



Department
for Work &
Pensions



Universal Credit: In-Work Progression Randomised Controlled Trial

Impact Assessment

September 2018

Research Report 966

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

Crown copyright 2018.

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>

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

First published September 2018.

ISBN 978-1-5286-0789-6

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

Executive summary

With the roll out of Universal Credit (UC) the Department for Work and Pensions (DWP) will, for the first time, work with individuals to help them stay in employment and, where appropriate, support and encourage them to increase their earnings. DWP is taking a 'Test and Learn' approach to developing its offer for in-work UC claimants. The In-Work Progression (IWP) Randomised Controlled Trial (RCT), which ran between April 2015 and March 2018, is the first major piece of evidence-building work to support the development of effective in-work services.

The IWP trial had three groups – two treatment groups and a comparison group – which received different levels of intervention:

- Frequent support group – Work Coach support and compliance checking through fortnightly Work Search Reviews (WSRs);
- Moderate support group – Work Coach support and compliance checking through eight-weekly WSRs;
- Minimal support group (the comparison group) – two 'light touch' telephone interviews (one on entry to the trial, the other eight weeks afterwards). These appointments were mandatory, although any actions resulting were voluntary.

The trial aimed to test whether increased Work Coach support and applying conditionality drove behaviours that led to earnings progression. This report concentrates on the quantitative impacts of the trial to see if there were any statistically significant differences between the three IWP groups. It is a more technical account of the summary report¹ and also supports the external evaluation report.² The impact analysis pre-specified two progression measures of interest:

- the earnings impact for each participant in each of the three groups at 52 weeks since their trial start date;
- the percentage of participants who have seen earnings progression of at least 10 per cent in each of the three groups at 52 weeks since their trial start date.

Full sample impact analysis results

We detected a small and positive statistically significant monetary progression impact for the Frequent support group versus the Minimal support group (the comparison group), and for the Moderate support group versus the Minimal support group at 52 weeks after starting the trial. This amounted to an additional progression of £5.25 per week for the Frequent support group and £4.43 per week for the Moderate support group compared to the Minimal support group at 52 weeks.

For participants who increased their earnings by at least 10 per cent since starting the trial, we detected small and positive statistically significant progression impacts for both the Frequent support group and the Moderate support group versus the Minimal support group at 52 weeks after starting the trial (an additional 2.90 and

¹ The Summary report can be found here <https://www.gov.uk/government/publications/universal-credit-in-work-progression-randomised-controlled-trial>

² The external evaluation conducted by Ipsos MORI can be found here <https://www.gov.uk/government/publications/universal-credit-in-work-progression-randomised-controlled-trial>

2.42 percentage points respectively). The impact for the Frequent support group occurs around week 12 and is sustained. The impact for the Moderate support group is sustained from around week 30.

Subgroups analysis

We investigated IWP groups by gender, age, Jobcentre type (live or full service) and region (England, Scotland and Wales). As with the full sample results, both the Frequent and Moderate support groups experienced small, positive earnings progression impacts over the Minimal support group within the subgroups.

The subgroup analysis did not find significant differences *between* subgroup impacts in general, indicating that the trial produced comparable, small and positive effects for all subgroups. The one exception indicated that at 52 weeks, participants in the 18-24 years subgroup saw a greater progression of £14.57 on average than those participants in the 25-34 years subgroup (for the Moderate support group versus the Minimal support group), which was statistically significant.

Sanctions analysis

We also analysed participants on the trial in Live Service who had been sanctioned. We excluded any sanction applied when the participant was not in the light touch conditionality regime since the sanction was not during the time they were on the trial.

The overall proportion of participants sanctioned across all three IWP groups over the course of the trial was 2.4 per cent. This value is not directly comparable with official statistics since different methodologies have been used: we produced one overall rate as opposed to monthly point-in-time measures; furthermore, the official statistics do not provide a sanction rate for the light touch conditionality regime.

Across all three IWP groups, failure to attend an interview with a Work Coach was the most common reason for being sanctioned. This accounted for 91.2 per cent of sanctions applied (or 443 sanctions). Consequently, the majority of sanctions applied were low-level sanctions, 91.6 per cent (or 445 sanctions).

The Minimal support (comparison) group which was subject to the least number of interviews received the lowest rate of sanction (1.5 per cent). Both the Frequent and Moderate support groups had higher sanction rates compared to the Minimal support group (3.1 per cent and 2.6 per cent respectively) and the differences compared to the Minimal support group were statistically significant.

Furthermore, the Frequent support group had a higher proportion of sanctions than the Moderate support group, but this difference was not statistically significant.

Recommendation

The results to date show a small statistical difference at the end of 52 weeks from the trial start date. We therefore recommend that we continue to track performance beyond the 52-week point to assess whether there is further impact of the intervention.

Contents

- Executive summary 3
- List of figures 7
- List of tables..... 9
- 1 Policy design and Trial delivery 11
 - 1.1 Background 11
 - 1.1.1 Policy 11
 - 1.1.2 Trial Design 12
 - 1.2 Trial Implementation 13
 - 1.2.1 The Compliance Function 13
 - 1.3 Trial Delivery..... 13
- 2 Impact analysis 15
 - 2.1 Methodology..... 15
 - 2.1.1 Data Sources 15
 - 2.1.2 Trial group randomisation..... 15
 - 2.1.3 Data matching..... 15
 - 2.1.4 Coverage of the sample used for the impact assessment..... 16
 - 2.1.5 Trial group balance..... 18
 - 2.1.6 Trial intervention appointments..... 18
 - 2.1.7 Earnings progression measures..... 18
 - 2.2 Results 19
 - 2.2.1 Content 19
 - 2.2.2 Progression at 52 weeks..... 19
 - 2.2.3 Progression at 26 and 78 weeks 27

Universal Credit: In-Work Progression Randomised Controlled Trial

- 3 Sanctions..... 29
 - 3.1 Methodology..... 29
 - 3.1.1 Previous sanctions analysis 29
 - 3.1.2 Data sources..... 29
 - 3.1.3 Data matching..... 29
 - 3.2 Results 30
 - 3.2.1 Total number of sanctions, levels and reasons..... 30
 - 3.2.2 Which IWP groups are more likely to be sanctioned?31
- 4 Conclusions and recommendations 36
- Annexes..... 38
 - Annex A: Trial group balance..... 38
 - Annex B: Frequency of intervention appointments..... 43
 - DWP administrative data measures..... 43
 - Further analysis of intervention appointments 44
 - Annex C: Methodology for earnings progression.....47
 - Earnings progression measure (£).....47
 - Earnings progression measure (per cent)..... 48
 - Earnings progression for subgroups..... 48
 - Statistical considerations 49
 - Annex D: Progression at 52 weeks impact assessment 50
 - Demographics 50
 - Further results51
 - Subgroup results 55
 - Annex E: Progression at 26 and 78 weeks61
- References 69

List of figures

Figure 1.1: Allocation to randomised control trial arms..... 12

Figure 2.1: Cumulative distribution of trial duration (weeks) for trial participants 17

Figure 2.2: Average earnings progression (£) with respect to trial start (week 0) 20

Figure 2.3: The difference (£) between the IWP treatment groups (Frequent and Moderate support groups) and the comparison group (Minimal support group)21

Figure 2.4: Percentage of trial participants who have increased earnings by at least 10 per cent since trial start (week 0)..... 23

Figure 2.5: The percentage point difference between IWP treatments (Frequent and Moderate support groups) and the comparison group (Minimal support group) for participants that have increased their earnings by at least 10 per cent since trial start (week 0). 23

Figure 2.6: Number of employment spells by IWP group..... 25

Figure 3.1: Average sanctions by IWP group 32

Figure 3.2: Average sanctions by IWP group and gender..... 32

Figure 3.3: Average sanctions for all participants with children 33

Figure 3.4: Average sanctions for all participants with children by gender..... 33

Figure 3.5: Average sanctions and failed to attend (FTA) by IWP group..... 34

Figure D1: Average weekly earnings for the three IWP groups over time51

Figure D2: Percentage of trial participants in employment over time for the three IWP groups 52

Figure D3: The percentage of trial participants that have decreased earnings by at least 10 per cent since trial start 52

Figure D4: Percentage of trial participants that have decreased earnings by at least 10 per cent since trial start – difference from the Minimal support group (comparison group) 53

Figure D5: Survival curves for first employment spells 53

Figure D6: Overall value of earnings for the three groups over time 54

Universal Credit: In-Work Progression Randomised Controlled Trial

Figure D7: Overall value of earnings for the three groups over time 54

Figure E1: Average earnings progression (£) to 26 weeks (sample size of 31,471).... 62

Figure E2: Average earnings progression (£) to 78 weeks (sample size of 12,706) ... 63

Figure E3: The difference (£) between IWP treatment groups (Frequent and Moderate) and the comparison group (Minimal) to 26 weeks 63

Figure E4: The difference (£) between IWP treatment groups (Frequent and Moderate) and the comparison group (Minimal) to 78 weeks 64

Figure E5: Percentage of trial participants who have increased earnings by at least 10 per cent since trial start (week 0) to week 26 (sample size of 31,471) 64

Figure E6: Percentage of trial participants who have increased earnings by at least 10 per cent since trial start (week 0) to week 78 (sample size of 12,706)..... 65

Figure E7: The percentage point difference between IWP treatment groups (Frequent and Moderate) and the comparison group (Minimal) for participants that have increased their earnings by at least 10 per cent since trial start (week 0) to week 26 65

Figure E8: The percentage point difference between IWP treatment groups (Frequent and Moderate) and the comparison group (Minimal) for participants that have increased their earnings by at least 10 per cent since trial start (week 0) to week 78 66

Figure E9: Percentage of trial participants who have decreased earnings by at least 10 per cent since trial start (week 0) to week 26 (sample size of 31,471) 67

Figure E10: Percentage of trial participants who have decreased earnings by at least 10 per cent since trial start (week 0) to week 78 (sample size of 12,706)..... 67

Figure E11: The percentage point difference between IWP treatment groups (Frequent and Moderate) and the comparison group (Minimal) for participants that have decreased their earnings by at least 10 per cent since trial start (week 0) to week 26 68

Figure E12: The percentage point difference between IWP treatment groups (Frequent and Moderate) and the comparison group (Minimal) for participants that have decreased their earnings by at least 10 per cent since trial start (week 0) to week 78 68

List of tables

Table 2.1: Reasons for sample attrition and number remaining	16
Table 2.2: Final sample sizes used for the impact assessment	17
Table 2.3: Significant differences (£) between IWP group findings – full sample participants in trial 52 weeks after random assignment (trial start)	19
Table 2.4: Significant differences (per cent) between IWP group findings – full sample participants in trial 52 weeks after random assignment (trial start).....	22
Table 2.5: Proportion of participants within each IWP group who had continuous employment over 52 weeks	25
Table 3.1: Levels of sanction	30
Table 3.2: Reason for sanction	31
Table 3.3: Failed to attend an interview	31
Table A1: Gender breakdown and comparison between IWP groups	38
Table A2: Age breakdown and comparison between IWP groups	39
Table A3: Jobcentre type breakdown and comparison between IWP groups	39
Table A4: Partner breakdown and comparison between IWP groups	39
Table A5: Partner on trial breakdown and comparison between IWP groups	40
Table A6: Region breakdown and comparison between IWP groups	41
Table B1: Relative difference in the number of interventions participants received....	43
Table B2: Time between appointments.....	44
Table B3: Statistics for inter-appointment time	44
Table B4: Appointments in two six-month periods	45
Table B5: Comparing average number of appointments in two six-month periods	46
Table D1: Participants in the In-Work Progression Trial by gender	50
Table D2: Participants in the In-Work Progression Trial by age	50

Universal Credit: In-Work Progression Randomised Controlled Trial

Table D3: Participants in the In-Work Progression Trial by Jobcentre type..... 50

Table D4: Participants in the In-Work Progression Trial by region51

Table D5: Average number of weeks in work by IWP group 55

Table D6: Average employment spell duration in weeks by IWP group 55

Table D7: Average employment spell duration in weeks by IWP group for two or more employment spells 55

Table D8: Significant differences between gender subgroup findings – male and female participants in the In-Work Progression trial 52 Weeks after random assignment (trial start) 57

Table D9: Significant differences between age subgroup findings – age of participants in the In-Work Progression trial 52 Weeks after random assignment (trial start)..... 58

Table D10: Significant differences between live and full service subgroup findings – live and full service participants in the In-Work Progression trial 52 Weeks after random assignment (trial start) 59

Table D11: Significant differences between geography subgroup findings – England, Scotland and Wales participants in the In-Work Progression trial 52 Weeks after random assignment (trial start) 60

1 Policy design and Trial delivery

1.1 Background

Once Universal Credit (UC) is fully rolled out, there will be around seven million households in receipt of UC, of which around three million will be in work.

This will be new territory for government, and a radically different context in which Department for Work and Pensions' (DWP) employment services will operate. For the first time, DWP can work with individuals to help them stay in employment and, where appropriate, support and encourage them to increase their earnings.

This section outlines the design and delivery of the In-Work Progression (IWP) trial as well as the monitoring and compliance which ran throughout the duration of the trial to ensure this was being delivered in line with policy expectations. More detail on the implementation and delivery can be found in the external evaluation report.³

1.1.1 Policy

This trial was a first step in a programme of wider testing to understand 'what works' in terms of helping support the progression of those in low pay. DWP wants to build on this trial, and the Autumn Budget 2017 committed £8 million over four years from 2018-19 to further develop the evidence base on progression.

For the purposes of this trial, DWP are testing the effect an active labour market regime has on earnings for claimants who are in low-paid work or low-income households and to what extent it:

- embeds the expectation that claimants take reasonable steps to increase their earnings in return for the support on offer;
- gives a clear understanding of what is required from claimants, regular engagement with a Work Coach and delivery of agreed actions in an individually tailored Claimant Commitment;
- coaches claimants to have conversations with their current employer, where possible, about opportunities for more, or better paid, work and where appropriate, look at wider opportunities for earnings progression;
- identifies barriers to progression, such as confidence and motivation, skills, or childcare, and directs them to support available;
- provides supportive but challenging conversations to help guide, steer and motivate claimants to realise their potential and free themselves from benefit dependency.

Undertaking research and evaluation of our policies enables us to understand more about what works as we continue to change and improve our services, ensuring we can help as many people as possible. This is especially the case when evidence is limited, as with IWP.

³ The external evaluation conducted by Ipsos MORI can be found here <https://www.gov.uk/government/publications/universal-credit-in-work-progression-randomised-controlled-trial>

1.1.2 Trial Design

The RCT was designed as a *three-arm trial*; this gave DWP the flexibility to not only test our main regime with in-work claimants, but understand the impact that varied degrees of support and conditionality might bring. Having this variability across the groups allowed DWP to consider the optimal level of support needed to help those in work to increase their earnings. This will ensure the development of any future in-work service would deliver value for money. Lack of evidence around the types of intervention which may assist individuals to increase their earnings was another important factor in the desire to test various scenarios to develop the evidence base further.

Although the primary measure of impact is an increase in earnings, DWP are also interested in intermediate/softer outcomes like whether this support influences behaviour so that people begin to look for alternative (better) work, increase their skills so they can compete for jobs in other sectors or even move into more sustainable jobs.

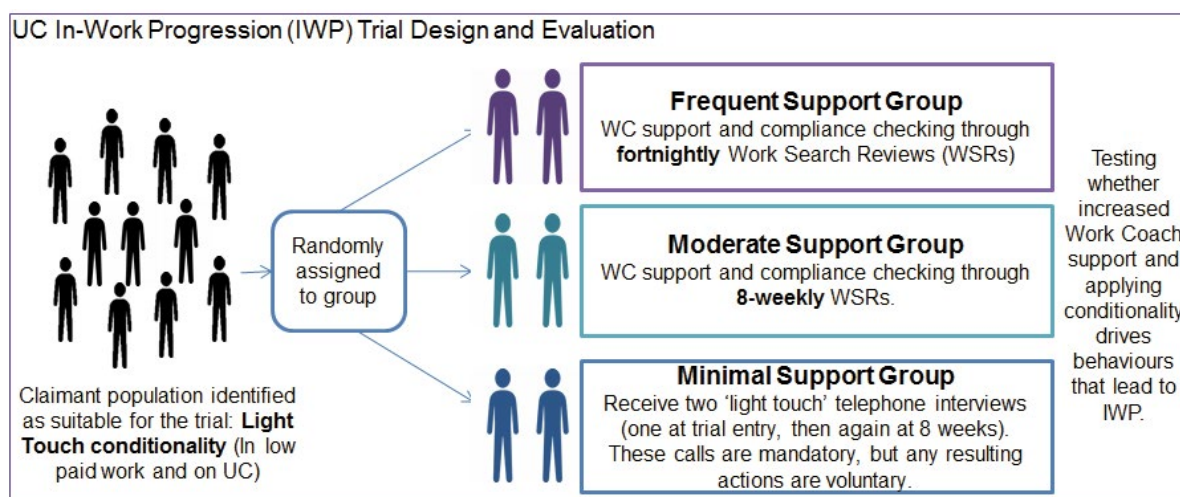


Figure 1.1: Allocation to randomised control trial arms

In parallel, a theory of change around the design of the trial intervention was developed based on three key assumptions, namely that:

- i) One-to-one support from a Work Coach, coupled with increased conditionality, may encourage individuals in the same way that it does for those out of work. Work Coaches would use their time with in-work claimants to have a 'quality' conversation, tailored around the needs of the individual and using local labour market intelligence and knowledge of the employer to consider a strategy for progression.
- ii) Encouraging individuals, where appropriate, to approach their respective employers to discuss training opportunities, career pathways, increasing hours or promotion opportunities would actively assist in their progression.
- iii) Employers would value a proactive approach on the part of their employees; and even if there were no opportunities for immediate progression, they would look favourably upon those individuals who appeared motivated and committed to expanding their current job roles or progressing within the organisation.

