions from such work would be indicative and contextual. That is, it would still not be able to properly test the conditional independence assumption. To do so would require a data source which includes all the data we currently use as well as some extra information on at least some of the unobserved eligibility criteria for all or a substantial majority of the sample. We are not aware that such a data source exists and if we were we would have sought to include it in our evaluation data already.

Part 3 - Assessing the accuracy and consistency of the evaluation data Deriving Employment Spells from RTI Earnings

The data on outcomes we use is from the RTI. This gives information about the payments reported for work done, the date of the payment and the amount paid. It does not specify directly the period in which the work was done, the wage rate, or number of hours worked. In consultation with HMRC, DWP has developed a method to use the RTI to estimate employment spells based on the earnings information it holds. This inevitably means making some assumptions. For example, we assume someone is employed continuously if there is a short break in their earnings of less than four weeks. If UC claimants have more employment spells that we wrongly count as continuous then we might over-estimate the impact of UC on employment outcomes. To explore this we re-ran the analysis but only defined someone as being in work in a given week if they had evidence of earnings in that particular week. Using this approach *Figure Ax* shows the impact of UC remains very similar to the original estimates.

Figure Ax



Outcome measures and the date of claim

The impact evaluation estimates what difference UC makes to the labour market outcomes of people after they make a new UC claim. We estimate what outcomes they would have achieved in the absence of UC by looking at the outcomes of a comparison group made up of similar people making equivalent new claims at the same time but who claim JSA instead of UC because of where they claim. We use the start of the claim:

- a) To identify an appropriate comparison group;
- b) And as the starting point from which to measure outcomes.

The date of claim that we use to identify the comparison group should be the date that the UC group would have claimed had they claimed JSA instead of UC. This reflects that we use the comparison group to estimate what would have happened to the UC group had they remained under the legacy system. This date might be before or after or at the same time as they claim under the UC regime. It depends on whether UC changes when people claim.

We want to measure outcomes from the same comparable point in the claim process. It seems reasonable that this should be when the treatment starts. Again, we want the date from which the treatment starts for the comparison group to be the date the treatment would have started for the UC claimants had they claimed JSA rather than UC. Identifying appropriate dates for both

UC and JSA claims is made difficult because the data we have about the claim process differs between UC and JSA.

Table 5 shows the date variables we have in the JSA and UC data. There are two dates that we can derive for both regimes when the claim:

- a) Was submitted; and
- b) Became effective.⁸

Somewhat confusingly the date the claim is submitted is called the `effective_start_date` in the UC data. Whilst this may be confusing we retain this variable name to make it clear which variables in the data we have used. It is important to remember that the `effective_start_date` is not the date that the UC claim became effective but the date that it was submitted.

Table 5 – Date Variables

Date	JSA	UC
Date the claim was submitted	<code>registered_date</code> or <code>event_date</code>	<code>effective-start-date</code>
Date the claim became effective/was paid from	<code>jsa_inflow_date</code> or <code>event_effective_date</code>	<code>first_assessment_period_start</code> *
Date the claim was cleared	<code>cleared_date</code>	
Date the claimant wanted the claim to start	<code>claim_start_date</code>	

* or `min_of_ap_start_cps` if `first_assessment_period_start` is missing

In the preliminary evaluation we identified the comparison group and measure outcomes for JSA claims using the date the claim became effective using the “**`jsa_inflow_date`**”. In contrast, for UC claims we use the date when the claim was submitted - “**`effective_start_date`**”.

In nearly all cases (97.5%) the date a UC was submitted was the same as the date when the claim became effective. In 2.5% of cases the claim was submitted **before** the claim became effective – in most of these cases the claim became effective 30 or 31 days after the claim was submitted. Only in 0.1% of cases did the submission date come **after** the effective date. In these cases the difference was typically just 1 day.

For JSA the date when the claim was submitted or registered on JSAPS was usually different from the date when the claim became effective. The submission date was **after** the effective date in 95% of all JSA claims in the sample (based on comparing the `jsa_onflow_date` with the `registered_date`). The mean difference was 9 days with a median of 5.

