

Evaluating the labour market impacts of Universal Credit: a feasibility study

July 2014

DWP ad hoc research report no. 6

A report of research carried out by the Institute for Fiscal Studies on behalf of the
Department for Work and Pensions

© Crown copyright 2014.

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/organisations/department-for-work-pensions/about/
research#research-publications](https://www.gov.uk/government/organisations/department-for-work-pensions/about/research#research-publications)

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

First published 2014.

ISBN 978-1-78425-290-8

Views expressed in this report are not necessarily those of the Department for Work and Pensions or any other Government Department.

Summary

The introduction of Universal Credit (UC) is arguably the most radical restructuring of the benefits system since the 1940s. The aim is to simplify the system and to strengthen work incentives. The importance of this policy to the Government's labour market strategy and anti-poverty strategy calls for a robust appraisal of its impact on work and incomes.

This report assesses what options are available for evaluating the labour market impacts of UC, and considers whether a credible quantitative evaluation is feasible and worthwhile based on a hypothetical roll-out scenario provided by DWP. The report identifies the main questions that would be interesting and potentially feasible to answer and considers alternative empirical approaches, methodological challenges and data availability.

Contents

Acknowledgements	8
The Authors	9
Executive summary	10
Evaluation of UC during its roll-out to new benefit claims.....	11
Evaluation of UC during the migration of existing claims	13
Long-run versus short-run impacts and the scope for disaggregation	14
Data	15
Conclusion: the most promising evaluation possibilities and the data available	16
1 Introduction.....	18
2 Outcomes of interest.....	20
2.1 Employment	20
2.2 Workless households.....	21
2.3 Working hours.....	21
2.4 Wages	21
2.5 Earnings or earned income	22
2.6 Family income and poverty.....	22
2.7 Benefit entitlement and take-up	22
3 What can we learn from the phasing in of Universal Credit for new claims?	23
3.1 Evaluation questions	24
3.1.1 Evaluation Question 1: Impact of UC on the outcomes of new claimants	24
3.1.2 Evaluation Question 2: Impact of UC on market-wide outcomes or the outcomes of specific groups	24
3.1.3 Evaluation Question 3: Substitution, displacement and other indirect impacts of UC	25
3.2 Sources of variation	26
3.3 Methods.....	28
3.3.1 Matching	28
3.3.2 Difference-in-differences.....	29

Evaluating the labour market impacts of Universal Credit: a feasibility study

3.4	Phase-by-phase assessment	32
3.4.1	Phase 1: Pathfinder	32
3.4.2	Phase 2: New JSA claims.....	36
3.4.3	Phase 3: New claims to WTC, CTC, and ISlp	39
3.4.4	Phase 4: New claims to ISc, ESA and HB.....	44
4	What can we learn from the migration of existing claimants to Universal Credit?	49
4.1	Sources of variation	50
4.2	Evaluation Question 1: Impact of migration (or notification of migration) to UC on the outcomes of existing claimants.....	50
4.2.1	Methodology	51
4.2.2	Data	52
4.2.3	Caveats and limitations	52
4.3	Evaluation Question 2: Substitution, displacement and other indirect impacts of UC	54
4.3.1	Methodology	55
4.3.2	Data.....	55
4.3.3	Caveats and limitations	56
4.4	Overall assessment of the evaluation possibilities offered by the migration of the stock of legacy benefit claimants	56
5	Overall assessment of evaluations that are possible in principle and alternative implementation strategies	57
5.1	Summary and assessment of each evaluation question.....	57
5.2	The challenge of isolating UC from other changes	66
5.3	Long-run versus short-run effects of UC.....	67
5.4	Disaggregating the estimated effects of UC	69
5.4.1	Disaggregation by sub-group	69
5.4.2	Disaggregation by cause	70
5.5	Alternative implementation strategies	71
5.5.1	Additional possibilities.....	73
5.5.2	Recommendations	74

Evaluating the labour market impacts of Universal Credit: a feasibility study

6	What evaluations will available data allow in practice?	76
6.1	Data requirements and sources: new claimants	76
6.1.1	Data sources and quality.....	76
6.1.2	Employment	79
6.1.3	Earnings	80
6.1.4	Hours	81
6.1.5	Family level outcomes: workless families, family earnings/income and poverty	81
6.2	Data requirements and sources: existing claimants	86
6.3	Possible new panel surveys being considered by DWP	88
6.4	Summary	89
7	Conclusions	91
Appendix A	Detail and discussion of possible evaluation methods.....	97
Appendix B	Summary of challenges for evaluation using the phasing in of UC for new claims	108
Appendix C	Evaluation design summary tables.....	111
Appendix D	Exploiting the associated reform of Pension Credit.....	133
References	135

List of tables

Table 1.1	Evaluation possibilities arising from the roll-out of UC to new claimants	59
Table 1.2	Evaluation possibilities arising from the migration of existing claim to UC	63
Table 6.1	Summary of evaluation questions and data availability for the evaluation of UC implementation hypothetical phases 1 to 4	82
Table 6.2	Summary of evaluation questions and data availability for the evaluation of UC implementation hypothetical phases 5 and 6.....	87
Table C.1	Evaluation design to assess the impact of Universal Credit (UC) on the outcomes of unemployed new claimants during hypothetical phases 1 or 2	111
Table C.2	Evaluation design to assess the impact of UC on benefit entitlement and labour market outcomes of potential future claimants during hypothetical phases 1 and 2.....	114
Table C.3	Evaluation design to assess the indirect effects of UC during hypothetical phases 1 and 2.....	116
Table C.4	Evaluation design to assess the impact of UC on benefit entitlement and labour market outcomes of potential future claimants during hypothetical phase 3	119
Table C.5	Evaluation design to assess the indirect effects of UC during hypothetical phase 3	122
Table C.6	Evaluation design to assess the impact of UC on the outcomes of new claimants to what would otherwise be ISc/ESA/HB	124
Table C.7	Evaluation design to assess the impact of UC on benefit entitlement and labour market outcomes of potential future claimants during hypothetical phase 4	127
Table C.8	Evaluation design to assess the indirect effects of UC during hypothetical phase 4	129
Table C.9	Evaluation design to assess the impact of UC on the outcomes of legacy benefit claimants migrated onto UC	130
Table C.10	Evaluation design to assess the indirect impacts of a large influx into UC	132
Table C.11	Evaluation design to assess the impact of UC on the outcomes of older couples on benefits	134

List of figures

Figure 3.1	The roll-out scenario (stylized)	27
Figure A.1	Timeline of events for panel DiD	99
Figure A.2	Timeline of events for repeated cross-section DiD	101
Figure A.3	Timeline of events for the repeated cross-section DiD exploring time variation; triple differences contrasts two such estimators, on areas <i>A</i> (treated after time t_A) and <i>B</i> (control area, not experiencing the policy reform during the time frame considered).....	104

Acknowledgements

The authors thank David Thompson at the Department for Work and Pensions (DWP) for providing valuable help throughout this project, and thank him and his colleagues at DWP for providing background information and comments on earlier drafts as well as seminar participants at the Institute for Fiscal Studies (IFS) and DWP's expert advisory panel on Universal Credit evaluation.

The Authors

Stuart Adam is a senior research economist at the Institute for Fiscal Studies (IFS). His research focuses on analysing the design of the tax and benefit system and he has written about many aspects of both tax and benefit policy. He was an author and editor of the Mirrlees Review, a major five-year review of the whole UK tax system. His past research also includes evaluating the pilots of the *Pathways to Work* reforms to incapacity benefits and assessing the likely impacts of Universal Credit and the coalition government's other welfare reforms on work incentives and labour supply.

Dr Monica Costa Dias is a senior research economist at the IFS and Co-Director of Programme Evaluation for Policy Analysis (PEPA), a node of the National Centre for Research Methods. Monica has taught empirical policy evaluation methods and microeconometrics widely, developed software for policy evaluation, and written accessible survey articles on alternative evaluation methods. She has extensively used a broad range of empirical evaluation methods in her research. Her current research interests include the determinants and dynamics of education, human capital and labour supply, the evaluation and design of education and employment policies in dynamic economies and the development of econometric methods for policy evaluation.

Dr Barbara Sianesi is a senior research economist at the IFS. She holds PhDs in Economics from Bocconi University, Milan, and from University College London, the latter one focusing on the evaluation of social programmes and educational qualifications. Barbara has been involved in the evaluation of numerous government programmes (e.g. the Employment Retention and Advancement Demonstration, the Adult Learning Grant and the Swedish Active Labour Market programmes), in the analysis of the consequences of mis-recorded treatment status, and in the estimation of the impact of different types and levels of education on individual labour market outcomes in the UK. She has published research in both the programme evaluation and the returns to education fields.

Executive summary

The introduction of Universal Credit (UC) is arguably the most radical restructuring of the benefits system since the 1940s. Between April 2013 and December 2017 it will gradually replace six major means-tested benefits and tax credits for those of working age: Income Support (IS), income-based Jobseeker's Allowance (JSA), Child Tax Credit (CTC), Working Tax Credit (WTC), income-related Employment and Support Allowance (ESA) and Housing Benefit (HB). The aim is to strengthen work incentives and to make the system both operationally simpler (therefore, reducing administrative costs, error and fraud, making it easier for people to claim their entitlements, and smoothing transitions into work) and simpler to understand (therefore, boosting take-up and making work incentives more transparent). The importance of this policy to the Government's labour market strategy and anti-poverty strategy calls for a robust assessment of its impact on employment, the number of workless households, hours worked, wage rates and poverty.

This report was largely written during the winter 2012-13 and reflects the best information available to us at that time. The analysis in this report is based on a hypothetical scenario for UC roll-out given to us by the Department for Work and Pensions (DWP). This scenario is not the final plan for UC but provides a very useful detailed picture of roll out which we can use to understand the feasibility of evaluation. Much of the analysis here can readily be adapted to slightly different roll-out plans that might be implemented.

In this scenario, the UC regime replaces new claims to existing benefits in four stages:

- 1 replacing new JSA claims for single unemployed claimants without children in certain initial 'Pathfinder' areas of North-West England;
- 2 replacing all new JSA claims;
- 3 replacing new claims to tax credits (TC) and IS for lone parents (ISlp); and
- 4 replacing new claims to IS for carers (ISc), ESA and HB.

Once all new claims to legacy benefits have been stopped, in this scenario the Government will start to transfer the stock of existing legacy benefit claimants to UC in subsequent phases.

The central challenge for any evaluation is identifying the counterfactual: finding a way to estimate what would have happened in the absence of a reform, usually by looking at what happens to a comparison group which is unaffected by it and assuming that the affected group would otherwise have experienced the same outcomes.

Ideally, we would like to estimate the impact of the full change from legacy benefit regime to UC regime on (short-run and long-run) labour market outcomes for the whole population. Unfortunately, in this scenario the way in which the policy is introduced across the country does not allow that. Specifically, an assessment of the effects of the whole regime change on aggregate variables such as employment or poverty rates is hindered by the facts that:

- UC is phased in over several years, precluding any sensible comparison of the situation before UC starts to be introduced with the situation when UC is fully in place; and
- UC is introduced for the whole working-age population across the whole country, so there is no group unaffected by the policy that can be used to identify how outcomes might have

been different if UC had not been introduced.

In this study we, therefore, devote our attention to looking at how this scenario's gradual roll-out of UC to different claimant groups in different areas at different times might provide opportunities to estimate its effects for certain groups.

Evaluation of UC during the scenario's roll-out to new benefit claims

We identify three evaluation questions to consider:

Q1. The impacts of the full UC regime versus the full legacy benefit regime for new claimants.

The focus here is on how the outcomes of people starting a claim are affected by the full UC regime as compared to the full legacy benefit regime.

Q2. The aggregate impacts of UC replacing some new benefit claims versus maintaining the full legacy benefit regime (or some hybrid regime with more limited UC coverage).

The focus here is on how the outcomes of the entire working-age population or of specific groups of interest such as those more likely to make a new claim (e.g. low earners, parents with young children, lone parents) are affected by the change in benefits system, from the full legacy benefit to a hybrid regime where UC has replaced new claims to certain legacy benefits.

Q3. The indirect impacts of UC replacing some new benefit claims versus maintaining the full legacy benefit regime (or some hybrid regime with more limited UC coverage) for those not directly affected by the reform.

The focus here is on how the outcomes of those who are not directly affected by the policy reform may change as a consequence of UC directly affecting the labour-related choices of a large group of individuals.

We consider the scope for addressing these questions for each phase of the scenario in turn.

Phase 1: The key problem with evaluating the Pathfinders is that UC as implemented may not be fully representative of UC as it will ultimately be implemented across the country: it may be deliberately implemented slightly differently, there may be teething problems or unfamiliarity, or there may be more attention and resources devoted to the Pathfinders. External validity is further reduced by UC in this phase only directly affecting one small and very specific group of claimants. That aside, the Pathfinder areas offer a very good opportunity to get reliable and potentially robust estimates of the impacts of starting a new UC claim under the full UC regime versus starting a similar JSA claim under the full legacy benefit regime for single unemployed claimants without children in the Pathfinder areas (Q1). Given the low volume of UC claims, aggregate effects are unlikely to be detectable (Q2) and we would not expect indirect effects to be of any relevance (Q3).

Phase 2: This phase provides the best opportunity for a 'clean' evaluation of the impact of starting a new UC claim under the full UC regime versus starting a similar JSA claim under the full legacy benefit regime for unemployed claimants (Q1). There are potential threats to the validity of these estimates; in particular:

- entry effects: UC availability may change the composition of new claimants under UC compared to the composition of new claimants under the equivalent legacy benefit. UC

Evaluating the labour market impacts of Universal Credit: a feasibility study

entails important changes in entitlement and generosity of the welfare system by level of earnings, particularly for those in mini-jobs, affecting incentives to start a new JSA-equivalent claim; and

- anticipation effects: The composition of new claimants in the comparison group might be affected if potential claimants to the legacy benefit react in anticipation of the foreseen introduction of UC in that area. Moreover, the outcomes of the comparison groups can also be affected if the comparison group reacts in anticipation of the imminent roll-out of phase 3, which might affect the incentives of new JSA claimants to move into work, choose particular working hours or make a new TC claim.

Still, such effects may be mild, and anticipation effects are particularly unlikely to be of any concern when making comparisons over time (rather than across areas) with a relatively large (say one year) time interval between the observations of treated and controls. Furthermore, one can test for their presence to some extent, as well as getting corroborative evidence on the other assumptions (common trends between treatment and comparison groups, and historical comparability of matched areas). More than one method can be applied using both cross-area and cross-time variation in implementation of UC to test the robustness of any effects uncovered, and impacts can be assessed over a relatively long-time horizon. Given these factors, we expect the reliability of answers to Q1 to be quite high. The parameter estimated is also highly policy relevant, given that by then UC would be an established policy and UC claimants would be representative of all unemployed claimants. As was the case in phase 1, however, an informative study of aggregate effects (Q2) or of indirect effects (Q3) looks less promising due to the limited volume of UC claims.

Phase 3: This is a crucial implementation phase in this scenario, when UC is extended to major target groups (low earners and non-working lone parents) and significant changes in entitlement are introduced for new claimants: eligibility will be extended to new groups (young low earners without children and those working less than the WTC hours threshold) while others will gain or lose entitlement because of the change to the means-test. Crucially, the characteristics that determine entitlement can be manipulated and individual decisions on whether to work, for what wage rate and how many hours may well respond to these new incentives. These changes in entitlement rules and their consequences for the composition of new claimants compromise the evaluator's ability to construct an adequate comparison group to such an extent that Q1 cannot be reliably addressed in this phase. We identify two major evaluation possibilities, which in practice can be implemented using a set of alternative methods:

- 1 The impact of UC replacing new JSA/TC/ISlp claims, versus the full legacy benefit regime or versus a hybrid regime in which UC has replaced new JSA claims only, on inflows into benefits and on future labour market outcomes, for the entire working-age population or some specific sub-groups of interest (e.g. all parents, lone parents, low earners, low educated – Q2);
- 2 The indirect impact of UC replacing new JSA/TC/ISlp claims versus a hybrid regime where UC only replaces new JSA claims, for those who started a UC claim (that would previously have been JSA) before the roll-out of phase 3 (Q3).

The limited volume of UC cases may be less important at this stage when assessing Q2 and Q3. In phase 3 it is crucial to understand the impact of extending eligibility to new groups, which makes Q2 a key evaluation question. Q2 deals appropriately with the consequences of changes in entitlement rules and how these may affect claim decisions by accounting for

all forms of effects from the reform (i.e. direct effects on claim decisions, direct effects on other labour market outcomes and indirect effects arising through market-wide effects). Q3 can here consider potential indirect impacts only on groups that have already claimed UC – but these are groups of great policy interest as well as disproportionately likely to be affected by such indirect effects.

Phase 4: In this phase of the scenario, Q1 can be addressed by comparing those making a new claim to UC after October in Year 2 of implementation only with those making an equivalent new claim for ISc, ESA or HB – before October Year 2. This phase provides less insight than previous phases as to the impact of UC on claimants for two main reasons:

- only those making such claims who had not already made other new claims that would bring them within the UC net would be included – a rather restricted and selected sample (especially for HB claims);
- the UC regime would be compared to hybrid regimes in which UC had replaced new JSA claims or new JSA/TC/ISlp claims (rather than the full legacy benefit regime), unless the comparison group made their claim a considerable time before October in Year 2 of implementation, which would render their comparability more questionable.

On the other hand, in this scenario with a one-month roll-out plan and full coverage of all new claims, the volume of UC cases might now be sufficient to detect aggregate impacts, while indirect effects might be non-negligible. This phase, therefore, offers potentially promising avenues to assess the aggregate effects of UC (Q2), and, for a restricted group, its indirect effects (Q3):

- The impact of UC replacing all new claims versus the full legacy benefit regime or versus a hybrid regime (in which UC replaced only new JSA claims or only new JSA/TC/ISlp claims), for the entire working-age population or some sub-group; this is a highly policy relevant parameter, capturing direct and indirect effects of the reform (Q2).
- The indirect impact of UC replacing all new claims versus a hybrid regime in which UC replaces only new JSA/TC/ISlp claims, for those who started a UC claim close to the end of phase 3 – or versus a hybrid regime in which UC replaces only new JSA claims, for those who started a UC claim close to the end of phase 2; this is also a policy relevant parameter, capturing potential substitution and displacement effects of expanding UC coverage on groups who recently started a UC claim.

Evaluation of UC during the scenario's migration of existing claims

We identify two evaluation questions to consider when assessing the evaluation possibilities offered by the migration of existing claims in this scenario:

- **Impact of (notification of) migration to UC on the outcomes of existing claimants:**
For existing claimants, the evaluation strategy can be carried out within area and time period for as long as there is some random (or pseudo-random) variation in which claimants are migrated. Specifically, if, in any given month, some people are moved over to UC while other, similar, people in the same area are (essentially randomly) not moved over, their subsequent outcomes can be compared. If there is a concern that notifying people of their forthcoming migration may induce a change in behaviour (not shared by the

comparison group who are notified later), then the evaluator could estimate the effect of notification itself rather than actual migration. The most reliable question would, therefore, focus on the ‘impact of notifying in a given month a long-standing legacy benefit claimant that migration under transitional protection will happen in three months’ time if the claim remains open versus not notifying them in that month (and eventually notifying later should the claim remain open), for long-term claimants whose circumstances have remained stable, in a world where a new claim triggers moving into UC’. But this is not a very relevant policy parameter, for several reasons: the treatment is notification of forthcoming migration under transitional protection (which will be irrelevant in the longer term); the counterfactual is a hybrid regime where the comparison group will, at some point, move into the full UC regime; and such impacts refer to an unrepresentative population, whose composition is influenced by the local roll-out history and individual preferences for UC and who would already be subject to UC if they changed their circumstances (‘natural migration’).

- **Indirect impacts of migration:** If, in some areas, a large volume of existing legacy benefit claims are moved to UC, while comparable areas have not started the migration of that type of claim yet, the migration phases in the hypothetical scenario also offer the possibility to assess potential substitution and displacement effects of a large and relatively sudden influx into UC on the outcomes of people who are not directly affected by the migration itself, but who are already under the full UC regime. At this stage it is not clear whether the required variation will be available; furthermore, such effects could be assessed only over a limited horizon, determined by how soon after completion of the (bulk of) the migration in the treated area the control area starts migrating its own stock.

Long-run versus short-run impacts and the scope for disaggregation

The impacts of UC that can be evaluated in this scenario are generally short term, in two main senses. First, they are impacts on short-run outcomes. It is typically only a few months until the comparison group is exposed to the policy and the impact of UC versus the full legacy benefit regime can be estimated. If it takes time for people’s behaviour and outcomes to respond fully to the change in incentives etc. they face, estimated short-run impacts may not be a good guide to the effect that UC has on them in the longer run. In some cases, slightly longer-run outcomes can be estimated; but these are usually less reliable and/or become estimates of the impact of UC versus a hybrid regime in which UC has replaced some parts of the legacy benefit regime, but not all of it.

Second, the estimates will typically represent the impact of UC as claimants experience it soon after implementation. Transitional protection is a feature built into the policy in the short-run, and it is hard to see how one could estimate the impact of UC once transitional protection has ceased to be relevant. More broadly, the Government might change aspects of UC over time – ironing out teething problems, for example – while people might become more familiar with the new system over time. Some of the evaluation exercises we examine will be closer to the long-run implementation of UC than others. But it is hard to find reliable comparison groups against which to assess later versions of UC. Additionally, only a small fraction of the population will be under UC in the short run, and the gradual national roll-out over claimant types means that it will take some years before everyone falls under the new system in this scenario. Since indirect effects depend crucially on the size of the affected population, it is not expected that the indirect effects identified in the shorter term and for specific groups are informative about the potential indirect effects of UC when fully implemented.

Where impacts of UC can be estimated, it should generally be straightforward to disaggregate the estimated impacts by sub-group according to observed variables. In contrast, there is little scope for disaggregating the impacts by cause or aspect of the policy (incentives, conditionality, simplification, etc) since the policy in this hypothetical scenario is planned to be implemented as a full package everywhere. Introducing variation in UC in different areas, for example, might allow for evaluation of the impact of that variation.

Data

This analysis shows that, in principle, the policy implementation scenario set out by DWP does offer some, albeit less than ideal, evaluation opportunities. In practice, however, the feasibility and value of these evaluations depend on what data is available to perform them. In this case the data have severe limitations.

The most promising survey data for this evaluation are the Labour Force Survey (LFS), a large representative survey which contains detailed information on labour market outcomes and background characteristics, and has the crucial advantage of following people over time, interviewing them in five successive quarters. But while the LFS could be valuable for addressing some of our questions, it has significant limitations. The sample size – while large by survey standards – is still too small to focus on new benefit claims. And the LFS is of little use for looking at impacts on existing legacy benefit claimants of being moved onto UC (versus not yet being moved at that time) since the dates of notification and migration will not be recorded precisely in the data. For most purposes, therefore, we are reliant on administrative data.

Until recently, the principal administrative data that are relevant for this evaluation was information held across Work and Pensions Longitudinal Study (WPLS) and UNITE databases. This has universal coverage of those who have claimed a benefit or tax credit since the late 1990s (including data from before and after their claim) and includes longitudinal information on benefit claim history, employment spell start and end dates and annual earnings. However, the data are limited. There is no information on hours of work for most of the sample; the employment and earnings data exclude self-employment and employment below the National Insurance Lower Earnings Limit (LEL); there is severe measurement error in employment start and end dates; and earnings are recorded only on an annual basis across all employments. This means that the only outcomes that can be examined are:

- **recorded employment**, meaning that impacts on self-employment and low-paid employment (both of significant policy interest for UC) cannot be assessed. Estimates of UC's impact on recorded employment could also be biased if changes in recorded employment are attributed to UC when they in fact result from a change in the LEL taking people's jobs in or out of WPLS/UNITE records; and
- **recorded earnings**, which could lead to biased estimates of the impact of UC on earnings if UC has an impact on earnings below the LEL (since earnings below the LEL must be set to zero before and after the reform, in the absence of any data). Furthermore, recording only annual earnings means that impacts within a fiscal year can only be assessed for specific groups for whom a given tax year's earnings can confidently be considered wholly post-treatment.

From April 2013, Her Majesty's Revenue and Customs (HMRC) began collecting Real Time Information (RTI) data. As compared to WPLS/UNITE, this dataset is much better suited to this evaluation. It has all the advantages of WPLS/UNITE data, but without some of the

Evaluating the labour market impacts of Universal Credit: a feasibility study

limitations: it includes data on all employment spells (including below the LEL), exactly apportioned earnings within the year, and records hours of work. The data are also expected to be more accurate and the linking of partners' data more reliable. A major drawback is the continued absence of information on self-employment.

A key recommendation of this report is that significant care and effort should be put into obtaining the best available data for evaluation purposes. This includes RTI data (and income tax self-assessment data for the self-employed, if possible) for the entire WPLS/UNITE sample (or a large, representative sub-sample), with the RTI and benefits data merged. Unless RTI (and self-assessment) data are collected for everyone (or for representative samples of those under UC and the legacy benefit regimes), its use is ruled out for most purposes: it is crucial for any evaluation study that data quality is consistent across the groups being compared.

Conclusion: the most promising evaluation possibilities and the data available

Despite the methodological hurdles, there are a number of policy-relevant questions that it appears the scenario provided by DWP would allow to be addressed relatively reliably if data allowed. Of course, we can only speculate on how important the various threats to the validity of the estimates will be: there is no a priori guarantee that, were these evaluation exercises carried out, the resulting estimates would be reliable. How reliable such estimates could be deemed in practice would instead hinge on the results emerging from the various ancillary and corroborating analyses that have been suggested to provide indicative evidence on the plausibility of what are inherently untestable assumptions. With this caveat in mind, we believe that the most promising evaluation questions, and the data that might be used to address them, are:

1 The impact of making a claim under the full UC regime versus the full legacy benefit regime on people starting what would previously have been a JSA claim. This is a very interesting treatment effect, comparing outcomes under the full new regime to those under the full legacy benefit regime for important groups of claimants who are exposed to UC for their first time. Indeed, these claims are expected to represent a majority of all new UC claims.

This question could be addressed well if RTI data were available for new legacy benefit claimants. Otherwise, the only option would be to proceed with WPLS/UNITE data, though that carries severe limitations in terms of the impacts that can be measured and potential biases in them.

2 The impact of UC replacing new JSA, TC and ISlp claims versus the full legacy benefit regime (or versus a hybrid regime in which UC has replaced only new JSA claims), on the entire working-age population or some policy-relevant sub-group thereof. This is a highly policy-relevant parameter in phase 3 of the scenario provided, a key implementation phase which extends UC to major target groups (low earners and non-working lone parents with pre-school children) and introduces significant changes in entitlement for new claimants. By taking into account the consequences of these changes in entitlement rules and how these may affect claim decisions, this parameter captures all effects of the reform – the direct effects on who claims UC (changes in who is entitled given their current behaviour, people changing their behaviour to become entitled, and take-up among those entitled), the direct effects on claimants' employment and other labour market outcomes, and the indirect effects arising through market

interactions.

This question could be addressed well if RTI data were available for, for example, previous benefit claimants. Otherwise an analysis could be attempted using LFS data, but detection would be more difficult with the smaller sample involved.

- 3 The indirect effects of UC replacing new JSA, TC and ISIp claims, versus a hybrid regime, on those who had already claimed UC.** The choice of reference population as those who have already made a UC claim, though driven by methodological considerations, actually focuses attention on a group that is both of high policy relevance, given their dependence on benefits, and disproportionately likely to experience any indirect effects of the expansion of UC to new types of claim. While this type of indirect effect can be assessed exploiting the roll-out of UC in both phases 3 and 4 of this scenario, phase 3 might be of particular policy interest. Specifically, in phase 3, the indirect effects are those of UC replacing new JSA/TC/ISIp claims, compared to a hybrid regime where UC has replaced only new JSA claims, on individuals who in the past have started a UC claim (which would previously have been to JSA).

This question can be straightforwardly addressed if RTI data were available for those already under UC.

- 4 More tentatively, the indirect (substitution and displacement) effects of moving a large group of existing legacy benefits onto UC, versus a hybrid regime in which a new claim would already be to UC ('natural migration'), on those already subject to the full UC regime.** The feasibility of addressing this evaluation question is more uncertain, as it is not yet known whether the required source of variation will be available (i.e. whether in some areas a large group of legacy benefit claimants will be moved to UC while comparable areas will not have started the migration of that type of claim yet). Additionally, substantial (and therefore detectable) indirect effects might be expected only from migrating some claimant types, specifically long-term TC claimants.

This could be addressed adequately using the LFS. RTI data, if available for everyone in the WPLS/UNITE sample or some sub-group of interest, offer a much bigger sample size.

In the hypothetical roll-out scenario we were asked to consider, a robust ex-post quantitative evaluation of the overall labour market impacts of UC would not be possible. There would, however, be some scope for feasible evaluations of certain short-run impacts for particular sub-groups. If this roll-out scenario were to be implemented, by far the biggest contribution to the feasibility of the evaluation would be to ensure that RTI data for the whole WPLS/UNITE sample are available to the evaluator.

1 Introduction

The introduction of Universal Credit (UC) is arguably the most radical restructuring of the benefits system since the 1940s. Between April 2013 and December 2017 it will gradually replace six major means-tested benefits and tax credits for those of working age: Income Support (IS), income-based Jobseeker's Allowance (JSA), Child Tax Credit (CTC), Working Tax Credit (WTC), income-related Employment and Support Allowance (ESA) and Housing Benefit (HB). The aim is to strengthen work incentives and to make the system both operationally simpler (therefore, reducing administrative costs, error and fraud, making it easier for people to claim their entitlements, and smoothing transitions into work) and simpler to understand (therefore, boosting take-up and making work incentives more transparent). The importance of this policy to the Government's labour market strategy and anti-poverty strategy calls for a robust assessment of its impact on employment, the number of workless households, hours worked, wage rates and poverty.

The report was largely written during winter 2012-13 and reflects the best information available to us at that time. The analysis in this report is based on a scenario for the provisional roll-out of UC given to us by the Department for Work and Pensions (DWP). This is not the final plan for roll-out. Although the scope and options for evaluation might change much of the analysis here can readily be adapted to different roll-out scenarios.

In this scenario, the UC regime will replace new claims to existing benefits in four stages:

- 1 replacing new JSA claims for single unemployed claimants without children in certain initial 'Pathfinder' areas of North-West England;
- 2 replacing all new JSA claims;
- 3 replacing new claims to tax credits (TC) and IS for lone parents (ISlp); and
- 4 replacing new claims to IS for carers (ISc), ESA and HB.

Once all new claims to legacy benefits have been stopped, the Government will start to transfer the stock of existing legacy benefit claimants to UC in subsequent phases.

The central challenge for any evaluation is identifying the counterfactual: finding a way to estimate what would have happened in the absence of a reform, usually by looking at what happens to a comparison group that is unaffected by it and assuming that the affected group would otherwise have experienced the same outcome. This study provides a guide to the possibilities for identifying a counterfactual and evaluating the impact of introducing Universal Credit. It does not, of course, address all the detailed questions that would have to be dealt with in the course of the evaluation itself.

The remainder of this report is structured as follows.

Section 2 examines the particular labour market outcomes of interest.

The core of the report consists of a detailed examination of the evaluation possibilities that each phase of the scenario allows in principle, looking first at the four phases of the roll-out for new claims (section 3) and then at the migration of existing claims (section 4). Section 3 starts by outlining the three broad evaluation questions we consider when assessing the evaluation possibilities during the roll-out of UC to new claims (section 3.1). It then discusses the sources of variation that could be used to address these evaluation questions (section 3.2) and describes the preferred evaluation methods for the evaluation questions and sources of variation described before (section 3.3). More detail on the methods can be found in Appendix A, which also contains a description of other, less promising, evaluation methods. The evaluation possibilities in each phase of the implementation for new claims are discussed in section 3.4. Specifically, we investigate what can be learnt from each phase to address each of the three questions and discuss the evaluation challenges that convincing studies will need to address. Throughout it is assumed that the required data are available. Details of each suggested evaluation study are tabulated systematically in Appendix C.

In section 4 we discuss what can be learnt from the migration of existing legacy benefits to UC. We start by discussing the sources of variation which will potentially be available for evaluation purposes (section 4.1), and then consider two evaluation questions exploiting those sources of variation (sections 4.2 and 4.3). We conclude with an overall assessment of the evaluation possibilities offered by the migration phases (section 4.4) in this scenario. Details of the two suggested evaluation studies are tabulated in Appendix C.

Section 5 is devoted to an overall assessment of evaluations that are possible in principle. It starts by summarising and assessing the evaluation possibilities examined in detail in sections 3 and 4. Specifically, in section 5.1 we consider each of the three evaluation questions we identified for each implementation phase, reaching an overall view as to which ones are the most promising (in terms of methodological reliability and policy relevance) evaluation exercises to pursue. Section 5.2 focuses on one wide-ranging challenge: how to isolate the impacts of UC from those of other changes happening around the same time. In sections 5.3 and 5.4 we give a systematic overview of the extent to which both short-run and long-run impacts can be estimated, and of the extent to which estimated impacts can be disaggregated by sub-group of the population and by aspect of the policy. In light of the challenges to evaluation posed by the scenario provided, in section 5.5 we discuss alternative implementation strategies that might facilitate the evaluation of the impacts of UC.

In Section 6 we turn to look at what data will actually be available, and the implications of that for which of the evaluation questions it will be possible to address in practice and for which of the outcomes of interest.

Section 7 concludes.

Additional and in-depth material is provided in appendices. Appendix A describes the main evaluation methods and discusses their applicability to evaluating the impacts of UC, in particular the assumptions they rely on, their potential weaknesses and ways to gain corroborative evidence on whether their underlying assumptions are likely to be met. In Appendix B we set out in general terms the evaluation challenges that arise when using the roll-out of UC to new claimants to evaluate its impact – these challenges are applied to each phase of the implementation in section 3.4. Appendix C contains 10 detailed evaluation design summary tables, one for each identified evaluation question in each implementation phase. Appendix D outlines evaluation possibilities arising from the associated reform of Pension Credit.

2 Outcomes of interest

We can think of a number of outcomes of interest. Those mentioned in the terms of reference are:

- 1 employment;
- 2 workless households;
- 3 working hours;
- 4 wages; and
- 5 poverty.

2.1 Employment

This is perhaps the most important outcome and that for which evaluation strategies are easiest to devise. However, assessing the impact of Universal Credit (UC) on aggregate employment rates may be difficult because in this scenario UC implementation will happen gradually over several years. Such a long phase-in period limits the scope of before-after comparisons as comparability cannot be ensured over long time periods and means earlier comparisons will only represent a small proportion of the population.

The alternative is to look at employment impacts for specific sub-groups. One could focus either on the impact on employment rates or on flows into and from employment. The choice of the outcome of interest is likely to depend on the group being studied.

For instance, when measuring the impact of UC replacing Jobseeker's Allowance (JSA) on new claimants, it may make sense to look at employment rates and flows into employment. In this case, outflows from employment are more difficult to study because they can only be observed for a selected subpopulation of the treated/comparison groups: those who have moved into employment, which in itself is an endogenous outcome. That is, to measure the impact of UC on the outflows from employment of new JSA/UC claimants, one needs to deal with two selection mechanisms: the flow into treatment (meaning flow into UC when making a new JSA claim) and the flow out of unemployment into employment. While both can be endogenously affected by UC, dealing with the latter can be notoriously more difficult, particularly shortly after the introduction of UC when lack of information on the new system can limit the extent of endogenous selection into benefits. Notwithstanding, the possibility of selection into treatment should be seriously considered when studying the impact of UC on new claimants, even in the early implementation stages. We recommend that empirical support on the comparability of the treated and comparison groups is carefully sought through ancillary analyses and the study of the impact of UC on inflows into benefits (see in particular section 3.1.2 and more generally the discussion in section 3; see also the last outcome of interest in this section).

By the same token, but for different groups, such as those making a new Tax Credit (TC)/UC claim while working, measuring the impacts of UC on employment rates and outflows from employment is simpler than measuring the impact of inflows into employment after a future unemployment spell.

Identifying impacts on employment rates or flows will be more difficult if the implementation of UC affects the inflow into what would otherwise be a legacy benefit. For instance, this can happen with the introduction of UC in place of new JSA claims if decisions to claim JSA/UC depend on the regime the individual is facing if starting a claim. However, we can gather empirical support for the comparability of groups exposed to different regimes and starting a claim that may trigger UC depending on time and place.

2.2 Workless households

Impacts on whether families (or households) have someone in work can be estimated in the same way, and subject to the same limitations, as individual employment. The only additional condition is that it must be possible to observe the employment status of both members of the couple (or all adults in the household) in the data.

2.3 Working hours

Just like employment, total hours worked can be observed for everyone, being zero for anyone out-of-work. Working hours can also be dealt with in the same way as employment. Effects on working hours measure a combination of the impact on employment and the impact on hours worked by those in work. One could look at average (mean) hours of work – including averaged over the zeros of non-workers – or at other features of the distribution of hours worked.

It is more difficult to assess the impact of UC on just the hours worked by those in work, since employment can itself be affected by the policy and so comparing the hours worked by those who work under UC with the hours worked by those who work under the legacy benefit regime may not be comparing like with like.

2.4 Wages

Looking at hourly wage rates suffers from the same additional selection problem mentioned above for working hours among workers: wages are only observed for workers and the composition of that group can be affected by the policy.

Identifying the impact of UC on individual wage rates would generally require additional assumptions – similar to those required to support a structural model that could help disentangle the mechanisms explaining the change in observed average wage rates (composition, reservation wage, skills formation...).

Without such additional assumptions, we could identify the impact on some average wage rate in the market, but such a measure would not be informative as to whether UC affected individuals' earnings capacity: even if UC did not change the wage rate that any individual could command, it might change the observed average wage because it changed the composition of who is working, and therefore, the observed distribution of wage rates.

2.5 Earnings or earned income

Not mentioned in the terms of reference, but (e.g. monthly) earnings are an alternative to (hourly) wages that can overcome the selection problem since they can be observed for everyone: positive among workers and zero for those out-of-work. The impact of UC on earnings arises through a combination of its impact on employment and wage rates.

2.6 Family income and poverty

Impacts on income can be studied in a way similar to employment or earnings. Insofar as impacts on income can be evaluated, one can also look at moments of the income distribution other than the mean, such as percentiles or probabilities of income being below a certain threshold – most obviously, impacts on the probability of equivalised income being below the poverty line.