1.2 Trial Implementation

Initially the trial was implemented in just ten areas of the country⁴, which largely mirrored the rollout of UC.

Staff across these areas were offered training and support through Learning & Development workshops, written guidance and hands-on support from the Integrity and Operational Support Managers (IOSMs) who were specifically put in place to provide assistance to the evaluation team in monitoring the trial and providing an operational compliance function.

In December 2015, the trial began to roll out to more jobcentres across the country⁵ and soon became a national trial across both live and full service jobcentres. Recruitment onto the trial ended in March 2017, with the delivery of the interventions ending on 31 March 2018.

1.2.1 The Compliance Function

Much of the rigor of trials as a way of testing the effectiveness of a given intervention is attributed to the compliance processes built in to the implementation and delivery of an RCT. The whole RCT design was therefore predicated on the concept that the delivery of a trial intervention, as far as possible, replicated the design intent. The only way to be assured of high fidelity was through compliance monitoring.

Having a dedicated compliance function, to monitor and challenge non-compliance, was crucial in delivering ongoing checks, assisting sites in delivering the RCT in line with expectations to maximise trial integrity and ensure robust findings.

During the development phase of the RCT, IOSMs roles were created for the first time within DWP to provide a compliance function as well as operational support and assistance; problem-solving technical issues and reaffirming the importance of compliance among jobcentre staff.

This function has included a small team, providing assurance visits and observing Work Coaches administering interventions; escalating concerns where non-compliance may jeopardise trial fidelity and communicating essential lessons and insight from interaction with operational colleagues. This has helped to feed into wider learning and build capability.

1.3 Trial Delivery

The claimant's trial journey started when their financial circumstances took them above the Administrative Earnings Threshold (AET) but below the Conditionality Earnings Threshold (CET), which triggers in-work claimants to be allocated to the 'Light touch' regime. The service centre is usually the first point of contact for claimants when their circumstances change, and this triggers action for Work Coaches within the claimant's local jobcentre to contact them directly to bring them

⁴ The jobcentres initially involved were Ashton, Bath, Hammersmith, Harrogate, Inverness, Oldham, Warrington, Wigan, Rugby, Shotton.

⁵ With the exception of two areas: Musselburgh and London Bridge.

Universal Credit: In-Work Progression Randomised Controlled Trial

onto the trial. They are randomly allocated to one of three groups, providing different degrees of support and conditionality.

- The **Frequent Support** group (also known as Group 2) – eight weeks after their initial appointment, claimants met with their Work Coach every fortnight to get support and review mandatory actions agreed in their claimant commitment.
- The **Moderate support** group (also known as Group 1) – claimants have the same set of requirements as the Frequent support group and access to work coaching, but they receive reviews every eight weeks, rather than every fortnight.
- The **Minimal Support** group (also known as Group 3) – this group received an initial telephone appointment to establish voluntary actions, and one follow up telephone appointment eight weeks later to consider progress.

The eligible claimant had an initial appointment at the jobcentre, where their Work Coach explained more about in-work progression and how the trial may help them – including an explanation of the mandatory aspects of the trial, completing a ‘claimant commitment’ as well as a short, voluntary baseline survey which gathered some basic metrics about the claimant’s current attitude to progression and any barriers they may face.

Although face-to-face contact was the ‘default mode’ of contact for Frequent and Moderate support groups, Work Coaches often used telephony as an alternative where it was impractical for claimants to attend in person due to work commitments or personal circumstances.

Trial delivery was focused around two key components – *Work Coach support and increased conditionality*. In terms of Work Coach support, the time spent with claimants aimed to focus on identifying barriers to progression, such as tackling motivation, confidence, and signposting them to additional support; and where appropriate to encourage dialogue with their employer around future opportunities.

Secondly, the increased conditionality was meant to embed the expectation that claimants take reasonable steps to increase their earnings in return for the support on offer. This was done through regular meetings and agreeing tailored actions through a ‘claimant commitment’, where both parties are assured that appropriate steps are discussed.

The more comprehensive IWP Evaluation report (Ipsos MORI, 2018) sets out more detail around delivery and the role of Work Coaches, and the help and support offered to claimants during the trial.

2 Impact analysis

This section reports upon the quantitative impact of the trial and presents the analysis that was performed. The analysis provides evidence as to whether there were any differences between the three In-Work Progression (IWP) Trial groups.

2.1 Methodology

2.1.1 Data Sources

The analysis was performed using data derived from Department for Work and Pensions (DWP) Universal Credit (UC) administrative databases. These databases provide details of:

- trial marker identification;
- characteristics of UC claimants;
- salaried earnings for UC claimants.

By matching trial participants' encrypted national insurance numbers to the administrative databases above, their earnings history could be tracked over time. Earnings analysis is based on taxable pay earnings data obtained from Her Majesty's Revenue and Customs (HMRC).

2.1.2 Trial group randomisation

Participants were recruited to the trial from 20th April 2015 to 31st March 2017. They were randomly assigned to one of the three IWP groups by using the last three digits of their national insurance number. The randomisation was 'hard-coded' into the operational system as a simple algorithm. We know, therefore, that the randomisation process worked reliably and robustly for anyone who started the trial, eliminating the possibility of interference with the trial group selection process that could bias the results.

2.1.3 Data matching

42,452 participants had been assigned a trial marker and therefore assigned to one of the three trial groups by a Work Coach. For various reasons outlined in the table below, we could not use all of these participants in the impact analysis. Table 2.1 below provides the reason for the sample attrition and number of sample cases dropped.

Reason for sample attrition	Number of participants dropped	Number remaining
		42,452
Participants didn't have a UC payment record	440	42,012
Participants not paid UC within month before trial start date	4,097	37,915
Participants' records could not be matched with DWP administrative data	521	37,394
Duplicated records	235	37,159
Participant with a missing trial start date	1	37,158
Participants for which no earnings data could be found	994	36,164
Out of period jobs: jobs with no pay in the period	429	35,735
Participants with trial start date earlier than 20th April 2015	17	35,718
Participants with no earnings on trial start date	4,010	31,708
Percentage	25.3%	74.7%

Table 2.1: Reasons for sample attrition and number remaining

From the table above, the main reasons for sample attrition were as follows:

- Participants were not paid UC within one month before the participant's trial start date. The main reasons for this included participants who were already working before their trial start date and earning too much to qualify for a UC payment and sanctions in place for failing to attend before starting the trial.
- Participants did not have earnings reported on their trial start date. This was because of non-working claimants brought onto the trial because they were partners of working claimants who were on the trial, or because of self-reported earners, either because of self-employment, or because the employer did not send earnings returns to HMRC.

Taking into account these exclusions, 31,708 trial participants could be used in the impact analysis.

2.1.4 Coverage of the sample used for the impact assessment

We undertook analysis at 26, 52 and 78 weeks (six, 12 and 18 months) since trial start, with the impact assessment at 52 weeks identified at the outset as our main focus and result. Table 2.2 below shows the sample sizes used to obtain statistical estimates at each of the time points – see final column. The final sample sizes used in the impact analysis were affected by: how old the cases were (dependent on when

participants joined the trial), and the number of cases that were trimmed from the sample to exclude outliers.

Impact assessment time point	Starting sample size	Number of cases with duration less than time point in first column	Sample size available for impact analysis	Number of cases trimmed	Final sample size used in the impact assessment
26 weeks	31,708	0 (0%)	31,708	237 (0.75%)	31,471
52 weeks	31,708	768 (2.4%)	30,940	231 (0.75%)	30,709
78 weeks	31,708	18,908 (59.6%)	12,800	94 (0.73%)	12,706

Table 2.2: Final sample sizes used for the impact assessment

The impact analysis at any of these time points only considers those participants whose duration on the trial is at least 26, 52 and 78 weeks. At the time of this analysis, all participants had been on the trial for at least 26 weeks – see third and fourth columns of Table 2.2. 768 participants (2.4 per cent) had not been on the trial for 52 weeks and the majority of participants (18,908, or 59.6 per cent) had not been on the trial for 78 weeks. Figure 2.1 further illustrates the overall age of cases for participants on the trial – there are three vertical reference lines at 26, 52 and 78 weeks – and is consistent with the percentages in the third column of Table 2.2. For example, at 52 weeks in Figure 2.1, 2.4 per cent of the sample (768 cases) had not reached 52 weeks of age.

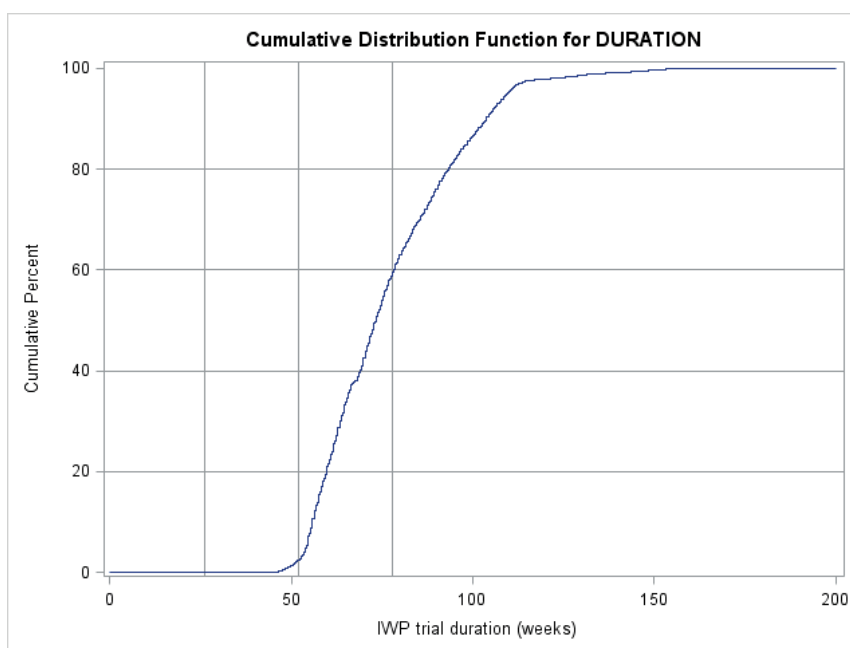


Figure 2.1: Cumulative distribution of trial duration (weeks) for trial participants

There was also attrition because of sample size trimming at each time point to exclude outlier values of reported earnings that could affect the statistical analysis for the impact assessment (shown in the fifth column of Table 2.2). We did not want to find significant differences between the IWP groups that were attributable to a few highly-paid individuals skewing the earnings progression distributions. Furthermore, other outlier values of reported earnings were attributable to data errors, which needed to be excluded. Trimming was kept to a minimum and the method applied

trimmed approximately 0.75 per cent of cases (spread evenly across the three IWP groups).

In summary, 31,471 cases (or participants) were used for the 26-week impact assessment, with 30,709 participants used for the 52-week impact assessment and 12,706 participants for the 78-week assessment. The consequence is that the impact analysis was run separately three times – one for each time point.

2.1.5 Trial group balance

An examination of the distributions of the information we have about the 31,471 participants in the three groups suggests that the groups are well-balanced. The difference between the mean values found for each group was less than 5 per cent of the standard deviation, for all variables tested. Please see Annex A for further information.

2.1.6 Trial intervention appointments

The frequency of appointments and support that participants in the IWP trial received throughout their time on the trial was investigated – see Annex B for the results. The main finding was that many participants received the right intensity of intervention (or appointments) according to their IWP group, but there was a fair amount of variation across the groups. The Frequent support group was defined to be the most intensive treatment group and participants in that group did indeed receive more appointments than participants in the other two groups.

2.1.7 Earnings progression measures

We defined our *earnings progression measures* as follows.

- A monetary earnings progression measure. We calculated an earnings change (progression) in £ GBP, whether positive (increase in earnings) or negative (loss of earnings) for *each* participant in each of the three groups over a given time interval (at 26, 52 and 78 weeks) since their trial start date. In advance of any analysis, we selected the main measure to be at the 52 weeks' time point, with the subsidiary measures at 26 and 78 weeks as previously stated in Section 2.1.4. This is statistically valid and produced unbiased mean progression estimates for each of the groups that were compared statistically (i.e. the impact).
- A percentage earnings progression measure. We calculated the percentage of participants who saw an earnings progression of at least + 10 per cent over a given time interval (at 26, 52 and 78 weeks) since their trial start date. Again, we could compare the three treatment groups statistically to determine the impact.

The method is outlined fully in Annex C.

2.2 Results

2.2.1 Content

This section summarises the results and findings from the IWP impact assessment. We focus on the results of the impact assessment at 52 weeks. We present the following analysis:

- the differences in average progression between IWP groups in £ GBP at 52 weeks from the time participants started the trial (that is, the impact);
- impacts for participants who have increased their earnings by at least 10 per cent since starting the trial;
- number of employment spells in the 52-week period – one employment spell is an unbroken period of employment;
- subgroup impacts for both progression measures.

Not all results have been included here. Annex D provides more breakdowns of the results of the progression at 52 weeks.

2.2.2 Progression at 52 weeks

Demographics

Annex D provides some demographic breakdowns of the sample for gender, age, jobcentre type (whether Live or Full Service) and geography.

The following two sections provide the results of the impact analysis in line with Section 2.1.7 where we defined the two progression measures.

i) Earnings Progression (£) at 52 weeks after random assignment (trial start)

The following table shows the full sample earnings progression of participants in £ GBP for the three IWP groups 52 weeks after random assignment and the resulting impacts.

Outcome	Frequent	Moderate	Minimal	Comparison	Impact	P-Value
<u>Earnings progression (£)</u>						
Full sample				Freq vs Mod	0.81	0.698
	10.44	9.63	5.20	Freq vs Min	5.25**	0.009
				Mod vs Min	4.43*	0.029

Table 2.3: Significant differences (£) between IWP group findings – full sample participants in trial 52 weeks after random assignment (trial start)

Source: Findings were based on Real Time Information (RTI) earnings data.

Notes: Statistical significance levels for the full sample are indicated as: *** = 0.1%; ** = 1%; * = 5%.

Universal Credit: In-Work Progression Randomised Controlled Trial

From Table 2.3 there are statistically different positive impacts for the two treatment groups (the Frequent and Moderate support groups) compared to the Minimal support group (the comparison group). At 52 weeks after trial start, the Frequent support group progressed £5.25 more on average than the Minimal support group in that week, i.e., £10.44 minus £5.20 ($p=0.009$). The Moderate support group progressed £4.43 more on average than The Minimal support group in that week, ($p=0.029$).

Another way to view the results above is in terms of confidence intervals, which offer further insight about statistical significance. Since the statistical question of interest is whether the average progressions differ between the IWP groups, the relevant confidence interval is for the difference in average progressions. For example, regarding the impact of sample members in the Frequent support group versus the Minimal support group, the difference in average progression is £5.25 ($p=0.009$) as stated. The 95 per cent confidence interval ranges from £1.29 to £9.17 (not shown). Although the full confidence interval is above zero – which is consistent with there being a significant difference between the IWP groups (or positive impact) – the confidence interval reveals uncertainty regarding the true difference in effects between groups. This uncertainty is around 75 per cent either side of the central estimate, which indicates that whilst the effect is positive, it's not particularly large.

We performed analysis to see whether the observed impacts in Table 2.3 were apparent before the 52-week point. To that end, the following chart shows the average earnings progression of participants in £ GBP for the three IWP groups as a function of time before or after the trial start date.

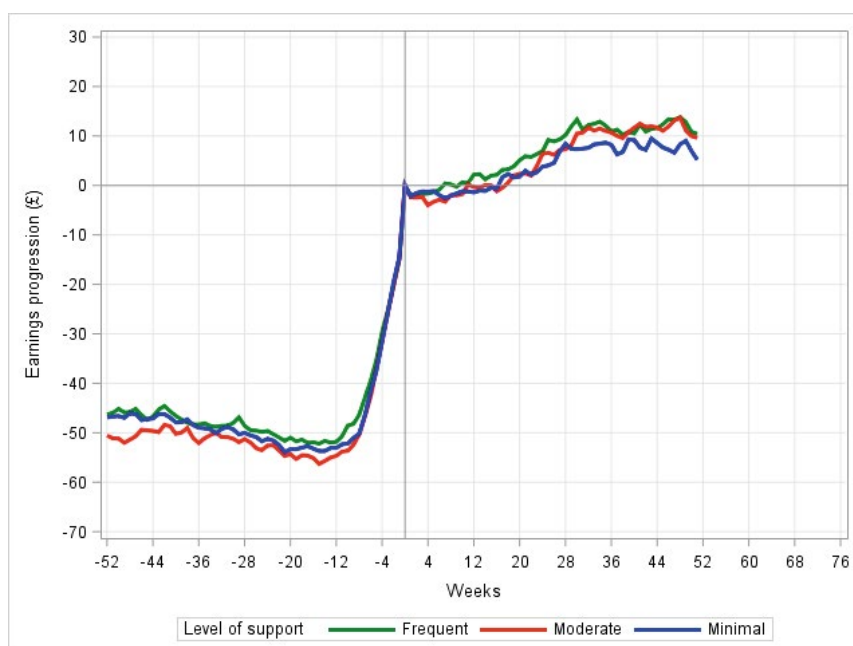


Figure 2.2: Average earnings progression (£) with respect to trial start (week 0)

From this, it appears that there are positive impacts for both treatment groups (the Frequent and Moderate support groups) versus the Minimal support group (comparison group) from week 30 onwards. To assess whether these impacts might be significant, it is more helpful to plot the impacts for the two treatment group lines together with their confidence intervals, and this is done in Figure 2.3.