⁸ For UC claims we proxy the date the claim became effective using the date of the 1st assessment period recorded in PMX which stands for? (or in CPS if the payment record in PMX is missing).

The difference between the registered_date and jsa_inflow_date reflects that when a claim gets submitted the date when payments are calculated from (which is called the Treat as Made Date TAM) is the Initial Date of Contact (IDOC). In most cases the jsa_inflow_date is the same as the IDOC and is nearly always before the registered_date because it typically takes a few days from the initial contact before the claim gets submitted. If the JSA claim is backdated the jsa_inflow_date will not be the same as the IDOC because the claim would be effective, or paid from, the date the claim is backdated to.

It is not possible to identify which claims are backdated with the data we have or can get access to. The jsa_inflow_date is usually before the registered_date because all non-backdated claims get paid from the IDOC rather than the date of submission. Under JSA claims can be backdated or paid in advance by a maximum of 3 months.

Under UC the claim is normally paid from the date the claim was submitted (effective_start_date). The effective_start_date is usually the same as the initial contact date for UC claims. We can depict the claim process for JSA and UC as in *Figure Axi*.

Figure Axi



So our analysis of the dates available in the data for JSA and UC claims suggests that the dates we are using are consistent and comparable for both groups. The only exception is that a small proportion of JSA claims could be backdated but we cannot separately identify these claims (or how long they have been backdated) in the data.

Defining a new claim

In the evaluation dataset we only include claims that make it through to payment. Since it typically takes longer for the first payment to be made under UC (where payments are monthly rather than weekly) this may mean that we exclude claims from the UC sample that would be included in the JSA sample. We also excluded 1 and 2 day duration claims from the JSA sample because these are not thought to be genuine new claims.

It is vital that the people in the UC and JSA samples are comparable. We want to define a new claim in exactly the same way as far as the data allows. The problem is that the data sources and processes for the two types of claims differ. It is consequently difficult to establish rules to select the samples that we are confident will lead to the same sorts of new claims being included in, and excluded from, both samples.

We have expanded the evaluation dataset to include all new claims that appear on either the UC or legacy systems. This means we capture all attempts that have been made to make a new UC or new JSA claim irrespective of whether these claims subsequently become genuine claims. However, it is unlikely that these two broad groups are comparable because we already sift out new JSA claims that we think would fail the UC eligibility criteria whereas some initial UC claims do get subsequently rejected because they fail the eligibility criteria.

We have experimented and compared different samples of new claims by defining a new claim in different ways. For example, for UC claims we have defined a new claim as:

- i) All new UC claims that appear on the system irrespective of whether they receive a UC payment or sign a claimant commitment;
- ii) All new UC claims that receive a UC payment;
- iii) All new claims that sign a claimant commitment;
- iv) All new claims that sign a claimant commitment *and* receive a UC payment.

When we define new UC and JSA claims in different ways and then model the probability of making a UC claim using all the observables there is virtually no difference in how well the models perform between the different samples. This implies that the samples of UC and JSA claims were broadly as comparable (in terms of the observables) irrespective of how we defined a new claim. Similarly, the estimated impact of UC is similar across different definitions of how we define the UC and JSA new claims. The impact did tend to be bigger when UC claims included all those that had signed a claimant commitment.

Since this sensitivity analysis was done we have added information into the evaluation dataset to identify which JSA claims have signed a claimant commitment and which have attended their initial job-seeker interview. Intuitively we would expect these JSA claims to be more comparable to new UC claims that have signed a CC because they have formally agreed to the conditionality requirements of their claim. Of course, there are some differences between the requirements under JSA and UC and so it is conceivable that there may still be some differences. That is, some people might be prepared to accept the JSA conditionality terms but not the UC terms and vice versa.

In future we will continue to conduct sensitivity analysis to check whether the impact estimates vary at all with how we define new claims for the treated and comparison groups.

Annex B: Independent NIESR Peer Review

Review of: *Estimating the labour market impacts of Universal Credit: Updated Analysis*
Universal Credit Analysis Division
November 2015

Rebecca Riley, National Institute of Economic and Social Research, 17 November 2015

This report by DWP (draft 6 November) provides an updated assessment of the labour market impacts of Universal Credit (UC). The core methodology used is the same as that adopted in an initial assessment of the early impacts of UC published in February 2015. As discussed in my review of the February 2015 report the main methodology seems an appropriate way to proceed given the available data. However, there were questions as to whether the impact estimates are wholly down to changes in claimants' behaviour in response to UC.