2.7 Benefit entitlement and take-up

Although not mentioned in the terms of reference for this study, this may nonetheless be an interesting outcome to explore. It also, of course, determines the population for whom the other outcomes discussed above will be studied when looking at the impact on those making a claim. So the measurement of the effect of UC on claim rates, as well as being of interest in its own right, can be considered an ancillary analysis for the study of other outcomes among those who make a new claim.

3 What can we learn from the phasing in of Universal Credit for new claims?

The hypothetical scenario for Universal Credit (UC) roll out provided by DWP that was available at the time of writing consists of six phases. The first four phases concern the flow and will gradually turn off new claims to legacy benefits, while the last two phases pertain to the stock and will manage the migration of existing legacy claims.

In this section we study the evaluation possibilities in phases 1 to 4. The section starts by discussing the evaluation questions that we think are both relevant and can be at least partially addressed. It then discusses the sources of variation created by the UC roll out scenario that can be used to address each of the evaluation questions. Under the premise that the necessary data is available, the remainder of the section then discusses the appropriate methodologies and specific issues that apply to each implementation phase.

Two features of the migration scenario should be highlighted at this stage. First, those claiming multiple benefits will be moved to UC if UC has replaced any of the benefits concerned. Therefore, for example, once UC has replaced new Jobseeker's Allowance (JSA) claims in an area, someone making a new claim to what would have been JSA and Housing Benefit (HB) in the legacy benefit regime will instead be given UC, even if HB has not yet been replaced by UC in that area. Similarly, existing claimants of both JSA and HB will be migrated alongside other claimants of JSA, not alongside claimants of HB only.

Second, once an individual has started receiving UC (by whatever route), the full UC regime will apply for him or her from then onwards. This implies that such an individual will cease to be exposed to the legacy benefit regime and instead face the full set of UC incentives, including those associated with moving into circumstances where the legacy benefit regime is still in place for new claimants.

Estimates of the impact of extending UC to a particular group of new claimants will depend on, amongst other things:

- the counterfactual against which it is compared – for example, whether UC is being compared with the full legacy benefit regime or with a 'hybrid' regime in which UC is already available in some circumstances, but not others (and often more than one choice of counterfactual is possible);
- how earlier phases of UC were rolled out, since that can influence the composition of those who make their first UC claim at this later stage; and
- the existing breadth of coverage of UC among other groups, which may determine the importance of indirect effects on those other groups.

Therefore, the effects that can be identified during phases 1 to 4 depend on the specific details of the implementation strategy and are not necessarily representative of what the impacts of UC will be in the longer run, when fully implemented.

As will be discussed at length below, the rapid succession of phases proposed in the scenario limits the scope of the analysis to short-term effects. If the impact of UC takes time to build up, the evaluation approaches discussed in this section will fail to capture them.

In everything that follows, references to ‘new claims’ include ‘natural migration’ – that is, changes in circumstances of existing claimants that trigger a move onto UC.

3.1 Evaluation questions

This section discusses the main evaluation questions we will consider when assessing the evaluation possibilities in implementing phases 1 to 4. The choice of questions is determined both by what is of interest and by the expected feasibility of gathering related empirical evidence (even if only partially addressing the question). All questions are stated on a general level, independent of some phase-specific issues (discussed below in section 3.4). For each question, we detail the data required to address it.

3.1.1 Evaluation Question 1: Impact of UC on the outcomes of new claimants

The question posed here is: What are the impacts of the full UC regime versus the full legacy benefit regime on the outcomes of individuals starting a new claim of a certain type?

The focus here is on how the outcomes of claimants being moved into UC when starting a new claim are affected by the full UC regime as compared to the full legacy benefit regime. The question can be separately addressed by type of claim and claimant characteristics such as age, gender, family demographics or education. Outcomes of interest may include employment rates, earnings, working hours, claim status, family income and poverty, all measured some time after the start of the claim.

To address this question in any of phases 1 to 4 would require data on all (or a large random sample of) new claims of a given type (e.g. new JSA claims by single adults without children and its equivalent under UC) that are treated differently across similar areas or time periods.

3.1.2 Evaluation Question 2: Impact of UC on market-wide outcomes or the outcomes of specific groups

The question posed here is: What are the impacts of introducing UC for some new benefit claims versus maintaining the full legacy benefit regime (or some hybrid regime with more limited UC coverage) on the outcomes of specific groups defined on pre-reform characteristics (such as the less educated, those with lower earnings capacity, or lone parents)? More generally, what are the market-wide impacts of introducing UC for some new benefit claims versus maintaining the full legacy benefit regime?

The focus here is on how the outcomes of particular populations are affected by the change in benefits system, from the full legacy benefit regime to a hybrid regime where UC has replaced new claims to certain benefits (which claims trigger UC entitlement varies over time, starting with those to income-based JSA). The population of interest consists of all those who could potentially be entitled depending on their behaviour, not only the new claimants who are the focus of question 1. One could consider the entire working-age population, and study the market-wide effects of the new regime, or focus on specific groups of interest

such as those more likely to make a new claim. Conditions to define the evaluation group of interest should be established before the reform, here understood as the introduction of a new phase, to avoid contamination by endogeneity (whereby the variable used to define the group is affected by the treatment itself). Outcomes of interest may include the size of the claimant count, inflow into benefits, employment rates, moments of the earnings distribution, moments of the hours' distribution, moments of the distribution of family income and poverty rates, all measured from the start of the policy reform onwards.

To address this question in any of phases 1 to 4 would require data on all (or a large random sample of) the population of interest – can be all working-age population, those with low formal education, lone parents, workers on low earnings, etc. – who are exposed to different benefit systems (depending on what can trigger a move into UC) across similar areas or time periods.

3.1.3 Evaluation Question 3: Substitution, displacement and other indirect impacts of UC

The question posed here is: What are the impacts of introducing UC for some new benefit claims versus maintaining the full legacy benefit regime (or some hybrid regime with more limited UC coverage) on the outcomes of those not directly affected by the change of regime?

The focus here is on how the outcomes of those who are not directly affected by the policy reform may change as a consequence of UC directly affecting the labour-related choices of a large group of individuals. Indirect effects are market-wide, induced by changes in labour market tightness or wage rates. For instance, an increase in labour supply among UC recipients could make jobs more difficult to find for other job seekers and/or lead to a fall in the prevailing wage rate. Therefore, even those who are generally untouched by the benefits system (possibly because they do not fulfil the means test) may be affected if, for instance, they experience a period of unemployment and find it more difficult to return to work due to increased competition for the available vacancies. Changes in market tightness may lead to substitution effects (a drop in the probability of finding a job among non-UC job seekers) or displacement effects (those currently in employment are displaced by new workers).

There is rarely a group that is 'not directly affected by the reform' as each phase is implemented universally within an area, exposing all working-age adults to the new set of incentives. Direct effects are responses to the new set of incentives that occur, even in the absence of aggregate effects on labour market tightness, wage rates or other market-wide characteristics. New claimants are directly exposed to the new set of incentives and react to it. But others may also be directly affected. For instance, existing claimants of another benefit may change their behaviour to qualify for or avoid UC; those presently not on benefits may find it more or less attractive to be on benefits under UC than under the legacy benefit regime and change their attitude to work accordingly. Perhaps these individuals will not become a claimant straight away, but the possibility of claiming in the future and the amount of insurance provided by UC may directly affect their behaviour in the present. That is, non-claimants are not necessarily unaffected directly by the reform. Their decisions may be influenced by changes in the relative reward to each of employment, working different hours or being on benefits, even if such decisions do not lead to a new claim at present. Therefore, it is difficult to disentangle behaviour responses from the indirect effects of the new regime among non-claimants.

Evaluating the labour market impacts of Universal Credit: a feasibility study

There are, however, some exceptions to this:

- During phases 1 and 2 of the scenario, one can identify clear groups that are not directly affected by the new phase. In phase 1, these include all parents and individuals in couples, since the Pathfinders only include single people without children. In phase 2, they include lone parents with children aged under five, people with disabilities and carers, who would receive Income Support (IS) or Employment and Support Allowance (ESA) rather than JSA if they did not work and who are, therefore, unaffected by the replacement of JSA by UC. One can study the impact on these groups' outcomes of introducing UC for other groups. However, it is unlikely that this analysis would provide any conclusive evidence on the potential indirect effects of UC since in phases 1 and 2 UC replaces only new claims to only one benefit; since this will involve a relatively small number of claims as a fraction of the population, indirect effects are likely to be negligible.
- In phases 3 and 4 of the scenario, groups not directly affected by the new phase are those who started a UC claim before the roll-out of the new phase. This is because once an individual has claimed UC, the full UC regime applies to him or her from then onwards. Hence those who started a UC claim before the introduction of phase 3, or of phase 4, are not directly affected by the reform extending UC to new claimant types. In contrast to phases 1 and 2, substitution and displacement effects are likely to be more substantial in these later phases which expand UC coverage to all new claim types.
- The migration of the stocks of existing claimants. We return to this point in section 4.

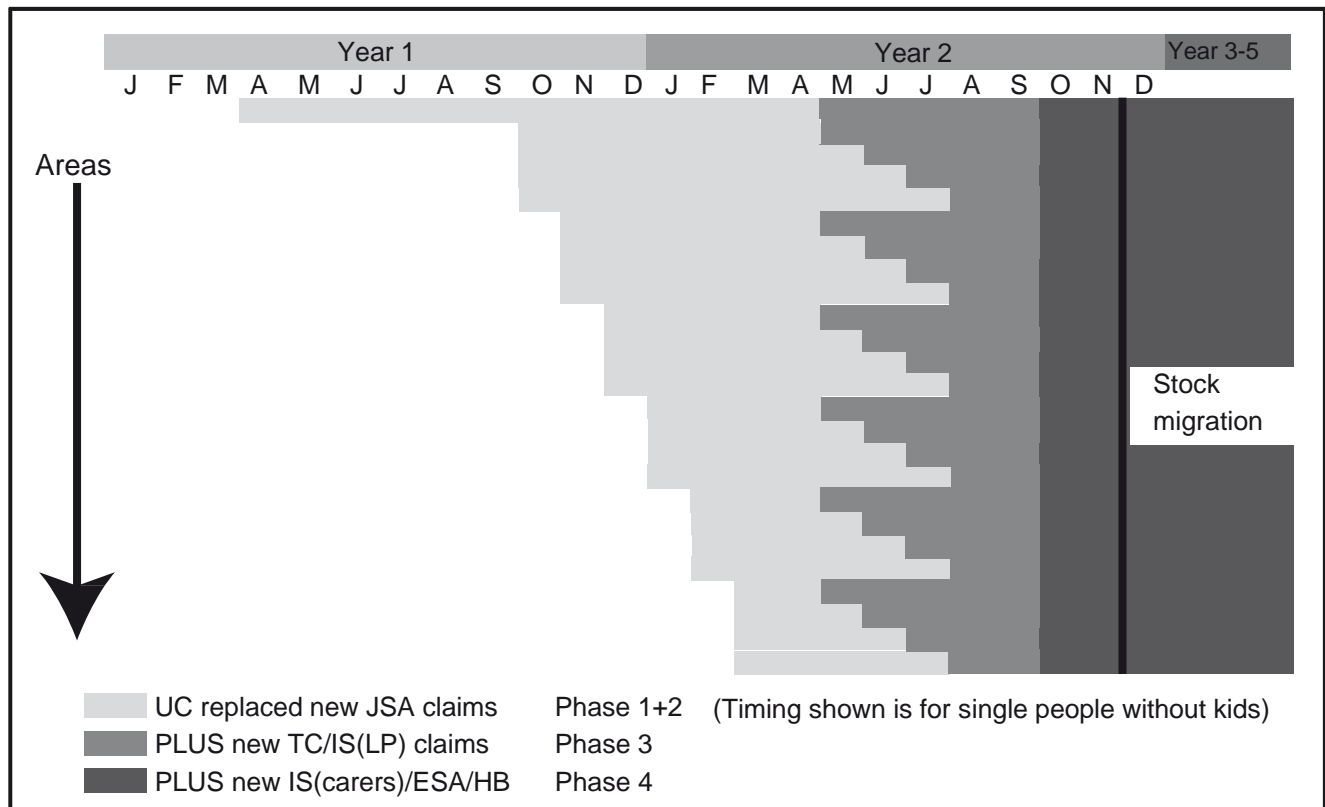
Outcomes of interest may include employment rates, earnings, working hours and claim status among non-claimants of some type, all measured some time after the introduction of the new phase under scrutiny.

To address this question in any of phases 1 to 4 would require data on all (or a large random sample of) people who are not directly affected by a new UC implementation phase and who are exposed to different benefits systems (depending on what can trigger a move into UC) across similar areas or time periods.

3.2 Sources of variation

Each implementation phase of the scenario examined would bring into UC new claims to additional benefits (see Figure 3.1), starting with income-based JSA for singles without children in Pathfinder areas in phase 1, all income-based JSA claims in phase 2, Tax Credit (TC) and Income Support for lone parents (ISlp) in phase 3 and Income Support for carers (ISc), ESA and HB in phase 4.

Figure 3.1 The provisional phase-in plan (stylized)



During the time it takes to extend UC to all new claimants of a given type there will be variation across areas in how similar claims are treated. Likewise, the introduction of the new policy at some point in time for each claimant type in each area implies that similar claims are treated differently over time.

That is, the hypothetical implementation scenario creates two sources of variation that can be useful to address the evaluation questions discussed above:

- 1 Some individuals will start a new claim in some areas under the UC regime while, at the same time, similar individuals starting a similar claim in comparable areas will be under the legacy benefit regime (geographical variation);
- 2 Some individuals will start a new claim at some moment in time under the UC regime while, in the same area, similar individuals have started a similar claim in an earlier period under the legacy benefit regime (time variation).

For instance, during phase 1 (the Pathfinders) new JSA claimants who are single and have no children will be moved into UC from April in Year 1 onwards (and be subject to the full UC system from then onwards) while (1) similar individuals starting similar claims in non- Pathfinder areas at the same time will be subject to the full legacy benefit and (2) similar individuals starting similar claims in Pathfinder areas prior to April in Year 1 were subject to the full legacy benefit regime. New claimants satisfying conditions 1 or 2 can be used to construct the missing counterfactual and estimate the impact of UC for new JSA claimants who are single and have no children.

For phases 1 to 3, both geographical variation and time variation in how similar claims are treated can be exploited to assess the impact of UC. However, phase 4 is implemented simultaneously in all areas, leaving only time variation to be exploited for the estimation of the impact of UC on new claims to ISc, ESA and HB.

Evaluating the labour market impacts of Universal Credit: a feasibility study

UC is a whole new benefit system that changes financial and non-financial incentives to claim benefits, work and earn more for the entire working-age population. Therefore, the sources of variation described above can potentially be used to identify the impacts of UC not only on the outcomes of those who become claimants, but also on the probability of becoming a claimant, on market-wide outcomes and on the outcomes of particular sub-groups perhaps more affected by the reform or perhaps not directly affected at all. In other words, these sources of variation can in principle be used to address all three of the evaluation questions discussed in the previous sub-section: the direct effects (changes in behaviour due to the change in incentives) and indirect effects (changes in outcomes due to market-wide effects on vacancy rates or wage rates) on those who become and who do not become UC claimants, and the effects on the wider working-age population, unconditional on claim status.

When exploiting time variation for evaluation purposes, the choice of the time period involves a trade-off between the time horizon for analysis and the reliability of estimates. The treated group, for whom one wishes to measure the impact of the UC reform, is observed after the reform (roll-out of new phase), while the comparison group is constructed from similar individuals observed some time before the reform, when UC was not yet available for them. On the one hand, a comparison group drawn longer before the time of the reform and who remain unexposed to UC via other routes can support the estimation of the same treatment effect parameter (UC versus, for instance, the pure legacy benefit regime) over a longer period and be less sensitive to changes in behaviour driven by the anticipation of an imminent policy reform (the parameters that can be identified and the anticipation effects are all discussed in detail below, in section 3.4). On the other hand, however, market conditions may change over even short periods of time (with time trends and seasonal variation) and make earlier comparison groups less reliable. Some robustness checking may be possible and is desirable, amounting to the comparison of the same parameter estimated on alternative counterfactuals (for alternative time periods).

3.3 Methods

We identify two evaluation methods that can be used to assess the impact of UC during the roll-out to new claimants: matching and difference-in-differences (DiD). The discussion below details some of the implementation issues and the relative strengths and weaknesses of each of these methods for the evaluation questions and sources of variation described above. More detail on the methods can be found in Appendix A.

3.3.1 Matching

Matching methods involve constructing the evaluation counterfactual – estimating what would have happened in the absence of UC – by reweighting the comparison group data to eliminate differences in observed characteristics between those exposed to a policy change, for whom one wishes to measure the impact (hereafter called the ‘treated’ group), and those who are not exposed to the policy change (the ‘comparison’ group). To minimise the risk of unobserved differences confounding the matching evaluation results, it is crucial to condition on a rich set of observed characteristics related to the labour market outcomes of interest. These may include demographics (age, sex, detailed family composition), education, past employment and benefit claiming history, family income (by source), and so on (see e.g. Heckman *et al.*, 1997).

In principle, one can consider using both geographic and time variation in how new claims are dealt with (some being moved into UC, others being kept under the legacy benefit regime) to draw the comparison group. For example, consider the evaluation of the impact of UC for singles with no children starting a new JSA claim in Pathfinder areas – this is Evaluation Question 1 for phase 1:

- using geographical variation means drawing the comparison group from similar individuals starting a new JSA claim at the same time in non-Pathfinder areas;
- using time variation means drawing the comparison group from similar individuals starting a new JSA claim in Pathfinder areas prior to the introduction of phase 1.

In practice, the use of time variation with matching may raise concerns over comparability in the presence of seasonal or cyclical variation. Choosing comparison groups observed closer to the treated group can minimise the risk of differential market conditions due to seasonal or business cycle variation. However, the horizon of the analysis is limited by the distance in time between treated and comparison groups, and the risk of anticipation effects is larger the closer the two groups are.

Geographical variation looks more promising but requires that comparison areas are selected to ensure that treated and comparison groups face similar labour market conditions. This implies a two-step matching approach for the evaluation of the impact of UC in phases 1 to 3:

- step 1: match treated and comparison areas on local labour market outcomes including the history of the outcome of interest, the history of UC roll-out and UC coverage (see e.g. Heckman *et al.*, 1997);
- step 2: match treated (in treated areas) and comparison (in comparison areas) individuals on a rich set of personal characteristics.

The assumption underpinning this matching methodology is that observationally similar individuals living in observationally similar areas would have experienced similar outcomes had UC not been introduced.

3.3.2 Difference-in-differences

DiD methods compare the change in outcomes among a group exposed to a policy reform (from before to after the reform) with a similar change in outcomes among a group not exposed to the policy reform. The typical before-after interval is 12 months, to minimise the scope for differential seasonal variation that could bias estimates. We consider two variants of this method, panel and cross-sectional DiD, and an extension to ‘triple differences’ that combines time and geographical variation. In all cases, the choice of the comparison group can be enhanced by prior matching on a rich set of observed characteristics.

Panel DiD

Panel DiD uses the same individuals in the ‘before’ period as in the ‘after’ period. One comparison group is selected, drawn from those not exposed to the policy reform in the ‘after’ period. Just as for matching, the selection of the comparison group should be carefully based on the observed characteristics of individuals including demographics (age, sex, detailed family composition), education, past employment and benefit claiming history, and family income by source. Panel DiD then compares the one-year difference in outcomes

Evaluating the labour market impacts of Universal Credit: a feasibility study

among the treated group with the one-year difference in outcomes among the comparison group.

The two sources of variation can be used with DiD:

- Geographical variation: the comparison group consists of matched claimants in similar areas not yet exposed to UC (that is, in areas implementing that phase of the reform later).
- Time variation: the comparison group consists of matched claimants in the same area some time before the policy reform.

Where geographical variation is used, it is important to choose the comparison areas carefully. This should be done based on two considerations:

- Areas with similar UC implementation histories and UC coverage, to minimise differential mechanical and other composition effects, whereby the composition of the treated population is affected by the past availability of UC to different claimants or by area-level characteristics.
- Areas with similar trends in relevant labour market variables, particularly the outcome of interest to minimise the risk of differential trends confounding the estimates of the impact of UC.

The comparison group can then be drawn from matched observations in earlier (pre-treatment) time periods, if exploiting time variation, or in matched areas, if exploiting geographical variation. Panel DiD can then be implemented by comparing the time change in the mean outcome observed for the treated group (from before to after treatment) with that for the comparison group over a similar time interval (as described in more detail in Appendix A).

The assumptions underpinning panel DiD are the following:

- 1 No differential trends between the treated and comparison groups.
- 2 No selection on transitory shocks that would cause differences in outcomes between treated and comparison groups even in the absence of treatment.

These assumptions cannot be tested directly. But indicative historical evidence on the likely importance of differential trends (assumption 1) can be gathered by computing the DiD estimates for earlier (wholly pre-reform) periods and assessing their time variation (see Blundell *et al.*, 2004). In the presence of past differential trends it is still possible to gauge their likely importance for the precision of the DiD estimates (see, in particular, Cameron *et al.*, 2008, and Bertrand *et al.*, 2004). Some empirical evidence on whether UC changes the selection mechanism into benefits (assumption 2) can also be gathered by, for example, comparing the distribution of observed variables among the treated with its historical counterpart (entry effects, discussed at length in section 3.4). Although not necessarily informative about how selection on the unobserved drivers of behaviour responds to the new incentives, the absence of significant changes in the distribution of observed variables for the group of interest with the policy reform is reassuring evidence that the same may apply to the distribution of unobserved variables. All this assumes the availability of long-term historical data.

Specific forms of differential trends – those that are uniform across areas when using time variation (so that a comparison of DiD estimates for different areas, some being exposed to the reform at a point in time, others not and used to identify the differential trend) or that are uniform over time when using geographical variation (so that DiD estimates over some

time period prior to the reform identify the differential trend – can be dealt with by the triple differences estimator described below (see Bell *et al.*, 1999).

Cross-sectional DiD

The time (before-after) differences in cross-sectional DiD use different samples, both for treated and comparison groups. That is, cross-sectional DiD requires the construction of three comparison groups: the treated before the reform, and comparisons both before and after. Just as for panel DiD, careful matching on individual characteristics to reproduce the composition of the treated group is recommended here as well.

Both time and geographical variation can be used with cross-sectional DiD, in all cases having to draw three comparison groups. For geographical variation to be used, and just as for the panel DiD described above, a careful choice (match) of comparison areas based on historical trends in the labour market outcome of interest (and other labour market characteristics if possible) and the history of UC implementation is recommended.

The assumptions underlying cross-sectional DiD are the following:

- 1 No aggregate trends with a differential impact on before-after differences for the treated and comparison groups;
- 2 No selection on transitory shocks leading to differential impacts on before-after differences for the treated and comparison groups in the absence of treatment;
- 3 Invariant composition of matched treated and comparison groups one year apart.

The same exploratory analyses of common trends described for panel DiD apply here as well.

Triple differences: DiDiD

Differential trends bias DiD estimates of the impact of UC, since the estimates conflate the impact of UC with group-level changes that would have happened independently of the reform. For instance, the common trends assumption is violated for DiD using time variation if the size of the seasonal effect depends on the business cycle or if there is a one-off macroeconomic shock. Likewise, if using geographical variation, different areas may not have exactly the same trends. In such cases, it may be useful to combine both time and geographical variation to strip out the bias induced by differential trends. This involves estimating the impact of UC using time variation in treated areas (which implement a new phase earlier) and in a matched sample of non-treated areas (which implement the phase later) and taking the difference between the two estimates to cancel out the impact of differential trends.

This method changes DiD assumption 1 to:

- 1 Same differential trends across treated and matched comparison areas for local treated and comparison groups selected on time variation in access to UC.

The main advantages of triple differences is that it can be used over a long period of time or contrasting different groups of areas to test the historical or geographic significance of the estimated effects. Instead of having to select a comparable period prior to the reform or a comparable group of areas, one can implement triple differences against each untreated earlier period or each sub-group of untreated areas and construct a set of comparable estimates in the absence of a reform. These identify the amount of variation that is observed

Evaluating the labour market impacts of Universal Credit: a feasibility study

in the data. To put it differently, they characterise the typical differences in trends between the groups being compared to estimate the impact of UC.

As compared to DiD, triple differences has two main drawbacks:

- i. it can reduce the precision of the estimated treatment effect; and
- ii. if confined to a single comparison, may introduce (more) bias if the comparison is not appropriate.

Triple differences can be combined both with panel and cross-sectional DiD.

3.4 Phase-by-phase assessment

This section discusses the evaluation possibilities in each phase of the hypothetical scenario for new claims, and the issues they raise. We investigate what can be learned from each phase to address each of the three questions discussed in section 3.1 above. We also discuss the evaluation challenges for each phase that convincing studies will need to address (Appendix B sets out these challenges in more general terms). These are partly driven by the scenario design and become more serious when assessing the impact of UC in later phases. Throughout this section it is assumed that the required data are available; data issues are discussed later, in section 6. Details of each suggested evaluation study are tabulated systematically in Appendix C.

3.4.1 Phase 1: Pathfinder

In the scenario provided by DWP the Pathfinder phase runs in four areas for the first six months of UC implementation and introduces UC in place of all new claims to JSA by single people without children in those areas.

Evaluation Question 1: Impact of starting a new UC claim under the full UC regime versus starting a new JSA claim under the full legacy benefit regime

In phase 1, this question can be addressed for those unemployed single people without children satisfying the JSA conditions who make a claim in the Pathfinder areas in the first six months (from April to October of Year 1).

The measurement of treatment effects relies on the comparison of two groups:

- **Treated:** single people without children living in Pathfinder areas who are unemployed and flow into UC sometime in the first six months.
- **Comparison:** can be drawn from two alternative populations: (i) single people without children living in areas similar to the Pathfinders who are unemployed and flow into JSA over the same period or (ii) single people without children living in Pathfinder areas who are unemployed and flow into JSA prior to the introduction of phase 1.

Both matching and cross-sectional DiD (including triple differences) can be used to estimate the impact of UC during phase 1. Panel DiD is less appropriate since being on JSA implies being unemployed at some point between the 'before' and 'after' observations. This condition then imposes strong restrictions on outcomes for the common trends assumption to be fulfilled. For instance, if the outcome of interest is the employment rate, the common trends assumption requires that, in the absence of UC, the difference in employment rates between

treated and comparison groups some time after unemployment exactly offsets the difference some time before unemployment had there not been UC. That is, if the employment rate for the treated group is 10 percentage points above that for the comparison group in the 'before' period, it should also be exactly 10 percentage points higher in the 'after' period had UC not been there. This would ensure that the estimated impact in the absence of UC would be zero. But since all in the treated and comparison groups experience unemployment between the 'before' and 'after' periods, this requires common trends demands that differential outflows from employment prior to the time of treatment and inflows after treatment for the treated and comparison groups exactly compensate each other. For more details, see Appendix A.2.1.

Evaluation Question 2: Impact of UC replacing new JSA claims for single people without children versus the full legacy benefit regime

This question can be addressed for different groups, including the entire working-age population, the population of single people without children, or some sub-groups of interest such as those with lower education, low earners or those with a recent history of benefit claiming; the crucial point is that the groups are chosen without regard to actual claim status, and so include both future claimants and non-claimants.

The measurement of treatment effects relies on the comparison of two groups:

- **Treated:** those exposed to the policy change – can be the entire working-age population or some sub-group of interest living in Pathfinder areas at the time phase 1 is introduced.
- **Comparison:**
 - i. matched individuals living in matched areas at the time phase 1 is introduced in Pathfinder areas; or
 - ii. matched individuals living in Pathfinder areas observed prior to the introduction of phase 1.

All methods described in section 3.3 can be used in this evaluation exercise.

Evaluation Question 3: Indirect effects of UC replacing new JSA claims by single people without children versus the full legacy benefit regime

During this phase, this question can be addressed for parents and individuals in couples (assuming that family composition does not change as a consequence of UC, as is likely to be the case in the short term). These groups are not directly exposed to UC in phase 1, so any impact on them must be indirect.

The measurement of treatment effects relies on the comparison of two groups:

- **Treated:** those not directly exposed to the policy change living in Pathfinder areas – can be the entire working-age population of parents or individuals in couples, or the sub-group of unemployed among these, or those starting a JSA claim at the time that similar claims by singles with no children are being treated under UC;
- **Comparison:**
 - i. matched individuals living in matched areas at the time phase 1 is introduced in Pathfinder areas or
 - ii. matched individuals living in Pathfinder areas observed prior to the introduction of phase 1.

Evaluating the labour market impacts of Universal Credit: a feasibility study

All methods described in section 3.3 can be used in this evaluation exercise.

Limitations and challenges in phase 1

We identify the following main evaluation challenges in phase 1. Unless otherwise stated, they apply to all three evaluation questions:

- **Length of evaluation period**

Comparison areas face the full legacy benefits regime only for the first six months of implementation; from then onwards in this hypothetical scenario UC starts to be extended nationwide to all new JSA claimants in phase 2. This implies that estimates relying on geographical variation can identify the impact of UC versus the full legacy benefit regime for between 6 and 11 months (depending on which area is matched to the treated area). After that period, the definition of the treatment effect that can be identified changes to the 'impact of UC versus a mixed regime where UC replaced new JSA claims after up to 11 months in the pure legacy benefit regime'.

Time variation may allow for some longer analysis of the impact of UC versus the full legacy benefit regime. However, comparison groups observed further away in time may constitute a less reliable counterfactual if the economic conditions change.

- **Entry effects**

The introduction of UC might potentially change the composition of claimants. If UC claimants differed from JSA claimants in terms of their unobserved characteristics, the two groups would not be comparable and answers to Q1 would be biased (an internal validity problem).¹ The composition of UC and JSA claimants could differ – including by people accelerating (or delaying) their claim to ensure it fell under the UC (or legacy benefit) regime – for several reasons: if the (financial or non-financial) incentive to claim UC were different from JSA; if people's level of understanding of JSA and UC were different; or if claiming UC when unemployed were more (or less) attractive than claiming JSA because at this stage it was the only way to get access to the rest of the UC regime (e.g. when moving back into work), which might be more (or less) attractive than other legacy benefits. Any of these could mean that the unobserved characteristics of UC and JSA claimants differed, posing a problem for Q1: attributing differences in outcomes between the two groups to the impact of UC would be wrong if the different compositions of the two groups meant that we might expect them to have different outcomes even without UC. Note that this is not a problem for Q2 and Q3, however, since those do not rely on comparing the claimant groups under the two regimes.

A related, but distinct, problem arises if the composition of UC claimants in phase 1 is different (in terms of observed or unobserved characteristics) from the composition of unemployed UC claimants in the longer-run. This could arise if people are not yet as familiar with the UC regime as they will become; if the policy itself is refined at a later stage; or if people are more (or less) inclined to claim UC than similar people will be in future because now, unlike later, it is the only way to get access to the rest of the UC regime. Any of these would mean UC claimants in phase 1 would not be representative of unemployed UC claimants later on, and the impacts estimated here might not carry over (an external validity problem). Unlike the internal validity problem mentioned above, this is a potential concern for all three evaluation questions.

¹ Differences in observed characteristics would not be a threat to internal validity since the differences could be controlled for through matching.

• Anticipation effects

There are two types of anticipation effects:

- 1 Similar to entry effects but now affecting the comparison group rather than the treated group, the composition of new JSA claimants can be affected if potential claimants react in advance to the foreseen local introduction of UC in place of new JSA claims to influence access to UC in the near future.
- 2 Anticipation may change the observed outcomes of the comparison group, who may react to the prospect of UC replacing new JSA claims in that area in the near future. For questions 2 and 3, the treated group too may react to the prospect of a new phase (phase 2 in this case) being locally rolled out in the near future.

Anticipation effects of the first type giving rise to observed compositional differences may limit the common support if part of the treated cannot be matched due to some sub-group of those that would normally make a new JSA claim (and would become part of the comparison group) deferring doing so to influence access to UC in the near future (question 1). They may also be a threat to internal validity for question 1 if they lead to unobserved compositional differences between matched individuals in Pathfinder and matched areas related to the outcome of interest.

For all questions, anticipation effects of the second type can be seen either as threats to the internal validity of the estimated treatment parameter of interest (the effect of UC versus the legacy benefit system), or as meaning that we estimate a slightly different (and probably less interesting) parameter. For question 1 one could reinterpret the estimated parameter as the effect of UC versus a legacy benefits system in which UC will be introduced shortly afterwards. For questions 2 and 3, the estimated parameter would be the effect of UC replacing JSA claims for single people without children and shortly afterwards for all claimants versus a legacy benefits system in which UC will be introduced shortly afterwards.

• Other threats to the representativeness of the findings

Extrapolating from the effects of the Pathfinders to the likely wider impact of UC would be problematic for other reasons: (i) the nature of the Pathfinder policy, which is not necessarily representative of the final policy as it may be deliberately implemented slightly differently, or there may be teething problems or unfamiliarity, or there may be more attention and resources devoted to the Pathfinders; (ii) the small number of sites, not necessarily representative of the whole country; and (iii) the specific small sub-group, as single JSA claimants without children represent only one to two per cent of JSA national caseload.

• Volume of UC claimants

One additional concern, mostly relevant to questions 2 and 3 only, relates to the small scale of UC during phase 1 even within the Pathfinder areas. As mentioned above, in this scenario the population directly exposed to UC in phase 1 is very small (less than 2 per cent of the local JSA caseload). Therefore, at this stage aggregate effects are unlikely to be detectable and we would expect indirect effects to be negligible. Any conclusions from this stage regarding the nature and size of aggregate and indirect effects are not informative on what these may be for a wider implementation of UC.

Overall, we would not expect entry and anticipation effects to be strong at this stage for two main reasons. First, UC has little impact on the incentive to make a new claim, since UC rates are the same as JSA rates and the conditionality regime little different. And second,

Evaluating the labour market impacts of Universal Credit: a feasibility study

unfamiliarity with the details (or even existence) of UC may preclude strong responses. Moreover, anticipation effects are unlikely to be of any relevance much before the start of phase 2, when (in this scenario) the forthcoming extension of UC to new JSA claimants across the UK (therefore, potentially affecting the comparison group) becomes more visible.

Since phase 1 precedes phase 2 by about 6 to 11 months, there will be some scope to evaluate its impact before anticipation effects may become an issue.

Empirical evidence on the importance of entry and anticipation effects can be gathered by assessing the impact on the number and observed characteristics of people making claims to JSA-equivalent benefits in Pathfinder and comparison areas (details in Appendix A.2.3). The threat to internal validity comes from changes in the unobserved characteristics of claimants rather than their observed characteristics; but the presence or absence of changes in observed characteristics might give a useful indication of whether unobserved differences are likely to have changed as well. Estimates of the impact of UC on inflows to JSA-equivalent claims (Evaluation Question 2) provide further empirical evidence on the importance of entry effects.

Overall assessment of evaluation possibilities for phase 1

This phase offers some of the cleanest evaluation possibilities because the full legacy benefit regime is still operating throughout most of the country, the phase extends over a relatively long period, and potential participants are less likely to act on information about the new regime, which could contaminate estimates. All three evaluation questions can be addressed using both geographical and time variation, but the reduced volume of UC claims precludes any meaningful analysis of questions 2 and 3.

More than one method can also be used in estimation. This wealth of possibilities allows for useful robustness checking. Some of the assumptions can be tested to some extent: no-entry-effects, no-anticipation-effects, common trends (for DiD) and historical comparability of matched areas (matching). Impacts can also be assessed over a relatively long time horizon – we would suggest just below a year if time variation is used, although with special care when drawing matched samples observed further away from the treated.

The main problem with evaluating the UC Pathfinders is external validity: in particular, that UC in the Pathfinders may not closely resemble UC as finally implemented across the country and that it will only directly affect one small group of claimants. That aside, the Pathfinder areas offer a good opportunity to get as close as possible to the parameter of interest for single unemployed claimants without children in those areas: that is, the impact of starting the equivalent of a JSA claim under the full Universal Credit regime versus starting a JSA claim under the full legacy benefit regime (question 1).

3.4.2 Phase 2: New JSA claims

In the scenario provided by DWP, phase 2 runs from 6 to 12 months after implementation begins (i.e. between October of Year 1 and March of Year 2) and rolls out the Pathfinder geographically across the country, by Jobcentre Plus districts a few at a time, for single people without children as well as for other JSA claimant types: lone parents and couples with and without children.

As with other phases, for each claimant group in a given Jobcentre Plus district at a given time, the regime is either switched on for everyone or for no-one. Note also that incentives for all potential claimants in a district change the moment UC is introduced for (actual)

claimants in that claimant group in that district.

Given the similarity of phases 1 and 2, the following analysis is very similar to that for phase 1 and we frequently refer back to the earlier discussion for details.

Evaluation Question 1: Impact of starting a new UC claim under the full UC regime versus starting a new JSA claim under the full legacy benefit regime

In phase 2, this question can be addressed for all individuals satisfying the JSA/UC conditions who claim the benefit 6-12 months after implementation begins.

The measurement of treatment effects relies on the comparison of two groups:

- **Treated:** individuals who make a new JSA-equivalent claim to UC some time between October of Year 1 and March of Year 2 in areas where phase 2 has been rolled out at that time (treated areas);
- **Comparison:** can be drawn from two alternative populations:
 - i. matched individuals flowing into JSA at the same time as the treated, but in matched areas not yet under UC; or
 - ii. matched individuals in treated areas flowing into JSA some time prior to the local roll-out of phase 2.

Both matching and cross-sectional DiD (including triple differences) can be used to estimate the impact of UC during phase 2. Just like for phase 1, panel DiD is less suitable.

Evaluation Question 2: Impact of UC replacing new JSA claims versus the full legacy benefit regime

This question can be addressed for different groups, including the entire working-age population, or some sub-groups of interest such as those with lower education, low earners or those with a recent history of benefit claiming.

The measurement of treatment effects relies on the comparison of two groups (unconditional on starting a new claim):

- **Treated:** those exposed to the policy change – can be the entire working-age population or some sub-group of interest observed at a time when phase 2 is implemented in that area;
- **Comparison:**
 - i. matched individuals in matched areas not yet phased in; or
 - ii. matched individuals in the same areas observed prior to the roll-out of phase 2.

All methods described in section 3.3 can be used in this evaluation exercise.

Evaluation Question 3: Indirect effects of UC replacing new JSA claims versus the full legacy benefit regime

During phase 2, this question can be addressed for groups that would receive IS or ESA, rather than JSA, if they did not work (such as lone parents with children under five, carers or disabled people) as UC does not directly affect them until phase 3 is introduced.² In the following, we use the case of lone parents with children under five as an example.