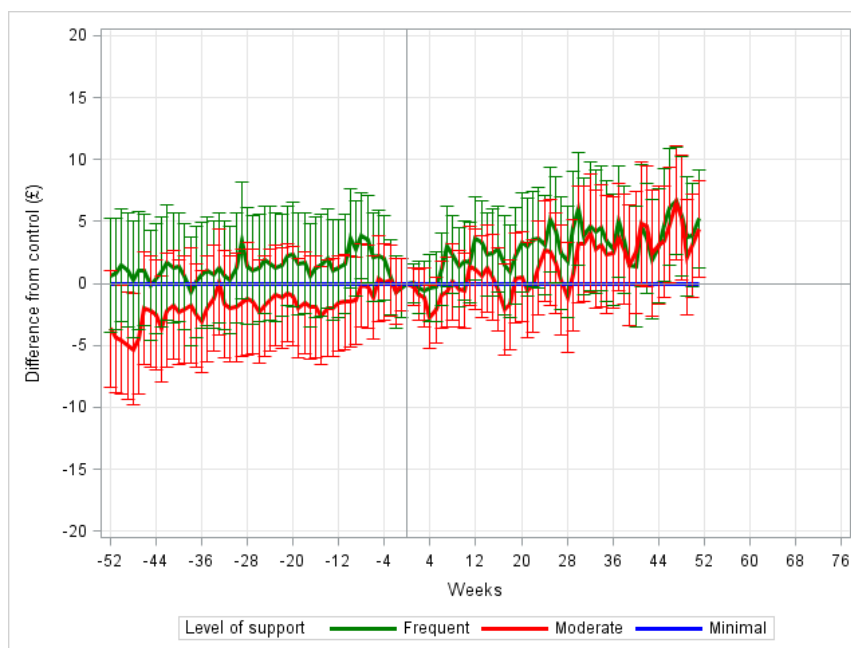


Figure 2.3: The difference (£) between the IWP treatment groups (Frequent and Moderate support groups) and the comparison group (Minimal support group)

Noting that there are dangers regarding violation of independence of multiple tests over time, we are looking to see in the chart above if the confidence intervals for the two IWP treatment groups (the Frequent and Moderate support groups) don't include the comparison group (the Minimal support group), indicating significant differences. Figure 2.3 suggests that the impact for the Frequent support group versus the Minimal support group, and the impact for the Moderate group versus the Minimal group, have a slight upward trend over time. However, the two treatment groups do not appear to be significantly different to the Minimal support group in the main over time, with differences only becoming apparent towards the end of the 52-week period. Further analysis is required from the 52-week point to see if the statistically significant impacts observed at 52 weeks from trial start are sustained over time.

In summary, we detected small and positive statistically significant **monetary** progression impacts for the Frequent support group versus the Minimal group, and for the Moderate group versus the Minimal group at 52 weeks after trial start (£5.25 and £4.43 progression in that week respectively). The sample needs to be tracked for longer to see if this is the start of a sustained impact.

ii) Earnings Progression (per cent) at 52 weeks after random assignment (trial start)

The following table shows the full sample earnings progression for participants who had increased their earnings by at least 10 per cent for the three IWP groups 52 weeks after random assignment and the resulting impacts.

Outcome	Frequent	Moderate	Minimal	Comparison	Impact	P-Value
<u>Earnings progression (%)</u>						
Full sample				Freq vs Mod	0.48	0.497
	45.37	44.89	42.47	Freq vs Min	2.90***	<0.001
				Mod vs Min	2.42***	0.001

Table 2.4: Significant differences (%) between IWP group findings – full sample participants in trial 52 weeks after random assignment (trial start)

Source: Findings were based on Real Time Information (RTI) earnings data.

Notes: Statistical significance levels for the full sample are indicated as: *** = 0.1%; ** = 1%; * = 5%.

From Table 2.4 it is clear that there are statistically significant small and positive impacts for the two treatment groups (the Frequent and Moderate support groups) versus the Minimal support group (comparison group). At 52 weeks after trial start, the Frequent support group is 2.90 percentage points higher than the Minimal support group, i.e., 45.37 per cent – 42.47 per cent ($p < 0.001$). That is, at 52 weeks, the Frequent support group had a greater percentage of participants than the Minimal support group that had increased their earnings by at least 10 per cent since starting the trial. The Moderate support group is 2.42 percentage points higher than the Minimal support group ($p = 0.001$).

Understanding these results in terms of confidence intervals offers further insight about statistical significance. For example, the difference between the means of the Frequent support group and the Minimal support group, (impact) is 2.90 per cent ($p < 0.001$) as stated. The 95 per cent confidence interval ranges from 1.54 per cent to 4.28 per cent. Although the full sample confidence interval is above zero – which is consistent with there being a significant difference between the IWP groups – the confidence interval once again reveals some uncertainty regarding the true difference in effects between groups. This uncertainty is around 45 per cent either side of the central estimate, indicating that whilst the effect is positive, it is not particularly large.

We performed analysis to see whether the observed impacts in Table 2.4 were apparent before the 52-week point. To that end, the following chart shows the percentage of participants who have increased their earnings by at least 10 per cent for the three IWP groups as a function of time before or after trial the trial start date.

Universal Credit: In-Work Progression Randomised Controlled Trial

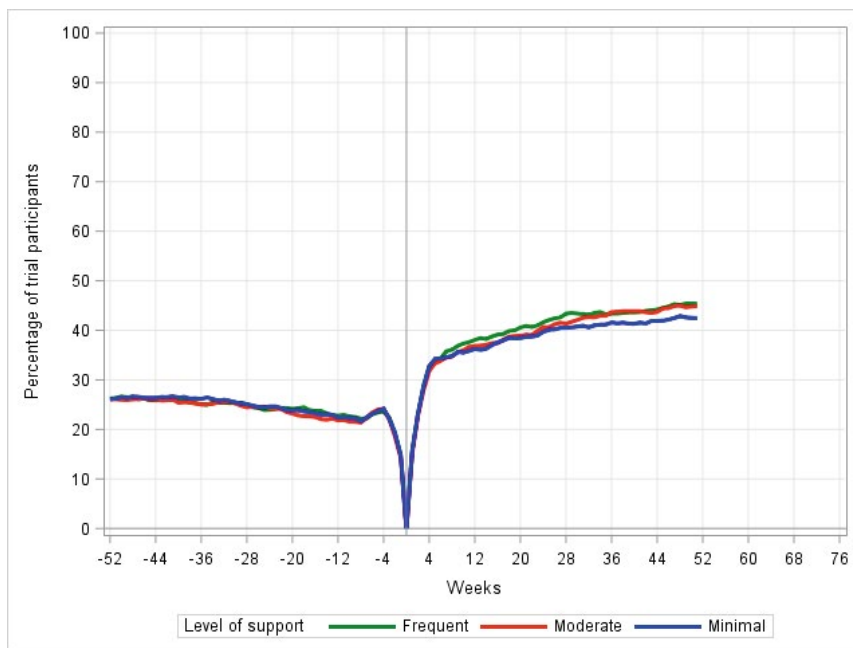


Figure 2.4: Percentage of trial participants who have increased earnings by at least 10 per cent since trial start (week 0)

From Figure 2.4, it appears that there are small, positive impacts for both treatments (the Frequent and Moderate support groups) versus the Minimal support group (the comparison) from around week 30 onwards. Figure 2.5 helps us understand whether these impacts are significant.



Figure 2.5: The percentage point difference between IWP treatments (Frequent and Moderate support groups) and the comparison group (Minimal support group) for participants that have increased their earnings by at least 10 per cent since trial start (week 0)

Figure 2.5 suggests that the impacts for the Frequent support group are statistically significant from around week 12, since the confidence intervals do not include the

Universal Credit: In-Work Progression Randomised Controlled Trial

comparison group line. Impacts for the Moderate support group start to become significant from around week 30.

In summary, for participants who have increased their earnings by at least 10 per cent since trial start, we detected statistically significant small, positive progression impacts for the Frequent support group and the Moderate support group at 52 weeks after trial start (2.90 and 2.42 percentage points respectively). The impact for the Frequent support group occurs around week 12 and is sustained. The impact for the Moderate support group is sustained from around week 30. The sample requires tracking for longer to see for how long progression is sustained.

Objectively, we also investigated the percentage of participants who had *decreased* their earnings by at least 10 per cent since trial start. See Annex D, Figures D3 and D4 for the detailed results. In summary, impacts for the Frequent and Moderate support groups at 52 weeks were significant, i.e.: they had less participants whose earnings had decreased by at least 10 per cent compared to the Minimal support group at 52 weeks. Over time, the Frequent and Moderate support groups do not appear to be significantly different to the Minimal support group in the main, indicating no statistical differences compared to the Minimal support group in general.

In conclusion, the earnings progression analysis at 52 weeks after trial start is encouraging. Both progression measures show statistically significant small, positive impacts at 52 weeks after trial start for both of the treatment (Frequent and Moderate support) groups versus the comparison (Minimal support) group. The largest impact belongs to the Frequent support group when compared to The Minimal support group: at 52 weeks, the difference in progression is £5.25 for that week and the percentage point impact of participants that increase earnings by at least 10 per cent is 2.90. Furthermore, the impacts are sustained before 52 weeks for the percentage of participants who had increased their earnings by at least 10 per cent, but not for the monetary progression measure. This sample also requires tracking for longer to see if the impacts are sustained.

Number of employment spells

The number of employment spells in the 52-week period from trial start was counted for each participant. One employment spell is an unbroken period of employment. The chart below shows the percentage of trial participants by IWP group falling into each of the employment spells categories.

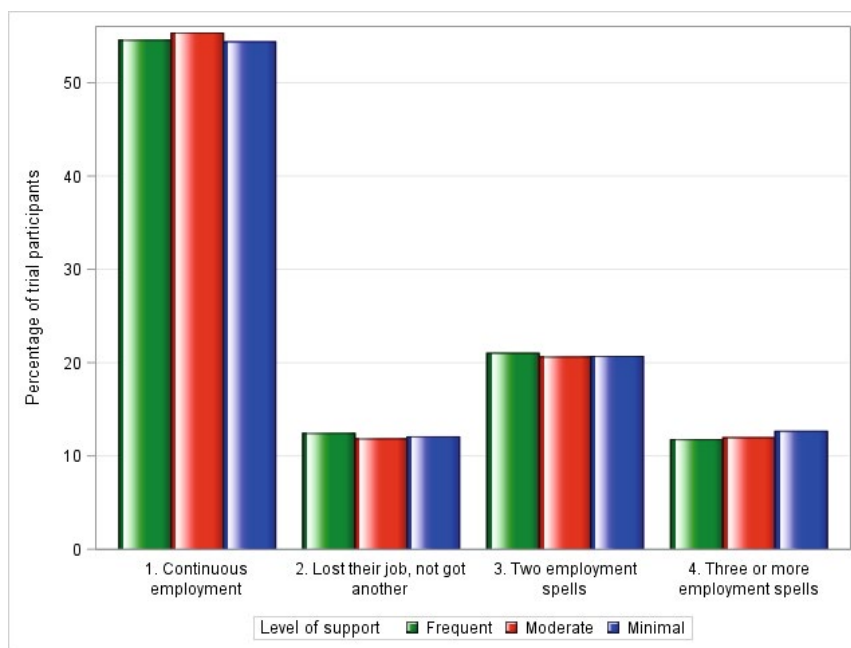


Figure 2.6: Number of employment spells by IWP group

A total of 16,834 trial participants (54.8 per cent) who had been on the trial for at least a year had no breaks in their employment. 3,734 (12.2 per cent) trial participants lost their job at some stage and did not get another during the 52-week period. 6,395 trial participants (20.8 per cent) had two employment spells during the 52-week period. A further 3,746 trial participants (12.2 per cent) had three or more employment spells in the 52-week period. During any employment spell, trial participants could have overlapping job spells with different employers and therefore more than one job at any time. However, it was not possible to deduce this from the data available.

The table below shows the proportion of trial participants within each IWP group who had continuous employment over 52 weeks.

	Sample size	Number of participants with continuous employment	Proportion of participants (%)
Frequent	9,733	5,317	54.6
Moderate	10,117	5,604	55.4
Minimal	10,859	5,913	54.5
Total	30,709	16,834	54.8

Table 2.5: Proportion of participants within each IWP group who had continuous employment over 52 weeks

Of the 16,834 trial participants who had continuous employment out of the total sample, 5,317 were in the Frequent support group, 5,604 were in the Moderate support group and 5,913 were in the Minimal support group. We compared the proportion of participants within each group who had continuous employment (fourth column of table above – 54.6 per cent, 55.4 per cent and 54.5 per cent respectively) and we did not find any statistical differences between the three groups at the 5 per cent significance level.

Subgroup analysis at 52 weeks after trial start

We performed subgroup analysis on the sample data. We chose to investigate the following subgroups:

- gender (two subgroups – male and female);
- age (four subgroups defined to align with the external evaluation – 18-24, 25-34, 35-44 and 45+);
- jobcentre type (two subgroups – Full service and Live service offices);
- geography (three subgroups – England, Scotland and Wales).

The trial was not designed to report differences between subgroups: subgroup analysis sub-divides the full sample size into smaller sample sizes leading to greater uncertainty around the estimates produced. We therefore kept the number of subgroups to a minimum: they were specified *a priori* before any analysis had been undertaken. The reason for choosing these subgroups was because it was hypothesised these subgroups could have effects on progression. Furthermore, by deciding the subgroup analysis *a priori*, we would not be ‘fishing’ for significant differences between subgroups by comparing all possible subgroups (generating hypotheses), since we would naturally expect to find some significant differences by chance alone if we performed enough tests.

We could not analyse other subgroups, for example, employment sector or job changers. The reason for this was that the data was not available to allow these subgroups to be created.

Each of the subgroups defined above were further sub-divided by IWP group. The differences between the impacts of the subgroups were analysed using the following measures (as before):

- earnings progression (£) at 52 weeks after random assignment (trial start);
- percentage of participants who had increased their earnings by at least 10 per cent at 52 weeks after random assignment.

The subgroup analysis is detailed in Tables D8 to D11 in Annex D. *General* subgroup results are summarised below (*specific* subgroup results follow later on in this section).

- As with the full sample analysis, the magnitudes of the impacts within subgroups (or within rows of a table) for the Frequent and Moderate support groups versus the Minimal support group (the comparison group) are small and positive for both progression measures.
- With one exception (see below for further details), the subgroup analysis did not find significant differences between subgroups. This indicates that the trial produced comparable effects for all subgroups.
- The subgroup analysis found examples where the finding for a subgroup of interest was statistically significant, the corresponding finding for the full study sample was in the same direction and was statistically significant, but the corresponding finding for the rest of the sample (that is, the other subgroup) wasn’t significant. Additionally, findings for the two subgroups were not statistically different from each other (that is, no between effect). An example of this phenomenon was the comparison of live service versus full service offices in Table C10.

With respect to the last bullet point above, the dilemma is whether to conclude that the trial benefits one subgroup while saying nothing about the other subgroup, or instead

to conclude that the trial produces comparable effects for both subgroups, since the full trial finding is statistically significant and subgroup findings are not statistically different from each other (Ref. 1). This report adopts the latter position.

More *specific* subgroup results are given below:

- Gender:
 - the results indicate that the trial increased progression for both the Frequent and Moderate support groups over the Minimal support group (the comparison) for both male and female subgroups at 52 weeks from trial start (see Annex D, Table D8).
- Age:
 - one statistically different progression result was discovered between age subgroups, namely, between the 18-24 years subgroup and the 25-34 years subgroup (for the Moderate support group versus the Minimal support group). The difference of the impacts is £14.57 ($p < 0.05$). See Annex D, Table D9. We found a greater progression for the younger age group, which should be seen as an exploratory result and could be tested by future research. Note that the progression impact for the 25-34 years subgroup was in a different direction (-£2.03) to the impact in the 18-24 years subgroup (£12.54).
- Jobcentre type:
 - the results indicate that the trial increased progression for both the Frequent and Moderate support groups over the Minimal support group (the comparison) for both the Live and Full service subgroups at 52 weeks from trial start (see Annex D, Table D10).
- Region:
 - the results indicate that the trial increased progression for both the Frequent and Moderate support groups over the Minimal support group (the comparison) for the England, Scotland and Wales subgroups at 52 weeks from trial start (see Annex D, Table D11). There is an extra caveat to be aware of, namely that the sample sizes for Scotland and Wales are much lower than the sample size for England.

It is noted that there are very large degrees of imprecision (i.e., large confidence intervals) of the computed subgroup central estimates. This arises because the full sample is being sub-divided into smaller groups that are being compared for statistical differences. The *minimum* uncertainty around the central estimates of the differences between the subgroup impacts was around 90 per cent. That is to say: a confidence interval that extends a minimum of 90 per cent either side of the central estimate.

2.2.3 Progression at 26 and 78 weeks

As stated previously, we pre-specified 52 weeks after trial start to be our main time point at which progression measures would be compared, with 26 and 78 weeks after trial start as subsidiary statistics. Annex E summarises the results at the 26- and 78- week points. Some charts for the progression at 26 and 78 weeks are also provided in Annex E; however, we do not produce any tables since there aren't any impacts of note.

Universal Credit: In-Work Progression Randomised Controlled Trial

We recommend that further analysis is undertaken in 2019 to repeat the 78-week analysis. By that time, every participant in the 26-week analysis (the full sample of 31,471) can have their earnings tracked to 78 weeks. The further tracking will indicate if the current trends of the 52 weeks analysis are continued.