This updated analysis provides a comprehensive set of sensitivity checks and makes revisions to the methodology in light of these checks. It also uses more developed and updated data and looks at longer term outcomes. Much as the initial assessment, the updated assessment appears to be carried out very carefully and competently.

The main finding is that new UC claimants were 8 percentage points more likely to be in work at some point in the 9 months after their claim start compared to the control group (71% of UC claimants versus 63% of JSA claimants).

There is also some analysis of differences in earnings between the two groups (UC claimants and the control group). Despite the substantially increased chance of having been in work at some point in the 9 months after claim start there is no statistically significant difference in the earnings of UC claimants compared to the control group.

This short review considers the extent to which the updated analysis is methodologically robust and appropriate. In particular this review considers the

- i. sensitivity analysis conducted and the implications and conclusions drawn from it;
- ii. changes made to the methodology since the preliminary evaluation was published in February 2015 largely in light of the sensitivity analysis;
- iii. options and recommendations for making further improvements to the analysis.

Sensitivity analysis and changes made to the methodology

The effect of UC on various labour market outcomes is identified by comparing the outcomes of new UC claimants in UC Pathfinder offices to the outcomes of a control group consisting of new JSA claimants in a set of comparator offices that were largely unaffected by UC. The analysis adopts a multi-stage matching approach intended to ensure that the control group mimics as closely as possible the behaviour of new UC claimants had they instead made their claim under the JSA regime. The stages are as follows:

1. In a first stage new Pathfinder eligible (PE) claims are selected. These are claims that should be UC claims if they occurred in the UC Pathfinders after UC roll out there, but that would not be UC claims if they occurred outside the Pathfinders.
2. In a second stage a set of comparator offices are chosen.
3. In a third stage claims in the Pathfinder offices are compared against claims in the comparator offices.

At each stage there are many choices that DWP analysts need to make to ensure that the impact estimates best reflect the effect of UC rather than something else.

Sensitivity analyses around many of these choices are outlined in section 4 of the report and detailed in Annex A. These focus on different stages of the matching process.

- One of the difficulties in conducting an evaluation of the labour market impacts of UC is that the claim process for UC claims is measured differently to that for JSA claims. This makes it difficult (in stage 1 of the match process) to define the start date of the claim and to decide what constitutes comparable claims across the UC and JSA groups. In the initial assessment report it was suggested that JSA claims were processed slightly faster than UC claims. As discussed in my previous review this measurement issue could potentially inflate the estimated employment impacts of UC. DWP analysts have further investigated differences in the measurement of start dates between UC and non UC claimants, discussed in Annex A, and find that the start dates chosen for the two types of claim are comparable. Additional analysis has been conducted, which suggests that the definition of what constitutes a new claim mostly has relatively little effect on the estimated impact of UC. In the revised methodology UC claims include everyone that signs a claimant commitment. This change is made to align comparability between JSA and UC claims, and to ensure new claims are consistent with start dates. This revision increases the estimated impact of UC slightly. The work to align measurement of claim durations between the two groups is clearly important and is on-going.
- Since the DWP's initial assessment a lot of additional work has been done on the way in which comparator areas are selected (stage 2). In carrying out the initial assessment a set of comparator offices was chosen for each Pathfinder office that minimised differences between the Pathfinder and comparator offices in benefit off-flow rates in the 2

years prior to the initial UC roll-out. The assumption was that this would also minimise differences in employment outcomes, the main outcome measure of interest. Further analysis has shown that a different set of comparator offices minimises differences in employment outcomes than those that minimise differences in benefit off-flow rates. The revised methodology uses comparator offices that minimise differences in employment outcomes at 120 days. These changes do not have much effect on the estimated impact of UC (see Figure 4). As Annex A notes, it is not possible to compare offices based on the employment measure actually used in estimating the impact of UC. This is because the RTI data on which these measures are based are not available in the pre-roll out period. It is also possible that different offices would be better comparators at different time horizons. For these reasons it is unclear whether these changes in methodology constitute a genuine improvement or not. But, intuitively these revisions to the methodology make a lot of sense and it is encouraging to see that the estimated UC impact is little affected by the precise choice of areas.