The measurement of treatment effects relies on the comparison of two groups:

- **Treated:** lone parents of children under five at the time phase 2 is rolled out in areas where this happens earlier;
- **Comparison:**
 - i. matched lone parents of children under five observed at the same time as the treated, but in matched areas where the roll-out of phase 2 happens later; or
 - ii. matched lone parents of children under five in the same treated areas observed prior to the roll-out of phase 2.

All methods described in section 3.3 can be used in this evaluation exercise.

Limitations and challenges in phase 2

The main limitations here are similar to those discussed for phase 1. Unless otherwise stated, they apply to all three evaluation questions:

- **Length of evaluation period**

The rapid progression of phase 2 throughout the country implies that the counterfactual under the full legacy benefit regime can only be measured for a limited period. In this scenario, comparisons across areas can only recover the impact of UC against the full legacy benefit regime on outcomes up to about four months from the claim start, after which the definition of the parameter that can be estimated changes. Using time variation can allow for the same parameter to be estimated over a longer period, although reliability may be compromised the longer before the introduction of phase 2 the comparison group flows into JSA.

- **Entry effects**

In this phase, the composition of new claimants may be affected by entry effects due to:

- i. the different incentives to claim UC versus legacy benefits;
- ii. the fact that the UC regime can only be accessed via a JSA-equivalent claim – and from a specific point in time; and
- iii. (un)familiarity with the new regime. Their consequences for internal and external validity are the same that were outlined for phase 1.

² Strictly speaking these groups could choose to start a new JSA claim if they fulfil the requirements. But we doubt this would be an empirically relevant phenomenon.

The first of these causes of entry effects – different incentives to claim UC versus legacy benefits – might be slightly stronger in phase 2 than in phase 1. This is because phase 1 applies only to single people without children, whose UC entitlement will always be the same as their JSA entitlement. Phase 2, on the other hand, also affects families with children, whose UC entitlement may be less than their combined JSA and child tax credit entitlement because the UC means test treats assets and unearned income (notably savings and spousal maintenance income) more harshly than the tax credit means test. If UC entitlements are lower, parents may be less inclined to claim UC than to claim legacy benefits, so UC claimants in phase 2 may have less unearned income than the legacy benefit claimants in the comparison group – a problem for internal validity if unearned income is not observed in the data. This is only a concern for question 1.

- **Anticipation effects**

Anticipation effects are of the same kind as those outlined for phase 1, but likely to be stronger because of the imminent introduction of phases 2 in this scenario (in comparison areas, where phase 2 is introduced later, or in treated areas before the introduction of phase 2 when using time variation) and 3. Notice in particular how the start of the phasing out of TC in phase 3 is likely to seriously affect the outcomes of the comparison group, as well as, for questions 2 and 3, of the treated group.

- **Volume of UC claimants**

The population under UC grows slowly over time as only new claims of certain types trigger a move into UC. Just like for phase 1, aggregate effects can be difficult to detect (question 2) at this stage and indirect effects can be immaterial (question 3). Results from this stage regarding the nature and size of aggregate and indirect effects are not informative on what these may be for a wider implementation of UC.

Overall, we would not expect entry and anticipation effects to be very strong at the start of this phase (though as mentioned, entry effects for some groups will be stronger than in phase 1). Anticipation effects may gain prominence as the start of phase 3 approaches when using geographical variation to construct the counterfactual. The use of time variation may reduce the importance of anticipation effects. Empirical evidence on the importance of entry and anticipation effects can be gathered by assessing the impact on the observed composition of inflows to JSA-equivalent benefits (details in Appendix A.2.3). Estimates of the impact of UC on inflows to JSA-equivalent claims (Evaluation Question 2) provide further empirical evidence on the importance of entry effects.

Overall assessment of evaluation possibilities for phase 2

Together with phase 1, this phase offers the cleanest design to get as close as possible to an evaluation of the full UC regime versus the full legacy benefit regime on new claimants, albeit only when either regime is entered via a JSA or equivalent claim. However, an informative study of aggregate or indirect effects looks less promising due to the limited volume of UC claims.

More than one method can be applied both across areas and across time to test the robustness of any effects uncovered. Some of the assumptions can be tested to some extent: no-entry-effects, no-anticipation-effects, common trends (for DiD) and historical comparability of matched areas (matching). Impacts can also be assessed over a relatively long time horizon in this scenario – we would suggest just below a year if time variation is used, although with especial care when drawing matched samples observed further away from the treated.

Evaluating the labour market impacts of Universal Credit: a feasibility study

Compared to phase 1, the treatment (i.e. the UC regime) should be closer to the ‘final’ version and less likely to suffer from any teething problems, and the UC claimants are now representative of all types of JSA claims (though strictly speaking of only those taking place in the six-month window October of Year 1– March of Year 2). Entry effects and the selection issues they engender are, however, likely to be more serious in this phase, but possibly still mild and less of a concern when using time variation. Additionally, impacts can be assessed for only a limited period of at most four months when relying on geographical variation.

3.4.3 Phase 3: New claims to WTC, CTC, and ISlp

In the scenario provided by DWP Phase 3 runs from May to July in Year 2. It gradually extends UC to replace new claims to TC and ISlp, geographically (and possibly by claimant type within the Jobcentre Plus district). As with the other implementation phases, within each Jobcentre Plus district the switch of regime will happen at a single point in time for each claimant group.

As compared with phases 1 and 2, this implementation phase is more challenging from an evaluation point of view for two reasons:

- By the time it starts, UC will have been operating in place of new JSA claims in all areas, reducing the usefulness of geographical variation for assessing the impact of UC against the full legacy benefit regime as all areas will be under a hybrid regime – except for those groups for whom JSA is not relevant even when they do not work (lone parents with children aged under five, ESA recipients and carers eligible for IS).
- Some important changes in entitlement are introduced during phase 3:
 - i. the hours requirements associated with Working Tax Credit (WTC) will be removed;
 - ii. entitlement will be extended to low earners under the age of 25 without children;
 - iii. the single 65 per cent withdrawal rate replacing a variety of different rates will mean that entitlement runs out at a lower income level for some groups and a higher income level for others; and
 - iv. the treatment of assets and unearned income for the UC means test, while similar to that under existing benefits, is harsher than that under existing tax credits.

These features limit the evaluation possibilities by introducing serious difficulties in the choice of the counterfactual.

In general, the evaluation exercise will become more difficult, and the questions that can be addressed narrower, as one moves further into the implementation of UC. The following discussion highlights the differences in analyses and the additional difficulties involved in evaluating the impact of UC during phase 3 when compared to phases 1 and 2.

Evaluation Question 1: Impact of starting a new UC claim under the full UC regime versus starting a new TC/ISlp claim under the full legacy benefit regime

The composition of new UC claimants in phase 3 is likely to differ substantially from the composition of the TC/ISlp claims they replace. Even if nobody's behaviour changes, new groups will gain entitlement (those in work but not meeting the WTC age or hours requirements) while others will gain or lose it as the reform to the means test changes the income level at which entitlement is exhausted. But more importantly, changes to entitlements (and work-search requirements, the ease of claiming support, etc. – all either actual or perceived) at different hours of work and levels of earnings will be quite different under UC, changing people's incentives to claim. Entitlement is partly the result of individual decisions and people's choices over whether to work, for what wage rate and how many hours may well respond to these new incentives. Similarly, take-up behaviour might change in response to these changes in understanding and (financial and non-financial) incentives to claim. These changes in behaviour, generating differences between UC and TC/ISlp in the unobserved characteristics of claimants, limit the evaluator's ability to construct an adequate comparison group to such an extent that this question cannot be reliably addressed in this phase (internal validity).

Evaluation Question 2: Impact of UC replacing new JSA/TC/ISlp claims versus the full legacy benefit regime or versus a hybrid regime where UC has replaced only new JSA claims

This question can be addressed for different groups, including the entire working-age population, or some sub-groups of interest such as those with lower education, low earners or those with a recent history of claiming benefits. It is robust to changes in entitlement rules and how these may affect the decision to make a new claim since it accounts for the overall direct effects of the reform, unconditional on claim status (i.e. effects on claim decisions, employment or other labour market outcomes, and indirect effects arising through market-wide effects).

The measurement of treatment effects relies on the comparison of two groups, unconditional on starting a new claim:

- **Treated:** those exposed to the policy change – can be the entire working-age population or some sub-group of interest observed at a time when phase 3 is implemented in that area;
- **Comparison group:** (i) matched individuals in matched areas not yet phased in or (ii) matched individuals in the same areas observed prior to the roll-out of phase 3 (or preferably prior to the roll-out of phase 1 or 2).

All methods described in section 3.3 can be used in this evaluation exercise.

When using time variation, one could choose a comparison group claiming prior to the first local roll-out of UC (phase 1 or 2, depending on area). In that case, one can identify the 'impact of UC replacing new claims to JSA/TC/ISlp versus the full legacy benefit regime' for as long a time period as the comparison group is under the full legacy benefit regime. The choice of the comparison period faces the trade-off between reliability of the estimates (market conditions are closer to those experienced by the treated in more recent periods) and the horizon over which the impact can be estimated. Anticipation effects may invalidate the counterfactual closer to when UC is first locally introduced, but we would not expect strong anticipation effects so early in the implementation stage. A word of caution

Evaluating the labour market impacts of Universal Credit: a feasibility study

if using matching with time variation: seasonal effects could invalidate the results, but their importance can be empirically assessed historically.

When using geographical variation, the parameter being estimated is the ‘impact of UC replacing all new claims to JSA/TC/ISlp versus replacing new claims to JSA only’. Comparison areas should be chosen to match the history of the outcome of interest, the UC implementation history and UC coverage (size of population under UC) at the time phase 3 is rolled out in the treated areas.³

An exception is lone parents with children under five, who would only access UC once phase 3 is rolled out (via TC/ISlp). For them, geographical variation can be used to recover the ‘impact of UC replacing all new claims to JSA/TC/ISlp versus the full legacy benefit regime’. In the absence of indirect effects, this is the same parameter as would be identified had UC not been implemented at all (i.e. in phases 1 and 2) in matched areas.

Evaluation Question 3: Indirect effects of UC replacing new JSA/TC/ISlp claims versus a hybrid regime where UC has replaced only new JSA claims

During phase 3, this question can be addressed for all (or particular sub-groups) of those who already started a UC claim (which would previously have been JSA) before the start of phase 3, since those who have already claimed UC are not directly affected by the extension of UC to replace new TC/ISlp claims.

The measurement of treatment effects relies on the comparison of two groups, which for this evaluation question in this phase depend on the chosen methodology.

For matching across areas:

- **Treated:** the stock of individuals who started a UC claim at any time before the roll-out of phase 3, observed at the time phase 3 is introduced in areas where this happens earlier;
- **Comparison:** the stock of matched individuals who started a UC claim at any time before the roll-out of phase 3, observed at the same time as the treated but in matched areas where the roll-out of phase 3 happens later.

For matching across time:

- **Treated:** the inflow into UC near the end of phase 2;
- **Comparison:** the matched inflow into UC just after the start of phase 2 in the same areas as the treated.

For DiD across areas:

- **Treated:** the inflow into UC near the end of phase 2 in areas where phase 3 is introduced earlier;
- **Comparison:** the matched inflow into UC near the end of phase 2 in areas where phase 3 is introduced later.

³ Obviously one can also explore time variation to estimate the ‘impact of UC replacing all new claims to JSA/TC/ISlp versus replacing new JSA claims only’, although it may be impossible to ensure that UC coverage is comparable when different time periods are being compared. Since coverage may affect the composition of claimants (even in the absence of a new reform) as well as the importance of indirect effects, it may alter the definition of the impact being estimated.

Therefore, matching across area, matching across time, and DiD across area (described in section 3.3) can all be used for this evaluation exercise. The scope for robustness checks is limited, however, as matching across areas considers impacts on those who started a UC claim at any time before the introduction of phase 3, while matching across time and DiD across area restrict attention to those who started a UC claim close to the end of phase 2.⁴ Although historical evidence on the comparability of time periods being contrasted can be obtained, matching across time is subject to potentially serious threats arising from differential seasonality (i.e. the treated and comparison groups in this case will have flowed onto UC at different times of year and might, therefore, expect to have different outcomes even without any indirect effects of phase 3 on them), and great care should be used when considering its use.

We therefore suggest relying on:

- matching across areas to recover the indirect effects of moving new claimants onto UC on those who already started a UC claim before the roll-out of phase 3 – unconditional on whether at the start of phase 3 they are still unemployed or have moved into work. Indirect effects for this population would, therefore, capture both substitution and displacement effects.
- DiD across areas to recover the indirect effects for those who started a UC claim (to what would have been JSA) just before the roll-out of phase 3. Indirect effects for this population will mainly reflect substitution effects since many, if not most, of those in the treated group will still be unemployed at the start of phase 3 and the duration of the analysis is limited by the rapid roll-out of this phase across areas.

Limitations and challenges in phase 3

These become more important with each phase.

- **Length of evaluation period**

When using geographical variation the horizon over which the effect of UC can be identified is much reduced in phase 3, to no more than two months due to the extension of phase 3 nationwide. When using time variation a longer period is available – with the usual trade-off between time horizon of outcomes and reliability of counterfactual.

- **Entry effects**

The changes to entitlements in phase 3 are much bigger than in phases 1 and 2, making entry effects much more problematic to the point where we doubt that question 1 can be addressed at all. In contrast, for questions 2 and 3 the entry effects, though similar to phases 1 and 2, are less severe and, therefore, less of a threat to external validity.

⁴ The treated group in these cases must be people starting a claim close to the end of phase 2. This is because the comparison group (for matching across time) or the 'before' group (for DiD across areas) must be people whose claim was late enough to fall within the UC regime (i.e. they made a claim after the introduction of phase 1/2 in their area). Since those comparison/before groups will themselves be potentially subject to indirect effects after the introduction of phase 3, clean outcomes can be observed only for a few months (between the start of their UC claim and the start of phase 3). Accordingly, the treatment groups must be those starting a claim close to the end of phase 2 so that the corresponding few months for them includes as much time as possible after phase 3 might have indirect effects on them.

Evaluating the labour market impacts of Universal Credit: a feasibility study

• Anticipation effects

While anticipation of the changes to be introduced with the roll-out of a new phase may affect the composition of the comparison group, this issue is mostly relevant for question 1, which we advised against addressing because of entry effects.

Anticipation of the changes to be introduced can also affect the outcomes of the comparison group for question 2 – as those facing the legacy benefit regime may react in anticipation of a new implementation phase (can be any of phases 1 to 3, depending on the choice of comparison group) being locally imminent. Likewise, the anticipation of an imminent change in incentives by the treated group (with the forthcoming introduction of UC for new ISc/ESA/HB claims) may affect outcomes when addressing question 2.

Anticipation effects may get stronger over time, especially in the proximity of a new reform. They can be more of a concern when using geographical variation given the rapid roll-out set out for this phase (at the time phase 3 is first rolled out, matched areas have at most three months left under the old system where UC has replaced only new JSA claims).

For question 2, anticipation effects can be a threat to internal validity, unless the treatment effect is reinterpreted as the effect of UC replacing new JSA/TC/ISlp claims and shortly afterwards all new claims versus the pure legacy benefit regime (or else a hybrid regime) in which UC will shortly replace more new legacy benefit claims.

For question 3 there can be no anticipation effects on outcomes as the population being examined is already under UC.

• Volume of UC claimants

A limited volume of claimants can be a problem for addressing questions 2 and 3. But with a quick within-area roll-out scenario and wider coverage of UC (due to UC replacing new claims to TC/ISlp in addition to JSA and the extension of benefits to new groups), the size of the UC caseload may grow enough at this stage to make aggregate and indirect effects large enough to detect.

As for phases 1 and 2, empirical evidence on the importance of entry and anticipation effects can be gathered by assessing the impact on the observed composition of inflows to TC/ISlp-equivalent benefits (details in Appendix A.2.3).

Overall assessment of evaluation possibilities for phase 3

This is a crucial implementation phase in the roll-out scenario provided, when UC is due to be extended to major target groups (low earners and non-working lone parents) and the most significant changes in entitlement are introduced for new claims. Unfortunately, the evaluation possibilities in phase 3 are strongly compromised by the changes in entitlement rules and their consequences for the composition of new claimants, which rule out addressing Evaluation Question 1.

We identify two major evaluation possibilities, which in practice can be implemented using a set of alternative methods (the first exploiting mostly time variation, the second relying more on geographical variation):

- 1 For the entire working-age population or some specific sub-groups of interest (e.g. all parents, lone parents, low earners, low educated, etc): assess the impact of UC replacing new JSA/TC/ISlp claims, versus the full legacy benefit regime or versus a hybrid regime in which UC replaces new JSA claims only, on inflows into benefits and on future labour market outcomes (employment, earnings, hours).

- 2 For those who started a UC claim (which would previously have been JSA) before the roll-out of phase 3: assess the indirect impact of UC replacing new JSA/TC/ISlp claims versus a hybrid regime in which UC replaces new JSA claims only.

A major limitation of the proposed evaluation exercises is the time horizon over which UC can be contrasted against the hybrid regime (as well as against the full legacy benefit regime in question 2). The use of geographical variation limits analysis to outcomes at most two months after the start of phase 3. For question 2 using time variation, with a comparison group drawn from new claims made some time before the start of phase 3, one can extend the evaluation period and this will also be less sensitive to anticipation effects, but at the cost of being less comparable to the treated group.

The limited volume of UC cases may be less important in this phase when considering questions 2 and 3.

3.4.4 Phase 4: New claims to ISc, ESA and HB

In the scenario provided by DWP phase 4 would be implemented in a single month nationwide, in October of Year 2. At this point, UC replaces new claims to ISc, income-related ESA and HB.

The evaluation possibilities for this implementation stage offer less potential. In particular, the meaning of the treatment effect parameters that can be identified becomes gradually more obscure as UC is extended throughout the country, gradually exposing the comparison group to more of the UC regime as successive phases are introduced so that UC is being compared with an evolving hybrid regime. Moreover, the new UC claimants of the type being phased in are a very restricted and unrepresentative group, even in the absence of anticipation behaviour to influence entitlement once the new phase is implemented (by, for instance, postponing a claim to ISc/ESA/HB within the legacy benefit regime until it triggers a UC claim instead). Specifically, the new UC claimants are those starting an ISc/ESA/HB-equivalent claim during or after October of implementation Year 2 who had not started a UC claim equivalent to JSA/TC/ISlp for some (area-dependent) time prior to the new claim. Therefore, mechanical composition effects, arising from the fact that UC has been available for some time to some claimant types, can seriously compromise the external validity of the results.

In this scenario phase 4 is introduced simultaneously across the country providing no geographical variation to exploit, so phase 4 only allows for comparisons across time.

Evaluation Question 1: Impact of starting a new UC claim under the full UC regime versus starting a similar new ISc/ESA/HB claim under the full legacy benefit regime or under a hybrid regime

In phase 4, this question can be addressed for all new claimants of ISc/ESA/HB-equivalent UC during or after October in implementation Year 2.

The measurement of treatment effects relies on the comparison of two groups:

- **Treated:** new ISc/ESA/HB-equivalent claims under UC during or after phase 4 implementation;
- **Comparison group:** matched ISc/ESA/HB claims in the same area starting some time before phase 4 implementation, under the full legacy benefit regime or a hybrid regime.

Evaluating the labour market impacts of Universal Credit: a feasibility study

The choice of the comparison period determines the nature of the identified treatment effect parameter. It will be the impact of UC for new ISc/ESA/HB versus:

- the full legacy benefit regime, when comparing UC claims with matched legacy benefit claims starting prior to the implementation of phase 1 or 2 in that area;
- a hybrid regime where UC has replaced only new JSA claims, when comparing UC claims with matched legacy benefit claims starting after the roll-out of phase 1 or 2, but before the roll-out of phase 3;
- a hybrid regime where UC has replaced only new JSA/TC/ISIp claims, when comparing UC claims with matched legacy benefit claims starting after the roll-out of phase 3 (but before phase 4 implementation).

All methods described in section 3.3 can be used in this evaluation exercise, but only using time variation. The reliability of matching will be reduced if there is significant seasonal variation in outcomes as estimates based on comparing groups observed at different points in time, particularly different months of the year, may attribute to UC what is really normal seasonal variation. However, the importance of seasonal effects can be assessed historically.

Evaluation Question 2: Impact of UC replacing all new claims versus the full legacy benefit regime or a hybrid regime

This question can be addressed for different groups, including the entire working-age population, or some sub-groups of interest such as those with lower education, low earners or those with a recent history of benefit claiming. It accounts for both the direct and the indirect effects of UC.

The measurement of treatment effects relies on the comparison of two groups, unconditional on starting a new claim:

- **Treated:** those exposed to the policy change – can be the entire working-age population or some sub-group of interest observed at a time when phase 4 is locally introduced;
- **Comparison group:** matched individuals in the same areas observed prior to the roll-out of phase 4, under the full legacy benefit regime or a hybrid regime.

Again the choice of the comparison period determines the nature of the identified treatment effect parameter. It will be the impact of UC replacing all new legacy benefit claims versus:

- the full legacy benefit regime, when comparing individuals observed after phase 4's implementation with matched individuals observed prior to the implementation of phase 1 or 2 in that area;
- a hybrid regime where UC has replaced only new JSA claims, when comparing individuals observed after phase 4's implementation with matched individuals observed after the roll-out of phase 1 or 2 but before the roll-out of phase 3;
- a hybrid regime where UC has replaced only new JSA/TC/ISIp claims, when comparing individuals observed after phase 4's implementation with matched individuals observed after the roll-out of phase 3 (but before phase 4's implementation).

All methods described in section 3.3 can be used in this evaluation exercise, but as with question 1, the reliability of matching will be reduced if there is significant seasonal variation in outcomes. The presence of seasonal variation can be assessed historically.

Evaluation Question 3: Indirect effects of UC replacing all new claims versus a hybrid regime where UC has replaced only new JSA claims or only new JSA/TC/ISlp claims

This question can be addressed for all (or particular sub-groups) of those who already started a UC claim (which would previously have been JSA/TC/ISlp) before the start of phase 4, since those who have already claimed UC are not directly affected by the extension of UC to replace new ESA/ISc/HB claims.

The measurement of treatment effects relies on the comparison of two groups, the treated and the comparison group. Depending on which treated and comparison groups are chosen, a different treatment effect is identified.

- **Treated (i):** the inflow into UC near the end of phase 3;
- **Comparison (i):** the matched inflow into UC in the same areas, who started their UC claim just after the start of phase 3.

Or

- **Treated (ii):** the inflow into UC near the end of phase 2;
- **Comparison (ii):** the matched inflow into UC in the same areas, who started their UC claim just after the start of phase 2.

The first treated and comparison groups define the indirect effects of UC replacing all new claims, versus a hybrid regime where UC has replaced only new JSA/TC/ISlp claims, on those who started a UC claim (which would previously have been JSA/TC/ISlp) close to the end of phase 3.

The second treated and comparison groups define the indirect effects of UC replacing all new claims, versus a hybrid regime where UC has replaced only new JSA claims, on those who started a UC claim (which would previously have been JSA) close to the end of phase 2.

Robustness checks are limited as only time variation can be used; historical evidence on the comparability of the time periods being contrasted can, however, be assessed.

Limitations and challenges in phase 4

These gain importance with each phase, and new challenges become relevant.

- **Length of evaluation period**

The extent to which UC can be compared with the full legacy benefit regime is very limited in phase 4. Even if drawing the counterfactual from matched observations one year earlier (bearing in mind that, in general, a longer gap separating treated and comparison observation raises comparability issues), one can estimate the effects of UC for at most five months before all areas are in phase 2. After that, the parameter that can be estimated has a different interpretation, depending on the benefits system faced by the comparison group. For question 3, the horizon is limited to five months (the time frame between the start of phase 3 and the start of phase 4) when using comparison group (i), and up to nine months when using comparison group (ii).

- **Entry effects**

Individuals may delay or accelerate a claim to affect whether it falls within the UC or legacy benefit regime, which will affect the composition of new claimants shortly after the roll-out of phase 4. They may also make different choices as to whether to make a claim at all under the new system.

Evaluating the labour market impacts of Universal Credit: a feasibility study

Entry effects may give rise to unrepresentative effects (external validity) and biased estimates (internal validity, only relevant for question 1, resulting from unobserved compositional differences related to the outcome of interest).

- **Mechanical composition effects**

The past roll-out history of UC influences the composition of new claimants. Some of the new claims to what would have been ISc/ESA/HB will be by individuals who had already claimed UC via other routes (specifically, via new claims to what would previously have been JSA/TC/ISlp) in the relatively recent past: the remainder are, therefore, only a sub-group of those making a claim to what would previously have been ISc/ESA/HB, and a sub-group that depends on how long UC has existing in place of the various types of new legacy benefit claim in that area. Mechanical composition effects may substantially alter the observed and unobserved composition of new claims to UC, affecting both the representativeness and the validity of the estimated effects. This problem is only relevant for question 1 as questions 2 and 3 uses observations on a wider population, unconditional on actual claim status in phase 4.

- **Anticipation effects**

Anticipation of imminent reforms (i.e. the new roll-out phase) may affect the composition of the comparison group for question 1 and the outcomes of the comparison group for questions 1 and 2. For question 2, the anticipation of the migration of the stock may affect the outcomes of the treated group. Anticipation effects for questions 1 and 2 may get stronger over time, especially in the proximity of a new reform (which, in this phase, includes the local introduction of phases 1, 2, 3 or 4, depending on the choice of the comparison groups). For question 3 there are no anticipation effects, as was the case in phase 3.

- **Volume of UC claimants**

A limited volume of claimants can be a problem for addressing questions 2 and 3. But with immediate coverage of all new claims, the UC caseload may be large enough at this stage to make indirect effects large enough to detect.

As for the previous phases, empirical evidence on the importance of entry and anticipation effects can be gathered by assessing the impact on the observed composition of inflows to ISc/ESA/HB-equivalent benefits (details in Appendix A.2.3).

Overview of evaluation possibilities for phase 4

Phase 4 is likely to provide less insight as to the impacts of UC than other phases for two main reasons:

- The new claims to ISc/ESA/HB that trigger UC are restricted to those who had not already started a UC claim through other benefits. This is a highly selected population that may be difficult to reproduce among the comparison group, which is observed before the start of phase 4, as UC will have been in place for a shorter time for them. This is a limitation for question 1 given its reliance on the population of claimants.

Evaluating the labour market impacts of Universal Credit: a feasibility study

- The choice of the comparison group relies on time variation. Comparison groups drawn earlier in time may be less comparable to the treated, while those drawn later will be more sensitive to anticipation effects and offer a shorter horizon over which the impact can be estimated. To compare UC against the full legacy benefit regime for questions 1 and 2 the comparison group would need to be drawn at least eight months prior to the start of phase 4 (more if one wishes to assess outcomes over a reasonable period), raising significant questions of comparability.

On the other hand, with a one-month roll-out and full coverage of all new claims, the volume of UC cases might now be sufficient to detect aggregate impacts, while indirect effects might be non-negligible:

- The impact of UC replacing all new claims versus the full legacy benefit regime or versus a hybrid regime (in which UC replaced only new JSA claims or only new JSA/TC/ISlp claims) for the entire working-age population or some sub-group;
- The impact of UC replacing all new claims versus a hybrid regime in which UC replaces only new JSA/TC/ISlp claims for those who started a UC claim close to the end of phase 3 – or versus a hybrid regime in which UC replaces only new JSA claims for those who started a UC claim close to the end of phase 2.

4 What can we learn from the migration of existing claimants to Universal Credit?

The hypothetical scenario provided by DWP set out two final phases, Phases 5 and 6, involving gradually transferring the remaining caseload of legacy benefit recipients to UC.

Claimants of some benefits will be migrated before others in this scenario. This will follow the same order as for the phasing in of Universal Credit (UC) for new claims: Jobseeker's Allowance (JSA) first, then Tax Credit (TC)/Income Support for lone parents (ISlp), and finally Income Support for carers (ISc)/Employment and Support Allowance (ESA)/Housing Benefit (HB). Claimants will be notified in advance, about two to three months prior to the migration of their spell.

The scenario involves the stock of existing JSA claimants to be gradually migrated to UC between December of implementation Year 2 and May of implementation Year 3, with notifications sent three months in advance (starting in September of implementation Year 2). The group being migrated is a population of the relatively long-term unemployed: they must have started a JSA claim at least nine months before migration (six months before notification), and in some cases more than two years before migration (for some single claimants without children in Phase 1 (Pathfinder) areas who are migrated to UC only in May of Year 3). Over this period, they must have remained out of UC: not started a new claim to what would previously have been HB or Child Tax Credit (CTC), for example, after new claims to those were replaced by UC.

The existing caseload of other benefit claimants starts to be migrated to UC once JSA claimants have all been migrated. The scenario provided involves starting to migrate TC and ISlp claimants first, followed by ISc, ESA and HB claimants. The migration happens over an extended period of over 2½ years from summer of implementation Year 3 to winter of Year 5. The scenario provides that the type of claimants being migrated will change over time, becoming more likely to be by claimants for whom UC brings no gains. Therefore, the groups in question become gradually more selected and atypical over time.

This section considers the evaluation possibilities offered by the migration of the stock in this hypothetical roll-out scenario. Given that the issues raised and solutions advanced for the evaluation of UC exploiting the stock migration apply equally to both phases 5 and 6, we do not differentiate between them in the following discussion.

Instead, we organise the discussion around the main evaluation questions, highlighting the methodological and interpretational challenges they raise, as well as their data requirements. First, though, we start by outlining the sources of variation created by the migration scenario that can be potentially exploited to learn about UC impacts.

4.1 Sources of variation

Within each area, a mix of different claims of a given type will be moved together, including complex and simpler claims. Until the migration to UC of all the existing claimants of a given type in a given area has been completed, there will be variation within that area on how similar claimants are treated.

It is also likely that some areas will migrate claims of a given type before others, so that areas migrating later can learn from the earlier experiences. This implies that similar legacy benefit claimants are treated differently across areas at a given time.

The proposed migration design, therefore, creates two potential sources of variation that can be useful to address the evaluation questions discussed below:

- 1 At a point in time, some legacy benefit claims are migrated to UC, while similar claims in the same area have not been migrated (yet).
- 2 In some areas a bulk of claims are moved to UC, while comparable areas have not started the migration of that type of claim yet.

The shorter the time period over which some claimant group is migrated within each area, the less scope there will be for evaluation of its effects. Some migration choices, such as earlier migration of those who gain financially from UC, or of particular demographic groups, or in certain geographical areas, may also undermine evaluation if they imply that shorter periods are allowed for the migration of each group.

The closer the policy can get to migrating people randomly (e.g. by National Insurance number) within each claimant type and area, over as long a period as possible, the greater the scope for evaluation.

4.2 Evaluation Question 1: Impact of migration (or notification of migration) to UC on the outcomes of existing claimants

One question of interest is:

- a What are the impacts of migrating a long-standing claim to UC under transitional protection in a given month versus not migrating it in that month (and eventually migrating it later should the claim remain open) for long-term claimants whose circumstances have remained stable, in a world where a new claim triggers moving into UC?

The focus here is on how the outcomes of claimants being migrated at a certain time are affected by the move to the full UC regime (albeit under transitional protection, whereby benefits are not allowed to fall in cash terms as a result of the migration) as compared to a hybrid regime where the comparisons will, at some point, move into the full UC regime – either if they start a new claim or, should their circumstances remain unchanged long enough, at the time of their own migration.

The first source of variation discussed in section 4.1 – at a point in time, some legacy benefit claims are migrated to UC, while similar claims in the same area have not been migrated (yet) – can be exploited to answer this question.

Evaluating the labour market impacts of Universal Credit: a feasibility study

The fact that people are notified in advance of their forthcoming migration may bias estimates of the average impact of migrating a certain type of claimant from legacy benefits to UC, if treated and comparison groups have had information for different amounts of time at the time the treated are migrated. If there is a concern that the notification itself may induce a change in behaviour (not shared by the comparison group who are notified later), then the evaluator could estimate the impact of being notified of forthcoming migration rather than actual migration.

An alternative question is, therefore:

- b What are the impacts of notifying in a given month a long-standing claimant that they will be migrated under transitional protection to UC in three months' time versus not notifying them in that month (and eventually notifying them later should the claim remain open) for long-term claimants whose circumstances have remained stable, in a world where a new claim triggers moving into UC?

For both questions, impacts can be assessed for long-term claims of a given type that start under the legacy benefit regime and survived for a certain number of months (depending on area) to reach the migration stage.

Impacts can further be disaggregated by claimant characteristics such as age, gender, family demographics or education. Outcomes of interest may include continuation of claim, individual employment rate, household employment rate, working hours, earnings, family income and poverty, all measured some time after the month of migration.

4.2.1 Methodology

Unlike the proposed evaluation design for new claims, the use of time variation is not appropriate for assessing the impact of migrating existing claimants:

- Similar claims reaching the same duration before and after the start of notification/migration in a given area will have been exposed to different policy environments over time and this may affect their composition (given observed characteristics). For example, people who have been on JSA for a year in January of implementation Year 3 will have had UC (rather than Working Tax Credit (WTC)) available for the last six to eight months of that period if they had moved into work; in contrast, people who have been on JSA for a year say in August of implementation Year 2 will have faced a regime which offered WTC almost throughout that spell if they moved into work. Faced with these two different offers, the groups who stay on JSA in each case may be different kinds of people. If so, their use for difference in differences (DiD) may introduce bias in the estimates of the impact of UC.
- Claimants that are notified/migrated later will, by virtue of that fact, have different characteristics at the point of notification/migration from those notified/migrated earlier. In particular, the group of claimants notified/migrated later will be, on average, of worse composition, as by definition it will not include claimants who were 'successful' in leaving the stock of long-term claimants.

Comparisons across areas are possible, provided there is some geographical variation in the notification/migration of stocks, but it may be hard to disentangle the treatment effect from differences in local labour markets when variation over time is unavailable.

Instead, exploiting variation within area and time period in who is notified or migrated is a more promising approach in this phase. Given their common claim history, we expect the

Evaluating the labour market impacts of Universal Credit: a feasibility study

stock of claimants of a given type in a given month to be a relatively homogeneous and very specific group. Controlling for claim duration and other observed characteristics further aids comparability of treated and controls.

For existing claimants, the evaluation strategy can, therefore, be carried out within area and time period for as long as there is some random selection of spells to be notified/migrated. Specifically, if, in any given month, some people are notified/moved over to UC while other, similar, people in the same area are (essentially randomly) not notified/moved over, their subsequent outcomes can be compared.

This strategy can best be implemented using matching methods, whereby claimants of a given type notified/migrated to the full UC regime during a given month are matched to claimants of the same type and in the same area who are not notified/migrated in that month. The required assumption is the conditional independence assumption (CIA) that the timing of notification/migration, given observed characteristics, in a given area is unrelated to labour market outcomes.

4.2.2 Data

- Data allowing for the identification of all (or a large random sample of) ongoing claims of a given type, including claim start dates, time of notification and time of migration.
- Data on outcomes of interest (individual and family labour market outcomes, claiming status or measures of wellbeing) measured some time after notification or migration.
- Individual and/or family background characteristics and labour market histories to support construction of the counterfactual.

4.2.3 Caveats and limitations

- **Violation of the identifying assumption**

The outlined approaches rely on variation in notification/migration dates among claimants of a particular type in a particular area being essentially random – a design feature which is yet to be decided.

Indeed, for non-JSA legacy benefit claims it has been envisaged to migrate claimants who gain from UC before those who require transitional protection. This selection may compromise the value for evaluation purposes of the variation introduced by the gradual migration of legacy benefit claimants. It means that those ‘not yet migrated/notified’ at a given time will be disproportionately composed of those who do not benefit from the move. Any evaluation exercise will be biased if the evaluator cannot construct a comparison group similar to those treated (migrated or notified) at that point in time – either because all those benefiting are migrated at once or because not enough information is observed to determine exactly who benefits from the migration. The assessment of the possibility of using a matching (or indeed regression) approach to correct for selection bias should always be rooted in an understanding of the sources of that bias. In this context, what the evaluator needs to understand is the process by which some people are selected for earlier migration than others, which will become known at some stage. If such factors are recorded in the data, and if there is still residual variation in the timing of migration conditional on such factors, a strong case could be made for the reliability of matching/regression methods.

Evaluating the labour market impacts of Universal Credit: a feasibility study

- **Length of evaluation period**

The impacts of what is essentially the timing of notification/migration can be assessed over an arbitrarily long time horizon. However, all claims will eventually be migrated within a fixed number of months (e.g. six months for JSA claims), implying that any differences in outcomes between those notified/migrated earlier and those notified/migrated later will most probably begin to vanish thereafter. Any potential impact is, therefore, likely to be present only in the shorter term.

- **Interpretation of the casual impact**

The 'ideal' parameter one could have hoped to identify would be the impact of being subject to the full UC regime versus remaining on the full legacy benefit regime for the population who remained under the legacy benefit regime after the flows have been phased in the way they were. The information content of the treatment effect parameter that can be identified at this stage is, however, limited by the nature of the treatment and of the comparison treatment. As to the treatment:

- i in possibly the most reliable case, treatment is being notified of the forthcoming migration to UC; and
- ii transitional protection means that being notified/migrated to UC need not represent the policy as it will eventually be in place.⁵

The interpretation of the estimated parameter is influenced also by the comparison treatment: a hybrid regime in which people in the comparison group;

- iii move to UC if they start a new claim; and
- iv are notified and moved into UC, possibly shortly after the treated group, if their circumstances remain unchanged.

- **External validity**

Effects are estimated only for long-term claimants, and only for those who have remained out of UC since the start of its roll-out in their area. These are potentially disproportionately composed of those not gaining from the new system.

In this scenario JSA claimants, for instance, must have started a JSA claim at least nine months before migration (six months before notification), and in some cases more than two years before migration (for some single claimants without children in phase 1 areas who are currently planned to be migrated to UC only in May of implementation Year 3), and must have remained out of UC over this period, in particular not started a new claim to what would previously have been HB or CTC, for example, after new claims to those were replaced by UC.

- **Anticipation effects (for question A)**

As mentioned above, the fact that people are notified in advance of their forthcoming migration may bias estimates of the average impact of migration. To the extent to which individuals anticipate potential benefits/losses associated with the migration, they may respond to being notified, in advance of their actual migration. Such behavioural responses might lead to compositional differences between those migrated in a given month and those not migrated in that month. If based on unobservables, such differences would bias the estimates of the impact of migration.