3 Sanctions

3.1 Methodology

The sanctions analysis in this section has been performed on Live Service In-Work Progression (IWP) Trial participants only as sanctions data on Full Service IWP trial participants was not available at the time of this analysis. It has been based on the same data source that produced the published benefit sanctions statistics latest release in May 2018 ([here](#) or Ref. 2.). Note that the results of the sanctions analysis presented here are not directly comparable with the official statistics, since different methodologies have been used. We produced one overall rate as opposed to monthly point-in-time measures. Furthermore, the official statistics do not provide a sanction rate for the light touch conditionality regime.

3.1.1 Previous sanctions analysis

We published previous information on IWP sanctions in March 2017 ([here](#) or Ref. 3).

3.1.2 Data sources

The analysis was performed using the following databases.

- DWP sanctions data – this is a frozen dataset for analytical use;
- Universal Credit (UC) administrative data;
- IWP marker data.

We matched the datasets above and this allowed us to perform analysis that would provide evidence about sanctions for Live Service participants on the IWP trial.

3.1.3 Data matching

The datasets created when DWP publishes UC sanctions statistics is made available for analysis. From the data we selected those IWP participants that received an 'adverse' decision, i.e., a decision that led to a sanction. This data was current up to January 2018, and is the latest available published data.

We merged IWP participants with the latest UC sanctions data described above using the encrypted national insurance numbers of the participants. We also added the IWP start and end dates and added conditionality regime data (for example, light touch, working enough, intensive) from the UC administrative data.

We were interested in participants who had non-compliance dates during the time they were actively on the trial, that is, in the light touch conditionality regime. (The non-compliance date relates to a date when the participant didn't fulfil the conditions set out in their UC claimant commitment that then led to a sanction.) We excluded from our analysis any sanction incurred at any time when the participant's

conditionality regime type was not light touch since the sanction was not during the time they were on the trial.

3.2 Results

3.2.1 Total number of sanctions, levels and reasons

Out of 20,207 Live Service participants, there were 486 sanctions applied to 468 trial participants while they were on the light touch regime and therefore on the trial. The table below shows the levels of sanctions recorded.

Sanction level	Frequent	Moderate	Minimal	Total
High	12	8	6	26
Medium	8	5	2	15
Low	180	164	101	445
Grand total	200	177	109	486
All participants in group	6417	6704	7086	20207
Percentage	3.1%	2.6%	1.5%	2.4%

Table 3.1: Levels of sanction

Table 3.1 shows that overall sanction levels are low, with the Frequent support group having the highest rate at 3.1 per cent, the Moderate support group having a sanction rate of 2.6 per cent and the Minimal support group having the lowest rate of 1.5 per cent. The overall rate is 2.4 per cent, that is, around 98 per cent of participants were not sanctioned. Across all IWP groups, 91.6 per cent of sanctions imposed were low level (445 sanctions out of the total of 486, see last column of Table 3.1). 5.3 per cent (26 sanctions) were at the higher level.

It should be noted that participants in the Minimal support group (the comparison group) will have had less mandatory interaction with Work Coaches, and therefore will have had less opportunity to be sanctioned.

Table 3.2 below shows the sanction reasons.

Reason for Sanction	Frequent	Moderate	Minimal	Total
Fail to apply for a job	1	0	0	1
Fail to comply with a work preparation requirement	2	0	0	2
Fail to undertake all reasonable work search action	8	5	2	15
Failed to Attend	178	164	101	443
Leaving employment Voluntarily	7	4	2	13
Lose pay Voluntarily	1	2	0	3
Loss of employment through Misconduct	3	2	4	9
Grand Total	200	177	109	486

Table 3.2: Reason for sanction

The most common reason for a sanction was ‘Failed to Attend’; across all IWP groups this reason lay behind 91.2 per cent of the sanctions applied and is the main driver of sanctions across all IWP groups. Failure to attend leads to a lower rate sanction. Higher rate sanctions are only imposed for loss of employment or a reduction in pay through misconduct, losing pay or employment voluntarily or refusing to apply for or accept a job.

Table 3.3 shows failure to attend percentages by IWP group.

	Frequent	Moderate	Minimal	Total
Failed to Attend	178	164	101	443
All participants in group	6417	6704	7086	20207
Percentage	2.8%	2.4%	1.4%	2.2%

Table 3.3: Failed to attend an interview

The failure to attend percentages by IWP group in Table 3.3 shows the same pattern of results as was seen for sanctions in Table 3.1, with those in the Frequent support group having the highest sanction and failure to attend rates.

3.2.2 Which IWP groups are more likely to be sanctioned?

As shown above, Table 3.1 shows both treatments (Frequent and Moderate support groups) had higher sanction rates than the Minimal support group, with the Frequent support group having a higher proportion of participants subject to sanction. Figure 3.1 below shows these statistics complete with 95 per cent confidence intervals. The confidence intervals show the uncertainty (or precision) of the central estimate since the central estimate is based on a sample.

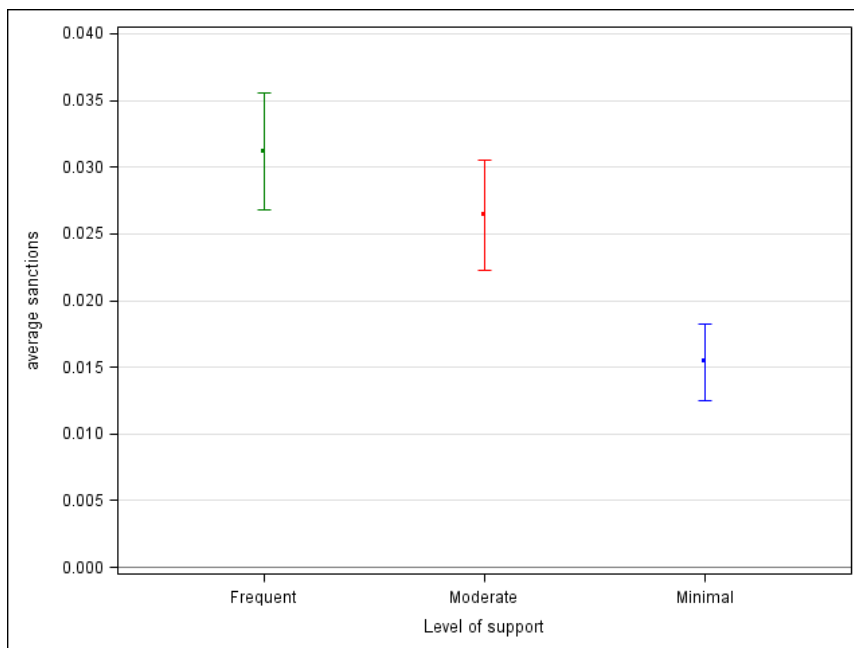


Figure 3.1: Average sanctions by IWP group

There were statistically significant differences between the Frequent support group and the Minimal support group (the comparison group) and between the Moderate support group and the Minimal group. There was no statistically significant difference between the Frequent and the Moderate support groups. The proportion of participants sanctioned in the Frequent support group is twice as high as for the Minimal group (the comparison group).

The chart below shows sanctions broken down by IWP group and gender.

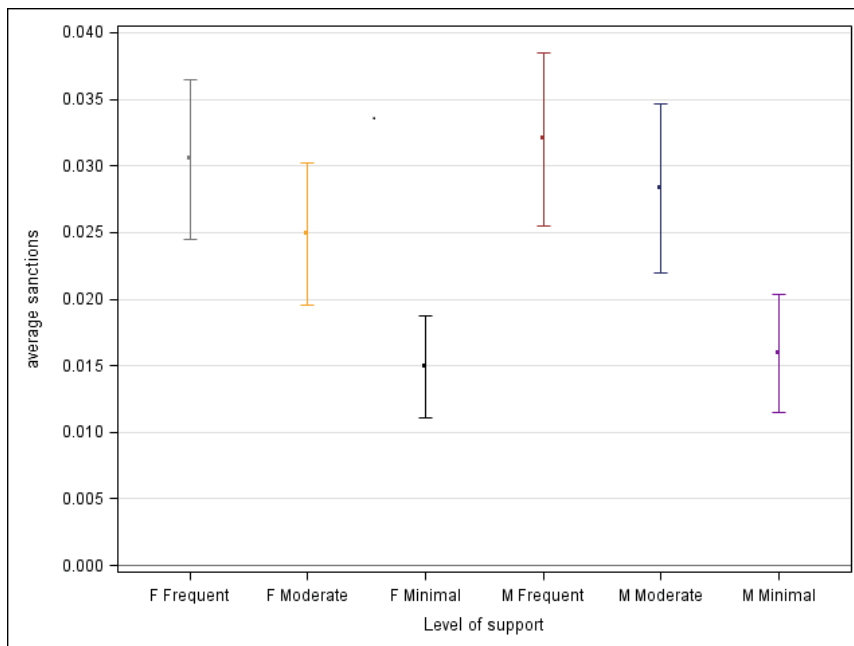


Figure 3.2: Average sanctions by IWP group and gender

Figure 3.2 shows that there were statistically significant differences in the sanctions rate between the Frequent and Minimal support groups and between the Moderate and Minimal groups for both females and males. For Males, the proportion of

participants sanctioned is higher for all three groups than for the comparable Female group.

Figure 3.3 below shows the sanction rate for participants where they have children.

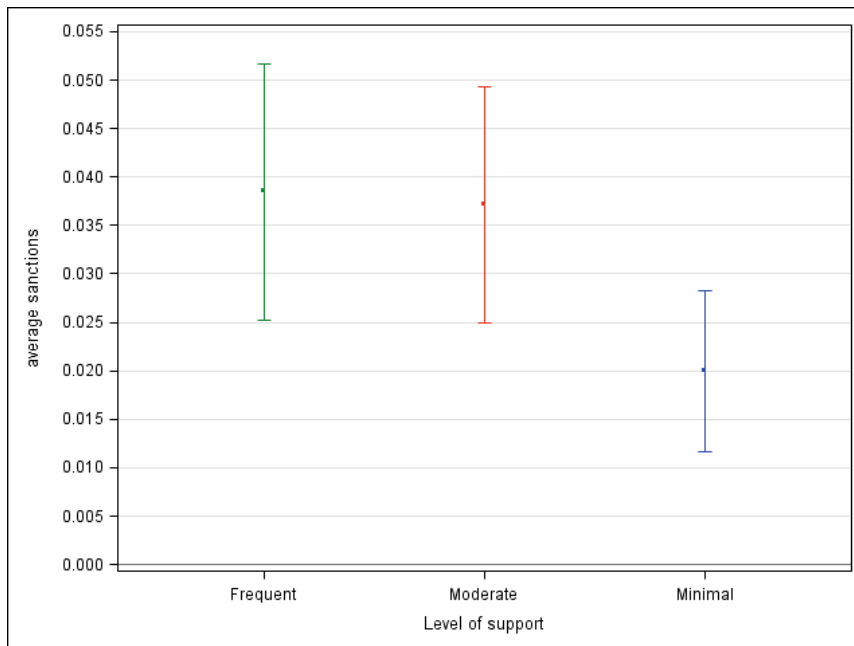


Figure 3.3: Average sanctions for all participants with children

Where participants have children, we found that there was a significant difference between the Frequent support group and the Minimal support group, and between the Moderate support group and the Minimal support group. No significant difference was observed between the Frequent support group and the Moderate support group.

For females with children (see Figure 3.4 below) we found that there were statistically significant differences between the Frequent support group and the Minimal support group and also between the Moderate and Minimal support groups.

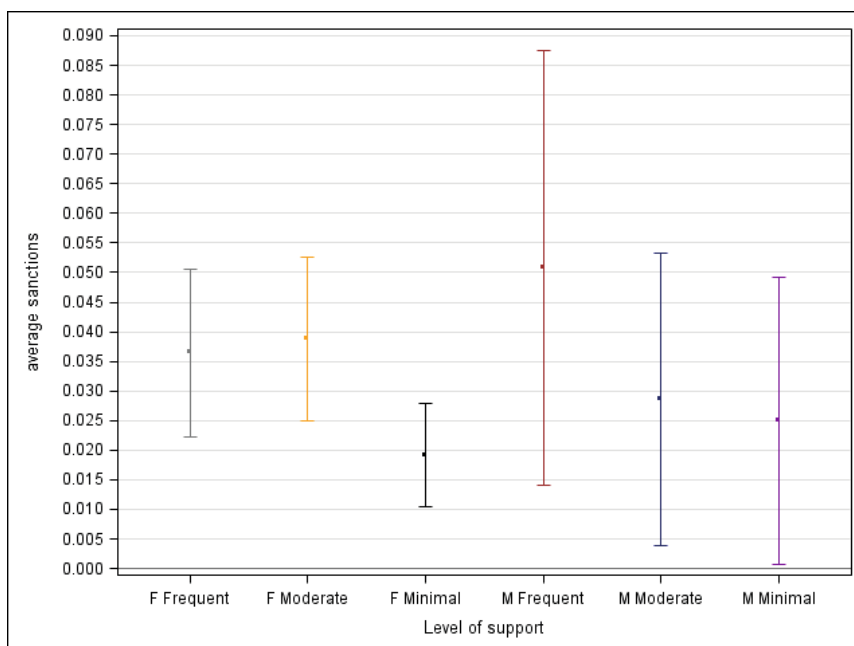


Figure 3.4: Average sanctions for all participants with children by gender

We did not detect any significant differences between the three IWP groups for males with children. We note that the confidence intervals for males with children are wider (or less precise) than those for females with children. This is due to the smaller sample sizes for males with children compared to females with children.

We saw the same pattern of results by IWP group for participants who failed to attend as we saw for all sanction cases. Figure 3.5 below shows the sanction proportions and the failure to attend proportions by IWP group.

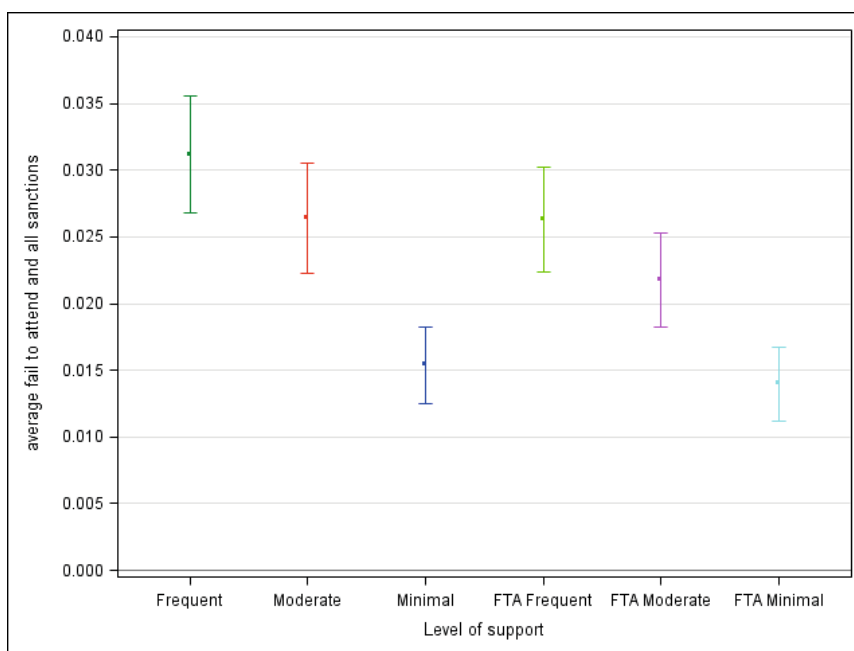


Figure 3.5: Average sanctions and failed to attend (FTA) by IWP group

For the failure to attend reason, we detected a statistically significant difference between the Frequent support group and the Minimal support group and between the Moderate and Minimal support groups, with no significant difference between the Frequent support group and the Moderate support group.

Sanctions Conclusions

For trial participants, the majority of sanctions imposed were low level, with 445 sanctions out of 486 at the low level. Twenty-six sanctions were at the high level, with the remainder of 15 at the medium level.

Across all three IWP trial groups, the most common reason for participants being sanctioned was a failure to attend an interview with their Work Coach. This accounted for 443 sanctions out of the total of 486 sanctions.

Across all three IWP groups, the overall sanction rate was 2.4 per cent (486 sanctions applied to 20,207 participants in Live Service). Both IWP treatment groups (Frequent and Moderate support groups) had a higher proportion of participants who were sanctioned compared to the Minimal support group (3.1 per cent, 2.6 per cent and 1.5 per cent respectively). Participants in the Minimal support group (the comparison group) had less mandatory interaction with Work Coaches, and therefore there was less opportunity to be sanctioned.

The differences for both the Frequent and the Moderate support groups versus the Minimal support group were found to be statistically significant. We found the same

Universal Credit: In-Work Progression Randomised Controlled Trial

pattern of results when we performed the analysis by gender. For participants with children, the Frequent and Moderate support groups had a higher rate of sanction than the Minimal support group.

The proportion of sanctions for the Frequent support group is higher than that for the Moderate group, however, this result was not statistically significant. The lower rate of sanctions for the Moderate support group compared to the Frequent group is consistent with the Moderate support group participants having fewer interactions with Work Coaches, and therefore fewer opportunities to get sanctioned.

In summary, participants in the Minimal support group, which had the least number of interviews had the lowest level of sanctions, whilst the two intervention groups, required to attend the most interviews, got a higher proportion of sanctions.