- The initial assessment matched claims in all Pathfinder offices against claims in all comparison offices. This resulted in a situation where some claims in one Pathfinder office were matched to claims from comparison areas for another Pathfinder office. Further analysis in this updated assessment suggests that claims in individual Pathfinder offices are better matched against claims in the set of comparison areas chosen specifically for that Pathfinder office. This is perhaps unsurprising given the extensive exercise to select the best comparison areas for each Pathfinder office. Therefore, the methodology has been revised to estimate UC impacts for each individual Pathfinder office separately. The aggregate impact estimate is then derived by weighting up the individual office level impact estimates. These changes are in my view a significant improvement to the methodology. They reduce the aggregate impact estimate by 1 percentage point.
- Matching comparison offices to Pathfinder offices on past outcomes doesn't necessarily guarantee that the comparison offices are good matches on current outcomes. It is possible that factors unrelated to UC might affect employment outcomes in the Pathfinder areas at the same time that UC is rolled out, e.g. the arrival of a large new employer, or that factors specific to the comparison areas cause changes in employment outcomes there. Such developments would bias the estimated UC impact. Much work has been done in the current report to assess the likelihood of such a bias. As highlighted in the report the use of several comparison areas for each Pathfinder office mitigates against the risk of diverging trends between the Pathfinder and comparison offices due to changes in the comparison areas (it is unlikely that all comparison areas would diverge in the same way). For similar reasons, the fact that the UC impact estimate is positive and significant for most of 10 Pathfinder offices also allays fears that there is something unrelated to UC and specific to the Pathfinders that lies behind the large positive impact estimates shown in the report.

DWP analysts have also compared outcomes for various non-UC groups across the Pathfinder and comparison offices. The objective is to probe further whether there are any Pathfinder specific shocks that might influence outcomes for UC claimants and hence bias the estimated impact of UC. The initial assessment published in February showed a comparison of PE non-UC claimants in the Pathfinders against PE claimants in the comparison areas. The difference in outcomes is zero between the two, which could suggest that there is nothing unusual going on in the Pathfinders at the same time as UC. But, this is not a strong test because the zero difference could be the outcome of an upward bias from a positive Pathfinder specific shock and a downward bias from unobservable differences between the two (for the reasons discussed below). Therefore the additional analysis of other non-UC groups is very useful. It suggests that differences in outcomes between non-UC groups in Pathfinder offices and comparison areas are either zero or small and positive. Because any positive effects are apparent also before the roll-out of UC they are unlikely to indicate any upward bias to the UC impact estimates arising from a positive Pathfinder specific shock.

- One of the key limitations to what can be done and to interpreting the estimated impact of UC as being wholly due to the policy stems from differences in the information available to identify the treatment and control groups (in stage 1). UC was initially rolled out to a select group of people. The administrative data allows DWP analysts to proxy these selection criteria to identify new PE claims in a set of comparator areas that can then be used as a control group. In order for the control group to constitute a good counterfactual for what might have happened to UC claimants, had they instead made a claim under the old regime, it is important that the PE selection criteria match the UC selection criteria closely in so far as they affect the outcome measure of interest.

Annex A sets out very clearly the particular issues that arise in identifying a suitable set of controls for the UC pathfinder impact evaluation with the available data. This is a very useful addition to the report. Table 2 there illustrates the range of factors that cannot be measured in the administrative data, but which nevertheless determine who can claim UC. Many of these unobserved factors will influence employment outcomes. This is only a problem if, as the report points out, after matching on observable differences (in stage 3):

- b) some unobservable differences remain.
- c) these residual differences in unobservables have an additional independent effect on expected outcomes over and above all the observables matched on in stages 1 and 3.