⁵ Note that, even though transitionally protected migrated claimants and non-migrated claimants both have the same current award and both face moving into UC if they start a new claim to what would previously have been a different legacy benefit, they do not face quite the same financial options (so it is not evaluating just the non-financial aspects of UC) because non-migrated claimants would still face the legacy benefit regime for changes in circumstances that did not trigger a new claim.

4.3 Evaluation Question 2: Substitution, displacement and other indirect impacts of UC

The question of interest is: What are the impacts of moving a large number of individuals from a legacy benefit into UC on the outcomes of those not directly affected by the migration?

The focus here is on how the outcomes of those who are not directly affected by the migration (but who are already subject to the full UC regime) may change as a consequence of a large influx into UC by individuals who were on a legacy benefit.

The channels through which such indirect effects may take place are informational in nature for all type of claimants, while for some groups, such as TC claimants, there would additionally be a real change in incentives. Specifically, consider the stock of long-term JSA claimants. These individuals already face the full incentives of the UC regime. The only channel which might give rise to changes in behaviour is informational in nature: once migrated, or indeed notified, they might become (even) more aware of what the incentives are under UC, potentially leading to changes in their behaviour in the labour market (e.g. they might become more willing to work in mini-jobs). For the long-term stock of TC claimants, in addition to such informational effects, the actual move to UC may directly affect incentives. This is because parents who have an ongoing claim of TC are not facing the UC incentives unless they either start a new claim to any other (non-TC related) benefit or stop their TC claim for some time before starting a new claim to UC of what would have been TC. Both of these channels show that it may be costly for them to move to UC prior to their migration. Such individuals do not, therefore, face the rest of the UC entitlement rules and incentives; for instance, should they decide to work in a mini-job, they would not be entitled to the work component of the TC; similarly, their budget constraint by hours worked is different from what they would face under the UC regime.

Whatever the channels (information and/or incentives), more individuals could now be looking for jobs. The labour market becoming tighter due to increased job competition may lead to substitution effects (whereby the other job seekers may find it harder to find a job), while effects on wage rates may cause displacement effects (those currently in employment are replaced by new workers willing to accept lower wage rates).

The second source of variation discussed in section 4.1 – in some areas a bulk of claims are moved to UC, while comparable areas have not started the migration of that type of claim yet – can be exploited to answer this question.

The methodological discussion that follows is the same for all claimant types, though of course effects could and should be assessed separately for each type.

Impacts can further be disaggregated by characteristics of the relevant population such as age, gender, family demographics or education. Outcomes of interest may include individual employment rate, household employment rate, working hours, earnings, family income and poverty, all measured once the (bulk of) the stock has been migrated.

4.3.1 Methodology

Depending on what types of effects one is trying to capture, the population of interest can be:

- job seekers not part of the legacy benefit stock to assess substitution effects;
- workers in low-wage jobs not part of the legacy benefit stock (i.e., not on TC) to assess displacement effects;
- the working-age population not part of the legacy benefit stock – or a more exposed sub-group such as the low educated – to also assess both substitution and displacement effects.

The treated within the reference population are those living in an area where a large bulk of legacy benefit claimants are migrated to UC; the comparisons within the reference population are those living in comparable areas where such ‘sudden’ influx of claimants will happen only later.

A natural way to assess the impact of a large influx into UC is to use matching methods. To exemplify, consider the stock of JSA claimants and the reference population as job seekers not on JSA. To assess the presence of substitution effects, one could match job seekers not in the JSA stock in an area that migrates the full JSA stock earlier to job seekers not in the JSA stock in areas that start the migration later. Matching could be done in two steps (first matching area, then individuals), or in one step (matching on local and individual characteristics at the same time). Crucial area-level variables to balance would be the time series of local employment rates and the history of UC roll-out. From then on, one could simply compare average employment rates of the treated and matched controls over time, until the comparison area starts migrating the stock.

As discussed for question 1 above, the gradual change in the benefits system precludes the use of time variation. Cross sectional (area- or time-based) DiD is, therefore, not appropriate as selection into unemployment may be affected by the benefits system, which changes during the period.

The use of panel DiD is not advised when one is conditioning on unemployment, as this type of conditioning imposes further restrictions on the common trend. By contrast, for reference populations other than job seekers this method might be appropriate for removing the fixed effect (in particular, the area fixed effect), and it might fruitfully be combined with matching.

4.3.2 Data

- Data allowing for the identification of all (or a large random sample of) the reference population (job seekers, workers or working-age population not part of the legacy benefit stock) in areas migrating the stock earlier and similar areas migrating it later.
- Data on outcomes of interest (individual and family labour market outcomes, claiming status or measures of wellbeing) measured once the bulk of the stock has been migrated.
- Individual and/or family background characteristics and labour market histories to support construction of the counterfactual.
- Historical data on local labour market outcomes (in particular, employment rates) to ensure comparability of areas being contrasted.

4.3.3 Caveats and limitations

This design is not expected to suffer from anticipation or entry effects. The most important caveats appear instead to be the following.

- **Availability of the required variation**

This design not only requires some variation over time when areas start migrating their stock of a given claim type, but it also requires that in the areas migrating earlier, as large as possible a number of claimants are migrated in as short as possible a period, while the comparable area hasn't started migration yet.

- **Length of evaluation period**

The assessment of the indirect effects of a large influx of legacy benefits into UC can only be assessed over a limited horizon, determined by how soon after completion of the (bulk of) the migration in the treated area the control area starts migrating its own stock.

- **Definition of the treatment effect**

The evaluation question this design is trying to answer concerns the impact of moving a large number of individuals from a legacy benefit into UC on the outcomes of those not directly affected by the migration itself, but indeed already under the full UC system. In other words, it concerns the impact of a large and 'sudden' migration to UC on a population already subject to UC.

4.4 Overall assessment of the evaluation possibilities offered by the migration of the stock of legacy benefit claimants

In the scenario provided, these phases target a very specific population (long-term legacy benefit claimants who managed to stay out of UC at the minimum for six/nine months prior to notification/migration). Using within-area variation in the time of notification/migration, we propose using matching to assess the impact of notifying/migrating an ongoing claim into the full UC system at a given point in time versus postponing migration. This is more the impact of the timing of notification/migration, by no means the impact of UC for this population. The definition of the estimated effect is further determined by:

- how soon the control group is notified/migrated if their claim continues;
- the fact that any new claims by the control group will also be to UC, possibly before they would be moved into UC through managed migration; and
- the transitional protection shielding claimants' benefits and implying that the individuals migrated do not face the full UC regime while their claim continues. More generally, migrating existing legacy benefits to UC is an inherently one-off enterprise, which further reduces the policy relevance of its effects.

Provided in some areas a bulk of claims are moved to UC, while comparable areas have not started the migration of that type of claim yet, these phases also offer the possibility to assess potential substitution and displacement effects of a large influx into UC – though only for a population which is already subject to UC. Such effects could be assessed only over a short time horizon.

5 Overall assessment of evaluations that are possible in principle and alternative implementation strategies

In this section we start by summarising and assessing the evaluation possibilities examined in detail in sections 3 and 4. Specifically, in section 5.1 we consider each of the three evaluation questions we identified, assessing the scope provided by the roll-out scenario for addressing each of them in each implementation phase and reaching an overall view as to which ones would be the most promising evaluation exercises to pursue in this hypothetical roll-out scenario. In section 5.2 we then discuss one wide-ranging challenge: how to isolate the impacts of Universal Credit (UC) from those of other reforms happening around the same time. In sections 5.3 and 5.4 we give a systematic overview of two other issues that have recurred through our discussion: first, the extent to which both short-run and long-run impacts can be estimated, and second, the extent to which estimated impacts can be disaggregated by sub-group of the population and by aspect of the policy. Finally, in section 5.5 we discuss alternative implementation scenarios that might facilitate the evaluation of the impacts of UC.

5.1 Summary and assessment of each evaluation question

This feasibility study has proposed a number of evaluation questions that the hypothetical scenario for UC implementation would allow one to address, provided some assumptions are met. In general, it is not possible to determine upfront whether these assumptions are likely to hold; they are mostly ‘identifying’ (hence untestable) assumptions, support for which can only be indirectly obtained through auxiliary analyses that could be done empirically as part of the actual evaluation itself.⁶ The feasibility study has, therefore, focused on what is feasible in principle and highlighted what checks and supporting exploratory or corroborating analyses would need to be undertaken to ensure that the estimates obtained are reliable.

As seen from the discussion in the previous sections, the implementation of UC in each phase of the hypothetical roll-out scenario raises specific evaluation challenges to internal validity, as well as restricting the type of evaluation questions that can be addressed in one or more dimensions: treatment, counterfactual and the reference population (external validity). The evaluation exercises become more difficult (internal validity), and the questions that can be addressed narrower (external validity), as one moves further into the roll-out of UC.

⁶ One cannot conclude from past evaluations how likely it is that some of the assumptions, like common trends, will hold because in evaluating UC we will be looking at different and specific sub-groups.

Evaluating the labour market impacts of Universal Credit: a feasibility study

Furthermore, the effects that can be identified during the roll-out of UC to new claims and the migration of existing claimants intrinsically depend on the specific details of the implementation strategy and are not necessarily representative of what the impacts of UC will be in the longer-run, when fully implemented (external validity).

Against this backdrop, Tables 1 and 2 summarise the treatment effect parameters (or evaluation questions) proposed and discussed in sections 3 and 4, together with a concise overall assessment of both their:

- **reliability** – the expected likelihood that internally valid, hence unbiased, estimates can be obtained. Evaluation exercises that allow the use of a variety of methods to identify the same treatment effect, and therefore, allow for an assessment of the robustness of the estimates, are considered more reliable *a priori*; and
- **relevance** – the expected representativeness, or external validity, of the estimates, and how interesting the effect identified is likely to be. Note that a major limitation in this dimension that applies to all the evaluation questions is the rather limited time horizon over which UC can be compared with the full legacy benefit regime in this implementation scenario.

We assess each of these on a six-point scale (highest, high, relatively high, relatively low, low, very low).

A third metric that still needs to be considered concerns data availability, an issue to which we return in section 6.

Table 1.1 Evaluation possibilities arising from the roll-out of UC to new claimants

Overall question	Phase-specific question	Reliability (internal validity)	Relevance (external validity)
Q1 Impact of the full UC regime versus the full legacy benefit regime on the outcomes of individuals starting a new claim of a certain type Reference population: new claimants	Phase 1: Impact of starting a new UC claim under the full UC regime: <ul style="list-style-type: none"> • versus starting a similar JSA claim under the full legacy benefit regime; • on single unemployed claimants without children in the Pathfinder areas in the 1st 6 months of implementation (phase 1). 	Highest: <ul style="list-style-type: none"> • Entry and anticipation effects not expected to be strong. • Can test some of the assumptions to some extent. • Robustness checks (more than one method can be applied both across areas and across time). 	Low: <ul style="list-style-type: none"> • Nature of policy (may not be final version, teething problems, unfamiliarity). • Small specific claimant group. • Only 4 Pathfinder sites.
	Phase 2: Impact of starting a new UC claim under the full UC regime: <ul style="list-style-type: none"> • versus starting a similar JSA claim under the full legacy benefit regime; • on unemployed claimants in the 2nd 6 months of implementation (phase 2). 	Highest: <ul style="list-style-type: none"> • Entry and anticipation effects possibly still mild and less of a concern when using time variation. • Can test some of the assumptions to some extent. • Robustness checks (more than one method can be applied both across areas and across time). 	High <ul style="list-style-type: none"> • More established policy. • All types of JSA claims.
	Phase 3: Impact of starting a new UC claim under the full UC regime: <ul style="list-style-type: none"> • versus starting a similar TC/ISIp claim under the full legacy benefit regime; • on new UC claimants between May and July of Year 2. 	None: <ul style="list-style-type: none"> • No valid comparison group available. 	Not applicable.

Continued

Evaluating the labour market impacts of Universal Credit: a feasibility study

Table 1.1 Continued

Overall question	Phase-specific question	Reliability (internal validity)	Relevance (external validity)
Q1 continued	<p>Phase 4: Impact of starting a new UC claim under the full UC regime:</p> <ul style="list-style-type: none"> • versus starting a similar ISc/ESA/HB claim under the full legacy benefit regime; or • versus starting a similar ISc/ESA/HB claim under a hybrid regime in which UC has replaced only new JSA claims or only new JSA/TC/IS(PLs) claims; • on carers, sick/disabled people and low income families who are entitled to ISc/ESA/HB under the legacy benefit regime, start a new claim on or after October of Year 2 and have not started a UC claim equivalent to JSA or TC/ISlp since UC replaced such claims in their area. 	<p>Relatively low:</p> <ul style="list-style-type: none"> • No robustness checks; relying on time variation. • Comparing UC with full legacy benefit regime less reliable. • Can test some of the assumptions to some extent (test historical comparability of treated and comparison groups). 	<p>Very low:</p> <ul style="list-style-type: none"> • Limited representativeness of treated group. • Less clean interpretation of treatment effects when counterfactual is a hybrid regime.
<p>Q2</p> <p>Market-wide impact (or impact on the outcomes of specific groups) of introducing UC for some new benefit claims versus maintaining the full legacy benefit regime (or some hybrid regime with more restricted UC coverage)</p> <p>Reference population: all those who may gain entitlement depending on their actions.</p>	<p>Phase 1: Impact of UC replacing new JSA claims by single people without children:</p> <ul style="list-style-type: none"> • versus the full legacy benefit regime; • on the entire working-age population or some sub-group. 	<p>Very low:</p> <ul style="list-style-type: none"> • Very small population directly exposed to UC makes aggregate effects unlikely to be detectable. 	<p>Very low:</p> <ul style="list-style-type: none"> • Nature of policy (may not be final version, teething problems, unfamiliarity). • Small specific claimant group. • Only 4 Pathfinder sites. • Low external validity of any results on aggregate effects.

Continued

Evaluating the labour market impacts of Universal Credit: a feasibility study

Table 1.1 Continued

Overall question	Phase-specific question	Reliability (internal validity)	Relevance (external validity)
Q2 continued	<p>Phase 2: Impact of UC replacing new JSA claims:</p> <ul style="list-style-type: none"> • versus the full legacy benefit regime; • on the entire working-age population or some sub-group. 	<p>Low:</p> <ul style="list-style-type: none"> • Small population directly exposed to UC, with effects 'diluted' across the whole market (and no obvious small sub-group to focus on), makes aggregate effects difficult to detect. 	<p>Low:</p> <ul style="list-style-type: none"> • Low external validity of any results on aggregate effects.
	<p>Phase 3: Impact of UC replacing new JSA/TC/ISlp claims:</p> <ul style="list-style-type: none"> • versus the full legacy benefit regime; or • versus a hybrid regime where UC has replaced only new JSA claims; • on the entire working-age population or some subgroup (in particular, lone parents with children under 5). 	<p>High:</p> <ul style="list-style-type: none"> • With quicker within-area roll-out, extended eligibility and combination of JSA and TC/ISlp claims, volume of UC cases might be sufficient to detect aggregate impacts. • Can test some of the assumptions to some extent. • Robustness checks (more than one method can be applied both across areas and across time). 	<p>High:</p> <ul style="list-style-type: none"> • Highly policy-relevant question (captures consequences of changes in entitlement rules and how these may affect claim decisions and accounts for all types of effects from the reform – direct and indirect). • External validity issues may remain (due to entry and anticipation effects).
	<p>Phase 4: Impact of UC replacing all new legacy benefit claims:</p> <ul style="list-style-type: none"> • versus the full legacy benefit regime; or • versus a hybrid regime where UC has replaced only new JSA claims or only new JSA/TC/ISlp claims; • on the entire working-age population or some sub-group. 	<p>Relatively low:</p> <ul style="list-style-type: none"> • With one-month roll-out and full coverage of all new claims, volume of UC cases might be sufficient to detect aggregate impacts. • Anticipation effects • Can test some of the assumptions to some extent. • No robustness checks; relying on time variation. • Comparing UC with full legacy benefit regime less reliable. 	<p>Highest:</p> <ul style="list-style-type: none"> • Highly policy-relevant question (captures direct and indirect effects).

Continued

Evaluating the labour market impacts of Universal Credit: a feasibility study

Table 1.1 Continued

Overall question	Phase-specific question	Reliability (internal validity)	Relevance (external validity)
Q3 Substitution, displacement and other indirect impacts of introducing UC for some new benefit claims versus maintaining the full legacy benefit regime (or some hybrid regime with more restricted UC coverage) on the outcomes of those not directly affected by the reform Reference population: those not directly affected by the reform.	Phase 1: Indirect effects of UC replacing new JSA claims by single people without children: <ul style="list-style-type: none"> • versus the full legacy benefit regime; • on parents and individuals in couples (the entire working-age population of parents or individuals in couples, or just the unemployed, or just those starting a JSA claim). 	Very low: <ul style="list-style-type: none"> • Very small population directly exposed to UC makes indirect effects unlikely. 	Very low: <ul style="list-style-type: none"> • Nature of policy (may not be final version, teething problems, unfamiliarity); • Small specific claimant group; • Only 4 Pathfinder sites; • Low external validity of any results on indirect effects.
	Phase 2: Indirect effects of UC replacing new JSA claims: <ul style="list-style-type: none"> • versus the full legacy benefit regime; • on lone parents with a child aged under 5, carers, disabled people. 	Very low: <ul style="list-style-type: none"> • Small population directly exposed to UC makes indirect effects likely to be immaterial. 	Low: <ul style="list-style-type: none"> • Low external validity of any results on indirect effects.
	Phase 3: Indirect effects of UC replacing new JSA/TC/ISlp claims: <ul style="list-style-type: none"> • versus a hybrid regime where UC has replaced only new JSA claims; • on those who already started a UC claim before the start of phase 3. 	Relatively high: <ul style="list-style-type: none"> • With quicker within-area roll-out and extended eligibility, low volume of UC cases might be less of an issue; • No anticipation effects • Can use matching across time or across area, as well as DiD across area; • However, limited scope for robustness checks, as DiD across area and matching across time restrict the population of interest to those who started a UC claim close to the end of phase 2. 	Relatively high: <ul style="list-style-type: none"> • Potential to capture substitution and displacement effects of expanding UC coverage; • However, only on a population which has already started a UC claim, and at a specific time.

Continued

Table 1.1 Continued

Overall question	Phase-specific question	Reliability (internal validity)	Relevance (external validity)
Q3 continued	<p>Phase 4: Indirect effects of UC replacing all new claims:</p> <ul style="list-style-type: none"> • versus a hybrid regime where UC has replaced only new JSA/TC/ISlp claims; • on those who started a UC claim close to the end of phase 3; <p>or</p> <ul style="list-style-type: none"> • versus a hybrid regime where UC has replaced only new JSA claims; • on those who started a UC claim close to the end of phase 2. 	<p>Relatively low:</p> <ul style="list-style-type: none"> • With one-month roll-out and full coverage of all new claims, low volume of UC cases might be less of an issue; • No anticipation effects; • Robustness checks are limited as can only use time variation; can use historical evidence on comparability of time periods being contrasted. 	<p>Relatively high:</p> <ul style="list-style-type: none"> • Potential to capture substitution and displacement effects of expanding UC coverage; • However, only on a population which has already started a UC claim, and at a specific time.

Table 1.2 Evaluation possibilities arising from the migration of existing claim to UC

Question	Reliability	Relevance
<p>(1) Impact of migrating a long-standing legacy benefit claimant to UC under transitional protection in a given month:</p> <ul style="list-style-type: none"> • versus not migrating them in that month (and eventually migrating them later should the claim remain open); • on long-term claimants whose circumstances have remained stable; • in a world where a new claim/natural migration already triggers a move to UC. 	<p>Relatively low:</p> <ul style="list-style-type: none"> • Anticipation effects from notification might be strong; • Requisite randomness of migration more likely for existing JSA than other claims; in any case, evaluator will know the sources of selection once migration plan finalised. 	<p>Low:</p> <ul style="list-style-type: none"> • Treatment is migration is under transitional protection; • Counterfactual is hybrid regime where the comparisons will, at some point, move into the full UC regime; • Unrepresentative population.

Continued

Evaluating the labour market impacts of Universal Credit: a feasibility study

Table 1.2 Continued

Question	Reliability	Relevance
<p>(2) Impact of notifying in a given month a long-standing legacy benefit claimant that migration under transitional protection will happen in 3 months' time if the claim remains open:</p> <ul style="list-style-type: none"> • versus not notifying them in that month (and eventually notifying later should the claim remain open); • on long-term claimants whose circumstances have remained stable; • in a world where a new claim/natural migration already triggers a move to UC. 	<p>Relatively high:</p> <ul style="list-style-type: none"> • Requisite randomness of notification more likely to hold for existing JSA than other claims; in any case, evaluator will know the sources of selection once migration plan finalised. 	<p>Very low:</p> <ul style="list-style-type: none"> • Treatment is notification is of forthcoming migration under transitional protection; • Counterfactual is hybrid regime where the comparisons will, at some point, move into the full UC regime; • Unrepresentative population.
<p>(3) Impact of moving a large number of individuals from a legacy benefit into UC:</p> <ul style="list-style-type: none"> • versus no such large influx into UC; • on those not directly affected by the migration but who are already subject to the full UC regime. 	<p>Uncertain:</p> <ul style="list-style-type: none"> • Requires that in some areas a bulk of claims are moved to UC, while comparable areas have not started the migration of that type of claim yet. 	<p>Relatively high:</p> <ul style="list-style-type: none"> • Potential to capture substitution and displacement effects of a large influx into UC; • However, only on a population which is already subject to UC.

Evaluating the labour market impacts of Universal Credit: a feasibility study

As can be seen from the tables, despite the methodological hurdles outlined in sections 3 and 4, there are a few policy-relevant questions that can be addressed with relatively high expected reliability. Of course, we can only speculate on how important the various threats to the validity of the estimates will be: there is no *a priori* guarantee that, were these evaluation exercises carried out, the resulting estimates would be reliable. How reliable such estimates could be deemed in practice would instead hinge on the results emerging from the various ancillary and corroborating analyses that have been suggested. With this caveat in mind, the most promising evaluation questions appear to be:

- 1 **The impact of making a claim under the full UC regime versus the full legacy benefit regime, on people starting what would previously have been a Jobseeker's Allowance (JSA) claim.** This is a very interesting treatment effect, comparing outcomes under the full new regime to those under the full legacy benefit regime for important groups of claimants who are exposed to UC for their first time. Indeed, these claims are expected to represent a majority of all new UC claims.
- 2 **The impact of UC replacing new JSA, Tax Credit (TC) and Income Support for lone parents (ISlp) claims versus the full legacy benefit regime (or versus a hybrid regime in which UC has replaced only new JSA claims), on the entire working-age population or some policy-relevant sub-group thereof.** This is arguably the most policy-relevant question that can be addressed – at least for groups (such as lone parents with children under five) for whom JSA is not relevant and for whom geographical variation can, therefore, be used to make comparisons of UC versus the full legacy benefit regime. Phase 3 in this scenario is a key implementation phase which extends UC to major target groups (low earners and non-working lone parents of pre-school children) and introduces significant changes in entitlements for new claimants. By taking into account the consequences of these changes in entitlement rules and how these may affect claim decisions, this parameter captures all effects of the reform – the direct effects on who claims UC (changes in who is entitled given their current behaviour, people changing their behaviour to become entitled, and take-up among those entitled), the direct effects on claimants' employment and other labour market outcomes, and the indirect effects arising through market interactions.
- 3 **The indirect effects of UC replacing new JSA, TC and ISlp claims (or all claims), versus a hybrid regime, on those who had already claimed UC.** The choice of reference population as those who have already made a UC claim, though driven by methodological considerations, actually focuses attention on a group that is both of high policy relevance, given their dependence on benefits, and disproportionately likely to experience any indirect effects of the expansion of UC to new types of claim. While this type of indirect effect can be assessed exploiting the roll-out of UC in both phases 3 and 4, phase 3 might be of particular policy interest. Specifically, in phase 3, the indirect effects are those of UC replacing new TC/ISlp claims, compared to a hybrid regime where UC has replaced only new JSA claims, on individuals who in the past have started a UC claim (which would previously have been to JSA).
- 4 More tentatively, the **indirect (substitution and displacement) effects of moving a large group of existing legacy benefits onto UC, versus not moving this large group and leaving them in a hybrid regime in which a new claim would move them to UC ('natural migration'), on those already subject to the full UC regime.** The feasibility of addressing this evaluation question is more uncertain, as it is not yet known whether the required source of variation will be available (i.e. whether in some areas a large group of legacy benefit claimants will be moved to UC while comparable areas will not have started the migration of that type of claim yet). Additionally, substantial (and therefore, detectable) indirect effects might be expected only from migrating some claimant types, specifically long-term TC claimants.

In the discussion so far we have completely abstracted from data availability issues. In section 6 we turn to consider what questions (including in terms of outcomes and populations) the available data would allow the evaluator to assess. But before that, we consider three further elements of what these evaluations might uncover: the extent to which UC can be separated from other changes happening at the same time, the scope to identify long-run and short-run effects, and the feasibility of disaggregating the estimated impacts.

5.2 The challenge of isolating UC from other changes

One particularly noteworthy difficulty for evaluating the impact of UC is how to separate the effects of UC from the effects of other changes happening around the same time. This is not really a separate issue from those considered so far – it is simply one threat to the assumptions on which difference-in-differences (DiD) and matching methodologies rely – but it merits a separate discussion in its own right because the problems it may cause are policy-induced, potentially serious, and not yet fully determined because they depend in part on future policy decisions.

Estimation of the impact of UC can be threatened/invalidated by any other change taking place during the evaluation period that may also affect the outcome of interest: it can be hard to separate what is the impact of UC from what is the impact of the other change. The confounding impact of such a change depends on the methodology being used. DiD results may be affected if the ‘before’ and ‘after’ periods are differentially exposed to the change, unless it affects treated and comparison groups in exactly the same additively separable way (i.e. it adds the same amount to outcomes for treated and comparisons). Both matching and DiD results can be affected if treated and control groups are differentially exposed to the change, react differently to it or if its impact is not additively separable. That is, between the benefit claim of the comparison (pre-UC) group and the outcome measurement of the treated (UC) group, unless the same change also happens (and has the same effect on outcomes) a year earlier and so is stripped out by the ‘before’ comparison. Hence, regular seasonal variation is not a problem, but one-off reforms will in general invalidate estimates of the impact of UC as the effects of UC cannot be separated from the effects of those other reforms.

The full set of policies that will be introduced during the UC roll-out period is subject to future government decisions, but about the actual timing of UC roll-out and about other reforms. For example, welfare reforms that had already been implemented or announced at the time this report was being written (winter 2012-13) which could be problematic for evaluation if coinciding with the roll-out of UC include:

- real-terms reductions in most working-age benefit rates (April 2013, April 2014 and April 2015);
- reductions in Housing Benefit (HB) for those under-occupying socially rented properties (April 2013);
- the overall benefit cap (April 2013); and
- ending new claims to In Work Credit (October 2013).

These reforms apply nationwide. If they have the same (additive) effects in all areas, then they might pose a problem for DiD across time, but not for matching or DiD across areas. However, in some cases the same reforms might have different effects in different areas. Uniform reductions in HB (or TC) across the country might have different effects in areas with different rent (or wage) levels; the benefit cap affects far more people in the highest-rent areas (notably London) than elsewhere. Furthermore, other reforms that have been, or are due to be, implemented differ explicitly by area. These include:

- replacing Council Tax Benefit (CTB) with locally-varying council tax support (CTS) schemes (April 2013, with local schemes potentially changing again each April thereafter); and
- uprating local housing allowance more rapidly in areas with high rent growth (April 2014 and April 2015);

Such reforms can compromise comparability across areas. For instance, CTB was replaced by locally varying CTS schemes in April 2013. Different local authorities have made widely varying choices as to the design of their new schemes (see <http://counciltaxsupport.org/the-story-so-far/>) which will have differential effects on incomes and work incentives, and, therefore, on the outcomes of interest for evaluating UC.

One can try to overcome the problem by matching areas along the relevant characteristics. In this example, the possible solution is to choose comparison areas that have similar CTS schemes at the time UC is introduced (and similar CTS scheme histories), although the scope to do so is limited by the small number of areas one can choose from – bearing in mind that we would be looking to match on many other area characteristics even in the absence of this reform.⁷ Similarly, matching on local rent levels and growth would be important to deal with reforms whose effects vary by those characteristics. Finally, it should be noted that these kinds of problems are not confined to policy reforms: other area-specific shocks (such as the closure of a large local employer) can also confound estimates of the impact of UC if not controlled for.

5.3 Long-run versus short-run effects of UC

We can distinguish between ‘long-run’ and ‘short-run’ effects in three distinct senses.

⁷ Local authorities’ decisions in April 2013 are not the end of the story. By the start of UC phase 2 (or phase 1 for LAs in the Pathfinder areas) LAs will also have to decide how their local CTS schemes will interact with UC (e.g. whether UC will count as income in the means test for CTS). This will directly affect how UC affects claimants, and LAs might again make varying decisions. LAs might then amend their schemes (differentially) in April 2014 – and indeed central government is giving many of them an incentive to change their schemes because the funding provided for councils whose schemes met certain criteria is being withdrawn after a year, so that local authorities (LAs) that satisfied the criteria in 2013–14 will have less incentive to do so in 2014–15. When looking at later evaluation periods, therefore, the evaluator might need to match on these subsequent choices as well as the scheme originally chosen in April 2013.

Evaluating the labour market impacts of Universal Credit: a feasibility study

First, the effects on people's outcomes a short or a long time after being exposed to UC. The evaluation strategies discussed in sections 3 and 4 can identify some of the impacts of UC versus the full legacy benefit regime only up to the point where the comparison group is exposed to UC. In each case we have indicated roughly how long that is in the hypothetical implementation scenario. However, the length will be sensitive to any adjustments to the actual roll-out schedule and sensitive to the empirical assessment of the reliability of the analysis (the latter is mostly relevant when exploiting time variation, which generally allows for longer horizons); but it is typically restricted to only a few months. The outcomes for treated and comparison groups can continue to be compared beyond that point – in principle, indefinitely – but for such comparisons the effect being identified is very different, and much less interesting: it is the impact of, for example, making a claim under the full UC regime versus making a claim under a regime which is the legacy benefit regime at first and then the UC regime after X months – in other words, the impact of being exposed to UC X months sooner, rather than the impact of being exposed to UC *per se*. Second, the impact of the policy when it is first implemented versus when there has been time for any operational 'teething problems' to be ironed out and for (actual and potential) claimants to have become familiar with the new system.

This is likely to be more of an issue for earlier phases of the roll-out, since one would expect operational 'teething' problems to be ironed out over time and for people to learn about the new regime more quickly if it already covers more of the rest of the population. It is not necessarily quite so simple, however. Later phases may involve new or different kinds of operational teething problems if, for example, they affect groups in more complex circumstances. Or groups exposed to UC later may be different in other respects that mean they take longer to become familiar with the new system – for example, if they and their peer group are in less regular contact with the benefits system.

This problem can be reduced by choosing a treatment group exposed to UC later in the implementation phase – either because it is composed of those making a new claim later or because some claimant types are brought into UC later. However, there is a trade-off here: the later the treatment group is observed, the shorter the horizon over which one can compare outcomes under the UC regime the full legacy benefit regime using geographical variation (as non-treated areas will be under the full legacy benefit regime for a shorter period) or the further away in time one needs to draw the comparison group when using time variation (possibly compromising the reliability of the estimates).

One particular way in which we know that the policy will be different in its early stages is the role of transitional protection. In the short run, no current benefit recipient will see their cash entitlement reduced by UC if their circumstances remain unchanged. This will become gradually less relevant:

- First, as time goes on, fewer and fewer people will still be in the same circumstances, receiving the same amounts of the same benefits as they were when UC was introduced, and so ever fewer people will be eligible for transitional protection.
- Second, existing entitlements will be protected only in nominal (cash) terms, falling relative to UC rates which will (presumably) be uprated in line with inflation each year, so transitional protection will become a decreasingly generous offer that ever fewer people will find beneficial (and to an ever decreasing extent).

Evaluating the labour market impacts of Universal Credit: a feasibility study

This is clearly a major issue for the stock of existing claimants: in the short run, transitional protection creates an incentive for existing legacy benefit recipients for whom UC would be less generous not to change their circumstances – an incentive that will dissipate in the long run. Any evaluation of phases 5 and 6 must, therefore, be clearly understood as an evaluation of migrating people from legacy benefits to UC with transitional protection, not just UC *per se*. It is far from clear how interesting such an evaluation is at all. Focusing just on the sub-group for whom UC is at least as generous as their existing benefit might be better, but it raises additional issues: (a) it is a more selected sub-group, so less interesting in an external validity sense; (b) we are not sure at this stage whether this group will be well-defined/exogenous in a way that would allow evaluation; (c) the details of the legacy benefits need to be carefully studied to understand their value to claimants once the migration of existing claimants has begun, since their value may depend on features other than the relative generosity of the benefit (such as its insurance value).

Transitional protection can also affect evaluation of the impact of UC on new claims. It enhances the desirability of legacy benefits for some. This could induce people to make a claim to legacy benefits before such claims are replaced by UC and it could also affect the outcomes of those claiming legacy benefits by making them more reluctant to change their circumstances to avoid being moved into UC.

It is hard to see any way to estimate an impact of UC that is long-term in the sense of estimating what the impact will be once transitional protection has ceased to be a relevant feature of the policy. Transitional protection will be a relevant feature for several years after UC is introduced, almost certainly long enough to preclude any meaningful comparison between post-transitional-protection UC and the full legacy benefit regime.

Third and finally, the short- and long-run impacts of UC could differ simply because in the short run the policy has not been fully rolled out: many of the evaluation strategies proposed in sections 3 and 4 apply while UC covers only part of the population. This has two consequences:

- The indirect effects of UC may not fully have developed (including substitution and displacement effects, effects on wages or other market-level adjustments to the new benefits system). Such effects may touch everyone in the economy, whether or not they make a claim, and whether or not UC directly affects them. Therefore, the impacts of UC, while it has limited coverage, do not necessarily represent the impacts it will have in the longer run for the same group.
- The gradual extension of UC to different claim types makes it accessible (and indeed avoidable) through particular routes only. The composition of UC claims of some type may be influenced by this design, and differ from the composition of UC claims of the same type in the longer run when all benefit claims will be treated under the same umbrella. If the effects of UC are heterogeneous, the short- and long-run effects may, therefore, differ in line with the changing composition of the group in question.

5.4 Disaggregating the estimated effects of UC

5.4.1 Disaggregation by sub-group

It is clearly of interest to estimate how the impact of UC varies between different sub-groups of the population – for example, by age, sex, ethnicity, disability, education, or employment/benefit status and history. Where good estimation strategies are available at all, such breakdowns should generally be straightforward: the estimation can simply be done separately for each

Evaluating the labour market impacts of Universal Credit: a feasibility study

sub-group of interest. However, there are three conditions to bear in mind.

First and most simply, the characteristic in question must be recorded (accurately) in the data.

Second, the characteristics in question must not have been influenced by the policy. In some cases (such as age and sex), the characteristic is fixed, so this is not an issue. Where the characteristic might potentially be affected by the policy (such as employment or benefit claim status), using sub-groups defined by **status before exposure to UC** can still be unproblematic. However, some attention should be paid to possible anticipation effects: in principle one could imagine someone deciding to anticipate or postpone starting a job or a claiming spell, for example because a new UC phase is soon going to be implemented in that area. With the possible exception of benefit/employment status, anticipation behaviour seems unlikely to be a significant problem in practice. However, evaluation studies should endeavour to gather supporting empirical information on the importance of anticipation effects by assessing the impact of UC on the observed composition of claimants, as described in the main analysis. Significant differences in the distribution of some conditioning variable (e.g. employment status) between those exposed and those not exposed to the UC reform would suggest that it may not be appropriate to condition on that variable.

Finally, the sample sizes for each sub-group must be large enough to identify the impact of UC with sufficient precision. This can be a serious issue with survey data. It is much less problematic with administrative data because the sample sizes are generally much larger. But it could be a problem in phase 1 of the evaluation, where UC is introduced for only a sub-group of new claims for a limited period in a small part of the country. And how much of a problem it is will also depend on how small the sub-groups are: estimating effects on groups defined broadly as 'non-white' or 'disabled' would be much easier than looking at particular small ethnic groups or people with very specific types of health problem, for example, or defining sub-groups based on a combination of dimensions (e.g. disabled black women aged 25-34). The simplest and least restrictive approach to disaggregating impacts by sub-group is to perform the estimation separately for each sub-group. But where small sample sizes leading to large standard errors is a concern, greater precision can be gained by performing a single estimation, allowing the effect of the policy (and perhaps of one or two other selected variables) to vary by sub-group but imposing the restriction that other characteristics affect outcomes in the same way for all sub-groups.

5.4.2 Disaggregation by cause

UC is changing several different aspects of people's experience of the benefits system: current entitlements, financial work incentives, the presentation/transparency of the system, work search conditions, the claims procedure, the timing of assessments and payments, and so on. Ideally it would be good to know the contributions of each of these aspects to UC's overall impact.

Qualitative assessments might provide indications of how claimants, staff and stakeholders perceive impacts from the different aspects of UC. However, quantitatively evaluating individual aspects of the policy is generally impossible under the hypothetical implementation scenario because all the different aspects are being delivered together: the full policy is being introduced in the same way everywhere.

Some groups will be more affected by some aspects of the policy than others: for example, for unemployed single claimants without children their current entitlements are largely unchanged (unless they have significant assets/unearned income) and their work search conditions are little changed either (certainly compared with the potential changes for working couples, who might in future be required to seek higher earnings as long as their combined earnings are below £416 per week). We might infer that any impacts estimated for a specific group are driven by the aspects of the policy to which it is immediately exposed – in this case financial incentives, transparency and procedural changes – and take differences across groups exposed to different aspects of the reform to isolate the impact of one specific policy dimension. This is difficult for three reasons:

- First, some individuals may react in advance to changes in dimensions of the system that do not affect them immediately, in preparation for potential future changes in circumstances.
- Second, the effects of different features of the UC reform may not be additively separable, meaning that the impact of a reform to, for example, both work search requirements and financial work incentives may not equal the sum of their individual effects.
- Third, we would expect any effects to be heterogeneous across groups, depending on group-specific circumstances that affect outcomes such as employment or earnings. The impact of a reform that changes financial incentives, transparency and procedures for single people without children might be quite different from the impact of exactly the same reform for lone mothers since, for instance, the latter face much higher fixed costs of work (such as childcare) and might typically be closer to, and better informed about, the benefits system. In such cases, comparing estimates between groups to isolate the impact of particular features of the UC reform would not be valid, even if additive separability holds.