4 Conclusions and recommendations

The In-Work Progression (IWP) Trial aimed to test whether increased Work Coach support and applying conditionality drove behaviours that led to earnings progression. We undertook an impact assessment for the trial participants in the three IWP groups at 52 weeks after starting the trial to see if there were any differences between the two treatment groups and the comparison group. We specified two earnings progression measures, one based on a monetary value, the other based on a proportion. The method is statistically valid and produced unbiased mean estimates for each of the groups that were compared statistically.

Full sample impact analysis results

We detected a small and positive statistically significant **monetary** progression impact for the Frequent support group versus the Minimal support group (the comparison group), and for the Moderate support group versus the Minimal support group at 52 weeks after starting the trial. This amounted to an additional progression of £5.25 per week for the Frequent support group and £4.43 per week for the Moderate support group compared to the Minimal support group at 52 weeks.

For participants who increased their earnings by at least 10 per cent since starting the trial, we detected small and positive statistically significant progression impacts for both the Frequent support group and the Moderate support group versus the Minimal support group at 52 weeks after starting the trial (an additional 2.90 and 2.42 percentage points respectively). The impact for the Frequent support group occurs around week 12 and is sustained. The impact for the Moderate support group is sustained from around week 30.

The sample requires tracking after 52 weeks to see how long progression is sustained.

Subgroups analysis

We investigated IWP groups by gender, age, Jobcentre type (live or full service) and region (England, Scotland and Wales). As with the full sample results, both the Frequent and Moderate support groups experienced small, positive earnings progression impacts over the Minimal support group within the subgroups.

The subgroup analysis did not find significant differences *between* subgroup impacts in general, indicating that the trial produced comparable, small and positive effects for all subgroups. The one exception indicated that at 52 weeks, participants in the 18-24 years subgroup saw a greater progression of £14.57 on average than those participants in the 25-34 years subgroup (for the Moderate support group versus the Minimal support group), which was statistically significant.

Sanctions analysis

We also analysed participants on the trial in Live Service that had been sanctioned. We excluded any sanction applied when the participant was not in the light touch conditionality regime since the sanction was not during the time they were on the trial.

The overall proportion of participants sanctioned across all three IWP groups over the course of the trial was 2.4 per cent. This value is not directly comparable with official statistics since different methodologies have been used. We produced one overall rate as opposed to monthly point in time measures; furthermore, the official statistics do not provide a sanction rate for the light touch conditionality regime.

Across all three IWP groups, failure to attend an interview with a work coach was the most common reason for being sanctioned. This accounted for 91.2 per cent of sanctions applied (or 443 sanctions). Consequently, the majority of sanctions applied were low level sanctions, 91.6 per cent (or 445 sanctions).

The Minimal support (comparison) group which was subject to the least number of interviews received the lowest rate of sanction (1.5 per cent). Both treatment groups, that is, the Frequent and Moderate support groups had higher sanction rates compared to the Minimal support group (3.1 per cent and 2.6 per cent respectively) and the differences compared to the Minimal support group were statistically significant.

Furthermore, the Frequent support group had a higher proportion of sanctions than the Moderate support group, but this difference was not statistically significant.

Recommendation

The results to date show a small statistical difference at the end of 52 weeks from the trial start date. We therefore recommend that we continue to track performance beyond the 52 week point to assess whether there is further impact of the intervention.

Annexes

Annex A: Trial group balance

This annex presents some comparisons of characteristics between the three In-Work Progression (IWP) groups (proportions across the three groups) to examine trial group balance. The tables below show the comparisons between groups for various participant characteristics such as gender, age, whether the participant was in a live or full service office, whether the participant has a partner, and region.

An examination of the distributions below about the 31,471 individuals suggests that the three IWP groups are well balanced. The difference between the mean values found for each group (denoted as ‘Absolute difference’ in the tables) was less than 5 per cent of the standard deviation for all characteristics (see last column in each of the tables). The maximum value is 4.1 per cent in the last column of Table A6.

Note that it isn’t sufficient to compare solely the ‘Absolute difference’ values in the tables below because we must normalise the difference first using the standard deviations to account for different variability of the characteristics.

Gender	Frequent (%)	Moderate (%)	Minimal (%)	Comparison	Absolute difference (signal %)	Standard deviation (noise %)	Signal to noise ratio (%)
Male	42.9	42.0	42.2	Freq vs Mod	0.9	69.9	1.3
				Freq vs Min	0.8	69.9	1.1
				Mod vs Min	0.1	69.8	0.2
Female	57.1	58.0	57.9	Freq vs Mod	0.9	69.9	1.3
				Freq vs Min	0.8	69.9	1.1
				Mod vs Min	0.1	69.8	0.2

Table A1: Gender breakdown and comparison between IWP groups

Universal Credit: In-Work Progression Randomised Controlled Trial

Age band	Frequent (%)	Moderate (%)	Minimal (%)	Comparison	Absolute difference (signal %)	Standard deviation (noise %)	Signal to noise ratio (%)
18-24	14.1	13.5	15.0	Freq vs Mod	0.6	48.8	1.2
				Freq vs Min	0.9	49.9	1.8
				Mod vs Min	1.5	49.5	3.1
25-34	28.9	28.5	29.6	Freq vs Mod	0.4	64.0	0.6
				Freq vs Min	0.7	64.3	1.1
				Mod vs Min	1.1	64.2	1.7
35-44	20.3	20.5	20.2	Freq vs Mod	0.2	57.0	0.3
				Freq vs Min	0.1	56.8	0.2
				Mod vs Min	0.3	56.9	0.6
45+	36.7	37.5	35.2	Freq vs Mod	0.8	68.3	1.2
				Freq vs Min	1.5	67.9	2.2
				Mod vs Min	2.3	68.0	3.4

Table A2: Age breakdown and comparison between IWP groups

Job-centre type	Frequent (%)	Moderate (%)	Minimal (%)	Comparison	Absolute difference (signal %)	Standard deviation (noise %)	Signal to noise ratio (%)
Live service	63.6	64.0	63.1	Freq vs Mod	0.4	68.0	0.5
				Freq vs Min	0.5	68.1	0.7
				Mod vs Min	0.8	68.1	1.2
Full service	36.4	36.0	36.9	Freq vs Mod	0.4	68.0	0.5
				Freq vs Min	0.5	68.1	0.7
				Mod vs Min	0.9	68.1	1.2

Table A3: Jobcentre type breakdown and comparison between IWP groups

Partner	Frequent (%)	Moderate (%)	Minimal (%)	Comparison	Absolute difference (signal %)	Standard deviation (noise %)	Signal to noise ratio (%)
No	82.4	81.8	82.6	Freq vs Mod	0.6	54.2	1.1
				Freq vs Min	0.2	53.7	0.3
				Mod vs Min	0.8	54.1	1.4
Yes	17.6	18.2	17.4	Freq vs Mod	0.6	54.2	1.1
				Freq vs Min	0.2	53.7	0.3
				Mod vs Min	0.8	54.1	1.4

Table A4: Partner breakdown and comparison between IWP groups

Universal Credit: In-Work Progression Randomised Controlled Trial

Partner on trial	Frequent (%)	Moderate (%)	Minimal (%)	Comparison	Absolute difference (signal %)	Standard deviation (noise %)	Signal to noise ratio (%)
No	95.2	95.5	95.5	Freq vs Mod	0.2	29.7	0.8
				Freq vs Min	0.2	29.8	0.7
				Mod vs Min	0.0	29.4	0.1
Yes	4.8	4.5	4.5	Freq vs Mod	0.2	29.7	0.8
				Freq vs Min	0.2	29.8	0.7
				Mod vs Min	0.0	29.4	0.1

Table A5: Partner on trial breakdown and comparison between IWP groups

Universal Credit: In-Work Progression Randomised Controlled Trial

Region	Frequent (%)	Moderate (%)	Minimal (%)	Comparison	Absolute difference (signal %)	Standard deviation (noise %)	Signal to noise ratio (%)
England, Central	14.8	14.9	12.9	Freq vs Mod	0.1	50.2	0.2
				Freq vs Min	1.9	48.8	3.9
				Mod vs Min	2.0	48.9	4.1
England, London and HC	20.9	20.8	21.1	Freq vs Mod	0.1	57.5	0.2
				Freq vs Min	0.2	57.6	0.4
				Mod vs Min	0.3	57.6	0.6
England, North East	12.2	12.5	13.2	Freq vs Mod	0.3	46.5	0.7
				Freq vs Min	1.0	47.1	2.1
				Mod vs Min	0.7	47.3	1.5
England, North West	28.4	28.7	28.7	Freq vs Mod	0.3	63.9	0.4
				Freq vs Min	0.2	63.9	0.4
				Mod vs Min	0.0	63.9	0.0
England, Southern	13.4	13.3	14.3	Freq vs Mod	0.1	48.1	0.2
				Freq vs Min	0.9	48.9	1.9
				Mod vs Min	1.0	48.8	2.1
Scotland	7.1	6.9	6.9	Freq vs Mod	0.1	36.1	0.4
				Freq vs Min	0.2	36.0	0.6
				Mod vs Min	0.1	35.9	0.2
Wales	3.2	2.9	2.9	Freq vs Mod	0.3	24.2	1.3
				Freq vs Min	0.3	24.3	1.1
				Mod vs Min	0.1	23.7	0.3

Table A6: Region breakdown and comparison between IWP groups

Note that in the examination of trial group balance above, we decided not to test whether there were any statistically significant differences in the means at trial start of a number of variables across treatment and comparison groups.

The reason for this is documented in literature, Ref. 4. In essence, we stated in Section 2 that the randomisation process worked reliably and robustly, eliminating the possibility of interfering with the trial group selection process that could bias the results: “Since allocation to groups has been properly randomised, any difference

Universal Credit: In-Work Progression Randomised Controlled Trial

between groups will necessarily be due to chance and performing a significance test to compare variables at the start of trial is to assess the probability of something having occurred by chance when we do know that it did occur by chance.”

Ref. 4 also states: “It is also worth noting that statistical significance is considered immaterial when considering whether any imbalance between the groups may have affected the results. In particular, it is wrong to infer from the lack of statistical significance that the variable in question did not affect the outcome of the trial, since a small imbalance in a variable highly correlated with the outcome of interest can be far more important than a large and significant imbalance for a variable uncorrelated with the variable of interest.”

Annex B: Frequency of intervention appointments

This is a summary of some statistics around the frequency of appointments that participants in the IWP trial received through their time on the trial. The main finding was that participants received the right intensity of intervention for their allocated group, and variability is high. Also, the Frequent support group was defined to be the most intensive treatment group and participants in that group did appear to receive more appointments than participants in the other two groups.

DWP administrative data measures

We defined two measures to identify whether participants received the correct amount of intervention across the IWP trial sample.

i) Relative difference in the number of interventions participants received

The relative difference measure in Table B1 below (fourth column) looked at how many more (or less) appointments participants received compared to what we expected them to receive as per the trial design/policy intent.

	In theory ⁶	In practice	Relative Difference ⁷
Frequent	26 meetings per year	21.0 meetings per year	- 19%
Moderate	6 meetings per year	9.8 meetings per year	+ 62%
Minimal	1 meeting	2.5 meetings	+ 150%

Table B1: Relative difference in the number of interventions participants received

Although the Frequent support group received fewer appointments than expected, that group was still the most intensive treatment group in practice. In practice, the Frequent support group still received over double the number of appointments on average compared to the Moderate group.

ii) Time between appointments

The inter-appointment measure looked at the amount of time between intervention appointments. It is not meaningful to include this measure for the Minimal support group as these participants only have one intervention appointment on the Trial. We introduced an interval comprising of +/- 1 week around the prescribed frequency of intervention to account for rescheduling of appointments.

⁶ The first, initial interviews when participants joined the trial have not been included for any of the groups.

⁷ These calculations are based on the time period from participants starting the trial up to their last recorded attended appointment.

	Appointments every 2 weeks	Appointments every 1-3 weeks
Frequent	54%	76%
	Appointments every 8 weeks	Appointments every 7-9 weeks
Moderate	28%	42%

Table B2: Time between appointments

NB: We are reporting on the percentage of appointments, not participants in the trial.

There is a wide variation in inter-appointment times (see range of inter-appointment intervals below). The table below suggests that the median and the mode of inter-appointment intervals are close to the prescribed level of support.

	Expected inter-appointment time	Average inter-appointment time (median)	Most frequent inter-appointment gap (mode)	Range of appointment intervals
Frequent	2	2.1	2	0 – 75.6
Moderate	8	7.7	8	0 – 71.9

Table B3: Statistics for inter-appointment time

NB: We are reporting on appointments, not participants in the trial.

Additional information

It is worth noting that different factors can influence whether a participant receives the right intensity of intervention:

- a) whether Work Coaches schedule meetings at the right intervals;
- b) whether participants can attend these meetings, need to rebook or fail to attend;
- c) whether the participant changes conditionality groups constantly which means that they cannot follow the prescribed level of support for the trial.

Further analysis of intervention appointments

We performed further analysis that investigated the number of appointments trial participants received in the first six months compared to the second six months. This was done to see if intervention appointments were applied consistently over time or whether they tailed off. This involved the following steps:

- selecting a sample of trial participants who had remained in the ‘Light Touch’ regime following the six months from their trial start date and counting the number of intervention appointments they received during that time;
- counting the number of intervention appointments for a proportion of trial participants who had naturally survived a second spell of six months in the ‘Light Touch’ regime;
- comparing the average number of intervention appointments per group for the three groups.

During the first six-month period, trial participants in the Frequent support group would be expected to receive 13 intervention appointments, the Moderate support group would be expected to receive around three intervention appointments, whilst the

Minimal support group participants would've only expected to receive one intervention appointment.

The only exception for the second six months is that the Minimal support group participants would not be expected to receive any further intervention appointments.

The table below shows the sample sizes associated with the two six-month periods together with the mean number of appointments and variability.

Measure	First six months			Second six months		
	Frequent	Moderate	Minimal	Frequent	Moderate	Minimal
Participants	1376	1568	1061	398	443	157
Mean	5.9	2.9	2.4	5.8	3.0	4.1
Variance	13.9	3.5	5.6	14.4	3.8	11.5
Range	22.0	15.0	14.0	17.0	12.0	14.0

Table B4: Appointments in two six-month periods

The table is discussed below.

- There are a much smaller number of participants that survived the second six-month spell compared to the first six months. This is consistent with the relatively high churn of participants in the Light Touch regime. It is important to note this when reading the following points below.
- The mean for the Frequent support group indicates that participants received around six intervention appointments on average in both the first and second six months. This is less than the number of appointments that would have been expected in each of the two six-month periods (13). One possible reason for this deficit is the relative logistical difficulty in arranging frequent fortnightly appointments because the participant is already working compounded by the available diary time of the Work Coach.
- The mean for the Moderate support group indicates that participants received the correct number of intervention appointments on average in both the first and second six months.
- The mean for the Minimal support group (the comparison group) indicates that participants received more than the one expected appointment on average.
- Variability is high (as shown in the variance and range rows).

In summary, for a sample of participants that remained in the Light Touch regime for a year, the Frequent support group participants were the most intensive group in receiving appointments, however, they did not receive the total expected number of appointments on average.

Furthermore, the Bootstrapping technique was used to compare the average number of appointments for the three groups statistically. In both six-month periods, the Frequent support group participants received significantly more appointments than the other two groups. The Minimal support group participants received significantly more appointments in the second six-month period compared to the first six-month period. See table below.

Universal Credit: In-Work Progression Randomised Controlled Trial

Outcome	Frequent	Moderate	Minimal	Comparison	Impact	P-Value
<u>No. of appointments</u>						
First six months	5.9	2.9	2.4	Freq vs Mod	3.0***	<0.001
				Freq vs Min	3.5***	<0.001†††
				Mod vs Min	0.5***	<0.001†††
Second six months	5.8	3.0	4.1	Freq vs Mod	2.9***	<0.001
				Freq vs Min	1.8***	<0.001†††
				Mod vs Min	-1.1***	<0.001†††

Table B5: Comparing average number of appointments in two six-month periods

Notes: Statistical significance levels for the first and second six months are indicated as: *** = 0.1%; ** = 1%; * = 5%. Statistically significant differences between the first and second six months are indicated as: ††† = 0.1%; †† = 1%; † = 5%.

Annex C: Methodology for earnings progression

This annex formally states the methodology used in the impact assessment for the earnings progression (impact). The method tests whether there are statistical earnings differences between the trial groups.

A quick review of quantitative methods to compare measurements over time (Ref. 5) reveals various methods. These methods can be complex and include mixed measures ANOVA and multilevel modelling.

In principle, regression methods are unnecessary because randomisation should ensure that participants in all three IWP groups are similar on average at the trial start date. Thus, simply comparing mean outcomes provides an unbiased impact estimate (Ref. 6). (The advantage of regression is that it allows for characteristics such as age and gender to be controlled for. This was investigated but did not add any further value.)

Since it is beneficial for any method to be transparent and simple, the advantage of the method below is that it is more easily understandable than the more complex regression approaches.

Earnings progression measure (£)

The method considered individual participants as the basic unit (the subject) and used the monthly earnings for each individual subject to construct a single number which summarised some aspect of that subject's earnings 'curve' (Ref. 7).

We decided to use an 'earnings difference', or more succinctly, a progression summary measure, d , in GBP defined as follows.

$$d_{mn} = e_{mn}^{t_1} - e_{mn}^{t_0}$$

where:

m is IWP group 1, 2 or 3 (1 = Frequent, 2 = Moderate, 3 = Minimal);

n is the n^{th} participant in IWP group 1, 2 or 3;

e^{t_1} is the earnings the participant received at time $t_1 > t_0$;

e^{t_0} is the earnings the participant received at time t_0 (their trial start date);

Progression can be a positive or negative number for any participant.