I agree with the position in the report that a. seems quite likely, even though many of the unobservable differences are probably picked up by matching on the observables (in stage 3). The report then suggests that b. "... seems less likely given we know the observables are important determinants of expected outcomes and their influence over expected outcomes is very likely to be correlated with the influence of

the unobservables.” I am not sure this is the case, i.e. that we can be confident that all residual unobserved differences have no additional independent effect on expected outcomes over and above all the observables matched on in stages 1 and 3. Of the PE group in the Pathfinder offices less than half claim UC; this is despite the fact individuals who qualify for UC in the Pathfinders do not have the option to stay on JSA. According to DWP figures the low take up cannot be attributed to some start-up effect where insufficient claims were processed as UC. The number of non UC PE claims in the Pathfinders does shrink over time, but not substantially. Following discussions with DWP it appears that the PE group vastly overestimates the number of claims that qualify for UC. This is because of all the factors that cannot be measured in the administrative data (column 2 of Table 2), but also because some of the factors that should be captured in the administrative data (column 1 of Table 2) appear to be mismeasured. Finally there are some PE Pathfinder claims that do qualify for UC, but which nonetheless are not UC. This may also reflect measurement error. Alternatively, if there is an element of voluntariness in selecting into UC, this may reflect the fact that UC will be more beneficial to those claimants that move into work. This latter issue would bias upwards any estimates of the impact of UC on employment.

Table 4 shows the UC impact estimate on the probability of having been in work at some point during the last 4 months when Pathfinder PE non UC claims are used as controls (column 1) and when comparison office PE claims are used as controls (column 2). This clearly shows that UC claims and non UC PE claims in the Pathfinders are more different on observable characteristics than UC claims in the Pathfinders and PE claims in comparison offices (observables better explain differences between the two groups in column 1 than in column 2, according to the R^2 before matching). Also, there is less common support between the two groups in the Pathfinder offices than between UC claims and PE claims in the comparison offices. This is not surprising because PE claims in comparison offices will include a mix of genuinely eligible and apparently eligible claims. Once balanced on observables the difference between UC claims and Pathfinder non UC PE claims is 1.5 times the difference between UC claims and comparison office PE claims. This is consistent with a situation where differences in unobservables, which we know to be greater between UC claims and non UC PE claims in the Pathfinders than between UC claims and PE claims in comparison offices, are not fully balanced and have an additional independent effect on expected outcomes over and above all the observables matched on; there may of course be other explanations for this pattern.

Given these data issues it remains important that there is a positive and statistically significant difference in outcomes between all Pathfinder PE claims and all comparison office PE claims. This suggests that there is a positive effect of UC on employment chances, although it may be smaller than the headline estimates suggest.

Options and recommendations for further improvement

My previous report highlighted some areas where further discussion and supplementary analysis might be helpful in further gauging the extent to which the impact estimates were likely to be capturing policy (UC) effects alone.

Much has been done in this updated analysis. In terms of taking the analysis forward I would highlight the following areas:

- There may be auxiliary evidence that can be provided to evaluate the assumptions being made and potential biases associated with the difficulty of identifying a UC eligible comparison group. One option is to test the assumptions being made directly using alternate data sources. For example, using a dataset that has information on both the observed co-variables (or similar co-variables) that are used in the matching process as well as on (key) unobserved factors in the administrative data, it would be possible to test in a regression framework what additional explanatory power the unobserved factors in the administrative data add to a model of employment outcomes that already includes the matching variables. If the unobserved factors add relatively little to the model then that supports the assumptions being made in the UC impact evaluation. The model may also allow DWP to gain a sense of the magnitude of any potential bias. However, such a dataset may not exist. Alternatively, it may be possible to gauge the potential biases that might arise from unmatched unobservables by other means. Two pieces of information are necessary. First, the effects on employment chances of such unobservables; second, the prevalence of these unobservable characteristics in the sample. The first of these may be gleaned from the literature. The second of these will be imprecise, because of the matching procedure, but may be informative nonetheless. For example, if we know certain characteristics to be relatively rare in the population or relatively unimportant in determining employment outcomes, then these are unlikely to bias much the UC impact estimate.
- The headline findings in the report suggest that UC increased the probability of being in work at some point in the last 9 months by 8 percentage points (approximately 12%). This seems like a large effect. In contrast, the impact of UC on the probability of being in work at 9 months is smaller, at 3 percentage points (approximately 6%). It is important to highlight also this second figure, because the contrast to the first figure suggests that the increase in employment chances with UC may be to temporary work. This would be consistent with other evidence discussed with DWP, which suggests that a high share of claimants are willing to accept short-term or temporary jobs. The survival analysis in the report might be extended to assess this, or it may be possible to evaluate differences between the treatment and control groups in exits from work ('returns' to benefit). In this context the earnings estimates presented in the report are also instructive. The difference in earnings between UC claimants and the control group is around 2%, although this is not statistically different. For this to be consistent with the 12% increase in the probability of being in work at