If it were possible to introduce only some aspects of the policy in certain areas and/or for certain claimants then, in principle, we might be able to evaluate the impact of different aspects separately. This might be worth considering, but in many cases the different aspects of the reform are intimately linked and would be hard to introduce separately. For example, it is difficult to see how one could arrive at the administrative and presentational integration associated with UC while retaining the financial incentive structure of the legacy benefit regime, or vice versa.

A less ambitious approach would be to introduce different variants of UC in different areas and/or for different claimants: for example, to have a single integrated benefit in all areas, but with different work-search conditions. In that case it might be possible to evaluate the effect of that variation. We do not pursue that further here, as it falls outside the scope of this report – to explore the feasibility of assessing the impact of UC as compared with the legacy benefit regime (rather than one variant of UC compared with another). But the feasibility of assessing the impact of UC by sub-group and cause would be much enhanced if there were some randomisation in how people were exposed to the different versions of the policy.

5.5 Alternative implementation strategies

The scenario we have based our analysis on is not the final roll-out plan which may turn out to be significantly different. At this stage we do not know what options could be seriously considered, so in the following we briefly explore hypothetical examples of how changing the strategy in different ways would change the feasibility and scope of evaluation

Evaluating the labour market impacts of Universal Credit: a feasibility study

exercises. In the following discussion we, therefore, uniquely focus on design issues, abstracting from data availability issues.

Changes that could increase the scope and reliability of the evaluation of UC include the following three designs.

• Design (A)

This design would roll out UC for all new legacy benefit claims in an area (or in some areas of the country) at the same time. This would be similar to phase 2, but encompassing all new claims to all the legacy benefits that UC will replace. This design would circumvent the methodological complications which arise in the current scenario from the fact that some benefit types get moved to UC before others. Ideally this would also include the simultaneous migration of the entire stock in those areas where UC is rolled out.

Advantages and disadvantages of this design are:

- Higher volume could allow for a more precise and representative estimation of aggregate and indirect effects – indeed, only under this design can one attempt to fully assess aggregate and indirect effects;
- No hybrid regime: the counterfactual could be drawn from the pure legacy benefit regime for all types of claim; how long impacts could be evaluated would depend on when the matched areas would start the roll-out;
- No mechanical compositional effects affecting the composition of claimants in later roll-out phases (question 1);
- No anticipation effects affecting the observed outcomes of the treated group (questions 2 and 3);
- Entry effects would remain, but might be negligible due to the lack of familiarity (in the early stages), as well as to the policy being fully implemented;
- Anticipation effects affecting the composition as well as the observed outcomes of the comparison group would depend on when the matched areas would start the roll-out.

• Design (B)

This design would keep an (ideally randomly selected) set of areas outside the roll-out of UC over a longer period, extending past phases 1 to 4. So, over phases 1 to 4 some areas would be kept operating under the legacy benefit regime for longer – ideally one or two years.

Advantages and disadvantages of this design are:

- No hybrid regime: the counterfactual could be drawn from the pure legacy benefit regime for all types of claim, and the evaluator would be capable of assessing the same treatment effects over a longer period;
- No anticipation effects for the comparison group affecting either the composition of the comparison group or the observed outcome;
- Mechanical effects, entry effects and (relevant for questions 2 and 3) anticipation effects affecting the observed outcomes of the treated group would remain.

A related, attractive option would be to randomly select individuals (e.g. via National Insurance number) and keep them under the full legacy benefit regime for a long period whilst UC is rolled out amongst everyone else.

Evaluating the labour market impacts of Universal Credit: a feasibility study

If only one area is (non-randomly) selected to be kept outside of UC for a longer time, it might still enhance the scope for evaluating the policy depending on how comparable the area is to the rest of the country where UC is introduced. If the area is large enough to cover smaller geographical units of sufficient diversity, then it might be possible to make comparisons between smaller geographical areas within the untreated area and similar treated areas in other parts of the country.

• Design (C)

In this design different areas would roll out UC to different types of claimants first (instead of all areas following the same sequence), while some areas would wait to start the roll-out. This design would amount to running the Pathfinder for different claimant types, where some areas would start the roll-out by moving, e.g. JSA claimants into UC first, while other areas would move TC claimants first.

Advantages and disadvantages of this design are:

- No hybrid regime: the counterfactual could be drawn from the pure legacy benefit regime for all types of claims; how long impacts could be evaluated would depend on when the matched areas would start the roll-out;
- No mechanical compositional effects;
- Entry effects would remain, but might be negligible due to the lack of familiarity;
- Anticipation effects affecting the observed outcomes of the treated group (questions 2 and 3) would remain;
- Anticipation effects for the comparison group would depend on when the matched areas start the roll-out.

5.5.1 Additional possibilities

One possible alternative scenario mentioned to us by DWP is to extend phases 1 and 2, which would entail extending coverage at a slower rate and postponing the start of phase 3.

This design change would improve the scenario we have examined in two respects for phases 1 and 2:

- No hybrid regime: the counterfactual could be drawn from the pure legacy benefit regime, and the evaluator would be capable of assessing the same treatment effects over a longer period; and
- Less serious threat from anticipation effects affecting the outcomes of the comparison group.

These gains for phases 1 and 2 would come at the expense of:

- Lower and more diluted volume would hinder a precise and representative estimation of aggregate and indirect effects for phases 1 and 2;
- Increased mechanical compositional effects for subsequent phases, as more of those who would have made a new claim to what would have been, e.g. TC would have already entered UC via a longer phase 2;
- Lower volume of UC claimants in subsequent phases (for the same reason), which can make aggregate and indirect effects harder to detect.

Evaluating the labour market impacts of Universal Credit: a feasibility study

The first two negative implications are not of great concern though. As we summarised in section 5.1, evaluation questions concerning the aggregate and the indirect effects of UC in phases 1 and 2 are already of very low reliability and low relevance. Similarly, mechanical roll-out effects in phase 3 are irrelevant, as we do not deem it feasible to address question 1 in this phase. We are also not particularly concerned about worsening any mechanical effects affecting the evaluation of the impacts of UC on claimants in phase 4, as we deem this evaluation question to be of relatively low reliability and already of very low relevance. However, extending phase 2 may compromise the feasibility and reliability of addressing question 2 for phase 3 for two reasons, which we identified as a highly policy-relevant parameter in the evaluation of UC. On the one hand, it reduces the scope to exploit time variation to reliably estimate the impact of UC versus the pure legacy benefit regime as the comparison group would have to be drawn further back in time. On the other hand, the reduction in the additional volume of UC claimants brought in during phase 3 (as more new TC/ISlp-equivalent claimants will be under UC already by the time this phase is locally rolled out) could, depending on its severity, change the nature of the impact on benefit take-up and make the impact on other outcomes more difficult to detect. In conclusion, given that phase 2 provides one of the two most promising opportunities to evaluate the impact of UC on new claims of a particular type, the option of extending phase 2 might indeed be worth considering, though weighted against its potential knock-on effects on UC volume in phase 3.

Finally, it has to be noted that the ‘once in UC, always in UC’ policy is key to the evaluation, though it comes at a price. This design feature implies that once an individual has entered the UC regime (by whatever route), the full new regime applies for him or her from then onwards – irrespective of whether UC has already been rolled out to specific new types of claims yet. This design feature is essentially binding for phases 1 and 2, as individuals starting what would have been a JSA claim now face the full range of UC incentives, including those conditional on employment, even before the roll-out of phase 3 to WTC claims. In the absence of such a feature, entry and anticipation effects would arguably be much smaller, if at all present, while mechanical entry effects affecting subsequent phases would by construction be absent. On the other hand, anticipation effects would now affect the outcomes of UC claimants as well as comparison individuals. Despite these few advantages, the removal of such a feature would, however, completely undermine the meaningfulness of any evaluation exercise exploiting variation in those early phases. The ‘once in UC, always in UC’ policy is absolutely key for a clean and meaningful definition of treatment as ‘being subject to UC’, with all its concomitant incentives and entitlement rules – irrespective of the current circumstances and choices of the individual.

As to the stock, a design feature that would be most helpful for the evaluation of the impact of notification of migration to UC during the phases 5 and 6 would be to ensure that the timing of migration within a given area is essentially random for a given claimant type. One potentially feasible option might be to use targeted migration for the bulk of volumes but to do random migration (e.g. based on NINO’s) in some areas. This may be feasible given the large volumes involved in the managed migration. Within an area some offices could migrate existing claims whilst other offices migrate them later, therefore, ensuring that some claimants are migrated and others are not yet migrated in a random way within the same labour market. Such design changes, though in principle feasible, might however, not be worthwhile given the low policy relevance and interest of this evaluation question.

5.5.2 Recommendations

The roll-out scenario provided to us by DWP is not conducive to a robust, wide-ranging evaluation of all the labour market impacts of UC. The scenario partly reflects operational and practical deliverability concerns. This section has suggested a few changes in design

Evaluating the labour market impacts of Universal Credit: a feasibility study
aimed at increasing the feasibility and scope of the evaluation exercises it allows.

Obviously, while combining designs (A) and (B) would prove most beneficial in terms of evaluation opportunities, it could increase operational and deliverability demands.

We conclude the section by suggesting the following two other roll-out modifications, which would considerably improve on the evaluation options offered by the current scenario, while imposing lower operational pressures than simply combining designs (A) and (B):

- Keep the current scenario, except that one or more areas would be left under the legacy benefit regime for as long a period as possible, ideally well after the introduction of phase 4 (design (B)). At that point, all new claims to all legacy benefits would be moved to UC in those areas at the same time (design (A) applied to the 'non-treated' areas). In addition to the advantages offered by design (B), this plan would allow for a full evaluation of UC in the omitted area(s) via design (A). All evaluation exercises in these areas would, however, exclusively rely on time variation, and no robustness checks would be possible.
- Extend the above to allow for two groups of areas being left under the legacy benefit regime for as long as possible, again ideally well after the introduction of phase 4 (design (B)). At that point, only one of the two groups of areas would roll-out UC to all new claims to all legacy benefits at the same time (design (A)), while the second group of areas would remain under the legacy benefit regime for longer. The presence of this second group of areas kept out of UC for longer would allow for a full evaluation of UC in the first group of areas exploiting both geographical and time variation and more than one method.

Extending phases 1 and 2, though at a cost, might be an interesting, though minor, option.

6 What evaluations will available data allow in practice?

6.1 Data requirements and sources: new claimants

The empirical strategies proposed to assess the impact of Universal Credit (UC) during phases 1 to 4, when new benefit claims are due to be gradually moved to UC, require data on similar individuals exposed to two different benefits systems, one treating some new types of benefit claims under UC, the other treating similar claims under the legacy benefit regime. Question 1 requires data on all or a sample of new claimants. Question 2 requires (all or sample) data on the entire working-age population or particular sub-groups of interest, perhaps those that may be more likely to become a claimant once the policy changes – e.g. those who have had some past benefit claim, lone parents or all parents, the least educated or those on low earnings, etc. Question 3 requires data on those who are not directly affected by the policy reform – e.g. lone parents with children aged under five during implementation phases 1 and 2. For the sample in question, the information required is:

- Data on outcomes of interest (individual and family labour market outcomes, claim status or measures of wellbeing) measured some time after the start of the new claim;
- Historical data on local labour market outcomes, together with the local history of roll-out and some measure(s) of UC coverage, to ensure comparability of areas or time periods being compared;
- Individual and/or family background characteristics and labour market histories to support construction of the counterfactual (ensure comparability between groups).

6.1.1 Data sources and quality

Among survey datasets, the Labour Force Survey (LFS) is the most promising. It is a large representative survey containing detailed information on labour market outcomes and background characteristics, and has the crucial advantage of following people over time, interviewing them in five successive quarters. But the sample size – while large by survey standards – is still too small to ensure sufficient representation of some sub-populations of interest, particularly when addressing question 1 and, to some extent, question 3.⁸ The LFS is also of little use for looking at impacts on existing legacy benefit claimants of being moved onto UC (versus not yet being moved at that time) since the dates of notification and migration will not be recorded precisely in the data. However, it can be of more use when addressing question 2 in phases 3 and 4, given the focus on the entire population and the

⁸ For instance, given the LFS sample size (around 20,000 people given a first interview each quarter and then followed for five quarters, so that in total 100,000 people are interviewed in each quarter), it will contain only a few dozen new Jobseeker's Allowance (JSA) claims from each month across the entire UK. Once that is subdivided between different types of claimant and between treated and non-treated areas, any statistical analysis looks impractical.

growing number of UC claimants (and we do, indeed, suggest exploring its use for this exercise). Yet, there may significant measurement error in the recording of benefit claiming, and the discretisation of time in quarters can be too coarse to address the evaluation questions when the UC roll-out moves at a relatively fast pace.

Data from cross-sectional surveys suffer from the additional problem of recording only a snapshot of information. These data support only some aggregate analyses as individual observations cannot be adequately matched on pre-treatment information, unaffected by the treatment itself. We, therefore, focus our attention on administrative data – including information on benefit history, employment spells and earnings – and discuss its strengths and weaknesses in relation to each of the outcomes of interest.

For all these reasons, the main data sources for evaluation of UC are likely to be administrative data.

The Work and Pensions Longitudinal Study (WPLS) covers the entire population of those who have made a benefit claim since the late 1990s, including all new claims to UC, making it an ideal candidate dataset. Along with the closely related UNITE database,⁹ it merges Department for Work and Pensions (DWP) data on individuals' benefit claim histories from the National Benefits Database (NBD) with Her Majesty's Revenue and Customs (HMRC) data on Tax Credit (TC) claiming histories, the start and end dates of employment spells (from P45/P46 forms) and individuals' total annual earnings (from P14 forms), and local authority data on Housing Benefit (HB) and Council Tax Benefit (CTB) claim histories (from the Single Housing Benefit Extract, SHBE) as well. The merging of datasets results in extremely rich data on benefit histories, going back more than ten years for everyone across the country who has claimed any of these benefits at any point over that period (though some important pieces of information are absent from the dataset, notably family income as used to implement the means test). However, while the merged-in employment and earnings data cover a similarly long time period, their quality is less good. The data do not contain any information on self-employment, nor does it record all employment spells when earnings fall below the National Insurance Lower Earnings Limit (LEL) – potentially important since the treatment of self-employment and the treatment of 'mini-jobs' are both significant features of the UC reform and ignoring them would be a notable limitation – and even for spells of employment above the LEL many start and end dates are missing in the data. Earnings are only available annually and are not recorded separately for different jobs during the year (important when we want to analyse outcomes within a short window, or look only at amounts earned after a person makes a benefit claim part-way through a year). There are no data on hours worked (except for tax credit claimants).

In contrast, the new Real Time Information (RTI) system introduced from April 2013 contains information on earnings and working hours collected in real time for all employment spells separately, as well as the respective start and end dates. The data are also expected to be more accurate than the WPLS. Its main drawback for evaluation purposes is the continuing lack of information on self-employment.

⁹ The data described here are in fact held across various versions of the WPLS and UNITE databases. In what follows we refer to their collective contents as WPLS/UNITE, though they are not currently a single dataset containing all of the data described.

Evaluating the labour market impacts of Universal Credit: a feasibility study

It has not yet been decided what sample of RTI data will be merged with the existing WPLS/UNITE data and made available to the evaluator: in particular, it is not yet clear whether RTI data will be made available only for UC claimants or for legacy benefit claimants as well. But RTI data is unlikely to be useful for most of the evaluation exercises discussed in this report unless it is made available for the entire WPLS/UNITE sample – both UC and legacy benefit claimants – or at least a representative sub-sample of it.

The absence of RTI data for the period preceding April 2013 is much less of a problem than the absence of RTI data for those not under UC would be, provided WPLS/UNITE data can be merged in and used to match individuals. Indeed, one of the main strengths of the matching methodology if RTI data are available for all new claimants (UC and legacy benefit) from April 2013 onwards is precisely that it is robust to changes in data quality from before to after treatment as long treated and comparison groups are similarly affected. When applying difference-in-differences (DiD), however, the use of RTI data to measure the impact of UC is only recommended from the moment they exist for both the treated and comparison groups before and after the treatment period (therefore, only for treatment groups sampled at least one year after RTI data is first collected, i.e. April 2014 or later, if exploiting geographical variation and using a one-year interval for the before-after comparison). How problematic this would be in practice for evaluating a particular phase of roll-out depends on when that phase of UC is rolled out: the later the roll-out, the less problematic it would be.

A crucial requirement for any evaluation study is consistency of data quality across the groups being compared. In particular, the effect of changing the quality of the outcome data only for the treated group after treatment cannot, in general, be disentangled from the impact of the treatment on the treated. For DiD methods, data need to be of consistent quality over time and (if using geographical variation) across areas. That is, changes in data collection mechanisms or data sources that differentially affect groups or time periods may bias the evaluation results. Matching methods are less demanding in this regard, this being possibly their main comparative advantage when applied to the evaluation of UC during the first four implementation phases if those are implemented soon after April 2013. Matching requires data quality to be constant across treated and comparison groups after treatment (on the outcome of interest) and before treatment (on matching information). But it deals effectively with changes in data quality over time for as long as treated and comparison groups are equally exposed to such changes.

The lack of self-employment information in both WPLS/UNITE and RTI reduces the relevance of the evaluation questions that can be addressed using these datasets as only a fraction of employment spells is observed. Moreover, it can also compromise the validity of the estimated impact of UC on earnings and related measures (family income, poverty and so on) if UC changes the distribution of earnings among the self-employed or the probability of becoming self-employed. One possibility would be to use income tax self-assessment data linked to the WPLS/UNITE to evaluate the impact of UC on self-employment. If no impact of UC on self-employment were found, we could have greater confidence in the relevance and reliability of impact estimates based only on employment spells.

We now discuss in more detail what the available data will permit and preclude for each of the outcomes of interest in turn. Table 3 at the end of the current section (6.1) provides a summary of the data availability for the evaluation questions related to the first four phases of the implementation of UC.

6.1.2 Employment

Pre-RTI WPLS/UNITE data cannot support a study of the impact of UC on all employment since not all employment is recorded – information is missing for the self-employed and for some employment spells where earnings are below the LEL. An alternative is to redefine the outcome to be measured as recorded employment and apply the evaluation strategies as described.

There are two major limitations to this analysis. The first is conceptual: as discussed above, the impact of UC on low-paid employment and on self-employment might differ from the impact of UC on other types of employment, and may be of particular policy relevance. However, such impacts cannot be analysed given the data coverage.

The second limitation is potentially serious. Changes to the LEL over time (relative to earnings growth in that part of the distribution) would affect the proportion of total employment that is recorded. Such a change in the LEL would, therefore, change recorded employment even if UC had no effect, and changes in reported employment over time might be attributed to UC when in fact they were caused by a reform to the LEL. Estimates of the impact of UC based on comparing inflow cohorts at different points in time would be biased if a change in LEL occurred between the points in time when treated and comparison groups make their claims. Moreover, different areas may be differently affected by changes in the LEL, depending on the local distribution of earnings. But this means that the impact of changes in the LEL on recorded employment may differ by area, contaminating estimates of the impact of UC based on geographical variation.

If RTI data is made available for both UC and legacy benefits claimants, it would allow for the estimation of the impact of UC on all employment (but still excluding self-employment). DiD methods require at least one full year of comparable data (to minimise the scope of confounding seasonal variation), implying that RTI data could be used for evaluation studies looking at those treated from April 2014 onwards. Matching methods can deal effectively with changes in data quality from before to after treatment for as long as common to the treated and comparison groups. Therefore, RTI data could be used for evaluation studies from the moment UC is first introduced in phase 1.

It is worth emphasising that RTI data on all new claimants would only be of interest to address question 1, on the impact of UC on new claimants. Questions 2 and 3 look at wider populations. To address those one would need RTI data on the entire population of interest from the start of UC implementation onwards, or perhaps a relevant subset such as those who have claimed a benefit in the past (the WPLS/UNITE sample).

Without RTI data for all in the treated and comparison groups, any of the evaluation designs described in the previous sections would be feasible, but only for the less interesting outcome of recorded employment and having to ignore the second limitation discussed above (on the potential differential impacts of changes in LEL). For DiD methods, even that would be problematic if the availability or quality of 'old-style' P45/P46-based employment data in the WPLS changed after the introduction of RTI.

Self-assessment data linked to WPLS/UNITE (and, possibly, RTI) would further allow for the identification of the impact of UC on self-employment on an annual basis, possibly providing empirical support to studies focusing on employment only if no significant impacts on self-employment can be detected.

6.1.3 Earnings

Only total annual earnings are recorded in the current WPLS datasets, and only for recorded employment. Therefore, one can only hope to measure the impact of UC on recorded earnings, a measure that suffers from the drawbacks discussed above with regard to recorded employment, aggravated by the fact that, where no earnings are recorded, zero earnings must be (perhaps wrongly) assumed. If UC affects the probability of working below the LEL and/or the distribution of earnings below that limit, setting to zero what is really a varying proportion of positive earnings drawn from two different distributions (pre- and post-UC) will bias the estimated impact in undetermined ways.

In addition, the annual frequency of recorded earnings will impose important restrictions on the measurement of the impact of UC on recorded earnings:

- Impacts within the fiscal year of the treatment can only be measured for specific subpopulations as described below.
- Impacts in future fiscal years can be measured for everyone, but will typically have a different interpretation since UC will eventually (and generally soon after the inflow of the treatment group) replace new claims of the same type by the control group. So an evaluation exercise will not measure the impact of the full UC regime versus the full legacy benefit regime. Instead, for instance under question 1, it will be the impact of ‘starting a new claim of a certain type under the full UC regime versus starting it under a regime where that type of new claims still falls within the legacy benefit regime but will soon fall under the UC regime’.
- Restricting attention to individuals observed during certain months around the turn of the fiscal year (March and April) would remove the problem of separating pre-treatment and post-treatment earnings in the data (without needing to restrict attention to people recently out of work as described below); but looking at people at those times is not always what the evaluation methodology calls for. The interpretation of the parameter estimated will again depend on when the comparison group becomes exposed to UC.
- More generally, the definition of the treatment effect being measured depends on the length of the time period over which earnings are observed after the time of treatment. The longer the period, the longer the comparison group may be exposed to a market where new claims will gain access to UC as the roll-out proceeds. Hence, the estimated effect can become gradually more influenced by policy changes affecting the comparison group.

Impacts on earnings for shorter periods within a fiscal year can be estimated for some specific sub-groups. For those out of work during the part of the fiscal year prior to exposure to UC, annual earnings will be a measure of how much they earn in the part of the fiscal year after exposure. And if earnings were available separately for each employment (which they currently appear not to be, though they must presumably be to HMRC, since each employer files a separate P14 form for each employee), then for those starting a new claim during a non-working period, future earnings could be separated from other earnings in the same fiscal year.

The availability of RTI information for new claims under both UC and legacy benefit regimes would relax the need to focus on particular subpopulations when addressing question 1. Having RTI information for past claimants – for example, for all individuals in the WPLS/ UNITE sample – would similarly expand the scope of what can be done to answer questions 2 and 3. The continuous-time earnings information exactly apportions earnings over

time within a fiscal year, therefore, allowing for earnings after treatment to be separately observed. Moreover, earnings are observed for all employment spells, including the lowest earners, therefore, improving the reliability of the analysis. But the lack of information for the self-employed still compromises reliability and relevance. Using Self-Assessment data could partly overcome this drawback (or provide supporting evidence to restrict attention to employment spells only) but only for assessing the impact of UC on an annual basis as must be the case when relying on WPLS/UNITE.

6.1.4 Hours

Working hours are not observed in the WPLS/UNITE data (except for TC claims, which are of limited value on their own as these cannot be compared to the hours worked by those not on TC). Impacts on working hours can, therefore, be estimated only if RTI data becomes available for treated and comparison groups.

6.1.5 Family level outcomes: workless families, family earnings/income and poverty

The impact of UC on family outcomes can be measured only if other adults in the family are observed and their employment and earnings recorded.¹⁰ The crucial information is the National Insurance number (NINO) of partners, and whether it can be used to match information on partners' employment and earnings over time regardless of the claiming status of the family. Occasional observations of family-level outcomes (say, when a new claim starts) are not sufficient for evaluation purposes since these are for specific populations selected on the basis of the outcomes of treatment.

Under the assumption that the partners' information is compiled continuously over time, existing WPLS/UNITE data can support the study of the effects of UC on family employment and family earnings under similar restrictions to those discussed above at the individual level but now applied at a family level:

- Employment – only recorded employment can be studied;
- Earnings – only for recorded employment; short-term effects (in the same fiscal year as treatment) only for the sub-group of families out-of-work during the months preceding the new claim in the same fiscal year (if only total annual earnings are observed).

Again, these restrictions could be lifted if RTI data were available for all adults in a claimant family independently of whether the claim was to UC or a legacy benefit. Self-assessment data could complement this analysis by bringing evidence on the impact of self-employment.

The measurement of poverty is more demanding as it requires observing family disposable income, including both earned and unearned income, as well as family demographics. The lack of information on unearned income in the data is one limitation: savings income is likely to be less important for these groups than for better-off households, but it is still not negligible and there are other sources of unearned income, such as maintenance, that may also be important for some of these groups. Furthermore, a poverty line would have to be calculated specially for the purpose since standard poverty lines (i.e. those used in DWP's Households Below Average Income publication) are based on household rather than family income.

¹⁰ Information on household members outside the benefit unit making the claim is never available in administrative data, so we cannot look at any outcomes at the household level; the most we can hope to look at is outcomes at the benefit unit ('family') level.

Table 6.1 Summary of evaluation questions and data availability for the evaluation of UC implementation scenario 1 to 4

Overall question	Phase-specific question	Data
Q1 Impact of the full UC regime versus the full legacy benefit regime on the outcomes of individuals starting a new claim of a certain type Reference population: new claimants.	Phase 1: Impact of starting a new UC claim under the full UC regime: <ul style="list-style-type: none"> • versus starting a similar JSA claim under the full legacy benefit regime; • for single unemployed claimants without children in the Pathfinder areas in the 1st 6 months of implementation (phase 1). 	<ul style="list-style-type: none"> • Survey (LFS): sample of new claimants too small. • WPLS/UNITE: problem in terms of recorded outcomes, but available for the whole population. • RTI: would overcome outcome issues of WPLS/UNITE if it were available for the new UC and legacy benefits claimants (except in what concerns to self-employment) • Self-assessment data for the self-employed could help establish the relevance and reliability of estimates that restricting attention to employment spells.
	Phase 2: Impact of starting a new UC claim under the full UC regime: <ul style="list-style-type: none"> • versus starting a similar JSA claim under the full legacy benefit regime; • for those unemployed satisfying the JSA conditions who have decided to claim such a benefit in the 2nd 6 months of implementation (phase 2). 	As above.
	Phase 3: Not possible – no reliable comparison group available.	n/a
	Phase 4: Impact of starting a new UC claim under the full UC regime: <ul style="list-style-type: none"> • versus starting a similar ISc/ESA/ HB claim under the full legacy benefit regime; or • versus starting a similar ISc/ESA/ HB claim under a hybrid regime in which UC has replaced only new JSA claims or only new JSA/TC/ IS(PLs) claims; • for carers, sick and low income families who are entitled to ISc/ ESA/HB under the legacy benefit, are starting a new claim on or after October of implementation Year 2 and have not started a UC claim equivalent to JSA or TC/ISlp since UC replaced such claims in their 	As above.

Continued

Table 6.1 Continued

Overall question	Phase-specific question	Data
<p>Q2</p> <p>Market-wide impact (or impact on the outcomes of specific groups) of introducing UC for some new benefit claims versus maintaining the full legacy benefit regime (or some hybrid regime with more limited UC coverage) for that benefit claims</p> <p>Reference population: all those who may gain entitlement depending on their actions.</p>	<p>Phase 1: Impact of UC replacing new JSA claims by single people without children:</p> <ul style="list-style-type: none"> • versus the full legacy benefit regime; • for the entire working-age population or some sub-group. <p>Phase 2: Impact of UC replacing new JSA claims:</p> <ul style="list-style-type: none"> • versus the full legacy benefit regime; • for the entire working-age population or some sub-group. 	<ul style="list-style-type: none"> • Survey (LFS): sample size too small given size of PF areas. • WPLS/UNITE: problem in terms of recorded outcomes; additionally, not available for the whole population, but could restrict attention to the (interesting) subpopulation of those who have ever claimed some benefit prior to some cut-off before the roll-out. • RTI: if available for WPLS/UNITE sample, would help overcome the issue with WPLS/UNITE on the subpopulation who has ever claimed some benefit prior to some cut-off (although misses effects on self-employment). <ul style="list-style-type: none"> - Survey (LFS): although sample size quite large, it is unlikely to be large enough to detect effects that are 'diluted' across many unaffected people, particularly as significant indirect effects of UC at this stage are unlikely. - WPLS/UNITE: problem in terms of recorded outcomes; additionally, not available for the whole population, but could restrict attention to the (interesting) subpopulation of those who have ever claimed some benefit prior to some cut-off before the roll-out. - RTI: if available for WPLS/UNITE sample, would help overcome the issue with WPLS/UNITE on the subpopulation who has ever claimed some benefit prior to some cut-off (although misses effects on self-employment).

Continued

Evaluating the labour market impacts of Universal Credit: a feasibility study

Table 6.1 Continued

Overall question	Phase-specific question	Data
Q2 Continued	<p>Phase 3: Impact of UC replacing new JSA/TC/ISlp claims:</p> <ul style="list-style-type: none"> • versus the full legacy benefit regime; or • versus a hybrid regime where UC has replaced only new JSA claims; • for the entire working-age population or some sub-group. 	<ul style="list-style-type: none"> • Survey (LFS): sample size quite large, maybe large enough to detect effects although these may still be quite 'diluted' across many unaffected people. • WPLS/UNITE: problem in terms of recorded outcomes; additionally, not available for the whole population, – but could restrict attention to the (interesting) subpopulation of those who have ever claimed some benefit prior to some cut-off before the roll-out. • RTI: if available for WPLS/UNITE sample, would help overcome the issue with WPLS/UNITE on the subpopulation who has ever claimed some benefit prior to some cut-off (although misses effects on self-employment). • Self-assessment data for the self-employed could help establish the relevance and reliability of estimates that restricting attention to employment spells.
	<p>Phase 4: Impact of UC replacing all new legacy benefit claims:</p> <ul style="list-style-type: none"> • versus the full legacy benefit regime; or • versus a hybrid regime where UC has replaced only new JSA claims or only new JSA/TC/ISlp claims; • for the entire working-age population or some sub-group. 	As above.
<p>Q3</p> <p>Substitution, displacement and other indirect impacts of introducing UC for some new benefit claims versus maintaining the full legacy benefit regime (or some hybrid regime with more limited UC coverage) on the outcomes of those not directly affected by the change</p> <p>Reference population: those not directly affected by the reform.</p>	<p>Phase 1: Indirect effects of UC replacing new JSA claims by single people without children:</p> <ul style="list-style-type: none"> • versus the full legacy benefit regime; • for parents and for individuals in couples (the entire working-age population of parents or individuals in couples, or the sub-group of unemployed, or of those starting a JSA claim). 	<ul style="list-style-type: none"> • Survey (LFS): sample sizes might be an issue given size of PF areas and restricted reference population. • WPLS/UNITE: problem in terms of recorded outcomes; additionally, not available for the whole population, but could restrict attention to the (interesting) subpopulation of those who have ever claimed some benefit prior to some cut-off before the roll-out. • RTI: if available for past LB claimants, would help overcome the issue with WPLS/UNITE on the subpopulation who has ever claimed some benefit prior to some cut-off (although misses effects on self-employment).

Continued

Table 6.1 Continued

Overall question	Phase-specific question	Data
Q3 Continued	Phase 2: Indirect effects of UC replacing new JSA claims: <ul style="list-style-type: none"> • versus the full legacy benefit regime; • for those who would not get JSA if not working (principally lone parents with children under 5, people with disabilities and carers). 	<ul style="list-style-type: none"> • Survey (LFS): sample sizes might be enough for the reference population. • WPLS/UNITE: problem in terms of recorded outcomes; additionally, not available for the whole population, but could restrict attention to the (interesting) subpopulation of those who have ever claimed some benefit prior to some cut-off before the roll-out. • RTI: if available for past LB claimants, would help overcome the issue with WPLS/UNITE on the subpopulation who has ever claimed some benefit prior to some cut-off (although misses effects on self-employment).
	Phase 3: Indirect effects of UC replacing new TC/ISlp claims: <ul style="list-style-type: none"> • versus a hybrid regime where UC has replaced only new JSA claims; • for those who started a UC claim before the start of phase 3. 	<ul style="list-style-type: none"> • Survey (LFS): sample sizes likely to be an issue for some methods, but might be enough for others. • WPLS/UNITE: problem in terms of recorded outcomes, but would be available for the whole population. • RTI: available for the whole reference population in the scenario examined, so easily the best option (although misses effects on self-employment). • Self-assessment data for the self-employed could help establish the relevance and reliability of estimates that restricting attention to employment spells.
	Phase 4: Indirect effects of UC replacing all new claims: <ul style="list-style-type: none"> • versus a hybrid regime where UC has replaced only new JSA/TC/ISlp claims; • for those who started a UC claim close to the end of phase 3; or <ul style="list-style-type: none"> • versus a hybrid regime where UC has replaced only new JSA claims; • for those who started a UC claim close to the end of phase 2. 	<ul style="list-style-type: none"> • Survey (LFS): sample sizes likely to be an issue. • WPLS/UNITE: problem in terms of recorded outcomes, but would be available for the whole population. • RTI: available for the whole reference population in the scenario examined, so easily the best option(although misses effects on self-employment). • Self-assessment data for the self-employed could help establish the relevance and reliability of estimates that restricting attention to employment spells.

6.2 Data requirements and sources: existing claimants

The empirical strategies proposed for assessing the impact of UC during phases 5 and 6 of the hypothetical roll-out scenario require data on similar individuals being differently exposed to the migration of stocks. Data requirements vary by evaluation question:

- Question 1 requires information on similar ongoing claims being differently treated within area and time period, some being migrated to UC at the same time that others remain on the legacy benefit regime. For each of these claims, one needs to observe the claim start date, the dates of notification and migration, future outcomes (e.g. employment), labour market and benefit histories and some background characteristics of claimants (to ensure comparability between groups).
- Question 2 requires information on the whole population already under UC or some sub-group of interest (e.g. job seekers) in similar areas, some being exposed to earlier migration of stocks into UC than others. For each individual, one needs to observe future outcomes (e.g. employment), labour market and benefit histories and some background characteristics of claimants (to ensure comparability between groups), and historical area-level information on the outcome of interest (and possibly other labour market variables), the history of UC roll-out and some measure(s) of UC coverage (to ensure area comparability).

WPLS/UNITE is in many ways an ideal candidate dataset for the measurement of effects on existing claimants (question 1). It covers the entire population of claimants and follows them over time, recording changes in circumstances as well as labour market and benefit status. In contrast, survey data will generally lack crucial claim-related information such as precise start, end, notification and migration dates. In the best-case scenario, claim status is recorded quarterly over a five-quarter period (in the Labour Force Survey (LFS)). Such discretisation (looking at five discrete points, one per quarter, rather than exact dates or even monthly data) is likely to be too coarse to allow for the study of the impact of migration to UC for claimant types that are moved over a relatively short period of time.

WPLS/UNITE can also be a rich source of information for the measurement of indirect effects on the subpopulation of those under UC whoever made a claim to benefits (or some sub-group of those). This group includes both those who made a past claim to UC and those not currently on legacy benefits. If further restricting attention to all those who have started a UC claim in the past, the RTI data will have information for all in the treated and comparison groups, even under plans examined in the current scenario to limit coverage to new UC claimants.

Survey data can allow for more general populations to be considered or to focus attention of specific demographic groups of interest when addressing question 2, but again the outcome information will not be as detailed.

However, any evaluation study based on current WPLS/UNITE data will suffer from the drawbacks described above:

- Only recorded employment can be studied – excludes self-employment and a proportion (possibly varying across time and areas) of low-paid work.

Evaluating the labour market impacts of Universal Credit: a feasibility study

- Earnings observed on an annual basis for recorded employment only – imposes further restrictions on the population for whom the impact can be studied to those with separable pre- and post-treatment earnings. Alternatively, one could study the impact of migration on declared earnings during the following fiscal year.
- Hours are not generally observed and so cannot be studied.
- Continuous observation of other adult family members is a crucial requirement for the assessment of impacts on family-level outcomes.

Evaluations based on RTI data would be more promising, but only if it were available for legacy benefit claimants as well as UC claimants.

Table 6.2 provides a summary of the data availability for the evaluation questions related to the last two phases of the implementation of UC.

Table 6.2 Summary of evaluation questions and data availability for the evaluation of UC scenario phases 5 and 6

Question	Data
<p>(S.1) Impact of migrating a long-standing legacy benefit claimant to UC under transitional protection in a given month:</p> <ul style="list-style-type: none"> • versus not migrating it in that month (and eventually migrating it later should the claim remain open); • for long-term claimants whose circumstances have remained stable; • in a world where a new claim (including 'natural migration') triggers moving into UC. 	<ul style="list-style-type: none"> • Survey (LFS): not fine enough to capture time of migration; duration of claim not recorded. • WPLS/UNITE: problem in terms of recorded outcomes, but would be available for the whole population. • RTI: would overcome outcome issues of WPLS/UNITE if it were available for the WPLS/UNITE sample (although misses effects on self-employment). • Self-assessment data for the self-employed could help establish the relevance and reliability of estimates that restricting attention to employment spells.
<p>(S.2) Impact of notifying in a given month a long-standing legacy benefit claimant that migration under transitional protection will happen in 3 months' time if the claim remains open:</p> <ul style="list-style-type: none"> • versus not notifying them in that month (and eventually notifying later should the claim remain open); • for long-term claimants whose circumstances have remained stable; • in a world where a new claim (including 'natural migration') triggers moving into UC. 	<ul style="list-style-type: none"> • Survey (LFS): no information on time of notification; duration of claim not recorded. • WPLS/UNITE: problem in terms of recorded outcomes, but would be available for the whole population. • RTI: would overcome outcome issues of WPLS/UNITE if it were available for the WPLS/UNITE sample (although misses effects on self-employment). • Self-assessment data for the self-employed could help establish the relevance and reliability of estimates that restricting attention to employment spells.
<p>(S.3) Impact of moving a large number of individuals from a legacy benefit into UC:</p> <ul style="list-style-type: none"> • versus no such large influx into UC; • for those not directly affected by the migration but who are already subject to the full UC regime. 	<ul style="list-style-type: none"> • Survey (LFS): sample sizes likely to be enough. • WPLS/UNITE: problem in terms of recorded outcomes, additionally is not available for the whole population; could restrict attention to the subpopulation who has ever claimed some benefit prior to some cut-off, or to some subgroup currently or ever on UC. • RTI is a better choice if available for all those under UC sample (although misses effects on self-employment). • Self-assessment data for the self-employed could help establish the relevance and reliability of estimates that restricting attention to employment spells.