Thus, for each of the three groups, we produced a vector that contained the difference between the first and last earnings values for a given time period (say t_0 = week 0 and t_1 = week 26) for each individual participant, $n = 1, \dots, l$ for the Frequent support group, $n = 1, \dots, j$ for the Moderate support group and $n = 1, \dots, k$ for the Minimal support group, where generally, $l \neq j \neq k$.

IWP Group	Frequent	Moderate	Minimal
Earnings difference (summary measure) for each individual participant	$\begin{pmatrix} d_{11} \\ d_{12} \\ \cdot \\ \cdot \\ d_{1l} \end{pmatrix}$	$\begin{pmatrix} d_{21} \\ d_{22} \\ \cdot \\ \cdot \\ d_{2j} \end{pmatrix}$	$\begin{pmatrix} d_{31} \\ d_{32} \\ \cdot \\ \cdot \\ d_{3k} \end{pmatrix}$
Measure of location (mean)	\bar{d}_1	\bar{d}_2	\bar{d}_3

Once the progression summary measure had been calculated for each subject, its values were treated as raw data for an appropriate statistical analysis. We could then test the following hypotheses by comparing the means $\bar{d}_1, \bar{d}_2, \bar{d}_3$ of the vectors, for example:

- for each group, is the mean difference greater than zero, i.e., do any of the groups show earnings progression?
- is one group different to another group?

Earnings progression measure (per cent)

The progression summary measure, d , above was altered to construct a progression measure based on proportions. We defined each participant to have progressed as follows.

$$prog_{mn} = \begin{cases} 1, & \text{if } e_{mn}^{t_1} \geq 1.1e_{mn}^{t_0} \\ 0, & \text{otherwise} \end{cases}$$

where:

$prog_{mn}$ is a binary 0, 1 variable where 0 indicates the participant did not progress and 1 indicates that the participant did progress by at least 10 per cent%;

$e_{mn}^{t_1}$ is the earnings at time t_1 for the n th participant of the Frequent, Moderate or Minimal support group (as before);

$1.1e_{mn}^{t_0}$ is the earnings at trial start for the n th participant of the Frequent, Moderate or Minimal support group (as before) inflated by 10 per cent.

Once again, three vectors were produced that contained a binary variable for progression for each participant. The mean of each vector, $\overline{prog}_1, \overline{prog}_2, \overline{prog}_3$, is thus bounded by 0 and 1 and therefore represents the percentage of participants who have increased their earnings by at least 10 per cent at a time t_1 after their trial start date.

Earnings progression for subgroups

The methodology described above was applied to test for statistical differences between subgroups.

Statistical considerations

The following statistical issues were considered for the impact analysis:

- We used the SAS statistical package version 9.4 to undertake all analysis;
- We used the bootstrapping and permutation tests to calculate confidence intervals and perform hypothesis tests (Refs. 8, 9 and 10);
- For all statistical tests, we used 2-sided p-values with $\alpha = 0.05$ level of significance and confidence intervals were calculated at the 95 per cent level;
- P-values are reported to three decimal places, p-values less than 0.001 reported as $p < 0.001$.

Annex D: Progression at 52 weeks impact assessment

This annex contains the remaining results of the analysis of progression at 52 weeks that was not presented in Section 2.2.2.

Demographics

The following tables show the breakdown of the whole IWP population (those participants that had taken part in the trial) and those included in the 52-week impact assessment in terms of gender, age and UC service.

Gender	Whole Trial Population		52-Week IA Population	
	Total	Percentage	Total	Percentage
Male	13,331	42.4	13,001	42.3
Female	18,140	57.6	17,708	57.7
Total	31,471		30,709	

Table D1: Participants in the In-Work Progression Trial by gender

Source: UCAD live service and full service reference data sets.

Age	Whole Trial Population		52-Week IA Population	
	Total ^{&}	Percentage [£]	Total ^{&}	Percentage
18-24	4,481	14.2	4,379	14.3
25-34	9,128	29.0	8,907	29.0
35-44	6,397	20.3	6,264	20.4
45 plus	11,464	36.4	11,158	36.3
Total	31,470		30,708	

Table D2: Participants in the In-Work Progression Trial by age

Source: UCAD live service and full service reference data sets.

Notes: [&]1 case with missing age.

UC	Whole Trial Population		52-Week IA Population	
	Total	Percentage	Total	Percentage
Live Service	20,003 ^{&}	63.6	19,462	63.4
Full service	11,468	36.4	11,247	36.6
Total	31,471		30,709	

Table D3: Participants in the In-Work Progression Trial by Jobcentre type

Source: UCAD live service and full service reference data sets.

Notes: [&]The live service value of 20,003 is different to the value of participants on live service indicated in the sanctions analysis in Section 3. This is partly because the counts come from two independent data sources and partly because there are a small number of participants on the trial that were subject to both live service and full service offers.

Universal Credit: In-Work Progression Randomised Controlled Trial

Region	Whole Trial Population		52-Week IA Population	
	Total ^{&}	Percentage	Total ^{&}	Percentage
England	28,320	90.1	27,652	90.1
Scotland	2,191	7.0	2,113	6.9
Wales	937	3.0	921	3.0
Total	31,448		30,686	

Table D4: Participants in the In-Work Progression Trial by region

Source: UCAD live service and full service reference data sets.

Notes: [&]23 cases with missing region.

Further results

Figure D1 below shows average weekly earnings for the three IWP groups up to 52 from trial start. The average weekly earnings for all three groups combined was £163.62 per week at the start of the trial.

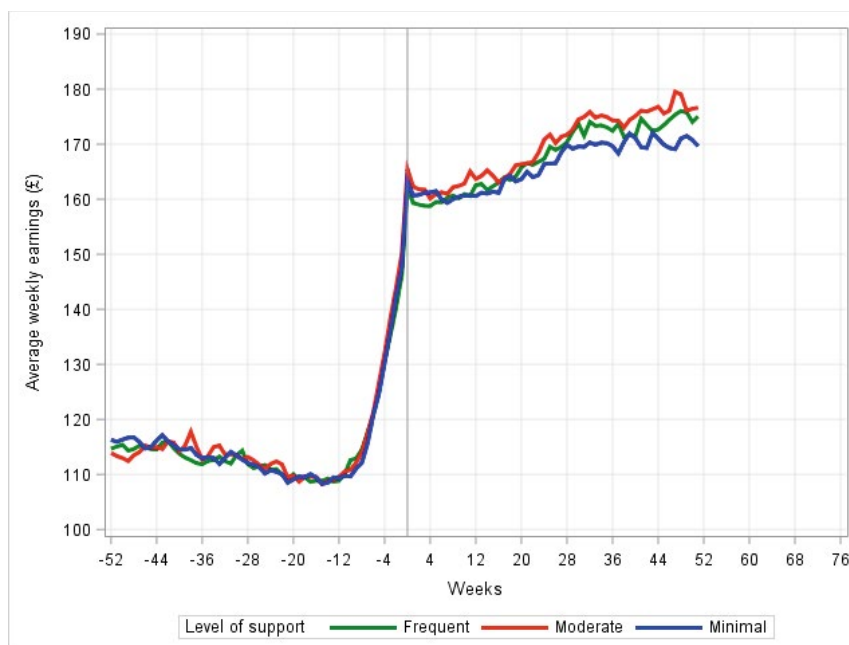


Figure D1: Average weekly earnings for the three IWP groups over time

Figure D2 below shows the percentage of trial participants in employment over time. Naturally, all participants are in employment at week 0 when they start the trial as it is a condition that you have to be in employment (and therefore in the Light Touch regime) to start the trial. At 52 weeks after trial start, 80 per cent of participants are in employment. Note that Figure D2 is not a survival curve – participants can have numerous employment spells.

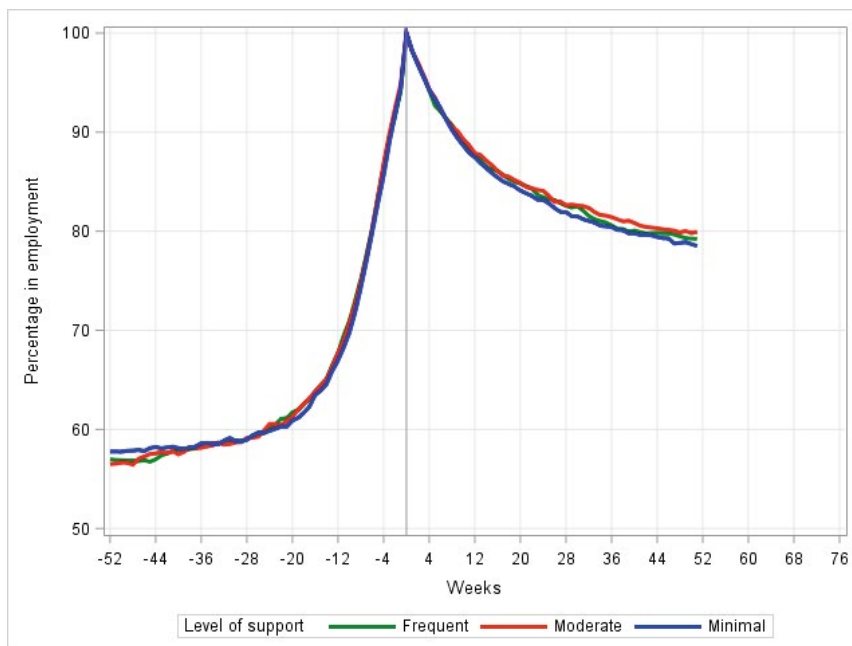


Figure D2: Percentage of trial participants in employment over time for the three IWP groups

Figure D3 below shows the percentage of trial participants that have decreased their earnings by at least 10 per cent since they started the trial. From a positive outcome point of view, we would like to see less participants in the treatment groups (Frequent and Moderate support groups) with decreased earnings than the comparison group (the Minimal support group).

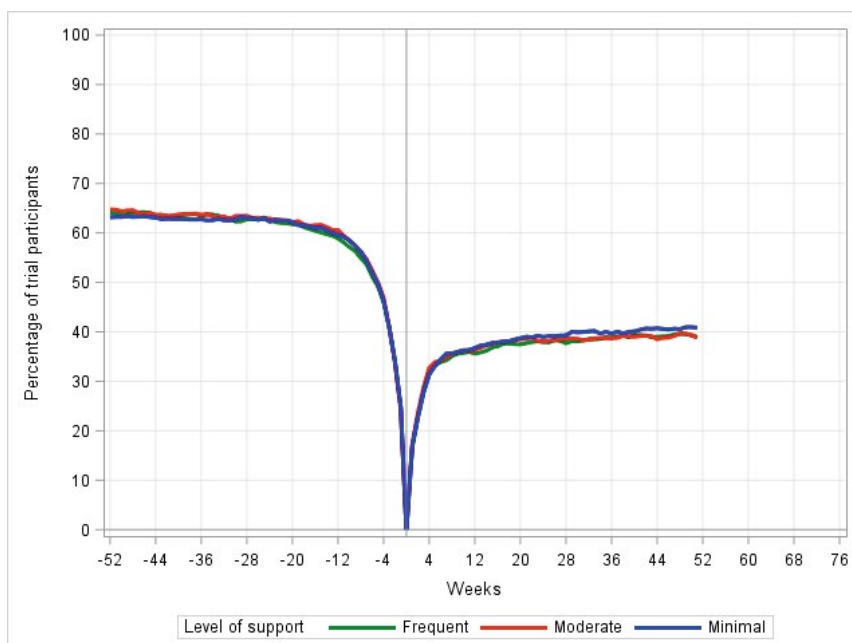


Figure D3: The percentage of trial participants that have decreased earnings by at least 10 per cent since trial start

To see if the differences in Figure D3 might be significant, Figure D4 plots the differences between the two treatment group lines and the comparison group. This shows there is no difference between the three IWP groups. There is some weak evidence that may suggest fewer participants in the treatment groups (Frequent and

Universal Credit: In-Work Progression Randomised Controlled Trial

Moderate groups) have decreased earnings compared to the comparison group (the Minimal support group), however, this hypothesis needs to be tested by tracking further in time to see if impacts sustain and/or increase.

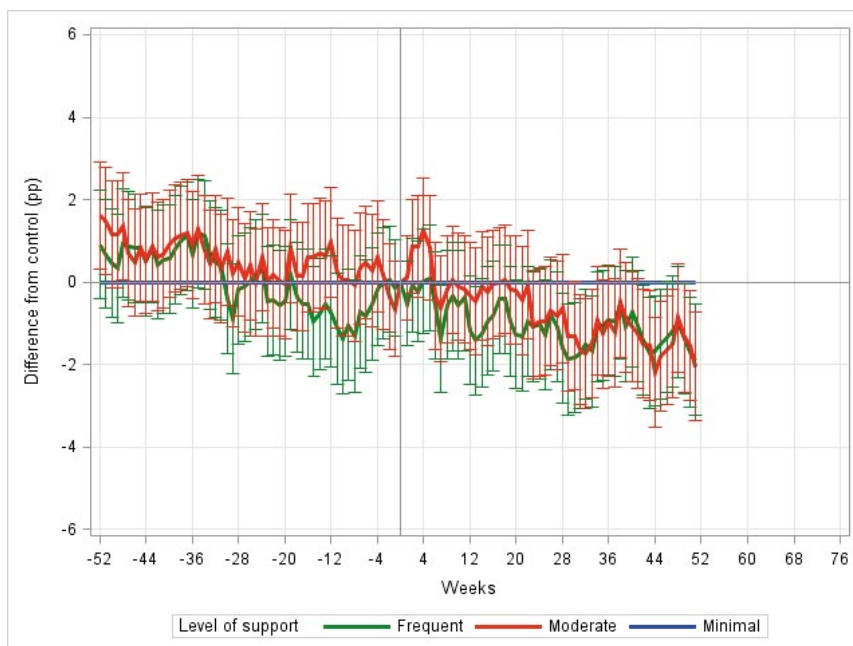


Figure D4: Percentage of trial participants that have decreased earnings by at least 10 per cent since trial start – difference from the Minimal support group (comparison group)

Figure D5 below shows the survival curves for the first employment spells for the three IWP groups. At 52 weeks after trial start, the average percentage of participants in employment across the three groups is 55 per cent (see also Table 2.5, Section 2.2.2). That is to say that 55 per cent of participants have had no break in their employment since starting the trial (although they may have changed jobs).

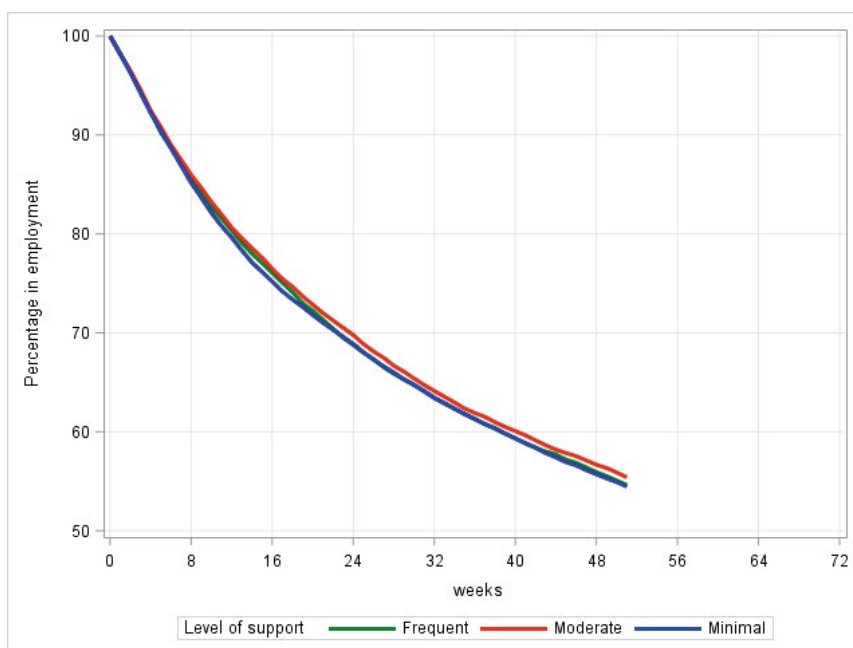


Figure D5: Survival curves for first employment spells

Universal Credit: In-Work Progression Randomised Controlled Trial

Figure D6 shows the overall value of earnings for the three groups over time. This statistic is analogous to calculating the area underneath each participant's earning 'curve' (or total earnings for that participant) between trial start and a particular point in time and then averaging these total earnings across all participants. From this, there may be a difference between groups in terms of the overall value of earnings.