some point in the last 9 months we would need to see either a reduction in hourly pay or a reduction in hours worked, conditional on being in work, for UC claimants relative to the control group.

- The report highlights correctly that the analysis of earnings is likely to be less robust than that of employment chances. I agree it will be interesting to see whether the estimated earnings effect changes once the matching of comparison areas in stage 2 includes earnings.

Conclusions

Overall, this report details a very carefully conducted evaluation of the early labour market impacts of UC. The substantial number of sensitivity checks that have been made invoke confidence in the matching process between Pathfinder offices and comparison offices and reduce concern that differences in developments between these two sets of offices affect the estimated treatment effect of UC. The individual level matching is complicated by the nature of the data that is available to DWP analysts. I have highlighted potential auxiliary analysis that may be feasible and that may help to confirm or otherwise the assumptions that DWP analysts necessarily have to make in undertaking this evaluation. I have also highlighted the importance of interpreting the headline figures in the context of some of the others figures presented in the report.

Annex C: The Gateway Conditions

1. Single Gateway

The Gateway Conditions that applied to the UC claims we focus on in this report are that the claimant must be:

- Aged at least 18 and under 60 years and 6 months.
- Single.
- A British citizen, resident in the UK throughout the last 2 years and not absent outside UK during those 2 years for continuous period of 4 weeks or more.

Fitness to work

- Not pregnant.
- Not have been pregnant within the last 15 weeks.
- Must not have a fitness for work note, unless determined since the note was issued that they do not have limited capability for work.
- Must not have applied for a fitness for work note.
- Claimant must declare they consider themselves fit for work
- Must not have been determined as having limited capability for work, unless it has subsequently been determined that the claimant does not have LCW.

Existing benefits

- Must not be entitled to old rules ESA, old rules JSA, IS, IB, SDA, DLA or PIP.
- Must not be waiting for a decision on a claim to old rules ESA, old rules JSA, IS, WTC, CTC or HB.
- Must not be awaiting an outcome of an application to revise (including a mandatory reconsideration) a decision of non-entitlement to old rules ESA, old rules JSA, IS or HB.
- Must not have an undecided appeal against non-entitlement to old rules ESA, old rules JSA or IS.

Income and capital

- Must declare that their earned income for the first month of the UC claim is not expected to exceed £338.;Capital must not exceed £6,000.ingle claimant is a member of a couple and their capital must not exceed £6,000
- A couples joint capital must not exceed £6,000

Housing

- Must not be homeless.

- Must not live in supported or exempt accommodation.
- Must not live in the same household as a member of the regular or reserve Armed Forces who is absent from the home on duty.
- Must not own or partly own their home.

Caring responsibilities

- Claimant must not have a child living with them for some or all of the time.
- Claimant must not have a qualifying young person living with them for some or all of the time.
- Must not be an adopter with whom a child is expected to be placed within the first 2 months of the claim.
- Must not be a foster parent.
- Must not be liable to pay child support maintenance.
- Must not be responsible for providing care to a person with a physical or mental impairment, unless in the course of paid or voluntary work.

Other requirements

- Must not be self-employed.
- Must not be in education or training of any kind.
- Must not have a person acting on their behalf.
- Must have a National Insurance Number.
- Must have an account with a bank, building society or the Post Office, or a current account with a credit union.