6.3 Possible new panel surveys being considered by DWP

DWP is considering launching two new panel surveys, the first to cover the Pathfinder and the second to start during the summer of 2014.

The first panel is intended to sample from new UC claimants in Pathfinder areas and similar JSA claimants in comparison areas, and to collect rich data on demographics, information, beliefs and attitudes. It could be a valuable source of information on people's understanding of the new rules and how it spreads, the importance of anticipation responses and, in general, the impacts of UC and possible mechanisms explaining such effects.

The second panel is currently at feasibility testing stage. It is planned to sample the population of new UC claimants and follow each respondent for four years, with in-depth interviews once a year and possible linkage to administrative data to provide a fuller profile (in continuous time and with more history). While this may be a rich and informative dataset, it is currently intended to provide qualitative information about UC claimants' experiences and progression over time, not to help with impact estimation, and the planned survey design renders it of little value for quantitative evaluation of the impact of UC as a whole. The main problem is that there is no data on claims outside the UC regime from which one could hope to draw the counterfactual.

A possibly useful alternative would be to draw a sample from the population of ongoing legacy benefit spells. Such data could inform an evaluation of the impact of migrating the stock of existing legacy benefit claimants to UC, as well as a study of the impacts of the new system on existing claims (both direct effects as incentives change for all, and indirect effects, as market conditions change the opportunities for all). To address the former, the sample would have to be large enough to ensure that enough spells of each type are still live by the time the migration of stocks starts and sufficient variation in the time of migration can be observed.

Compared to administrative data, the advantage of a panel survey is the richer information about individuals and families that could be obtained. Adding additional background characteristics obtained from a panel survey to the WPLS/UNITE data would be somewhat beneficial, but it would not overcome the main shortcomings of the current WPLS/UNITE data, in that in those data only a subset of employment spells can be observed, earnings are recorded on an annual basis and systematically only when above the LEL. A panel survey could only solve that problem if the panel survey, rather than administrative data, were used to measure outcomes (with administrative data providing background information such as benefit histories), since the survey need not restrict attention to jobs paying above the LEL, for example. But – aside from possible concerns over sample sizes – panel surveys have a big disadvantage relative to administrative data for measuring outcomes such as employment and benefit receipt, in that people are only interviewed occasionally and so we will only observe employment etc. at particular points in time (say, annually) rather than having continuous spell data as found in WPLS/UNITE.¹¹

¹¹ The panel survey could include retrospective questions about spells since the last interview, but such retrospective data are typically much less reliable.

Given that the scenario we have been considering means that we would often be looking at short-run outcomes over very particular periods, occasional snapshots of outcomes might be of little use except in the (rare) event of a fortuitous relationship between (for example) the benefit inflow date, the start date of a subsequent roll-out phase and the survey interview date for both treatment and comparison groups. A much more promising resource for evaluation purposes would be to make use of RTI data for the entire WPLS/UNITE sample.

6.4 Summary

Coming back to the evaluation questions, this is how well the data allows us to address them:

Q1: Impact of full UC regime versus full legacy benefit regime for new claimants

RTI, if made available for all new claimants to UC and legacy benefits, would be ideal. It lacks information on self-employment outcomes, but the importance of this omission for the results could be assessed (on an annual basis) if self-assessment data were available. WPLS/UNITE is the only alternative, but the quality of recorded outcomes severely limits the evaluation results. The LFS is unlikely to be of much use as the sample size would be prohibitively small when only a small fraction of the population each month starts a UC claim.

Q2: Impact of UC replacing new JSA claims, new JSA/TC/Income Support for lone parents (ISlp) claims or all claims versus full legacy benefit (or hybrid) regime, for entire working-age population or some sub-group

RTI, if made available for the WPLS/UNITE sample, would allow for estimates of the impact of UC on the subpopulation of those who have ever been a claimant (again, excluding impacts on self-employment). This is perhaps an interesting group of individuals, closer to the benefits system than others. The same population can be looked at using the WPLS/UNITE, but the quality of the recorded outcomes limits analysis here as well. LFS data could be used; the detection of any effects might be impossible during at least phases 1 and 2, when the population of UC claimants is very small and indirect effects are unlikely to be significant, but the LFS may become a more useful data resource for the evaluation of UC during phases 3 and 4.

Q3: Indirect effects of UC replacing new JSA/TC/ISlp claims or all claims, versus a hybrid regime, on those who had already claimed UC

RTI can be used for this question as both the treated and the comparison groups have started a UC spell sometime in the past (again, excluding impacts on self-employment). LFS has the advantage of containing more detailed demographic information, but it can be difficult to identify the population of interest in case the UC spell has happened prior to the observation window or in the presence of measurement error in reporting benefits. Moreover, sample sizes of the population of interest can be too small.

Indirect effects are unlikely to be of any significance during phases 1 and 2 as the dimension of the treated group is crucial for the formation of market-wide effects.

S.1 (stock) Impact of migrating a long-standing legacy benefit claimant to UC under transitional protection in a given month on long-term claimants whose circumstances have remained stable in a world where a new claim triggers moving into UC

Give the prior notification of claiming spells to be migrated, we are sceptical as to the possibility of addressing this question reliably. If migration to UC has any impact on existing claimants, it should start to be realised once notification is received, making it particularly difficult to construct an adequate counterfactual to those that remain on legacy benefits by the time of migration. None of the available datasets is expected to be capable of dealing with the selection process resulting from the notification, unless individuals do not respond to it.

Under the assumption that existing claimants do not react to notification (and question S2 provides some ancillary information on this), RTI would be the most useful data source if it is available for the WPLS/UNITE sample (but excludes impacts on self-employment). In the scenario provided by DWP, however, it is not clear whether that will be the case, or whether only WPLS/UNITE data will be available for all existing claimants. This dataset has limitations

in terms of recorded outcomes, only partly recording employment and earnings, the latter on an annual basis only, and having no information on hours. The LFS is not an alternative in this case as its quarterly frequency is too coarse to capture time of migration (which is the treatment in this question) and its information not detailed enough to characterise the ongoing claiming spells (as, for instance, elapsed duration is unknown).

S.2 (stock) Impact of notifying in a given month a long-standing legacy benefit claimant that migration under transitional protection will happen in three months' time if the claim remains open on long-term claimants whose circumstances have remained stable in a world where a new claim triggers moving into UC

RTI would be the most useful data source if available for the whole WPLS/UNITE sample (though it still precludes estimating impacts on self-employment). Under the current plans, however, it is not clear whether that will be the case, or whether only WPLS/UNITE data will be available for all existing claimants. This dataset has limitations in terms of recorded outcomes, only partly recording employment and earnings, the latter on an annual basis only, and having no information on hours. The LFS is not an alternative in this case as notification is not recorded (which is the treatment in this question) and its information not detailed enough to characterise the ongoing claiming spells (as, for instance, elapsed duration is unknown).

S3 (stock): Indirect effects of moving a large group into UC on those already subject to the full UC regime

The LFS should be large enough to support the analysis of the effects of UC during the migration of stocks as all those not currently claiming legacy benefits will be under UC. For the study of the indirect effects of migrating the stocks in legacy benefits, the LFS has the advantage of representing the entire population not on legacy benefits, independently of past claimant history. The outcomes are measured on a quarterly basis, possibly a slightly lengthy interval if the comparison areas migrate existing claimants soon after the treated areas. The RTI data can be used to capture indirect effects on those who have already claimed UC, and overcome (for that group) the problems induced by the relatively coarse time division of the LFS (although excluding impacts on self-employment).

7 Conclusions

The central challenge for any evaluation is identifying the counterfactual: finding a way to estimate what would have happened in the absence of a reform, usually by looking at what happens to a comparison group which is unaffected by it and assuming that the affected group would otherwise have experienced the same outcomes.

Ideally, we would like to estimate the impact of the full change from legacy benefit regime to Universal Credit (UC) regime on (short-run and long-run) labour market outcomes for the whole population. Unfortunately, for roll-out under the scenario we have been asked to consider the way in which the policy is being introduced across the country does not allow that. Specifically, an assessment of the effects of the whole regime change on aggregate variables such as employment or poverty rates is hindered by the facts that, in this scenario:

- UC is being phased in over several years, precluding any sensible comparison of the situation before UC starts to be introduced with the situation when UC is fully in place; and
- UC is being introduced for the whole working-age population across the whole country, so there is no group unaffected by the policy that can be used to identify how outcomes might have been different if UC had not been introduced.

In this study we have, therefore, devoted our attention to looking at how the gradual roll-out of UC to different claimant groups in different areas at different times might provide opportunities to estimate its effects for certain groups.

For the roll-out of UC to new benefit claims, we identified three evaluation questions to consider:

- Q1. The impacts of the full UC regime, versus the full legacy benefit regime, on new claimants.**
- Q2. The aggregate impacts of UC replacing some new benefit claims, versus maintaining the full legacy benefit regime (or some hybrid regime with more limited UC coverage), on the entire working-age population or of specific groups of interest defined on pre-reform characteristics (e.g. low earners, parents with young children, lone parents).**
- Q3. The indirect impacts of UC replacing some new benefit claims, versus maintaining the full legacy benefit regime (or some hybrid regime with more limited UC coverage), on those not directly affected by the reform.**

Given the hypothetical scenario provided, we have concluded that Q1 could potentially be addressed relatively cleanly in phases 1 and 2 of the roll-out (the replacement of new Jobseeker's Allowance (JSA) claims with UC): entry effects and anticipation effects are likely to be mild and their likely presence could be assessed, and several different evaluation strategies could be pursued, allowing for some robustness checking. Addressing Q1 in phase 3 (when UC replaces new Tax Credit (TC) and Income Support for lone parents (ISlp) claims) looks unfeasible since UC claimants are likely to differ substantially from TC/ISlp claimants, making it impossible to construct an adequate comparison group. And in phase 4 (when UC replaces other new claims), addressing Q4 might be possible, but does not look appealing since this would be a highly unrepresentative

Evaluating the labour market impacts of Universal Credit: a feasibility study

group, only (less reliable) time variation could be used and claimants quite far apart would have to be compared to attempt an estimate of the impact of UC versus the full legacy benefit regime.

In contrast, phases 1 and 2 are likely to be of little value in addressing Q2 and Q3, since the relatively small number of new claims involved make detecting aggregate or indirect impacts unlikely. Yet Q2 might be well addressed in phase 3, by which time the volume of new UC claims will be greater. This is an extremely interesting evaluation possibility: in the scenario examined, phase 3 is a crucial implementation phase, in which UC is made available to major target groups (low earners and non-working lone parents) and the most significant changes in entitlement are introduced, and for this phase it is particularly valuable that Q2 (unlike Q1) takes account of effects of the reform on the composition of claimants and indirect effects of the reform, as well as the direct effects on those who would have claimed under both UC and the legacy benefit regime. Again both assessing assumptions and checking robustness seem possible. Addressing Q2 in phase 3 looks particularly promising for those groups for whom JSA is not relevant (notably lone parents with children aged under five), since in those cases geographical variation (not just time variation) can be used to compare UC with the full legacy benefit regime. Addressing Q2 in phase 4 looks less promising since only time variation will be available and the most reliable comparisons would be less interesting ones, between UC and a hybrid regime in which UC has replaced new claims to some legacy benefits but not others.

Indirect effects of replacing new legacy benefit claims with UC (Q3) are best addressed in phases 3 and 4. In those phases it should be possible to look at the impact of introducing UC for more claims on those who are already within the UC regime and who, therefore, should not be directly affected by an expansion of its coverage.

When it comes to the migration of existing legacy benefit claimants to UC, one could consider estimating the impact of notifying people of their forthcoming migration to UC on the outcomes of those existing claimants. This might be feasible if there is some random (or pseudo-random) variation in which claimants are migrated. However, it is not clear that such an estimate would be very informative as to the underlying impact of UC. Those moved onto UC would already have been subject to UC if their circumstances changed to trigger a new claim or 'natural migration', while the comparison group (those not yet migrated in that area) might know that they are to be migrated soon afterwards if their claim remains open. More fundamentally, the migration of existing legacy benefit claimants is inherently one-off: never again will there be a large stock of legacy benefit claimants to be migrated. And the migration will be under transitional protection, which will be important in the short-run but irrelevant in the longer run.

A second question about the migration of existing claims might potentially be more interesting: the indirect impact of moving a large group of people from legacy benefits to UC on those who are already subject to the UC regime. The same broad caveats apply as to the exact nature of the impact being identified here, but they are perhaps less troubling if the goal is simply to assess whether market-wide impacts are significant. And if there any indirect impacts to be observed, the rapid transfer of a large number of people from legacy benefits to UC is surely the likeliest time to find them. The feasibility of this exercise, though, depends on whether in practice some areas do migrate a large group of claimants before other, similar, areas do so.

Taking all the different phases of the roll-out together, we believe that, in the scenario set out by DWP, the most promising evaluation questions are:

- 1 **The impact of making a claim under the full UC regime versus the full legacy benefit regime on people starting what would previously have been a JSA claim.**
- 2 **The impact of UC replacing new JSA, TC and ISlp claims versus the full legacy benefit regime (or versus a hybrid regime in which UC has replaced only new JSA claims), on the entire working-age population or some policy-relevant sub-group thereof.**
- 3 **The indirect effects of UC replacing new JSA, TC and ISlp claims, versus a hybrid regime, on those who had already claimed UC.**
- 4 More tentatively, the **indirect (substitution and displacement) effects of moving a large group of existing legacy benefits onto UC, versus a hybrid regime in which a new claim would already be to UC ('natural migration'), on those already subject to the full UC regime.** The feasibility of addressing this evaluation question is more uncertain, as it is not yet known whether the required source of variation will be available (i.e. whether in some areas a large group of legacy benefit claimants will be moved to UC while comparable areas will not have started the migration of that type of claim yet).

Of course, we can only speculate on how important the various threats to the validity of the estimates will be: there is no *a priori* guarantee that, were these evaluation exercises carried out, the resulting estimates would be reliable. How reliable such estimates could be deemed in practice would instead hinge on the results emerging from the various ancillary and corroborating analyses that have been suggested to provide indicative evidence on the plausibility of what are inherently untestable assumptions.

There are several features that all these possible evaluations have in common.

Where impacts of UC can be estimated, it should generally be straightforward to disaggregate the estimated impacts by sub-group according to observed variables. In contrast, there is little scope for disaggregating the impacts by cause or aspect of the policy (incentives, conditionality, simplification, etc) since the policy is planned to be implemented as a full package everywhere. Introducing variation in UC in different areas, for example, might allow for evaluation of the impact of that variation.

A limitation common across all of the evaluation exercise is that the impacts that can be evaluated are generally short term, in two main senses:

- First, they are impacts on short-run outcomes. It is typically only a few months in this scenario until the comparison group is exposed to the policy and the impact of UC versus the full legacy benefit regime can be estimated. In some cases slightly longer-run outcomes can be estimated; but these are usually less reliable and/or become estimates of the impact of UC versus a hybrid regime in which UC has replaced some parts of the legacy benefit regime but not all of it.
- Second, the estimates will typically represent the impact of UC as claimants experience it soon after implementation. Transitional protection is a feature built into the policy in the short-run, and it is hard to see how one could estimate the impact of UC once transitional protection has ceased to be relevant. More broadly, the Government might change aspects of UC over time – ironing out teething problems, for example – while people might become more familiar with the new system over time. Some of the evaluation exercises we

Evaluating the labour market impacts of Universal Credit: a feasibility study

examine will be closer to the long-run implementation of UC than others. But it is hard to find reliable comparison groups against which to assess later versions of UC. Additionally, only a small fraction of the population will be under UC in the short run, and the gradual national roll-out over claimant types means that it will take some years before everyone falls under the new system. Since indirect effects depend crucially on the size of the affected population, it is not expected that the indirect effects identified in the shorter term and for specific groups will be informative about the potential indirect effects of UC when fully implemented.

Another important limitation that applies across the whole evaluation is the difficulty of distinguishing the effects of UC from the effects of other changes – most obviously policy reforms – that happen around the same time. Any outcome-relevant change that happens at an inopportune moment can confound comparisons across time. Comparisons across areas are more robust – only threatened if reforms have differential effects across areas – but there are several reforms planned (or already implemented) to which that description could apply. The problem can be solved if comparison areas can be found that are affected in the same way as the treatment areas, but that might not be easy. In any case, the possibility of confounding reforms is one that any evaluation would need to consider – if only to be aware that estimates might actually be capturing the combined impact of UC and some other reform happening at the same time.

Finally, any evaluation is only as good as the data used to conduct it. In this case, the data could be a major constraint.

The most promising survey data for this evaluation are the Labour Force Survey (LFS), a large representative survey which contains detailed information on labour market outcomes and background characteristics, and has the crucial advantage of following people over time, interviewing them in five successive quarters. But while the LFS could be valuable for addressing some of our questions, it has significant limitations. The sample size – while large by survey standards – is still too small to focus on new benefit claims. And the LFS is of little use for looking at impacts on existing legacy benefit claimants of being moved onto UC (versus not yet being moved at that time) since the dates of notification and migration will not be recorded precisely in the data. For most purposes, therefore, we are reliant on administrative data.

Until recently, the principal administrative data that are relevant for this evaluation was information held across Work and Pensions Longitudinal Study (WPLS) and UNITE databases. This has universal coverage of those who have claimed a benefit or tax credit since the late 1990s (including data from before and after their claim) and includes longitudinal information on benefit claim history, employment spell start and end dates and annual earnings. However, the data are limited. There is no information on hours of work for most of the sample; the employment and earnings data exclude self-employment and employment below the National Insurance Lower Earnings Limit (LEL); there is severe measurement error in employment start and end dates; and earnings are recorded only on an annual basis across all employments. This means that the only outcomes that can be examined are:

- **recorded employment**, meaning that impacts on self-employment and low-paid employment (both of significant policy interest for UC) cannot be assessed. Estimates of UC's impact on recorded employment could also be biased if changes in recorded employment are attributed to UC when they in fact result from a change in the LEL taking people's jobs in or out of WPLS records; and

- **recorded earnings**, which could lead to biased estimates of the impact of UC on earnings if UC has an impact on earnings below the LEL (since earnings below the LEL must be set to zero before and after the reform, in the absence of any data). Furthermore, recording only annual earnings means that impacts within a fiscal year can only be assessed for specific groups for whom a given tax year's earnings can confidently be considered wholly post-treatment.

From April 2013 Her Majesty's Revenue and Customs (HMRC) began collecting Real Time Information (RTI) data. As compared to WPLS/UNITE, this dataset is much better suited to this evaluation. It has all the advantages of WPLS/UNITE data, but without some of the limitations: it includes data on all employment spells (including below the LEL), exactly apportions earnings within the year, and records hours of work. The data are also expected to be more accurate and the linking of partners' data more reliable. A major drawback is the continued absence of information on self-employment. Most importantly, however, it is not yet clear whether RTI data will be available to the evaluator for all benefit claimants, or just for UC claimants.

A key recommendation of this report is that significant care and effort should be put into obtaining the best available data for evaluation purposes. This includes RTI data (and income tax self-assessment data for the self-employed, if possible) for the entire WPLS/UNITE sample (or a large, representative sub-sample) and merged to it. Unless RTI (and self-assessment) data are collected for everyone (or for representative samples of those under UC and the legacy benefit regimes), its use is ruled out for most purposes: it is crucial for any evaluation study that data quality is consistent across the groups being compared.

Returning to the four evaluation questions we identified as most promising, the data that might be used to address them are:

1 The impact of making a claim under the full UC regime versus the full legacy benefit regime on people starting what would previously have been a JSA claim.

This question could be addressed well if RTI data were available for new legacy benefit claimants. Otherwise, the only option would be to proceed with WPLS/UNITE data, though that carries severe limitations in terms of the impacts that can be measured and potential biases in them.

2 The impact of UC replacing new JSA, TC and ISlp claims versus the full legacy benefit regime (or versus a hybrid regime in which UC has replaced only new JSA claims), on the entire working-age population or some policy-relevant sub-group thereof.

This question could be addressed well if RTI data were available for, for example, previous benefit claimants. Otherwise, an analysis could be attempted using LFS data, but detection would be more difficult with the smaller sample involved.

3 The indirect effects of UC replacing new JSA, TC and ISlp claims, versus a hybrid regime, on those who had already claimed UC. This question can be straightforwardly addressed if RTI data were available for those already under UC.

4 More tentatively, the indirect (substitution and displacement) effects of moving a large group of existing legacy benefits onto UC, versus a hybrid regime in which a new claim would already be to UC ('natural migration'), on those already subject to the full UC regime.

This could be addressed adequately using the LFS. RTI data, if available for everyone in the WPLS/UNITE sample or some sub-group of interest, offer a much bigger sample size.

Evaluating the labour market impacts of Universal Credit: a feasibility study

In the hypothetical roll-out scenario we were asked to consider, a robust ex-post quantitative evaluation of the overall labour market impacts of UC would not be possible. There would, however, be some scope for feasible evaluations of certain short-run impacts for particular sub-groups. If this roll-out scenario were to be implemented, by far the biggest contribution to the feasibility of the evaluation would be to ensure that RTI data for the whole WPLS/UNITE sample are available to the evaluator.

Appendix A

Detail and discussion of possible evaluation methods

This appendix describes the main evaluation methods and discusses their applicability to evaluating the labour market impacts of Universal Credit (UC), in particular the assumptions they rely on, their potential weaknesses and ways to gain corroborative evidence on whether their underlying assumptions are likely to be met.

The methods of regression discontinuity design and instrumental variables are presented last as they are not appropriate for the case at hand. By contrast, most of the discussion is devoted to matching and difference in differences (DiD) methods, since they are the preferred choices for most of the evaluation questions one can hope to answer on the basis of the variation arising from the various implementation phases.

A.1 Regression and matching methods

Matching and regression methods represent a natural choice to assess the migration of the stock of existing Jobseeker's Allowance (JSA) claimants from legacy benefits to UC, so we discuss them in that context.

Consider the stock of JSA claimants (perhaps of a given type), the migration of whom happens gradually over some period. We are interested in the impact of being moved to the full UC regime versus remaining on JSA. Comparing the outcomes of those migrated earlier to the contemporaneous outcomes of those not yet migrated would suffer from selection bias if there are any observed or unobserved differences in the composition of the two groups which have an independent effect on outcomes; the issue is how that selection bias can be avoided.

Matching methods focus on observed differences and construct the counterfactual by selecting, from the pool of not-yet-migrated claimants, a comparison group that 'looks the same', in terms of observed characteristics, to the claimant group being migrated to UC. Matching is indeed the best available method for selecting a matched (or re-weighted) comparison group in which the distribution of observed characteristics is as similar as possible to the one in the claimant group migrated to UC.

To ensure comparability, claimants migrated to UC for whom no suitable non-migrated claimant can be found (i.e. who fall outside of the so-called 'common support') are excluded from the analysis.

The estimate of the impact of being moved to UC is then obtained by comparing the mean labour market outcomes of the migrated group with the mean outcomes of the appropriately matched/re-weighted comparison group. This can be done for each month of migration and for each claimant type.

By simply comparing means, matching techniques are non-parametric and, therefore, not subject to several potential misspecification biases that can affect standard parametric methods such as Ordinary Least Squares (OLS) regression, in which the outcome is regressed on a treatment dummy and on the observed variables, generally entered linearly. In particular, this type of regression may misspecify the outcome equation under JSA as it imposes linearity; it may use the imposed (linear) functional form to extrapolate outside the common support and, therefore, *de facto* compare observationally different claimants; and it restricts the impact of UC to be the same for all claimants, therefore, incurring in further bias in the presence of heterogeneous impacts. On the other hand, by imposing a (linear) structure, OLS allows one to obtain more precise estimates, a feature which is likely to be particularly attractive when disaggregating the analysis and, therefore, reducing sample sizes. Furthermore, the boundary between OLS and matching is a thin one, as the restrictions typically imposed in regression models can be significantly weakened. In particular, in the regression model one can allow the impact of UC to vary according to each observed characteristic. Previous research has shown that such a fully interacted regression model can often produce results that are very similar to the ones obtained by matching (Blundell, Dearden and Sianesi, 2005).

Regression and matching methods are, therefore, best considered and experimented with in tandem. Both, however, are based on the assumption that one can observe all differences between claimants migrated earlier and claimants migrated later that affect outcomes. Any unobserved differences between the groups being compared will wrongly show up as an impact of UC. The success and reliability of estimates based on matching or regression approaches will, therefore, crucially depend on the extent, nature and quality of the data we have on claimants' characteristics. Conditioning on long and detailed labour market histories has been widely recommended as an effective way to control for selection bias in the evaluation of labour market interventions (see e.g. Dehejia and Wahba, 1999, Heckman and Smith, 1999, Heckman *et al.*, 1998, Heckman *et al.*, 1999, and Froehlich, 2004; for caveats, however, see Dolton and Smith, 2011, and Sianesi, 2012¹²).

A.2 Difference in differences

DiD is a strong candidate for comparisons during the migration of flows given the strategy of gradual implementation over time, by claimant type and across areas. Consequently, we discuss DiD for new benefit claims that will gradually be brought within UC.

In the following discussion, we assume, as recommended, that DiD is combined with matching, where treated and comparison groups are matched on a set of observed characteristics. This can be done parametrically or non-parametrically. The latter is preferred as it allows for more flexibility in the functional forms. Just as for matching, it is important to carefully select the variables and observation periods to be included in the conditioning set. One wants to control for a rich set of observed variables to reduce the scope for selection on unobserved variables – although DiD can still cope with some selection on unobserved

¹² Dolton and Smith (2011) find that while controlling for pre-programme outcome measures – and doing so in a flexible way – helps in reducing selection bias, important unobserved factors still remained unaccounted for. Exploiting a randomised experiment as a benchmark, Sianesi (2013) also finds that other unobserved characteristics are important, with the additional finding that labour market histories – summarised in any of a number of ways – do not make the slightest difference in reducing selection bias.

variables for as long as these are constant over time, as explained below. For the evaluation of the impacts of employment-related policies on labour market outcomes, it is generally recommended that careful control is taken of employment histories prior to inflow. In contrast, any information observed after inflow should be excluded, since part of the effect can come through impacts on such variables. Moreover, information on variables that can be influenced to affect eligibility needs to be carefully considered.

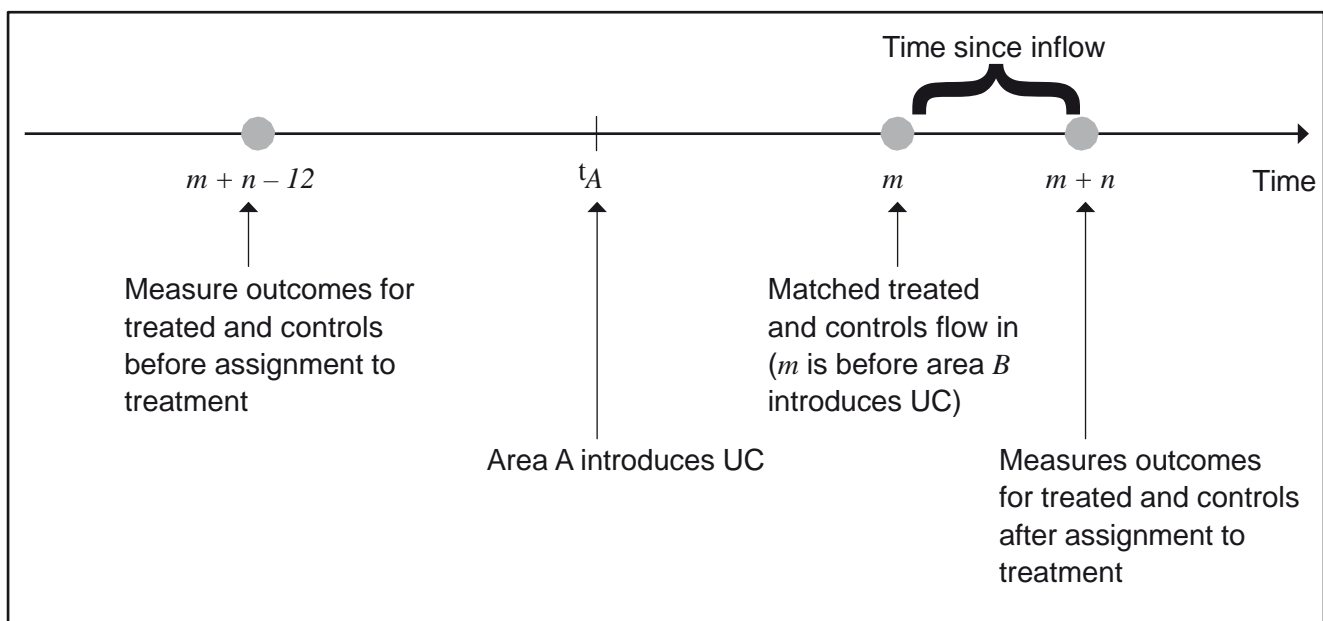
To describe the DiD research design, suppose two groups of geographical areas are exposed to different policy regimes at some point in time. For simplicity of exposition, consider the problem of evaluating the impact of replacing new JSA claims with UC, as per phases 1 and 2 of the implementation strategy. The gradual geographic roll-out of the policy for this type of claim implies that, from some point in time, t_A , area A replaces new JSA claims with UC while area B does so only from $t_B > t_A$ onwards. Area A is a ‘treated’ area as it introduced UC earlier, while area B is a ‘control’ area. *Treatment* is defined as being assigned to UC when flowing into the equivalent of a JSA claim during some month m ($= t_A, \dots, t_B - 1$).

DiD exploits the time dimension in the data. It relies on having either panel data, where the same individuals are followed over time and observed before and after treatment, or repeated cross-section data, where samples are drawn from the same population before and after the policy change in treated areas. Below we consider the two possible comparisons separately.

A.2.1 Panel versus cross-sectional DiD

Panel DiD compares new claimants in area A and month m ($= t_A, \dots, t_B - 1$), called the *treated*, with similar new claimants in areas B during the same month m , called the *controls*. The outcome of interest is measured n months after inflow, in month $m + n$. It is also measured prior to inflow – say one year before $m + n$, in month $m + n - 12$. The DiD method then compares the change in average outcomes in treated areas from $m + n - 12$ to $m + n$ with the change in average outcomes in control areas over the same time frame. Figure A.1 illustrates the timeline of events.

Figure A.1 Timeline of events for panel DiD



Evaluating the labour market impacts of Universal Credit: a feasibility study

The choice of n (time since inflow) is unrestricted although comparisons are cleaner for months $m + n$ falling before the introduction of UC in area B (at time t_B). Some individuals in the control group may move into UC after t_B , changing the definition of the treatment effect being estimated.

Under the following two conditions, DiD identifies the average impact of treatment (being assigned to UC instead of JSA when starting a new benefit claim) on the treated (those starting a new claim in area A and month m):

- 1 **Common trends:** individuals in treated and control areas respond similarly to aggregate shocks. This means long-term secular average outcomes across the two areas follow parallel paths.
- 2 **No selection on transitory unobserved variables:** this means that the reform does not lead individuals to claim JSA/UC differentially between areas in ways that cannot be controlled for and that create transitory differences in outcomes. 'Ashenfelter's dip' is the typical example of how this assumption can be violated. It describes the case where individuals experiencing a transitory dip in earnings are more likely to become treated. Since the dip is transitory, recovery from it would be expected regardless of treatment, and such effect can be incorrectly attributed to treatment.

DiD deals with permanent differences between groups in average untreated outcomes by subtracting out the before-after difference in outcome measures. It also deals with aggregate shocks (economy-wide changes in outcomes, unrelated to the policy), as long as these are common across areas. But DiD does not deal effectively with differences between treated individuals and controls that vary over time. Matching prior to DiD solves this problem insofar as the differences are in observed characteristics, but time-varying unobserved differences between treated individuals and controls in characteristics that affect outcomes create an unavoidable bias in DiD estimates.

In general, two conditions may preclude the use of panel DiD. The first is the obvious case where adequate panel data is not be available. But it may also be that the same outcome is not observable.

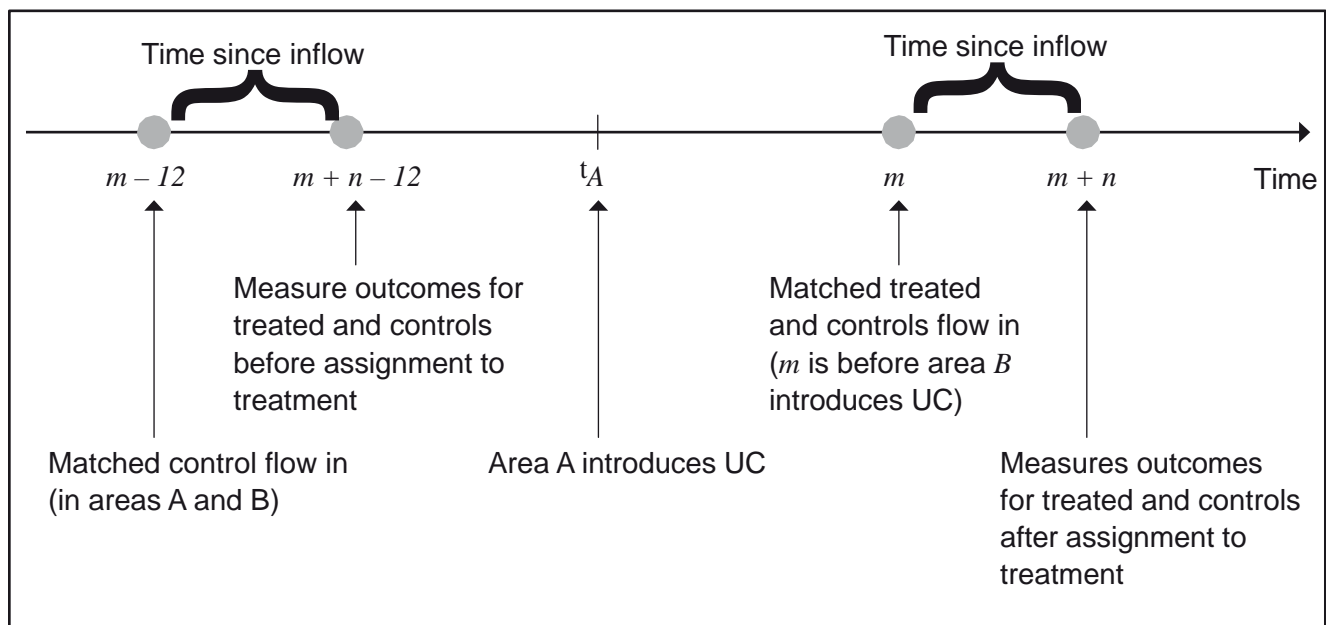
The latter is particularly relevant where the outcome measure is a flow. For instance, suppose we wanted to estimate the effect of UC on outflows from unemployment n months after the start of a new claim. To strip out the effect of any pre-existing differences between claimants in the two areas using panel DiD would require treated and controls, who are all observed starting a new claim in month m , to start a new claim in month $m - 12$ as well. Only in this case could we observe the same outcome in the 'before' period $m + n - 12$. But, of course, the set of people who start a new claim both in month $m - 12$ and again in month m is a very selected sample and perhaps not of particular interest.

One could define the outcome of interest to be the employment rate (rather than a flow) n months after the start of a claim and contrast it with employment rates one year earlier. However, this can still be problematic. The employment rate in month $m - 12$ (a year before the start of a benefit claim) is determined by both inflow and outflow rates into employment. On the other hand, employment rates shortly after registration for unemployment-related benefits depend almost exclusively on the rate of flows into employment. It is conceivable that two areas with similar cyclical variation in employment rates have different underlying employment dynamics, implying different cyclical variation in terms of flows. Put differently, ensuring that the employment rates in treated and control areas follow the same secular trends does not necessarily imply that the same is true for the flows into employment.

For example, two areas might both have high employment rates, but in one area people frequently move in and out of work while in the other people rarely do so. We would then expect the high-turnover area to see more new benefit claimants be in work n months after their claim. Subtracting the difference in these people's employment rates a year before the reform would not prevent us from wrongly attributing the difference in subsequent employment rates to the policy.

Repeated cross-section DiD uses two additional control groups, drawn from new JSA claims at time $m - 12$ in areas A and B . The DiD method with repeated cross-section data uses pre-treatment inflows and outcomes to compute the change in average outcomes in treated and control areas. Figure 2 illustrates the research design for this implementation of DiD.

Figure A.2 Timeline of events for repeated cross-section DiD



In addition to the two assumptions required for panel DiD, the use of repeated cross-section requires one additional assumption:

3 Similar composition of inflows one year apart: within area, the matched samples of inflows at times m and $m + n$ have the same distribution of outcome-relevant unobserved characteristics.

The need for this additional assumption when using cross-sectional DiD means that panel DiD is preferable where it is available.

A.2.2 Sources of variation: choice of comparison groups for DiD

The introduction of UC for new claims (phases 1 to 4) offers three sources of variation:

A Same claimant types treated differently across areas. This is the source of variation used above to discuss DiD. At a given point in time, a new benefit claim will be to UC in one area while it will be to a legacy benefit in another area. For panel DiD, the control group is drawn from the same claimant type in areas where UC has not yet been introduced for these claims. For repeated cross-section DiD, two additional control groups are drawn from the same population of claims in the two areas one year earlier. The underlying assumptions are conditions 1, 2 and 3 above (the last for cross-sectional DiD only), all conditional on observed characteristics.

B Same claimant types within area treated differently across time. This uses variation in treatment around the time a new phase of UC is implemented in each area. In area A , new claims are to a legacy benefit if they happen before t_A but to UC if they happen after t_A . Therefore, one can draw the control group from the same type of spells starting prior to t_A . Let (m_T, m_C) denote the start of the new claim for the treated and control groups, respectively. Therefore, $m_C < t_A < m_T$.

Panel DiD compares the change in average outcomes for the treated group in an interval around m_T , say from $m_T + n - 12$ to $m_T + n$ (where n is months since inflow), to the change in average outcomes for the control group in a similar time interval around m_C , from $m_C + n - 12$ to $m_C + n$. The repeated cross-section DiD relies on two additional comparison groups, drawn from similar claims starting in a similar months one year prior to the introduction of UC, at times $m_T - 12$ and $m_C - 12$ for comparisons with treated and controls, respectively.