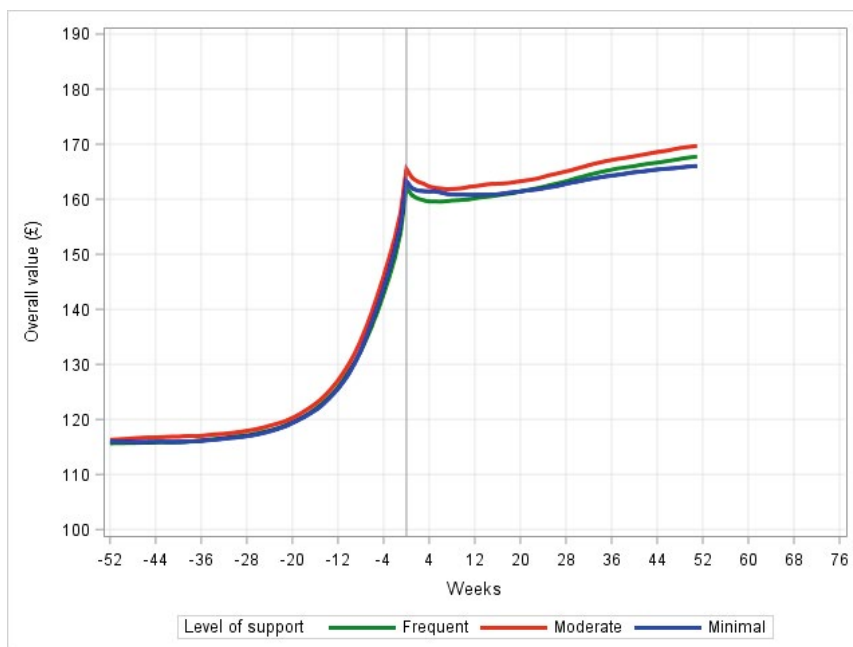


Figure D6: Overall value of earnings for the three groups over time

To see if the differences in Figure D6 above might be significant, figure D7 shows the differences between the two treatment groups (the Frequent and Moderate support groups) and the comparison group. There are no statistically significant differences between the two treatment groups. A difference appears to be occurring between the Moderate support group and the Minimal support group but, again, this needs to be tracked for longer.

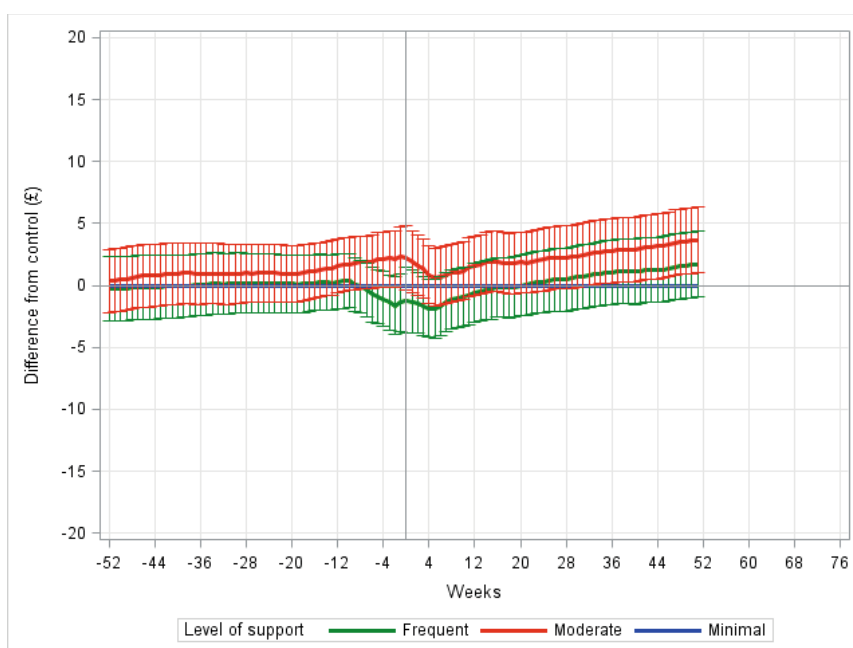


Figure D7: Overall value of earnings for the three groups over time

The table below shows the average number of weeks in work for the three groups:

IWP Group	Number of weeks in work	95% lower confidence interval	95% upper confidence interval
Frequent	44.1	43.8	44.4
Moderate	44.3	44.0	44.5
Minimal	43.9	43.6	44.1

Table D5: Average number of weeks in work by IWP group

The table below shows the average employment spell duration in weeks:

IWP Group	Employment spell duration	95% lower confidence interval	95% upper confidence interval
Frequent	37.0	36.6	37.3
Moderate	37.1	36.8	37.5
Minimal	36.7	36.4	37.0

Table D6: Average employment spell duration in weeks by IWP group

The table below shows the average employment spell duration in weeks for two or more employment spells:

IWP Group	Employment spell duration	95% lower confidence interval	95% upper confidence interval
Frequent	18.9	18.5	19.2
Moderate	18.7	18.4	19.0
Minimal	18.4	18.1	18.7

Table D7: Average employment spell duration in weeks by IWP group for two or more employment spells

Subgroup results

The following tables show the various subgroup findings. They provide the average earnings progression in both £ GBP and the proportion progressing 10 per cent by subgroup for the IWP group comparisons (e.g. the Frequent support group compared to the Moderate support group).

To avoid repetition, the footnotes for each table are.

- Source: findings were based on HMRC Real Time Information (RTI) earnings data;
- Statistical significance levels for the full sample and individual subgroups are indicated as: *** = 0.1%; ** = 1%; * = 5%.
- Statistically significant differences between the subgroups are indicated as: ††† = 0.1%; †† = 1%; † = 5%.

Universal Credit: In-Work Progression Randomised Controlled Trial

It is important to note that we have undertaken multiple hypothesis tests. The more hypothesis tests that are tested, the greater the risk that we obtain an impact estimate that is statistically significant by chance when the true impact is zero. The approach we have chosen is to reduce the number of confirmatory hypothesis tests conducted by stating all tests in advance of any analysis undertaken. We dismissed another approach, namely, the Bonferroni adjustment for multiple hypothesis testing since this approach tends to reduce statistical power considerably.⁸

The table below shows the gender subgroups findings. The monetary measure shows a significant difference at the 5 per cent level of £5.99 for the female subgroup between the Frequent support group and the Minimal group (shown by the ‘*’ symbol). The proportion progressing by 10 per cent measure shows significant differences for both male and female subgroups for the Frequent support group versus the Minimal support group and the Moderate support group versus the Minimal support group.

There are no significant differences between males and females (absence of ‘†’ symbols).

⁸ Bloom Howard S, and Michalopoulos Charles, November 2010, When is the Story in the Subgroups? Strategies for Interpreting and Reporting Intervention Effects for Subgroups. MDRC.

Universal Credit: In-Work Progression Randomised Controlled Trial

Outcome	Frequent	Moderate	Minimal	Comparison	Impact	P-Value
<u>Earnings progression (£)</u>						
Full sample	10.44	9.63	5.20	Freq vs Mod	0.81	0.698
				Freq vs Min	5.25**	0.009
				Mod vs Min	4.43*	0.029
Male subgroup	10.67	10.21	6.45	Freq vs Mod	0.46	0.887
				Freq vs Min	4.21	0.177
				Mod vs Min	3.75	0.223
Female subgroup	10.28	9.21	4.29	Freq vs Mod	1.06	0.716
				Freq vs Min	5.99*	0.028
				Mod vs Min	4.93	0.061
<u>Earnings progression (%)</u>						
Full sample	45.37	44.89	42.47	Freq vs Mod	0.48	0.497
				Freq vs Min	2.90***	<0.001
				Mod vs Min	2.42***	0.001
Male subgroup	45.35	45.04	42.60	Freq vs Mod	0.32	0.781
				Freq vs Min	2.75**	0.010
				Mod vs Min	2.44*	0.019
Female subgroup	45.39	44.79	42.38	Freq vs Mod	0.60	0.525
				Freq vs Min	3.00***	0.001
				Mod vs Min	2.40**	0.007

Table D8: Significant differences between gender subgroup findings – male and female participants in the In-Work Progression trial 52 Weeks after random assignment (trial start)

The table D9 below shows the age subgroups findings. For both the monetary and proportion progressing by 10 per cent measures, there are significant differences within the 18-24 years subgroup (the Frequent support group versus the Minimal support group and the Moderate support group versus the Minimal support group) and within the 35-44 years subgroup (the Frequent support group versus the Minimal support group). There are additional significant differences with the proportion progressing by 10 per cent measure within the 25-34 years subgroup (the Frequent support group versus the Minimal support group) and within the 45 years plus subgroup (the Moderate support group versus the Minimal support group).

There is one significant difference between age subgroups, namely between the 18-24 years subgroup and the 25-34 years subgroup (the Moderate support group versus the Minimal support group) denoted by the '†' symbols. The difference of the impacts is £14.57 (£12.54 minus -£2.03).

Universal Credit: In-Work Progression Randomised Controlled Trial

Outcome	Frequent	Moderate	Minimal	Comparison	Impact	P-Value
<u>Earnings progression (£)</u>						
Full sample	10.44	9.63	5.20	Freq vs Mod	0.81	0.698
				Freq vs Min	5.25**	0.009
				Mod vs Min	4.43*	0.029
18-24 subgroup	25.53	25.98	13.44	Freq vs Mod	-0.45	0.938
				Freq vs Min	12.09*	0.024
				Mod vs Min	12.54*	0.014†
25-34 subgroup	10.80	6.50	8.53	Freq vs Mod	4.30	0.250
				Freq vs Min	2.27	0.536
				Mod vs Min	-2.03	0.577†
35-44 subgroup	6.10	5.39	-2.98	Freq vs Mod	0.72	0.888
				Freq vs Min	9.09*	0.042
				Mod vs Min	8.36	0.060
45 plus subgroup	6.75	8.40	3.60	Freq vs Mod	-1.65	0.634
				Freq vs Min	3.15	0.346
				Mod vs Min	4.80	0.154
<u>Earnings progression (%)</u>						
Full sample	45.37	44.89	42.47	Freq vs Mod	0.48	0.497
				Freq vs Min	2.90***	<0.001
				Mod vs Min	2.42***	0.001
18-24 subgroup	51.56	53.04	47.60	Freq vs Mod	-1.48	0.441
				Freq vs Min	3.97*	0.032
				Mod vs Min	5.45**	0.002
25-34 subgroup	45.74	44.50	43.15	Freq vs Mod	1.24	0.329
				Freq vs Min	2.60*	0.040
				Mod vs Min	1.35	0.296
35-44 subgroup	44.94	42.00	39.85	Freq vs Mod	2.94	0.052
				Freq vs Min	5.08***	0.001
				Mod vs Min	2.14	0.153
45 plus subgroup	42.93	43.81	41.24	Freq vs Mod	-0.89	0.450
				Freq vs Min	1.69	0.149
				Mod vs Min	2.57*	0.026

Table D9: Significant differences between age subgroup findings – age of participants in the In-Work Progression trial 52 Weeks after random assignment (trial start)

Universal Credit: In-Work Progression Randomised Controlled Trial

Table D10 below shows the Full and Live service subgroup findings. For both the monetary and proportion progressing by 10 per cent measures, there are significant differences within the Live Service subgroup (the Frequent support group versus the Minimal support group and the Moderate support group versus the Minimal support group). There is a further significant difference with the proportion progressing by 10 per cent measure within the Full service subgroup (the Frequent support group versus the Minimal support group).

There are no significant differences between Live and Full Service (absence of '†' symbols).

Outcome	Frequent	Moderate	Minimal	Comparison	Impact	P-Value
<u>Earnings progression (£)</u>						
Full sample	10.44	9.63	5.20	Freq vs Mod	0.81	0.698
				Freq vs Min	5.25**	0.009
				Mod vs Min	4.43*	0.029
Live service subgroup	12.68	11.41	6.05	Freq vs Mod	1.27	0.627
				Freq vs Min	6.64**	0.009
				Mod vs Min	5.37*	0.032
Full service subgroup	6.57	6.49	3.76	Freq vs Mod	0.08	0.993
				Freq vs Min	2.82	0.397
				Mod vs Min	2.74	0.406
<u>Earnings progression (%)</u>						
Full sample	45.37	44.89	42.47	Freq vs Mod	0.48	0.497
				Freq vs Min	2.90***	<0.001
				Mod vs Min	2.42***	0.001
Live service subgroup	46.94	46.88	43.76	Freq vs Mod	0.05	0.948
				Freq vs Min	3.18***	<0.001
				Mod vs Min	3.13***	0.001
Full service subgroup	42.67	41.38	40.29	Freq vs Mod	1.29	0.266
				Freq vs Min	2.38*	0.036
				Mod vs Min	1.09	0.331

Table D10: Significant differences between Live and Full Service subgroup findings – Live and Full Service participants in the In-Work Progression trial 52 Weeks after random assignment (trial start)

The table below shows the region subgroup findings. Note that the England sample size far outnumbers those of Scotland and Wales (27,652, 2,113 and 921 respectively) and the sample sizes for Scotland and Wales are low meaning any statistical tests will have low power and an inability to detect significant differences if they exist in reality.

Universal Credit: In-Work Progression Randomised Controlled Trial

Twenty-three cases had missing region values and were not included in the analysis below.

For both the monetary and proportion progressing by 10 per cent measures, there are significant differences within the England subgroup (the Frequent support group versus the Minimal support group and the Moderate support group versus the Minimal support group).

There are no significant differences between regions (absence of '†' symbols).

Outcome	Frequent	Moderate	Minimal	Comparison	Impact	P-Value
<u>Earnings progression (£)</u>						
Full sample	10.44	9.63	5.20	Freq vs Mod	0.81	0.698
				Freq vs Min	5.25**	0.009
				Mod vs Min	4.43*	0.029
England	10.76	9.76	5.45	Freq vs Mod	0.99	0.647
				Freq vs Min	5.31*	0.011
				Mod vs Min	4.31*	0.043
Scotland	7.06	3.98	3.56	Freq vs Mod	3.08	0.696
				Freq vs Min	3.50	0.650
				Mod vs Min	0.42	0.951
Wales	9.20	18.86	2.18	Freq vs Mod	-9.66	0.410
				Freq vs Min	7.02	0.547
				Mod vs Min	16.68	0.150
<u>Earnings progression (%)</u>						
Full sample	45.37	44.89	42.47	Freq vs Mod	0.48	0.497
				Freq vs Min	2.90***	<0.001
				Mod vs Min	2.42***	0.001
England	45.51	44.84	42.41	Freq vs Mod	0.67	0.376
				Freq vs Min	3.10***	<0.001
				Mod vs Min	2.43***	0.001
Scotland	42.29	43.72	43.84	Freq vs Mod	-1.43	0.583
				Freq vs Min	-1.55	0.554
				Mod vs Min	-0.12	0.968
Wales	48.24	49.14	41.38	Freq vs Mod	-0.89	0.839
				Freq vs Min	6.86	0.085
				Mod vs Min	7.76	0.053

Table D11: Significant differences between geography subgroup findings – England, Scotland and Wales participants in the In-Work Progression trial 52 Weeks after random assignment (trial start)

Annex E: Progression at 26 and 78 weeks

This Annex presents some of the results from the impact assessment for results at the 26 and 78-week points. As has been stated earlier in the main report, the analysis at the 52-week point is the main set of results, with the results for the 26 and 78 weeks being subsidiary.

Progression at 26 weeks

At 26 weeks after trial start, the progression in monetary terms for the Frequent, Moderate and Minimal support groups is £5.25, £3.76 and £1.92 respectively. No statistically significant impacts between the groups were detected.

Regarding the other progression measure, the percentage of participants who had increased their earnings by at least 10 per cent were 41.86 per cent, 40.51 per cent and 39.96 per cent for the Frequent, Moderate and Minimal support groups respectively. The impact between the Frequent support group and The Minimal support group (1.90 percentage points) was significant ($p < 0.01$, with confidence limits of 0.55 and 3.23 percentage points). Furthermore the impact was sustained from around week ten. In addition, no statistically significant impacts were found between groups for the percentage of participants who had *decreased* earnings by at least 10 per cent.

For the subgroups analysis, we did not detect any statistically significant differences *between* any of the subgroups for either of the progression measures.

Progression at 78 weeks

A key point to note for the progression at 78 weeks is that the majority of trial participants (60 per cent) haven't been on the trial for 78 weeks, resulting in a much smaller sample size of 12,706 available for analysis (c.f., 30,709 sample size for the progression at 52 weeks) – see Section 2.1.4, Table 2.2 and Figure 2.1. Consequently, the sample making up the analysis at 78 weeks is naturally biased towards the older duration cases, of which the majority of participants were in Live Service since Full service offices were not rolled out to the trial till later.

The sample size has reduced by some 18,000 participants, or 59 per cent compared to the 52-week progression analysis. Analysis reveals that the confidence intervals for the 78-week progression analysis are around 60 per cent wider than those for the 52-week analysis.

The reduced sample of 12,706 has much lower statistical power to detect any significant impacts at 78 weeks after trial start.

- For the full sample of 12,706, we did not detect any statistically significant impacts between the three IWP groups for either of the two progression measures.
- For the subgroup analysis, we did not detect any statistically significant impacts *within* subgroups.
- For the subgroup analysis, we did not detect any statistically significant impacts between subgroups bar a couple of examples. The examples are given below for transparency. However, it should be noted that we do not

Universal Credit: In-Work Progression Randomised Controlled Trial

place confidence in them because of the extremely low precision of the central estimates (a consequence of the reduced sample size of 12,706 that has been further sub-divided into subgroups):

- We detected one statistically significant impact *between* subgroups for the *monetary* measure, namely, between the 18-24 years subgroup and the 25-34 years subgroup (for the Moderate support group versus the Minimal support group), also observed in the 52-week analysis. Note that the impact for the 25-34 years subgroup was in a different direction (-£7.95) to the impact in the 18-24 years subgroup (£12.66). The difference of the impacts is £20.61 ($p < 0.05$). However, the 95 per cent confidence interval ranges from £1.09 to £40.97 indicating extremely low precision of the central estimate of £20.61 (the confidence interval extends around 100 per cent either side of the central estimate).
- We detected one statistically significant impact *between* subgroups for the *percentage of participants to increase their earnings by at least 10 per cent*, namely, between Scotland and Wales (for the Frequent support group versus the Moderate support group). Note that the impact for the Wales subgroup was in a different direction (8.10 percentage points) to the impact in the Scotland subgroup (-6.34 percentage points). The difference of the impacts is 14.44 percentage points ($p < 0.05$). However, the 95 per cent confidence interval ranges from 0.84 percentage points to 28.30 percentage points indicating extremely low precision of the central estimate of 14.44 percentage points (the confidence interval extends around 100 per cent either side of the central estimate).