2. Family Gateway

The current Gateway Conditions are that claimants or joint claimants must be:

- Aged at least 18 and under 60 years and 6 months
- A British citizen, resident in the UK throughout the last 2 years and not absent outside UK during those 2 years for continuous period of 4 weeks or more

Fitness to work

- Not pregnant
- Not have been pregnant within the last 15 weeks
- Must not have a fitness for work note, unless determined since the note was issued that they do not have limited capability for work
- Must not have applied for a fitness for work note
- Claimant must declare they consider themselves fit for work
- Must not have been determined as having limited capability for work, unless it has subsequently been determined that the claimant does not have LCW.

Existing benefits

- Must not be entitled to old rules ESA, old rules JSA, IS, IB, SDA, DLA or PIP
- Must not be waiting for a decision on a claim to old rules ESA, old rules JSA, IS, WTC, CTC or HB
- Must not be awaiting an outcome of an application to revise (including a mandatory reconsideration) a decision of non-entitlement to old rules ESA, old rules JSA, IS or HB
- Must not have an undecided appeal against non-entitlement to old rules ESA, old rules JSA or IS

Income and capital

- Single claimants must declare that their earned income for the first month of the UC claim is not expected to exceed £338
- Couple claimants must declare that they do not expect earned income to exceed £338 for each member of the couple, and that their joint income is not expected to exceed £541
- Single claimant's capital must not exceed £6,000
- If single claimant is a member of a couple, their capital must not exceed £6,000
- A couples joint capital must not exceed £6,000

Housing

- Must not be homeless
- Must not live in supported or exempt accommodation
- Must not live in the same household as a member of the regular or reserve Armed Forces who is absent from the home on duty
- Must not own or partly own their home

Caring responsibilities

- Claimant must not have a child or qualifying young person living with them for some or all of the time who is certified as blind or severely sight-impaired, receiving DLA/PIP or who is looked after by the Local Authority (apart from respite care)
- Claimant must not have a qualifying young person living with them for some or all of the time
- Must not be an adopter with whom a child has been placed within the last 12 months or is expected to be placed within the first 2 months of the claim
- Must not be a foster parent
- Must not be liable to pay child support maintenance
- Must not be responsible for providing care to a person with a physical or mental impairment, unless in the course of paid or voluntary work

Other requirements

- Must not be self-employed
- Must not be in education or training of any kind
- Must not have a person acting on their behalf
- Must have a National Insurance Number

Must have an account with a bank, building society, the Post Office, or a credit union

Bibliography

- Caliendo, M., Mahlstedt, R., Mitnik, O. A., 2014. Unobservable, but Unimportant? The Influence of Personality Traits (and Other Usually Unobserved Variables) for the Estimation of Treatment Effects. *IZA Working Paper* (Draft from 12 May 2014). Available at: https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=COMPIE&paper_id=28.
- Department for Work and Pensions, 2015. *Estimating the early labour market impacts of Universal Credit – Early analysis*. DWP ad hoc research report no. 19. Available at: <https://www.gov.uk/government/publications/universal-credit-estimating-the-early-labour-market-impacts>.
- Department for Work and Pensions, 2012. *Universal Credit – Impact Assessment*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220177/universal-credit-wr2011-ia.pdf.
- Department for Work and Pensions, 2014a. *Universal Credit at Work*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/368805/uc-at-work.pdf.
- Department for Work and Pensions, 2014b. *Evaluating the Impact of Universal Credit on the Labour Market in Live Service and the North West Expansion*. DWP ad hoc research report no. 7. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/334312/evaluating-impact-universal-credit-labour-market-live-service-north-west.pdf.
- Institute for Fiscal Studies, 2014. *Evaluating the labour market impacts of Universal Credit: a feasibility study*. DWP ad hoc research report no. 6. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/333912/evaluating-labour-market-impacts-universal-credit-feasibility-study.pdf.
- Rosenbaum, P.R., Rubin, D.B., 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1), pp.41-55.
- Thomas, A., 2008. Evaluating the econometric evaluations of active labour market programmes using administrative data: evidence from the Jobseeker's Allowance pilots. *Department for Work and Pensions Working Paper No 50*. Available at: <http://webarchive.nationalarchives.gov.uk/20130314010347/http://research.dwp.gov.uk/asd/asd5/WP50.pdf>.