The assumptions underlying comparisons across time within areas are (all conditional on observed characteristics): (1) common yearly macro trends across groups defined by month of inflow within area; (2) no selection on time-varying idiosyncratic unobserved variables; and (3) similar composition of inflows one year apart (cross-sectional DiD only).

C Different claimant types treated differently within area and time. At least during the implementation of phases 1 and 2, UC will first replace new JSA claims by single people without children and only later extended to other types of JSA claims.

It is difficult to argue that individuals who gain access to the legacy benefit through different routes or face different circumstances (e.g. the presence/absence of children, of a partner or of Housing Benefit (HB)/Child Tax Credit (CTC)) are comparable. In these cases, matching would not resolve differences in unobserved variables and, conceivably, some of the remaining differences concern time-varying characteristics and invalidate DiD. Therefore, we strongly suggest that all comparisons take place within claimant group, avoiding the variation described in (C). In the following, we therefore ignore design (C). If the evaluator could make a convincing case that, given additional observable characteristics they can condition on, differential trends for different claimant groups in the same area are unlikely, the methodologies outlined below can be straightforwardly adapted.

A.2.3 Supporting information on evaluating the plausibility of the DiD assumptions

- 1 **Common trends.** Identification of the average impact of UC using DiD relies on the assumption that treatment and control groups experience common trends or, in other words, the same macro shocks (assumption 1 above). The common trends assumption holds across groups when the paths of average outcomes are parallel over time. If this is not the case, DiD will not consistently estimate the average impact of UC on the treated group. Differential trends might arise in the evaluation of UC if, say, treated and controls operate in different labour markets. For example, unemployment in urban and rural areas may be subject to different cyclical fluctuations. Whether the common trends assumption holds during the relevant period cannot be tested, but strong suggestive evidence as to its plausibility can be obtained by testing whether there were common trends across the relevant areas in previous periods. The extent to which areas are comparable with respect to aggregate trends should be carefully studied using historical data.
- 2 **Sample selection at inflow.** A problem of sample selection at inflow occurs when the introduction of UC affects the composition of new claimants. It can happen if UC changes who is entitled, awareness of entitlement, or the (financial or non-financial) incentive to claim, relative to legacy benefits. Changing who starts to claim benefits in these ways is indeed part of the aims for UC; but these effects on benefit inflows will then result in the violation of DiD assumption 2 and cause bias in DiD estimates. Although assumption 2 is untestable – one can never gather direct information to assess it – it may be possible to collect suggestive evidence on the impact of UC on the distribution of observed variables at inflow. This can be done by applying the DiD methodology to information at inflow on a set of observed characteristics, which may include demographics, employment history, pre-claimant status or claimant type.

Composition of inflows at different points in time. Repeated cross-section DiD relies on one extra assumption, namely that the composition of inflow samples is invariant across time within area, at least for the selected pre- and post- treatment inflow periods (assumption 3). The evidence that could be examined to gauge the plausibility of this assumption is similar to that supporting assumption 2, but comparing inflows within area, one year apart.

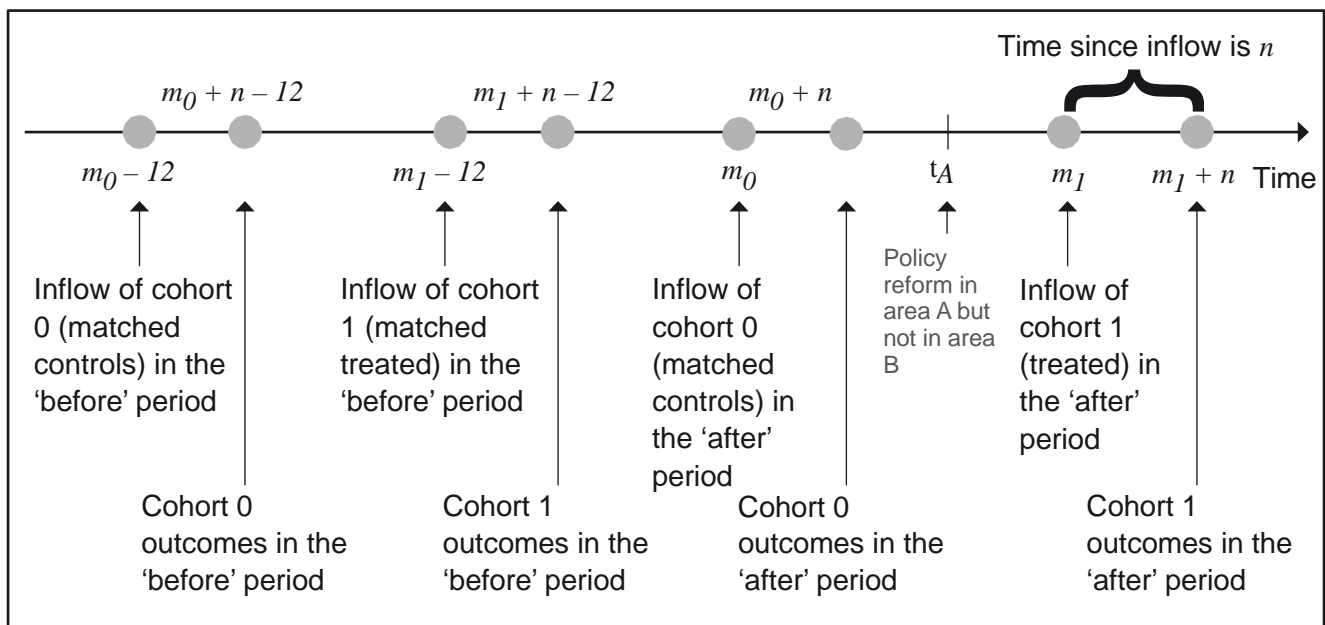
A.2.4 Triple differences

Triple differences (DiDiD) is a simple extension of the DiD method intended to minimise the bias induced by the violation of the differential trends assumption. The idea is simply to apply similar DiD estimators to two (sets of) areas, exploring time variation in the implementation of the reform. In the first area, the reform is introduced before the treated are observed for the second time, while in the second area neither the treated nor the controls are ever exposed to the reform during the observation period. The first estimate is the typical DiD described above. Under the DiD assumptions it will identify the impact of the reform on the treated. The second estimate is a placebo, identifying differential trends in case the DiD assumption is violated. Under the assumption of common differential trends across areas, the difference between the two DiD estimates identifies the impact of treatment on the treated.

Figure A.3 shows the timeline of events for each of two areas: area A, where the reform happens at time t_A , and area B, where no reform is implemented during the observation period. A DiD estimator is computed for each of these areas, using time variation. The treated group flows in at time m_1 . It is matched to three control groups, all having outcomes observed prior to the reform: the ‘treated before’, flowing in some time earlier (say one year),

at $m_1 - 12$, the 'controls after', flowing in at m_0 and the 'controls before', flowing in at $m_0 - 12$. The time m_0 may be before or after $m_1 - 12$. The DiD estimator for area B identifies the differential trends in case the DiD assumption of common trends is violated. Under the assumption that treated and controls face the same differential trends in areas A and B over the considered time frame, the difference between the two DiD estimates identifies the impact of treatment on the treated. In effect, the DiD assumption that there are no differential trends (other than the reform) between the 'before' and 'after' groups is replaced by an assumption that any such difference in trends between the 'before' and 'after' groups is the same in the treated area (A) as in the placebo area (B).

Figure A.3 Timeline of events for the repeated cross-section DiD exploring time variation; triple differences contrasts two such estimators, on areas A (treated after time t_A) and B (control area, not experiencing the policy reform during the time frame considered)



A.3 Regression Discontinuity Design

The idea behind the regression discontinuity design (RDD) approach is that the rules determining treatment are often arbitrary and, therefore, provide random variation in treatment assignment. Specifically, in this design exposure to treatment is determined by the individual's standing on an observed, continuous variable, say income, according to a known, deterministic rule: if income falls below a known threshold, the individual is exposed to the treatment, and if income falls above the threshold, they are not. This design exploits the fact that some individuals are excluded from participation because they marginally failed to meet the eligibility criteria. As in a neighbourhood of the cut-off treatment status is as good as randomly assigned, identification is made possible by comparing individuals arbitrarily close to the cut-off point who did and did not receive the treatment. In the presence of treatment impacts, we would expect to observe a step-like discontinuity of the outcome of interest at the threshold determining participation.

This is a very strong design, comparable to running a randomised experiment, albeit locally at the threshold. Its main disadvantages are that impacts are identified only locally for the sub-group at the threshold and that data requirements in terms of sample sizes are very demanding, as estimation is purely based on individuals close to the cut-off.

The design would be irremediably compromised if:

- a discontinuities other than the treatment of interest spuriously take place at the cut-off;
- b there is selective manipulation of the rule by individuals; or
- c if substitution effects take place between marginal participants and marginal non-participants.

Finally it has to be kept in mind that to meaningfully define marginal individuals, the selection variable has to be continuous.

With these caveats in mind, let us consider the case for using RDD in the evaluation of UC, in particular whether and how the design can be fruitfully applied to the case of natural migration. Natural migration relates to a change in circumstances which gives rise to a new claim (as opposed to a change in the existing claim) and which, therefore, moves the individual into UC. The following list of triggers for natural migration broadly ranks them based on how easy it is for individuals to affect them:

- 1 Start working at least 16 (or 24 or 30) hours per week.
- 2 Stop working.
- 3 Claimant's partner becomes entitled to JSA or starts working at least 24 (or 30) hours per week.
- 4 Income increases.
- 5 House move between (rather than within) local authorities.
- 6 Become lone parent.
- 7 Become sick or no longer sick.
- 8 Birth of first child.
- 9 Lone parent's youngest child reaches five.
- 10 No longer a carer.
- 11 Partner joins/leaves household.

The first five changes in circumstances appear to be under an individual's control to varying degrees. The first four in particular are behavioural responses: individuals with incentives to start a new claim under UC rather than continuing their legacy benefit claim could easily change such circumstances to access the new system. Such manipulation would likely violate the RDD assumption.

Evaluating the labour market impacts of Universal Credit: a feasibility study

The remaining changes in circumstances (6 to 11) appear by contrast to be less – though to some degree still¹³ – under an individual's control (assuming, of course, no fraud in reporting). The problem here is that apart from number nine (lone parent's youngest child reaches five), they are not continuous, indeed they are binary variables. There is, therefore, no way to define arbitrarily close, hence comparable, individuals. For example, looking at birth of the first child, one would compare individuals with one child to individuals with no children, when of course such sharp difference in family circumstances is bound to have an independent effect on labour market outcomes such as employment.

Finally, let us consider the discontinuity arising from a lone parent's youngest child reaching five. Not only is this circumstance not under the parent's control, but age is a continuous variable, as required by the RDD. One could, therefore, envisage comparing, within area (Pathfinder area or district rolling out UC), lone parents whose youngest child turns five – who are moved to UC under what would have been JSA and are, therefore, subject to the full UC regime – to lone parents whose youngest child turns four – who remain in their Income Support (IS) claim under the full legacy benefit regime. The problem of this comparison is that the underlying design is invalidated by (a), i.e. there is a discontinuity other than the natural migration to UC which spuriously takes place at the five-year-old cut-off: the child is now in full-time education, which is most likely to have an independent impact on the lone parent's labour market outcomes, e.g. labour supply. This means the design could not disentangle discontinuities due to the switch to UC from discontinuities that exist regardless of the reform, in this case, from entering compulsory schooling at age five.¹⁴

The problem with the RDD as applied to the natural migration conditions is, therefore, that such conditions can either be manipulated (behavioural response), are discrete or give rise to multiple discontinuities.

A separate and potentially more promising avenue to apply the ideas of RDD to the discontinuity at age 60 arises from the reform of Pension Credit and is discussed separately in Appendix D.

¹³ The most obvious one perhaps is that a couple's decision to move into the same accommodation is likely to be influenced to some degree by the effect this has on their joint entitlement to benefits/UC. The same is probably true of the caring decision – while the need of a relative or neighbour for care is beyond plausible influence, the decision to become a full-time carer is again likely to be influenced to some degree by the financial implications.

¹⁴ Exploiting variation in the age at which children enter full-time education varies by area, and is often younger than five, would not help. Apart from the very demanding data requirements, the horizon over which one could assess impacts would be very limited and indeed one might argue that given how close the comparison child is to starting school, anticipation effects might invalidate the comparison group.

A.4 Instrumental variable methods

Even in the presence of rich background data for cross-sectional matching methods, the concern always remains that we might still miss some of the determinants of labour market outcomes that differ between claimants migrated earlier and claimants migrated later, or between the inflows into new claims across regions. The method of instrumental variables (IV) can account for such unobservables, therefore, yielding unbiased estimates of the impact of UC. What is needed to implement this method is an ‘instrument’, that is an event or a variable which does affect which regime an individual is in, but which at the same time is unrelated to the unobserved determinants of claimant’s labour market outcomes, such as unobserved innate ability or work commitment.

A general caveat concerning the interpretation of estimates using IV methods is that when the impact of UC differs across individuals, the estimated impact is a so-called local average treatment effect (LATE): the average impact for the specific sub-group of the population who are induced to be exposed to UC because of the change in the instrument. Depending on the instrument, this might be or might not be a policy relevant group. Care in any case should, therefore, be taken before viewing the impact estimated using IV as an indicator of the average impact of the reform for all claimants or for all ‘treated’ claimants.

In any case, the method of IV cannot be used for the evaluation of the UC reform as no defensible instrument – something that affects the probability of being exposed to UC while not directly impacting on outcomes – is indeed available. District fully determines exposure to UC for new claimants, but can’t certainly be argued to provide a source of exogenous variation in exposure.

Appendix B

Summary of challenges for evaluation using the phasing in of UC for new claims

The evaluation strategies discussed in this report face some challenges that are briefly discussed here and referred to in the phase-by-phase assessment of Section 3.4.

The first two – entry effects and anticipation effects – are driven by behavioural responses to the implementation of Universal Credit (UC).

B.1 Entry effects

Entry effects relate to the composition of the treatment group. In the scenario examined, the implementation of UC in phases, each limited to specific groups of claimants, may affect the composition of benefit claims of a certain type relative to what they would have been: (i) under the legacy regime, or (ii) had UC been fully implemented.

Issue (i) raises concerns about the comparability of the treated group (claimants in areas under some new UC phase) and the comparison group (similar claimants not exposed to the new phase). This is a threat to internal validity: estimates of treatment effects will be biased if UC claimants differ from equivalent legacy benefit claimants in terms of unobserved characteristics that affect the outcomes of interest.¹⁵ In such cases one would expect the two groups' outcomes to differ regardless of treatment status, and such differences in outcomes would be erroneously attributed to the effects of UC.

Issue (ii) means that UC claimants during the new phase are not representative of claimants of a similar type later on, once UC is fully implemented. It implies that, even if all outcome-relevant differences between the composition of the treatment and comparison groups can be controlled for (so there is no internal validity problem), results from the evaluation study may not carry over to the more policy-relevant setting where UC fully replaces the legacy system – an external validity problem.

¹⁵ Effects on the observed characteristics of the treated group do not pose the same kind of threat to internal validity, since differences in the observed characteristics of treated and comparison groups can be eliminated during the matching process.

B.2 Anticipation effects

Anticipation effects result from responses in advance of a policy reform. Claimants already under UC are not exposed to anticipation effects preceding a new implementation phase, as their incentives remain unchanged. But other benefit claimants and potential future claimants may react in anticipation of an imminent change in the benefit regime.

The potential presence of anticipation effects raises two main concerns: first, that the composition of the comparison group might be affected; and, second, that the outcomes of those in either treated or comparison groups might be affected.

The first concern is directly parallel to the entry effects discussed above, but applied to the composition of the comparison group rather than the treated group. Anticipation effects causing observed differences between the treated and comparison groups may limit the common support to a subset of the treated, hence compromising the external validity of the evaluation exercise (analogous to issue (ii) for entry effects above). More seriously, the internal validity of estimates of the treatment effect can be violated if reactions in advance of a reform lead to compositional differences in unobserved characteristics that affect the outcome of interest (analogous to issue (i) for entry effects above), resulting in biased estimates of the treatment effect.

The second concern may affect all who never claimed UC in the past, for as long as their behaviour can affect whether they move onto UC in the future. Since the treated in question 1 are UC claimants, the potential effect of anticipation on outcomes applies only to the comparison group. But since the wider treated groups considered in questions 2 and 3 include some who will never have claimed UC, anticipation effects in those cases can affect the outcomes of the treated groups as well.

The impact of anticipation on the outcomes of interest changes the definition of the parameter that can be identified to the impact of the reform on the treated group when individuals may react in advance to a forthcoming extension of UC. In other words, the estimated effects will be specific to the particular implementation strategy and may have limited value to the understanding of how UC affects the outcomes of particular groups if fully implemented. If interpreted as estimates of the main parameter of interest – the impact of the reform on the treated group – such estimates may be biased (an internal validity problem).

Other challenges to the evaluation exercise are mechanical consequences of the implementation strategy, independent of individual responses to the reforms:

- **Length of evaluation period**

At most stages of implementation, a rapid extension of UC to new areas and new claimant groups imposes strong limitations on what can be estimated at different durations after inflow into treatment. The definition of the treatment effect that can be identified changes as treated and comparison groups become exposed to different benefits systems.

- **Other threats to the representativeness of the findings**

Some of the most useful variation for evaluation purposes happens early in the scenario examined. But this is also a time when people may still be unfamiliar with the new rules, officials may be learning about their practical and operational details and there is some room for fine-tuning, and while transitional protection is still an important feature of the policy. For these reasons, results for the early implementation stages, even if reliable, might provide limited information about the effects that UC will have once it has fully bedded down.

Evaluating the labour market impacts of Universal Credit: a feasibility study

- **Volume of UC claimants**

One additional concern is that the small number of additional UC claims generated in treated areas during some phases of the implementation may impair the chances of identifying a significant impact. This is mostly relevant to questions 2 and 3 during phases where UC is extended to a small additional share of the population.

- **Mechanical composition effects**

This is mostly relevant in later implementation stages, when a significant proportion of those who would otherwise have claimed legacy benefits instead receive UC because they have claimed UC in the past through some other route (and therefore, remain within the UC regime thereafter). Mechanical composition effects may substantially alter the observed and unobserved composition of new claims, affecting both the representativeness and the validity of the estimated effects. This problem is only relevant for question 1.

Appendix C

Evaluation design summary tables

Table C.1 Evaluation design to assess the impact of Universal Credit (UC) on the outcomes of unemployed new claimants during scenario phases 1 and 2

Sources of variation/ potential comparisons	Some individuals will start a UC claim (and still be subject to the UC regime if their circumstances change thereafter) in some areas whilst: (A) similar individuals in comparable areas will, at the same time, start Jobseeker's Allowance (JSA) claim under the legacy benefit regime and be subject to the full legacy benefit regime. (B) similar individuals in the same areas at a previous date started a JSA claim under the legacy benefit regime and were subject to the full legacy benefit regime.
Outcome variables	Individual employment rate, household employment rate, working hours, earnings, family income and poverty, n (>0) months after inflow.
Evaluation question	Impact of the full UC regime versus the full legacy benefit regime; more specifically, impact of starting the equivalent of a JSA claim under the full UC regime versus starting a JSA claim under the full legacy benefit regime.
For whom	For those unemployed satisfying the JSA conditions who claim the benefit in the first 6 months of implementation from April to October of Year 1 (Pathfinder areas only, for single people without children) or in the second six months of implementation, from November of Year 1 to March of Year 2 (everyone else except a small tail of special cases) and for whom the
Methodology	
Cross-sectional DiD (A)	Same claimant type treated differently across areas.
Treated	The inflow into UC of unemployed people in month m (m =April of Year 1–February of Year 2) in areas where UC has already been phased in.
Comparison	Matched inflow into JSA in month m in similar/matched areas not yet under UC.
Before period	Matched inflow into JSA in month $m-12$ in the treated areas and the same not-yet-treated areas.
Assumptions	<ol style="list-style-type: none"> 1 No differential aggregate trends between treated and not-yet-treated areas for new JSA claimants (given observed characteristics). 2 ('Ashenfelter's dip') No selection on transitory shocks that leads individuals to claim JSA differentially between treated and not-yet treated areas (given observed characteristics). 3 Invariant composition of inflow cohorts one year apart, within area (given observed characteristics).
Cross-sectional DiD (B)	Same claimant type treated differently across time within areas.
Treated	The inflow into UC of unemployed people in month m (m =April of Year 1 to March of Year 2) in areas where UC has already been phased in.
Comparison	Matched inflow into JSA within same areas some time prior to the local roll-out of phase 1 or 2 (explore the use of alternative comparison group for robustness checking).

Continued

Evaluating the labour market impacts of Universal Credit: a feasibility study

Table C.1 Continued

Before period	Matched inflow into JSA in the same areas 12 months prior to the inflow of treated and comparison groups.
Assumptions	<ol style="list-style-type: none"> 1 No differential aggregate trends for inflow cohorts in periods being compared (given observed characteristics). 2 ('Ashenfelter's dip') No selection on transitory shocks that leads individuals to claim JSA differentially between October and November inflow cohorts (given observed characteristics). 3 Invariant composition of October and November inflow cohorts one year apart (given observed characteristics).
Panel DiD	Not recommended for the evaluation of the impact of UC on unemployed starting a JSA claim since this condition (being found as unemployed at a point in time between the 'before' and 'after' periods) imposes overly restrictive assumptions in terms of common trends.
Matching (A)	Same claimant type treated differently across areas.
Treated	The inflow into UC of unemployed people in month m (m =in the first 11 months of implementation) in areas where UC has already been phased in.
Comparison	Matched inflow into JSA in month m in similar/matched areas not yet under UC.
Assumptions	<ol style="list-style-type: none"> 1 <i>Conditional independence assumption (CIA)</i>: no differential composition between treated and control groups (given observed characteristics). 2 No differential aggregate outcomes/shocks between treated and matched not-yet-treated areas for new JSA claimants (given observed characteristics).
Matching (B)	Same claimant type treated differently across time.
Treated	The inflow into UC of unemployed people in month m (m =in the first 11 months of implementation) in areas where UC has already been phased in.
Comparison	Matched inflow into JSA within same areas some time prior to the local roll-out of phase 1 or 2 (explore the use of alternative comparison group for robustness checking).
Assumptions	<ol style="list-style-type: none"> 1 <i>Conditional independence assumption (CIA)</i>: no differential composition between treated and control groups (given observed characteristics). 2 No differential aggregate outcomes/shocks by time for the inflow cohorts (given observed characteristics). This effectively means no differential seasonal effects.
Additional note	This matching method is Regression Discontinuity Design (RDD) in the limit, when time of inflow for the treated and comparison cohorts becomes closer.
For how long	Outcomes assessed e.g. monthly, from the first month after inflow onwards. Impacts can be assessed from the first month of implementation April 2013 till when JSA claims are turned off for the comparison group (March of Year 2 at the latest) – at most for 12 months if keeping time comparisons limited to cohorts observed for no more than one year apart. After this, estimates will identify a different effect.
Ancillary analyses	
Common trends (DiD)	Historical evidence on common trends for the outcome variable of interest for the treated and control groups. Major differences in cyclical variation, especially if closer in time to the evaluation period, would suggest the groups are not fit for DiD analysis.
Common level of outcomes (matching)	The common aggregate shock assumption of matching (A) and (B) can be checked historically, and indeed the areas should be matched on the history of outcomes.

Continued

Table C.1 Continued

Entry effects	Inflows in the treated and not-yet-treated areas, or in the treated areas before and after the move to UC, might have been affected as individuals change their take-up behaviour or change behaviour to influence entitlement. In such cases, the composition of the treated under any of the proposed approaches will be different from the composition of claimants in not-yet-treated areas or in the treated areas before the introduction of UC. We can test for entry effects by testing for the impact of UC on inflow as detailed in Appendix A.2.3.
Limitations/issues	
Parameter of interest	We can measure the impact of UC versus the full legacy benefit regime only for as long as UC is not phased in for the comparison group. For phases 1 and 2, this amounts to a maximum of 11 months after treatment. After that the impact that can be measured has a different interpretation.
External validity	<p><i>Entry effects:</i> Impacts assessed only for those starting a new JSA/UC claim in the first year of implementation in the areas where UC has been phased in for their type of claim. This can be a peculiar group given that access to UC is limited to specific types of claims and individuals can influence entitlement in ways that will not be possible/desirable once UC is fully implemented. Therefore, the treated might not be representative of the typical UC claimants that would have started a JSA claim under the legacy benefit regime.</p> <p><i>Anticipation effects:</i> Choosing the pre-treatment period a year in advance should confidently rule out any impact of the future move to UC. But comparisons across time/area may be affected by anticipation effects if the comparison group foresees the imminent change in incentives and responds by changing claiming behaviour.</p> <p>Outcomes for the comparison group may be affected by anticipation effects as the introduction of phase 3 approaches. Since UC will replace new claims to Working Tax Credit (WTC) soon afterwards, new JSA claimants have an incentive to move into work less quickly (if UC would be more generous than WTC), or more quickly (if UC would be less generous than WTC), than they would under a full legacy benefit regime.</p> <p><i>Short-run – teething and familiarity:</i> impacts can be measured only shortly after UC is implemented, when potential claimants are still unfamiliar with the new incentives and the offices/Department for Work and Pensions (DWP) deal with the initial practical implementation adjustments.</p> <p><i>Volume:</i> the impact that can be measured is specific to the implementation strategy and does not embed the indirect effects that could take place once UC is fully implemented as coverage still very limited at this stage. Since indirect effects affect all, they may in particular change the impact of UC on claimants.</p>
Internal validity	<p><i>Entry effects:</i> may also compromise the internal validity of estimates if affect the composition of the treated groups in unobserved ways.</p> <p><i>Anticipation effects:</i> may also compromise the internal validity of estimates if affect the composition of the comparison group in unobserved ways.</p>
Data required	<p>Data allowing one to identify all (or a large random sample of) claims to JSA/UC during the first year of implementation; earlier to allow testing the common-trends/common-outcomes assumption.</p> <p>Data on labour market outcomes/histories and background characteristics of claimants.</p> <p>Data on labour market status before inflow to test for entry effects.</p>

Evaluating the labour market impacts of Universal Credit: a feasibility study

Table C.2 Evaluation design to assess the impact of UC on benefit entitlement and labour market outcomes of potential future claimants during scenario phases 1 and 2

Sources of variation/ potential comparisons	Some areas will replace new JSA claims with UC whilst: (A) similar areas at the same time treat a similar new claim under the legacy benefit regime. (B) the same areas in a previous period treat the new claim under the legacy benefit regime.
Outcome variables	Fulfilling UC entitlement rules for phases 1 or 2; individual employment, household employment, working hours, earnings, family income and poverty measured in month n (>0) of phases 1 or 2.
Evaluation question	Impact of the full UC regime versus the full legacy benefit regime; more specifically, impact of the reform on the probability of becoming entitled to UC (or on some moments of other outcomes of interest like the probability of future employment) versus the full legacy benefit regime (the specific question depends on the control group). It aggregates both direct and indirect effects on all, whether or not directly affected by the reform.
For whom	All working-age population, workers in low-paid jobs, lone parents or all parents, people without children, the less educated, those with a past history of benefit claiming, etc.
Methodology	
Cross-sectional DiD (A)	Geographical variation in the treatment of JSA claimants.
Treated	All working-age population or some sub-group of interest observed just before the local introduction of phase 1 or 2 (some time in the first 11 months of implementation).
Comparison	Matched individuals in matched areas where phase 1 or 2 is introduced later.
Before period	Matched individuals in the same areas one year earlier.
Assumptions	<ol style="list-style-type: none"> 1 No differential aggregate trends between treated and not-yet treated areas for the groups being considered (given observed characteristics). 2 ('Ashenfelter's dip') Invariant composition of treated versus control groups with respect to transitory shocks (given observed characteristics). 3 Invariant composition of the treated/comparisons groups one year apart, within area (given observed characteristics).
Cross-sectional DiD (B)	Time variation in the treatment of JSA claimants, within areas.
Treated	All working-age population or some sub-group of interest observed just before the local introduction of phase 1 or 2 (some time in the first 11 months of implementation).
Comparison	Matched individuals observed in the same areas during an earlier month.
Before period	Matched individuals in the same areas one year earlier.
Assumptions	<ol style="list-style-type: none"> 1 No differential aggregate trends across similar groups observed in the same areas at different points in time (given observed characteristics). 2 ('Ashenfelter's dip') Invariant composition of the treated versus control groups with respect to transitory shocks (given observed characteristics). 3 Invariant composition of the stock of potential future claimants one year apart, within area (given observed characteristics).
Panel DiD	DiD methodologies (A) and (B) can also be implemented longitudinally, with the 'before period' being 12 months before the outcomes of treated and comparison groups are measured.
Assumptions	In both cases (A) and (B) assumption 3 can be dropped.

Continued

Table C.2 Continued

Matching (A)	Geographical variation in the treatment of JSA claimants.
Treated	All working-age population or some sub-group of interest observed just before the local introduction of phase 1 or 2 (some time in the first 11 months of implementation).
Comparison	Matched individuals in matched areas where phase 1 or 2 is introduced later.
Assumptions	<ol style="list-style-type: none"> 1 <i>Conditional independence assumption (CIA)</i>: no differential composition between treated and comparison groups (given observed characteristics). 2 No differential aggregate outcomes/shocks between treated and matched not-yet-treated areas for the group of interest (given observed characteristics).
Matching (B)	Time variation in the treatment of JSA claimants, within area.
Treated	All working-age population or some sub-group of interest observed just before the local introduction of phase 1 or 2 (some time in the first 11 months of implementation).
Comparison	Matched individuals in the same areas observed some time before treated, prior to the roll-out of phase 1 or 2.
Assumptions	<ol style="list-style-type: none"> 1 <i>Conditional independence assumption (CIA)</i>: no differential composition between treated and control groups (given observed characteristics). 2 No differential aggregate outcomes/shocks by time for the group of interest (given observed characteristics). This effectively means no differential seasonal effects.
For how long	<p>Outcomes assessed, e.g. monthly, from the first month after implementation of phase 1 or 2 onwards.</p> <p>Impacts can be assessed from the start of implementation till when JSA claims are turned off for the comparison group (March of Year 2 at the latest) – at most for 12 months if keeping time comparisons limited to cohorts observed for no more than one year apart. After this, estimates will identify a different effect.</p>
Ancillary analyses	
Common trends (DiD)	Historical evidence on common trends for the outcome variable of interest for the treated and control groups. Major differences in cyclical variation, especially close in time to the evaluation period, suggest the groups are not fit for DiD analysis.
Common level of outcomes (matching)	The common aggregate outcome/shock assumption of matching (A) and (B) can be checked historically, and indeed the areas should be matched on the history of outcomes.
Entry effects	Since this design is unconditional on claiming status, it is unlikely to be much affected by entry effects. However, entry effects resulting from the implementation strategy condition the identified effect and are a threat to external validity.
Limitations/issues	
Parameter of interest	We can measure the impact of UC versus the full legacy benefit regime only for as long as UC is not phased in for the comparison group. For phases 1 and 2, this amounts to a maximum of one year if exploiting time variation. After that, the impact that can be measured has a different interpretation.

Continued

Evaluating the labour market impacts of Universal Credit: a feasibility study

Table C.2 Continued

External validity	<p><i>Entry effects:</i> unlikely to change the composition of treated and comparison groups. However, may affect composition of those ending up making a new benefit claim and, through that process, affect the estimated impact for the entire population or some sub-group of interest.</p> <p><i>Anticipation effects:</i> Comparisons across time may be affected by anticipation effects if the treated or comparison groups foresee the imminent change in incentives (for the comparison group this amounts to the introduction of UC for new JSA claims, for the treated the extension of UC to new Tax Credit (TC)/Income Support for lone parents (ISlp) claims) and respond by changing behaviour. This behaviour is induced by the implementation strategy and, therefore, not externally valid.</p> <p><i>Short-run – teething and familiarity:</i> impacts can be measured only shortly after UC is implemented, when potential claimants are still unfamiliar with the new incentives and the offices/DWP deal with the initial practical implementation adjustments. This may affect the size of the impact and coverage of UC for the specific groups it becomes available in ways that would not be realised in the longer-run and affect both the direct and indirect effects of UC.</p> <p><i>Volume:</i> small coverage means that indirect effects are unlikely to be of any relevance, even if strong in the longer run under full implementation of UC.</p>
Data required	<p>Data allowing one to identify all (or a large random sample of) the stock of potential future claimants of different types just before the local introduction of phase 1 or 2, and earlier cohorts to allow for time comparisons and the testing of common trends, common levels and entry effects.</p> <p>Data on labour market outcomes, labour market histories and background characteristics of claimants.</p>

Table C.3 Evaluation design to assess the indirect effects of UC during scenario phases 1 and 2

Sources of variation/ potential comparisons	<p>Some areas will replace new JSA claims with UC whilst:</p> <p>(A) similar areas at the same time treat a similar new claim under the legacy benefit regime.</p> <p>(B) the same areas in a previous period treat the new claim under the legacy benefit regime.</p>
Outcome variables	Individual employment, household employment, working hours, earnings, family income and poverty measured in month n (>0) of phases 1 or 2.
Evaluation question	Impact of UC replacing new JSA claims versus the full legacy benefit regime on individuals not directly affected by UC during phases 1 and 2. Measures only indirect effects.
For whom	<p>Parents and couples without children during phase 1; those for whom JSA not relevant (e.g. lone parents with children under five) during phase 2</p> <p>Among them, sub-groups of interest: those out-of-work, workers on low-paid-jobs, those with a past history of benefit claiming, etc.</p>

Continued

Table C.3 Continued

Methodology	
Cross-sectional DiD (A)	Geographical variation in the treatment of JSA claimants.
Treated	Parents and couples without children (phase 1) or lone parents of children under five (phase 2) observed just before or at the start of the local introduction of phase 1 or 2, respectively.
Comparison	Matched individuals in matched areas where phase 1 or 2 is introduced later.
Before period	Matched individuals in the same areas one year earlier.
Assumptions	<ol style="list-style-type: none"> 1 No differential aggregate trends between treated and not-yet treated areas for the groups being considered (given observed characteristics). 2 ('Ashenfelter's dip') Invariant composition of treated versus control groups with respect to transitory shocks (given observed characteristics). 3 Invariant composition of the treated/comparisons groups one year apart, within area (given observed characteristics).
Cross-sectional DiD (B)	Time variation in the treatment of JSA claimants, within areas.
Treated	Parents and couples without children (phase 1) or lone parents of children under five (phase 2) observed just before or at the start of the local introduction of phase 1 or 2, respectively.
Comparison	Matched individuals observed in the same areas during an earlier month.
Before period	Matched individuals in the same areas one year earlier.
Assumptions	<ol style="list-style-type: none"> 1 No differential aggregate trends across similar groups observed in the same areas at different points in time (given observed characteristics). 2 ('Ashenfelter's dip') Invariant composition of the treated versus control groups with respect to transitory shocks (given observed characteristics). 3 Invariant composition of the treated/comparisons groups one year apart, within area (given observed characteristics).
Panel DiD	DiD methodologies (A) and (B) can also be implemented longitudinally, with the 'before period' being 12 months before the outcomes of treated and comparison groups are measured.
Assumptions	In both cases (A) and (B) assumption 3 can be dropped.
Matching (A)	Geographical variation in the treatment of JSA claimants.
Treated	Parents and couples without children (phase 1) or lone parents with children under five (phase 2) observed just before or at the start of the local introduction of phase 1 or 2, respectively.
Comparison	Matched individuals in matched areas where phase 1 or 2 is introduced later.
Assumptions	<ol style="list-style-type: none"> 1 <i>Conditional independence assumption (CIA)</i>: no differential composition between treated and comparison groups (given observed characteristics). 2 No differential aggregate outcomes/shocks between treated and matched not-yet-treated areas for the group of interest (given observed characteristics).
Matching (B)	Time variation in the treatment of JSA claimants, within area.
Treated	Parents and couples without children (phase 1) or lone parents with children under five (phase 2) observed just before or at the start of the local introduction of phase 1 or 2, respectively.
Comparison	Matched individuals in the same areas observed some time before treated, prior to the roll-out of phase 1 or 2.

Continued

Evaluating the labour market impacts of Universal Credit: a feasibility study

Table C.3 Continued

Assumptions	<p>1 <i>Conditional independence assumption (CIA)</i>: no differential composition between treated and control groups (given observed characteristics).</p> <p>2 No differential aggregate outcomes/shocks by time for the group of interest (given observed characteristics). This effectively means no differential seasonal effects.</p>
For how long	<p>Outcomes assessed, e.g. monthly, from the first month after implementation of phase 1 or 2 onwards.</p> <p>Impacts can be assessed from the start of implementation till when JSA claims are turned off for the comparison group (March of Year 2 at the latest) – at most for 12 months if keeping time comparisons limited to cohorts observed for no more than one year apart. After this, estimates will identify a different effect.</p>
Ancillary analyses	
Common trends (DiD)	Historical evidence on common trends for the outcome variable of interest for the treated and control groups. Major differences in cyclical variation, especially close in time to the evaluation period, suggest the groups are not fit for DiD analysis.
Common level of outcomes (matching)	The common aggregate outcome/shock assumption of matching (A) and (B) can be checked historically, and indeed the areas should be matched on the history of outcomes.
Entry effects	Since this design applies to individuals not yet entitled to JSA, entry effects are unlikely.
Limitations/issues	
Parameter of interest	<p>We can measure the impact of UC versus the full legacy benefit regime only for as long as a new implementation phase extends access to UC to the treated or comparison groups for those becoming entitled (and therefore, potentially affecting them directly). Therefore, except for those groups for whom JSA is not relevant (e.g. lone parents with children under five), who can be followed up to the start of phase 3, other groups can only be studied during phase 1. This amounts to a maximum of 13 months for lone parents drawn in Pathfinder areas in the month before the start of implementation when using matching with geographical variation.</p>
External validity	<p><i>Entry effects</i>: unlikely to change the composition of treated and comparison groups or their behaviour.</p> <p><i>Anticipation effects</i>: Comparisons closer to the time UC becomes available to the treated or comparison groups can be affected by the anticipation of a change in incentives, therefore, changing the nature of the identified impact (external validity).</p> <p><i>Short-run – teething and familiarity</i>: may affect the size of the impact and coverage of UC for the specific groups it becomes available in ways that would not be realised in the longer-run and, therefore, affecting the indirect effects on others as well.</p> <p><i>Volume</i>: small coverage means that indirect effects are unlikely to be of relevance, even if strong in the longer run under full implementation of UC.</p>
Data required	<p>Data allowing one to identify all (or a large random sample of) the stock of parents and couples without children (phase 1) or lone parents with children under five (phase 2) just before or shortly after the local introduction of phase 1 or 2, and earlier cohorts to allow for time comparisons and the testing of common trends, common levels (and entry effects if desired).</p> <p>Data on labour market outcomes, labour market histories and background characteristics of claimants.</p>

Table C.4 Evaluation design to assess the impact of UC on benefit entitlement and labour market outcomes of potential future claimants during scenario phase 3

Sources of variation/ potential comparisons	Some areas will replace new TC/ISlp claims with UC whilst: (A) similar areas at the same time treat a similar new claim under the legacy benefit regime; (B) the same areas in a previous period treat the new claim under the legacy benefit regime.
Outcome variables	Fulfilling UC entitlement rules for phase 3, claiming benefits, individual employment, household employment, working hours, earnings, family income and poverty measured in month n (>0) after roll-out of phase 3.
Evaluation question	Impact of the full UC regime versus the full legacy benefit regime; or Impact of the full UC regime versus a mixed regime. More specifically, impact of UC replacing new JSA/TC/ISlp claims on the probability of becoming entitled to benefits, claiming benefits, or future labour market outcome versus the full legacy benefit regime or versus a mixed regime in which UC replaces only new JSA claims.
For whom	The entire working-age population, those not on benefits, new parents, lone parents, workers on low-paid jobs, etc.
Methodology	
Cross-sectional DiD (A)	Regional variation in treatment of new TC/ISlp spells (identifies the impact of the full UC regime versus a mixed regime).
Treated	The entire working-age population or some sub-group of interest observed just before the start of phase 3 in areas where phase 3 is first introduced.
Comparison	Matched individuals in matched areas where phase 3 is introduced later.
Before period	Matched individuals in the same areas one year earlier.
Assumptions	1 No differential aggregate trends between treated and not-yet treated areas for potential future claimants of the type being considered (given observed characteristics). 2 ('Ashenfelter's dip') Invariant composition of transitory shocks by treatment status – meaning whether or not exposed to phase 3 (given observed characteristics). 3 Invariant composition of treated and comparison groups one year apart (given observed characteristics).
Cross-sectional DiD (B)	Time variation in treatment of new TC/ISlp spells, within area (can identify the impact of the full UC regime versus the full legacy benefit regime or versus a mixed regime, depending on the time of inflow for the control group).
Treated	The entire working-age population or some sub-group of interest observed just before the start of phase 3.
Comparison	Matched individuals in the same areas during an earlier month, prior to the roll-out of phase 3.
Before period	Matched individuals in the same areas one year earlier.

Continued

Evaluating the labour market impacts of Universal Credit: a feasibility study

Table C.4 Continued

Assumptions	<ol style="list-style-type: none"> 1 No differential aggregate trends for the treated and comparison cohorts (given observed characteristics). 2 ('Ashenfelter's dip') Invariant composition of transitory shocks by treatment status – meaning whether or not exposed to phase 3 (given observed characteristics). 3 Invariant composition of treated and comparison groups one year apart, within area (given observed characteristics).
Panel DiD	DiD methodologies (A) and (B) can also be implemented longitudinally, with the 'before period' being 12 months before the outcomes of treated and comparison groups are measured.
Matching (A)	Geographical variation in the treatment of TC/ISlp claimants.
Treated	All working-age population or some sub-group of interest observed just before the local introduction of phase 3.
Comparison	Matched individuals in matched areas where phase 3 is introduced later.
Assumptions	<ol style="list-style-type: none"> 1 <i>Conditional independence assumption (CIA)</i>: no differential composition between treated and comparison groups (given observed characteristics). 2 No differential aggregate outcomes/shocks between treated and matched not-yet-treated areas for the group of interest (given observed characteristics).
Matching (B)	Time variation in the treatment of TC/ISlp or JSA/TC/ISlp claimants, within area.
Treated	All working-age population or some sub-group of interest observed just before the local introduction of phase 3.
Comparison	Matched individuals in the same areas observed some time before treated, prior to the roll-out of phase 3 and, possibly, prior to the local roll-out of UC.
Assumptions	<ol style="list-style-type: none"> 1 <i>Conditional independence assumption (CIA)</i>: no differential composition between treated and control groups (given observed characteristics). 2 No differential aggregate outcomes/shocks by time for the group of interest (given observed characteristics). This effectively means no differential seasonal effects.
For how long	<p>Cleanest comparisons: for n months, until the control group is exposed to a new implementation phase:</p> <ul style="list-style-type: none"> • Geographical variation only allows for outcomes to be measured for up to two months before phase 3 is introduced in control areas. • Time variation allows for control groups to be selected from earlier inflows if comparability is justified, therefore, leaving more scope for a longer analysis.
Ancillary analyses	
Common trends	Historical evidence on common trends for the outcome variable of interest for the treated and control groups. If there are major differences in cyclical variation, especially close in time to the evaluation period, that would suggest the groups are not fit for DiD analysis.
Common level of outcomes (matching)	The common aggregate shock assumption of matching can be checked historically, and indeed the areas should be matched on the history of outcomes.

Continued

Table C.4 Continued

Entry effects	<p>These are effects changing the composition of the stocks of potential future claimants at the start of phase 3. Since this design is unconditional on claiming status, entry effects are unlikely to be a problem. But the same may not hold if restricting attention to those not on benefits as composition may depend on past local history of roll-out and foreseen changes in incentives.</p> <p>Careful analysis of compositional differences between treated and control groups and how these change over time can provide suggestive evidence of the presence or absence of entry effects.</p>
Limitations/issues	
Parameter of interest	The definition of the estimated parameter changes with time as the roll-out of UC advances.
External validity	<p><i>Entry effects:</i> unlikely to change the composition of treated and comparison groups if unconditional on treatment status but the same not true when looking at impacts on those not on benefits. Moreover, may affect composition of those ending up making a new benefit claim under the new regime and, through that process, affect the estimated impact for the entire population or any sub-group of interest.</p> <p><i>Anticipation effects:</i> Comparisons across time may be affected by anticipation effects if the treated or comparison groups foresee the imminent change in incentives (for the comparison group this amounts to the introduction of UC for new JSA/TC/ISlp claims, for the treated the extension of UC to new ISc/ESA/HB claims) and respond by changing behaviour. This behaviour is induced by the implementation strategy and, therefore, not externally valid.</p> <p><i>Short-run – teething and familiarity:</i> this problem is likely to become gradually less important.</p> <p><i>Volume:</i> small coverage means that indirect effects are unlikely to be of any relevance, even if strong in the longer run under full implementation of UC. This problem is likely to become gradually less important.</p>
Data required	<p>Data allowing one to identify all (or a large random sample of) the working-age population or some sub-group of interest in April of Year 2 or earlier to allow for time comparisons and the testing of common trends, common outcomes and entry effects.</p> <p>Data on labour market outcomes, labour market histories and background characteristics of claimants.</p>

Evaluating the labour market impacts of Universal Credit: a feasibility study

Table C.5 Evaluation design to assess the indirect effects of UC during scenario phase 3

Sources of variation/ potential comparisons	Some areas will replace new TC/ISlp claims with UC whilst: (A) similar areas at the same time treat a similar new claim under the legacy benefit regime; (B) the same areas in a previous period treat the new claim under the legacy benefit regime.
Outcome variables	Individual employment, household employment, working hours, earnings, family income and poverty measured in month n (>0) of phase 3.
Evaluation question	Impact of UC replacing new TC/ISlp claims versus a hybrid regime where UC has replaced only new JSA claims on individuals not directly affected by the reform (i.e. phase 3). Measures only indirect effects.
For whom	Those who started a UC claim (to what would have been JSA) any time before the start of phase 3 (Matching A), or just before the roll-out of phase 3 (Matching B and DiD A). Among them, sub-groups of interest: for Matching A, e.g. those out-of-work or workers on low-paid-jobs.
Methodology	
Cross-sectional DiD (A)	Geographical variation in the treatment of TC/ISlp claimants.
Treated	Inflow into UC in treated areas observed just before the start of the local introduction of phase 3.
Comparison	Matched inflow into UC at the same time as the treated, in matched areas where phase 3 is introduced later.
Before period	Matched inflow into UC in treated and non-treated areas just after the start of phase 2 (in both treated and comparison areas).
Assumptions	<ol style="list-style-type: none"> 1 No differential aggregate trends between treated and not-yet treated areas for the groups being considered (given observed characteristics). 2 ('Ashenfelter's dip') Invariant composition of treated versus control groups with respect to transitory shocks (given observed characteristics). 3 Invariant composition of the treated/comparisons groups between the start and the end of phase 2, within area (given observed characteristics).
Matching (A)	Geographical variation in the treatment of TC/ISlp claimants.
Treated	The stock of individuals who started a UC claim at any time before the roll-out of phase 3, observed at the time phase 3 is introduced in areas where this happens earlier.
Comparison	The stock of matched individuals who started a UC claim at any time before the roll-out of phase 3, observed at the same time as the treated, but in matched areas where the roll-out of phase 3 happens later.
Assumptions	<ol style="list-style-type: none"> 1 <i>Conditional independence assumption (CIA)</i>: no differential composition between treated and comparison groups (given observed characteristics). 2 No differential aggregate outcomes/shocks between treated and matched not-yet-treated areas for the stock of individuals who started a UC claim before the roll-out of phase 3 (given observed characteristics).

Continued

Table C.5 Continued

Matching (B)	Time variation in the treatment of TC/ISlp claimants, within area.
Treated	The inflow into UC near the end of phase 2 in areas where phase 3 is introduced earlier.
Comparison	The matched inflow into UC just after the start of phase 2 in areas where phase 3 is introduced earlier.
Assumptions	<p>1 <i>Conditional independence assumption (CIA)</i>: no differential composition between treated and control groups (given observed characteristics).</p> <p>2 No differential aggregate outcomes/shocks by time for the inflow into UC during phase 2 (given observed characteristics). This effectively means no differential seasonal effects.</p>
For how long	<p>Outcomes assessed, e.g. monthly, from the first month after implementation of phase 3 onwards.</p> <p>Impacts can be assessed for three months, from May to July of Year 2, until the comparison group too is subject to the same reform as the treated group.</p>
Ancillary analyses	
Common trends (DiD)	Historical evidence on common trends for the outcome variable of interest for the treated and control groups. Major differences in cyclical variation, especially close in time to the evaluation period, suggest the groups are not fit for DiD analysis.
Common level of outcomes (matching)	The common aggregate outcome/shock assumption of matching (A) and (B) can be checked historically, and indeed the areas should be matched on the history of outcomes.
Entry effects	Since this design applies to individuals not directly affected by phase 3, entry effects are unlikely.
Limitations/issues	
Parameter of interest	We can measure the indirect effects of UC replacing all new claims to JSA, ISlp and TCs only versus a hybrid regime and only until the comparison group is subject to the same reform. This amounts to a maximum of only three months.
External validity	<p><i>Entry and anticipation effects</i>: unlikely to change the composition of treated and comparison groups or their behaviour.</p> <p><i>Short-run – teething and familiarity</i>: this problem is likely to become gradually less important.</p> <p><i>Volume</i>: With relatively quick within-area roll-out and extended eligibility, low volume of UC cases might be less of an issue.</p>
Data required	<p>Data allowing one to identify all (or a large random sample of) the inflow into UC prior to the introduction of phase 3.</p> <p>Data on labour market outcomes, labour market histories and background characteristics of claimants.</p>

Evaluating the labour market impacts of Universal Credit: a feasibility study

Table C.6 Evaluation design to assess the impact of UC on the outcomes of new claimants to what would otherwise be ISc/ESA/HB

Sources of variation/ potential comparisons	Some individuals will start a UC claim whilst similar individuals in the same areas at a previous date start the same type of claim under a different regime that varies with time and area: (A) full legacy benefit regime if prior to local implementation of phase 1 or 2; (B) UC for new JSA claims if after local implementation of phase 1 or 2; (C) UC for new claims to JSA/TC/ISlp if after local implementation of phase 3.
Outcome variables	Individual employment rate, household employment rate, working hours, earnings, family income and poverty, n (>0) months after inflow.
Evaluation question	Impact of the full UC regime versus the full legacy benefit regime. or Impact of the full UC regime versus a mixed regime. More specifically, impact of starting the equivalent of an Income Support for carers (ISc)/Employment and Support Allowance (ESA)/Housing Benefit (HB) claim under the full UC regime versus starting a similar claim under the full legacy benefit regime or under a mixed regime in which UC has replaced new claims to JSA or JSA/TC/ISlp (depending on the control group).
For whom	Low-income carers, sick and disabled people and renters, not already claiming UC, who start a claim during or after October of implementation Year 2 and qualify for ISc/ESA/HB under the legacy benefit (and now qualify for UC
Methodology	
Cross-sectional DiD (A)	Not applicable.
Cross-sectional DiD (B)	Same claimant type treated differently across time within areas (can identify the impact of the full UC system versus the full legacy benefit regime or versus a mixed system, depending on the time of inflow for the control group).
Treated implementation Year 2	The inflow into a UC claim equivalent to ISc/ESA/HB in or after October of
Comparison	Matched inflow into ISc/ESA/HB in the same areas during some earlier month.
Before period	Matched inflow into ISc/ESA/HB in the same areas one year earlier.
Assumptions	1 No differential aggregate trends for treated versus comparison inflow cohorts (given observed characteristics). 2 ('Ashenfelter's dip') No selection on transitory shocks that leads individuals to claim ISc/ESA/HB differentially across time the periods being compared (given observed characteristics). 3 Invariant composition of inflow cohorts one year apart (given observed characteristics).
Longitudinal DiD	DiD methodology (B) can also be implemented longitudinally, with the 'before period' being 12 month before the outcomes of treated and comparison groups are measured.
Assumptions	Assumption 3 can be dropped.
Matching (A)	Not applicable.
Matching (B)	Time variation in the treatment of ISc/ESA/HB claimants, within area (can identify the impact of the full UC system versus the full legacy benefit regime or versus a mixed system, depending on the time of inflow for the control group).
Treated 2.	The inflow into a UC claim equivalent to ISc/ESA/HB in or after October of Year

Continued

Table C.6 Continued

Comparison	Matched inflow into ISc/ESA/HB in the same areas during some earlier month.
Assumptions	<p>1 <i>Conditional independence assumption (CIA)</i>: no differential composition between treated and control groups (given observed characteristics).</p> <p>2 No differential aggregate outcomes/shocks by time for the group of interest (given observed characteristics). This effectively means no differential seasonal effects.</p>
For how long	<p>Outcomes assessed, e.g. monthly, from November of implementation Year 2 onwards.</p> <p>Cleanest comparisons while control groups face full legacy benefit regime (i.e. before phase 1 or 2 was locally introduced), having ensured comparability. So if the control group were those starting an ISc/ESA/HB claim in November of Year 1, then in an area where UC replaced new JSA claims in March of Year 2, outcomes for up to three months from the claim date could be assessed. Choosing a control group from earlier inflows allows for impacts to be estimated over a longer period after treatment.</p> <p>Comparisons after control groups are exposed to UC, even if UC is restricted to claims equivalent to JSA or JSA/TC/ISlp, identifies different parameters: the impact of UC on the treated versus a control group with a time-varying increasing proportion of individuals under UC.</p>
Ancillary analyses	
Common trends	Historical evidence on common trends for the outcome variable. Major differences in cyclical variation, especially if closer in time to the evaluation period, would suggest the groups are not fit for DiD analysis.
Common level of outcomes (matching)	The common aggregate shock assumption of matching can be checked historically, and indeed the areas should be matched on the history of outcomes.
Entry effects	<p>These can be due to:</p> <p>1 <i>Selection at entrance</i>: If people delay (or accelerate) a claim so that it falls within UC (or the legacy benefit regime), the composition of treated and/or control groups could be affected, invalidating the comparison. This is a more likely to be a problem the smaller is the gap between implementation of phase 4 and the inflows chosen as treatment and control groups.</p> <p>2 <i>Past roll-out of UC for treated and control groups</i>: The inflow to what would be new ISc/ESA/HB claims under the legacy benefit regime can be affected by the UC implementation history both mechanically and through choice in response to the past changes in incentives. Since only time variation can be exploited for this phase, treated and comparison groups will have experienced a different roll-out history and, therefore, may be compositionally different.</p> <p>We can test for entry effects on observed variables by testing for the impact of UC on inflow as detailed in Appendix A.2.3.</p>
Limitations/issues	
Parameter of interest	The impact of UC against a mixed regime may not be as interesting as the impact of UC against the legacy benefit regime since the control group may find alternative ways of moving into the system (or may act to avoid it). However, the impact of UC versus the full legacy benefit regime can be measured only for as long as UC is not phased in for the comparison group.

Continued

Table C.6 Continued

External validity	<p><i>Entry effects:</i> Impacts assessed only for people who do not start a UC claim before phase 4 and who then make a new claim to UC that would otherwise have been to ISc/ESA/HB. This may be an unrepresentative group, particularly because some of the new claimants will already be under UC, having gained access through some earlier claim to another benefit (mechanical composition effects).</p> <p>Moreover, mechanical composition effects are also likely to affect the comparison group, depending on when it is they are drawn, potentially limiting the common support.</p> <p><i>Anticipation effects:</i> Comparisons across time may be affected by anticipation effects if the comparison group foresees the imminent change in incentives and responds by changing claiming behaviour.</p> <p>Outcomes for the comparison group may be affected by anticipation effects as they foresee the introduction of phase 4.</p> <p>Possibly less relevant at this stage.</p> <p><i>Short-run – teething and familiarity:</i> these are likely to become gradually less important.</p> <p><i>Volume:</i> the impact that can be measured is specific to the implementation strategy and does not embed the indirect effects that could take place once UC is fully implemented as coverage still very limited at this stage. Since indirect effects affect all, they may in particular change the impact of UC on claimants. Also becoming gradually less relevant.</p>
Internal validity	<p>Differential entry effects on unobserved variables may lead to biased estimates of the impact of UC during this phase.</p>
Data required	<p>Data allowing for the identification of all (or a random sample of) new claims to ISc/ESA/HB and equivalent UC claims over time; the longer the time coverage the more evidence can be brought to support or reject the DiD and matching assumptions.</p> <p>Data on labour market outcomes, labour market histories and background characteristics of claimants.</p>

Table C.7 Evaluation design to assess the impact of UC on benefit entitlement and labour market outcomes of potential future claimants during scenario phase 4

Sources of variation/ potential comparisons	Some areas will replace new ISc/ESA/HB claims with UC whilst the same areas in a previous period treat the new claim under the legacy benefit regime.
Outcome variables	Fulfilling UC entitlement rules for phase 4, claiming benefits, individual employment, household employment, working hours, earnings, family income and poverty measured in month n (>0) after roll-out of phase 4.
Evaluation question	Impact of the full UC regime versus the full legacy benefit regime. or Impact of the full UC regime versus a mixed regime. More specifically, impact of UC replacing new JSA/TC/IS/ESA/HB claims on the probability of becoming entitled to benefits, claiming benefits, or future labour market outcome versus the full legacy benefit regime or versus a mixed regime in which UC replaces only new JSA or JSA/TC/ISlp claims.
For whom	The entire working-age population, those not on benefits, new parents, lone parents, workers on low-paid jobs, etc.
Methodology	
Cross-sectional DiD (A)	Not available.
Cross-sectional DiD (B)	Time variation in treatment of new ISc/ESA/HB spells, within area (can identify the impact of the full UC regime versus the full legacy benefit regime or versus a mixed regime, depending on the time of inflow for the control group).
Treated	The entire working-age population or some sub-group of interest observed just before the start of phase 4.
Comparison	Matched individuals in the same areas during an earlier month, prior to the roll-out of phase 4.
Before period	Matched individuals in the same areas one year earlier.
Assumptions	<ol style="list-style-type: none"> 1 No differential aggregate trends for the treated and comparison cohorts (given observed characteristics). 2 ('Ashenfelter's dip') Invariant composition of transitory shocks by treatment status – meaning whether or not exposed to phase 4 (given observed characteristics). 3 Invariant composition of treated and comparison groups one year apart, within area (given observed characteristics).
Longitudinal DiD	DiD methodology (B) can also be implemented longitudinally, with the 'before period' being 12 months before the outcomes of treated and comparison groups are measured.
Assumptions	Assumption 3 can be dropped.
Matching (A)	Not available.
Matching (B)	Time variation in the treatment of ISc/ESA/HB or TC/IS/ESA/HB or JSA/TC/IS/ESA/HB claimants, within area.
Treated	All working-age population or some sub-group of interest observed just before the local introduction of phase 4.
Comparison	Matched individuals in the same areas observed some time before treated, prior to the roll-out of phase 4 and, possibly, prior to the local roll-out of UC.
Assumptions	<ol style="list-style-type: none"> 1 <i>Conditional independence assumption (CIA)</i>: no differential composition between treated and control groups (given observed characteristics); 2 No differential aggregate outcomes/shocks by time for the group of interest (given observed characteristics). This effectively means no differential seasonal effects.

Continued

Evaluating the labour market impacts of Universal Credit: a feasibility study

Table C.7 Continued

For how long	Cleanest comparisons: for n months, until phase 4 is introduced for the control group. Comparability can be more difficult for earlier comparison groups.
Ancillary analyses	
Common trends	Historical evidence on common trends for the outcome variable of interest for the treated and control groups. Major differences in cyclical variation, especially close in time to the evaluation period, would suggest the groups are not fit for DiD analysis.
Common level of outcomes (matching)	The common aggregate shock assumption of matching can be checked historically, and indeed the areas should be matched on the history of outcomes.
Entry effects	<p>These are effects changing the composition of the stocks of potential future claimants at the start of Phase 4. Since this design is unconditional on claiming status, entry effects are unlikely to be a problem. But the same may not hold if restricting attention to those not on benefits as composition may depend on the past local history of roll-out and foreseen changes in incentives.</p> <p>Careful analysis of compositional differences between treated and control groups and how these change over time can provide suggestive evidence of the presence or absence of entry effects.</p>
Limitations/issues	
Parameter of interest	The definition of the estimated parameter changes with time as the roll-out of UC advances.
External validity	<p><i>Entry effects</i>: unlikely to change the composition of treated and comparison groups if unconditional on treatment status, but the same is not true when looking at impacts on those not on benefits. Moreover, may affect composition of those ending up making a new benefit claim under the new regime and, through that process, affect the estimated impact for the entire population or any sub-group of interest.</p> <p><i>Anticipation effects</i>: Comparisons across time may be affected by anticipation effects if the comparison group foresees an imminent change in incentives (for the comparison group this may amount to the start of any of the phases 1 to 4) and respond by changing behaviour. This behaviour is induced by the implementation strategy and, therefore, not externally valid.</p> <p><i>Short-run – teething and familiarity</i>: this problem is likely to become gradually less important.</p> <p><i>Volume</i>: small coverage means that indirect effects are unlikely to be of any relevance, even if strong in the longer run under full implementation of UC. This problem is likely to become gradually less important.</p>
Data required	<p>Data allowing one to identify all (or a large random sample of) the working age population or a sub-group of interest in October of Year 2 or earlier to allow or time comparisons and the testing of common trends, common outcomes and entry effects.</p> <p>Data on labour market outcomes, labour market histories and background characteristics of claimants.</p>

Table C.8 Evaluation design to assess the indirect effects of UC during scenario phase 4

Sources of variation/ potential comparisons	Some areas will replace new ISc/ESA/HB claims with UC whilst the same areas in a previous period treat the new claim under the legacy benefit regime.
Outcome variables	Individual employment, household employment, working hours, earnings, family income and poverty measured in month n (>0) after the roll-out of phase 4.
Evaluation question	Impact of UC replacing all new claims versus a hybrid regime where UC has replaced (i) only new JSA/TC/ISlp claims or (ii) only new JSA claims on individuals not directly affected by the reform (i.e. by phase 4). Measures only indirect effects.
For whom	Those who started a UC claim (to what would have been JSA/TC/ISlp) before the roll-out of phase 4. Among them, sub-groups of interest, e.g. those out-of-work or workers on low-paid-jobs.
Methodology	
Matching (B)	Time variation in the treatment of new ISc/ESA/HB claimants, within area.
Effect (i)	Indirect impacts of UC replacing all new claims versus a hybrid regime where UC has replaced only new JSA/TC/ISlp claims for those who started a UC claim (to what would have been JSA/TC/ISlp) close to the end of phase 3.
Treated (i)	The inflow into UC near the end of phase 3.
Comparison (i)	The matched inflow into UC in the same areas, who started their UC claim just after the start of phase 3.
Effect (ii)	Indirect impacts of UC replacing all new claims versus a hybrid regime where UC has replaced only new JSA claims for those who started a UC claim (to what would have been JSA) close to the end of phase 2.
Treated (ii)	The inflow into UC near the end of phase 2.
Comparison (ii)	The matched inflow into UC in the same areas, who started their UC claim just after the start of phase 2.
Assumptions	1 <i>Conditional independence assumption (CIA)</i> : no differential composition between treated and control groups (given observed characteristics). 2 No differential aggregate outcomes/shocks by time for the inflow into UC (i) during phase 3 or (ii) during phase 2 (given observed characteristics). This effectively means no differential seasonal effects.
For how long	Outcomes assessed, e.g. monthly, from the first month after implementation of phase 4 onwards. Impacts can be assessed for five months (the time frame between the start of phase 3 and the start of phase 4 when using comparison group (i); and the longest evaluation horizon when using comparison group (ii)).
Ancillary analyses	
Common level of outcomes (matching)	The common aggregate outcome/shock assumption of matching (B) can be checked historically.
Entry effects	Since this design applies to individuals not directly affected by phase 4, entry effects are unlikely.
Limitations/issues	
Parameter of interest	We can measure the indirect effects of UC replacing all new claims only versus a hybrid regime and for a maximum of five months.

Continued

Table C.8 Continued

External validity	<p><i>Entry and anticipation effects:</i> unlikely to change the composition of treated and comparison groups or their behaviour. May however, affect the composition of those ending up making a new benefit claim under the new regime in phase 4 and, through that process, affect the estimated impact for reference population.</p> <p><i>Short-run – teething and familiarity:</i> this problem is likely to become gradually less important.</p> <p><i>Volume:</i> With one-month roll-out and full coverage of all new claims, low volume of UC cases might be less of an issue.</p>
Data required	<p>Data allowing one to identify all (or a large random sample of) the inflow into UC prior to the introduction of phase 4.</p> <p>Data on labour market outcomes, labour market histories and background characteristics of claimants.</p>

Table C.9 Evaluation design to assess the impact of UC on the outcomes of legacy benefit claimants migrated onto UC

Sources of variation/ potential comparisons	Some claims will be moved into UC at a point in time while similar claims in the same area will remain legacy benefit claims for at least one additional month.
Outcome variables	Continuation of claim, individual employment rate, household employment rate, working hours, earnings, family income and poverty, n (>0) months after migration.
Evaluation question (A)	<p>Impact of migrating a long-standing claim to UC in month m versus not migrating in month m (and eventually migrating later if the claim remains open).</p> <p>This comparison is within a policy regime characterised by two features that are important in interpreting any estimated treatment effect:</p> <ul style="list-style-type: none"> • non-treated claimants will in any case move into UC if they take any action leading to a new benefit claim, or at the latest when they reach their migration date; • the migration happens with transitional protection, whereby benefits are not allowed to fall in cash terms as a result of the migration.
For whom	Legacy benefit claims that start under the legacy benefit regime and survive for at least a number of months (e.g. nine months for JSA claims, depending on area) to reach the migration date.
Evaluation question (B)	<p>Impact of notifying a claimant in month m of migration to happen in n months' time if claim remains open versus not notifying in month m (and eventually notifying later if the claim remains open).</p> <p>This comparison is within a policy regime characterised by two features that are important in interpreting any estimated treatment effect:</p> <ul style="list-style-type: none"> • non-treated claimants will in any case move into UC if they take any action leading to a new benefit claim, or at the latest when they reach their migration date; • the migration happens with transitional protection, whereby benefits are not allowed to fall in cash terms as a result of the migration.

Continued

Table C.9 Continued

For whom	Legacy benefit claims that start under the legacy benefit regime and survive for at least a number of months (e.g. six months for JSA claims, depending on area) to reach the notification date.
Methodology	
Matching methodology	Same claimant type treated differently within area at time m .
Treated (A)	Claimants migrated to the full UC regime during month m .
Comparison (A)	Matched claimants of the same type in the same area not yet migrated in month m .
Treated (B)	Claimants notified of migration forthcoming in n months' time during month m .
Comparison (B)	Matched claimants of the same type in the same area not yet notified in month m .
Assumptions	<i>Conditional independence assumption (CIA)</i> : no differential composition between treated and control groups (given observed characteristics); applies to both (A) and (B).
For how long	Unlimited, but short-term impacts more likely to reveal differences in outcomes induced by migration or notification: e.g. for JSA, within six months all claims will be migrated/notified, and any differences in outcomes between those migrated earlier and those migrated later will presumably fade thereafter.
Limitations/issues	
Parameter of interest	<p>In what follows, we discuss the impact of migration, but it can all be applied straightforwardly to the impact of notification.</p> <p>This approach identifies the impact of migrating an ongoing claim into the full UC system at a given point in time versus postponing migration. It is more the impact of the timing of migration than the impact of UC.</p> <p>All migration is to be completed in a given time period (six months for JSA). Furthermore, any new claims by the control group will also be to UC, possibly before they would be moved into UC through managed migration. Therefore, the control group will quickly be moved into UC itself, possibly implying any impacts may be short-lived, fading quickly over time.</p> <p>Transitional protection affects the nature of the treatment effect parameter. It may mean that incentives remain unaltered for those being migrated, at least in the short-run. With transitional protection, the estimated effect is the average effect of 'migrating a long-standing claim at time m, keeping the award constant if UC implies a drop, versus not migrating at time m'.</p>
External validity	Impacts assessed only for long-term claimants. These are potentially disproportionately composed of those not gaining from the new system.
Other issues	For evaluation question (A) only, the impact of migration (rather than notification): At treatment time m , treated and control groups will have been exposed to systematically different information about their forthcoming migration (always earlier for the treated) and for different periods of time (treated will have known for longer). To the extent that individuals anticipate potential benefits/losses associated with the migration, they may respond to being notified, in advance of the actual migration. Such behavioural responses might lead to compositional differences between treated and comparison groups, which matching may not be able to eliminate.
Data required	
	<p>Data allowing for the identification of all (or a large random sample of) ongoing JSA claims including claim start dates, time of notification and time of migration.</p> <p>Data on labour market outcomes, labour market and claimant histories and background characteristics of claimants.</p>

Table C.10 Evaluation design to assess the indirect impacts of a large influx into UC

Sources of variation/ potential comparisons	In some areas a bulk of claims of a given type are migrated to UC, while comparable areas have not started the migration of that type of claim yet.
Outcome variables	Individual employment rate, household employment rate, working hours, earnings, family income and poverty, n (>0) months after the bulk of the stock has been migrated.
Evaluation question	Indirect effects (substitution and displacement) of moving a large number of individuals from a legacy benefit into UC on the outcomes of those not directly affected by the migration (but who are already subject to the full UC regime).
For whom	Relevant population (or subpopulation) of job seekers, workers or working-age population not on the stock of legacy benefits.
Methodology	
Matching methodology	Large influx into UC only in some areas.
Treated	Relevant population not on the legacy benefit stock in an area that migrates the (bulk of the) stock earlier.
Comparison	Matched individuals from the relevant population not on the legacy benefit stock in an area that starts the migration of the stock later.
Assumptions	<i>Conditional independence assumption (CIA)</i> : no differential composition between treated and control groups (given observed characteristics).
For how long	Limited by how soon after completion of the (bulk of) the migration in the treated area the control area starts migrating its own stock.
Limitations/issues	
Parameter of interest	This approach identifies the substitution and displacement effects of a large influx into UC for a population which is already subject to UC. It is the impact of a large and 'sudden' migration to UC rather than the impact of UC itself.
Data required	
Data allowing for the identification of all (or a large random sample of) the reference population (job seekers, workers or working-age population not part of the legacy benefit stock) in areas migrating the stock earlier and similar areas migrating it later.	
Data on labour market outcomes, labour market histories and background characteristics of the reference population.	
Historical data on local labour market outcomes.	

Appendix D

Exploiting the associated reform of Pension Credit

For very specific groups, one can think of exploiting other policy changes to infer the impact of Universal Credit (UC). One example is the planned change in entitlement to Pension Credit (PC) for couples. Under current rules, couples are eligible for PC if either partner is aged 60 or older. When UC is introduced, PC eligibility will be restricted to couples where both partners are aged 60 or older. This means that some couples with the elder partner turning 60 in the month preceding the reform will gain entitlement to PC at that time while similar couples with an elder partner one month younger will only become entitled to UC when the younger spouse reaches age 60.

In principle, one could use this variation to estimate the impact of UC versus starting on PC for couples close to retirement age who start a new PC claim the moment the oldest spouse reaches 60. This is a very specific population group, therefore, defining a very specific treatment effect.

There are some potential problems with this analysis:

- Data issues: is the information available to identify the flow into PC?
- Selection issues: PC is more generous than UC, and this introduces selection problems. A particular couple may decide to claim PC if entitled but would behave in a different way (labour market-wise) and not claim UC if PC not available. Put differently, the identification of the correct counterfactual is difficult because it requires (untestable) assumptions about the choices that couples claiming PC would make if only UC was available to them.

In the presence of sufficiently rich data to allow for the identification of UC and PC claims, one could consider further restricting the population studied to couples claiming UC in the month preceding the 60th birthday of the elder partner, and use the policy reform to assess the impact of UC on inflow into PC and subsequent labour market outcomes. Table 7 details the evaluation design.

Table C.11 Evaluation design to assess the impact of UC on the outcomes of older couples on benefits

Sources of variation/ potential comparisons	PC reform in month m : a couple in which the elder partner reaches 60 in month m (and where the younger partner is under 60) is not entitled to PC until both spouses turn 60, while a similar couple with a one-month-older elder partner becomes entitled to PC in month $m-1$.
Outcome variables	Claim status, individual employment rate, household employment rate, working hours, earnings, family income and poverty, n (≥ 0) months after the elder partner's 60th birthday.
Evaluation question	Impact of being entitled to UC at the time the elder partner turns 60 versus being entitled to PC at that time.
For whom	Couples claiming UC in the month preceding the 60th birthday of the oldest spouse and in which the younger partner is more than one month younger than the elder partner.
Methodology	
Matching methodology	Same UC claimant type treated differently over time in terms of PC eligibility.
Treated	Couples whose oldest spouse turns 60 in month m (after policy reform) claiming UC in month $m-1$.
Comparison	Matched couples whose oldest spouse turns 60 in month $m-1$ (before policy reform) claiming UC in month $m-2$.
Assumptions	<i>Conditional independence assumption (CIA)</i> : no differential composition between treated and control groups (given observed characteristics). This means that the composition of the UC claimant group among couples in which elder partner is 59 years and 11 months old is unaffected by the entitlement rules to PC, at least around the time of reform (again, conditional on observed characteristics).
For how long	Unlimited: impacts can be measured in the short- and long-run, data permitting, even after both spouses turn 60.
Limitations/issues	
Parameter of interest	What is estimated is the effect of an 'intention to treat', since not all of those entitled to UC at $m-1$ will claim it. The average effect of treatment on the treated may be recovered by a local instrumental variable method (LATE).
External validity	Impact assessed only for a very narrow population – albeit one that might be of interest – and may not be generalisable to other groups.
Anticipation effects	Having been announced in advance, the reform may induce responses in anticipation of the prospect of losing entitlement to PC and affect the composition of treated and control groups in ways that cannot be re-balanced using observed variables. More specifically, some people foreseeing not being entitled to PC in month m (when they would have been under the old regime) may change their labour market choices in advance and become ineligible to UC when the elder partner is 59 years and 11 months old – therefore, not being found among the treated group, but being found among the controls.
Data required	Data allowing for the identification of UC and PC claims, including the characteristics of the partner and the family. Data on labour market and claim outcomes, labour market and claiming histories and demographic characteristics of claimants.

References

- Bell, B., Blundell, R. and Van Reenen, J. (1999), 'Getting the Unemployed Back to Work: An Evaluation of the New Deal Proposals', *International Tax and Public Finance* 6: 339-360.
- Bertrand, M., Duflo, E. and Mullainathan, S. (2004), 'How Much Should We Trust Differences-in-Differences Estimates?', *Quarterly Journal of Economics* 119: 249-275.
- Blundell, R., Dearden, L., and Sianesi, B. (2005), 'Evaluating the Effect of Education on Earnings: Models, Methods and Results from the National Child Development Survey', *Journal of the Royal Statistical Society A*, 168, 3, 473-512.
- Blundell, R., Costa Dias, M., Meghir, C. and Van Reenen, J. (2004), 'Evaluating the employment effects of a mandatory job search program', *Journal of the European Economic Association* 2(4): 569-606.
- Cameron, C., Gelbach, J.B., and Miller, D.L. (2008), 'Bootstrap-Based Improvements for inference with clustered errors', *The Review of Economics and Statistics* 90(3): 414-427.
- Dehejia, R. and Wahba, S. (1999), 'Causal effects in non-experimental studies: re-evaluating the evaluation of training programs', *Journal of the American Statistical Association*, 94, 1053-1062.
- Dolton, P. and Smith, J. (2011), 'The Impact of the UK New Deal for Lone Parents on Benefit Receipt', IZA Discussion Paper No.5491.
- Froehlich, M. (2004), 'Programme evaluation with multiple treatments', *Journal of Economic Surveys*, 18, 181-224.
- Heckman, J., Ichimura, H., Smith, J. and Todd, P. (1998), 'Characterising Selection Bias Using Experimental Data.' *Econometrica* 66, 1017-1098.
- Heckman, J., Ichimura, H. and Todd, P. (1997), 'Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program.' *Review of Economic Studies*, 64: 605-654.
- Heckman, J., LaLonde, R. and Smith, J. (1999). 'The Economics and Econometrics of Active Labor Market Programs', in Orley Ashenfelter and David Card (eds.), *Handbook of Labor Economics, Volume 3A*, 1865-2097.
- Heckman, J. and Smith, J. (1999), 'The Pre-Programme Dip and the Determinants of Participation in a Social Programme: Implications for Simple Programme Evaluation Strategies.' *Economic Journal* 109, 313-348.
- Sianesi, B. (2013), 'Dealing with randomisation bias in a social experiment exploiting the randomisation itself: The case of ERA', IFS Working Paper No.W13/1.