The main charts are now shown for the 26 and 78 weeks analysis: This annex isn't as exhaustive as Annex D that contains the 52-week results.

The following two charts show the earnings progression of participants in £ GBP for the three IWP groups as a function of time before or after the trial start date, up to 26 and 78 weeks.

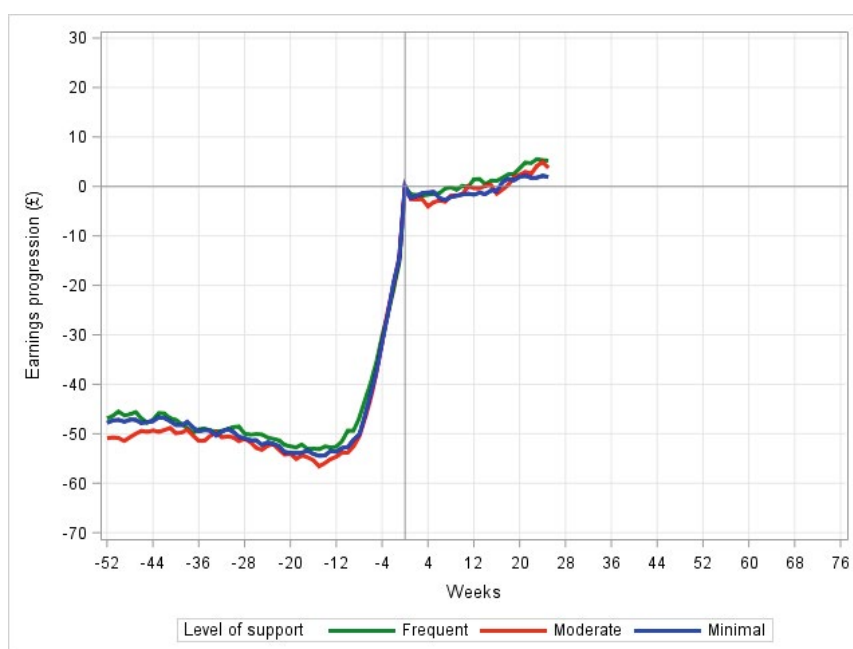


Figure E1: Average earnings progression (£) to 26 weeks (sample size of 31,471)

Universal Credit: In-Work Progression Randomised Controlled Trial

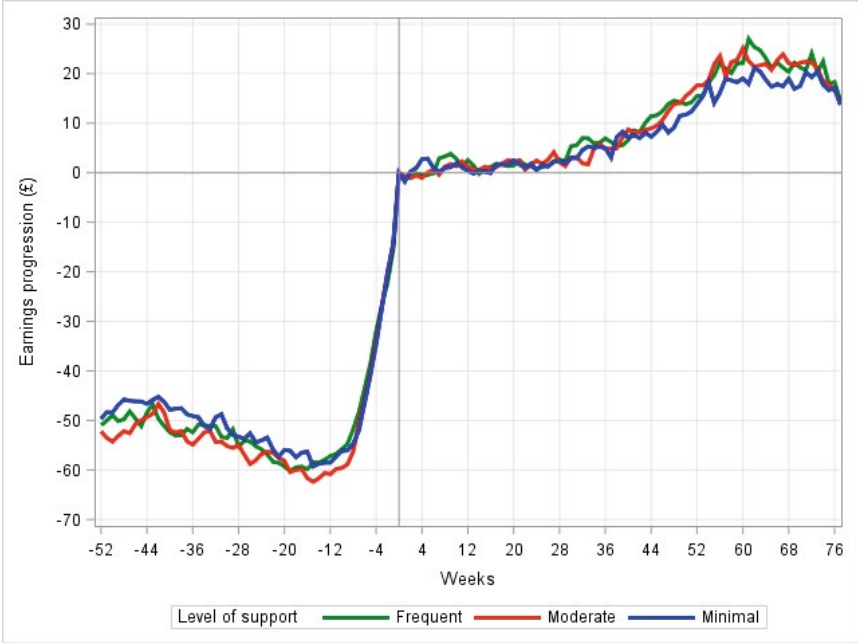


Figure E2: Average earnings progression (£) to 78 weeks (sample size of 12,706)

Note the different sample sizes used in the analysis – this is because all participants have been on the trial for 26 weeks but the majority of participants have not been on the trial for 78 weeks (see Section 2.1.4). To assess whether any differences might be significant, the two charts below show the differences between the two treatment group lines and the comparison group.

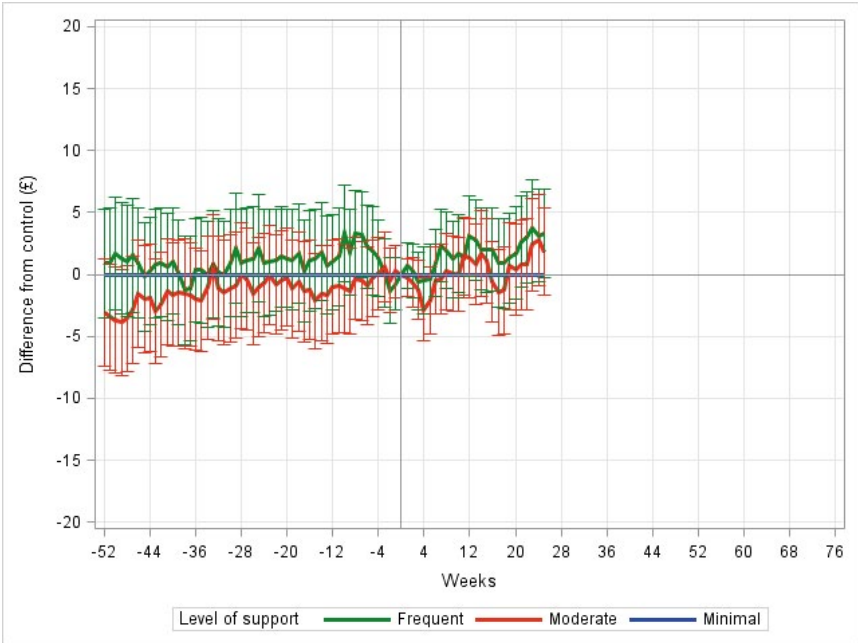


Figure E3: The difference (£) between IWP treatment groups (Frequent and Moderate) and the comparison group (Minimal) to 26 weeks

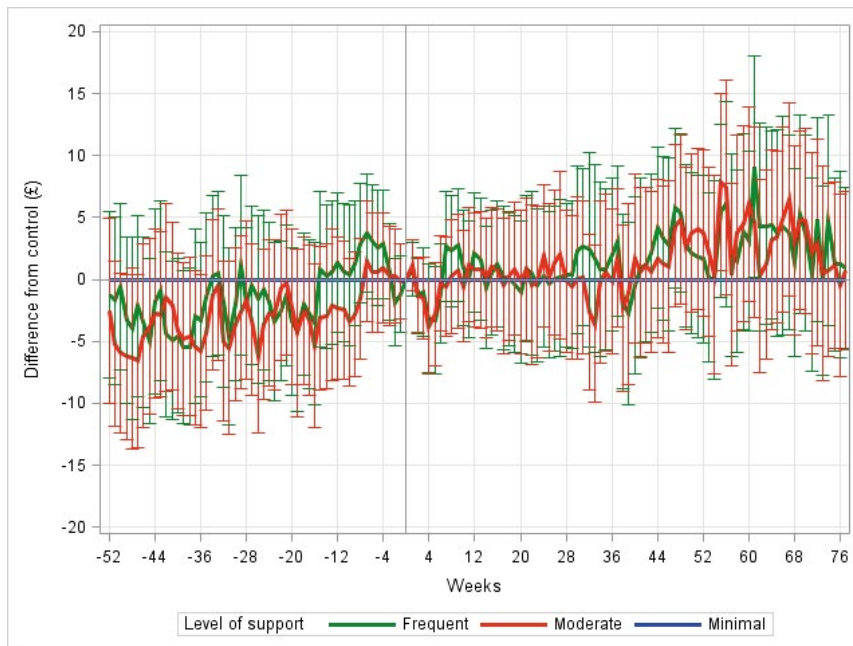


Figure E4: The difference (£) between IWP treatment groups (Frequent and Moderate) and the comparison group (Minimal) to 78 weeks

Note the confidence intervals are wider (i.e., less precision) in the 78-week analysis owing to the reduced sample size (31,471 for the 26-week analysis versus 12,706 for the 78-week analysis). The two charts above do not indicate any significant differences (impact) in monetary progression at either 26 weeks or 78 weeks.

The following two charts show the percentage of trial participants who have increased their earnings by at least 10 per cent since trial start to 26 and 78 weeks respectively.

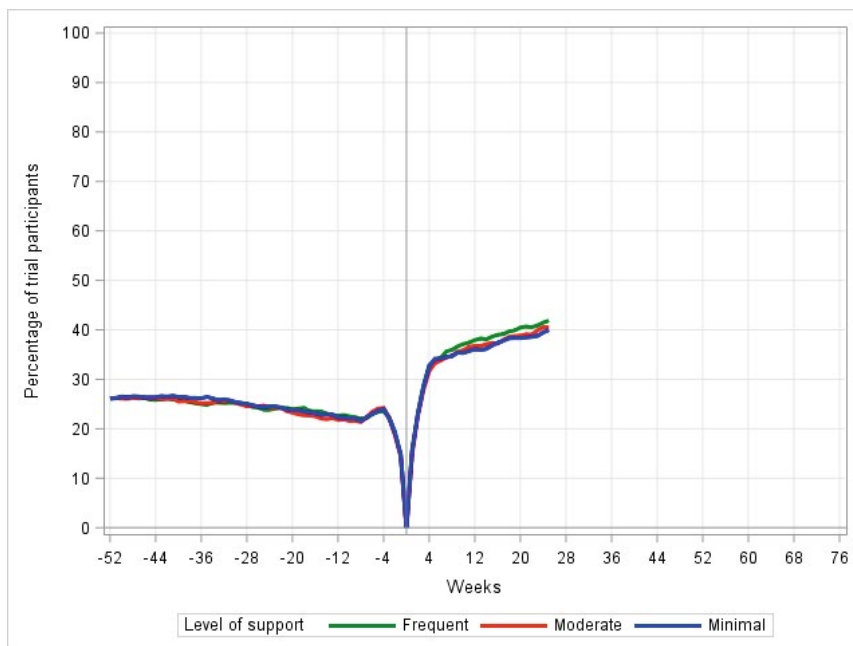


Figure E5: Percentage of trial participants who have increased earnings by at least 10 per cent since trial start (week 0) to week 26 (sample size of 31,471)

Universal Credit: In-Work Progression Randomised Controlled Trial

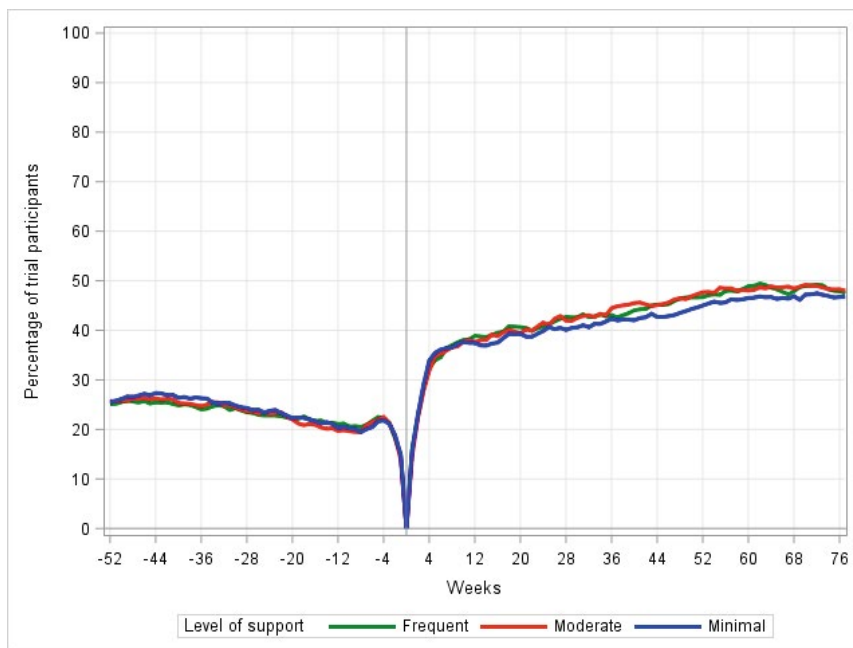


Figure E6: Percentage of trial participants who have increased earnings by at least 10 per cent since trial start (week 0) to week 78 (sample size of 12,706)

To assess whether any differences might be significant, it is more helpful to plot the differences between the two treatment group lines and the comparison group, and this is done in the following two charts.

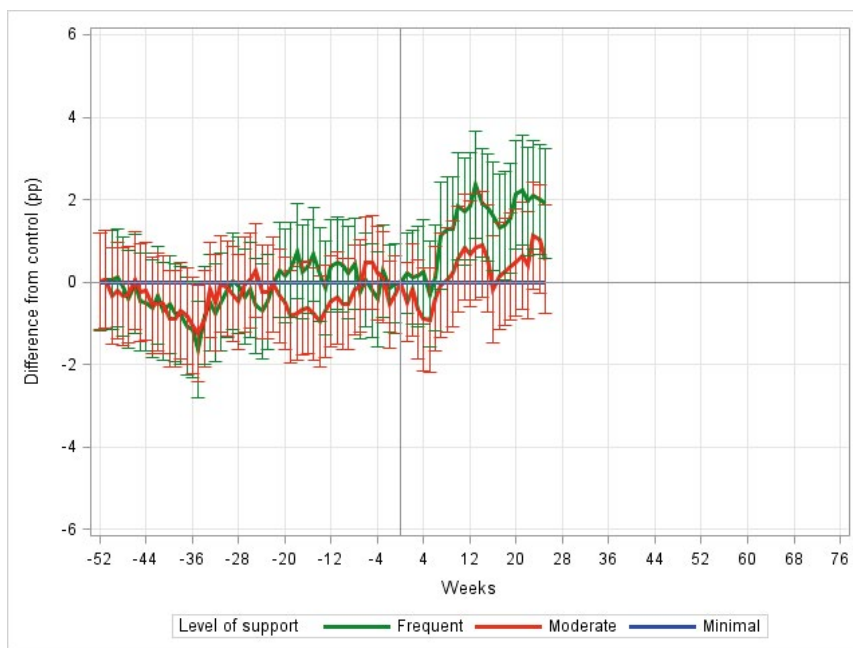


Figure E7: The percentage point difference between IWP treatment groups (Frequent and Moderate) and the comparison group (Minimal) for participants that have increased their earnings by at least 10 per cent since trial start (week 0) to week 26

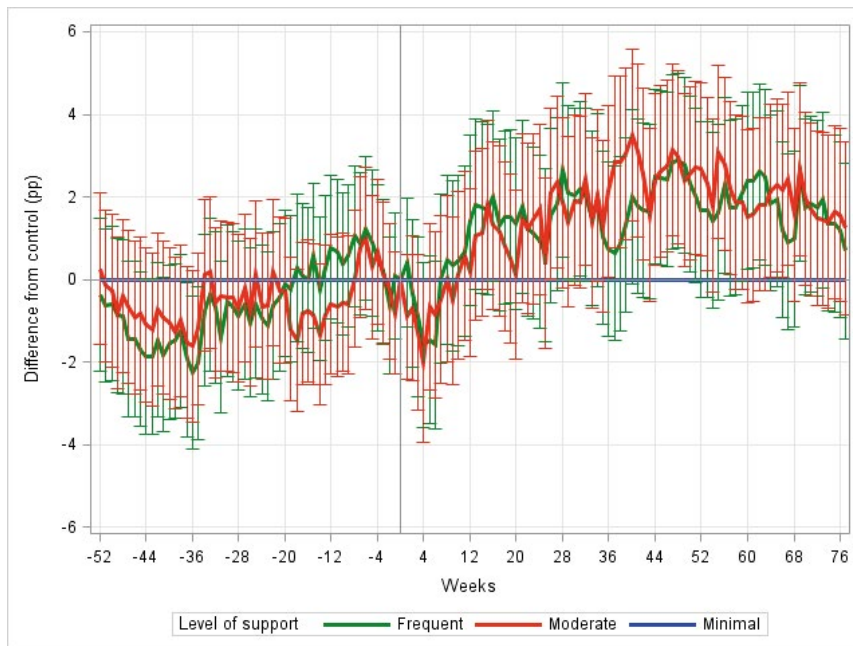


Figure E8: The percentage point difference between IWP treatment groups (Frequent and Moderate) and the comparison group (Minimal) for participants that have increased their earnings by at least 10 per cent since trial start (week 0) to week 78

At 26 weeks (Figure E7), the impact between the Frequent support group and the Minimal support group (1.90 percentage points) was significant ($p < 0.01$, with confidence limits of 0.55 and 3.23 percentage points). Furthermore, the impact appears to have been sustained from around week ten. At 78 weeks (Figure E8), the impacts do not appear to be significant. This could be due to the smaller sample size in which the noise is swamping any progression signal.

The following two charts show the percentage of trial participants who have decreased their earnings by at least 10 per cent since trial start to 26 and 78 weeks respectively.

Universal Credit: In-Work Progression Randomised Controlled Trial

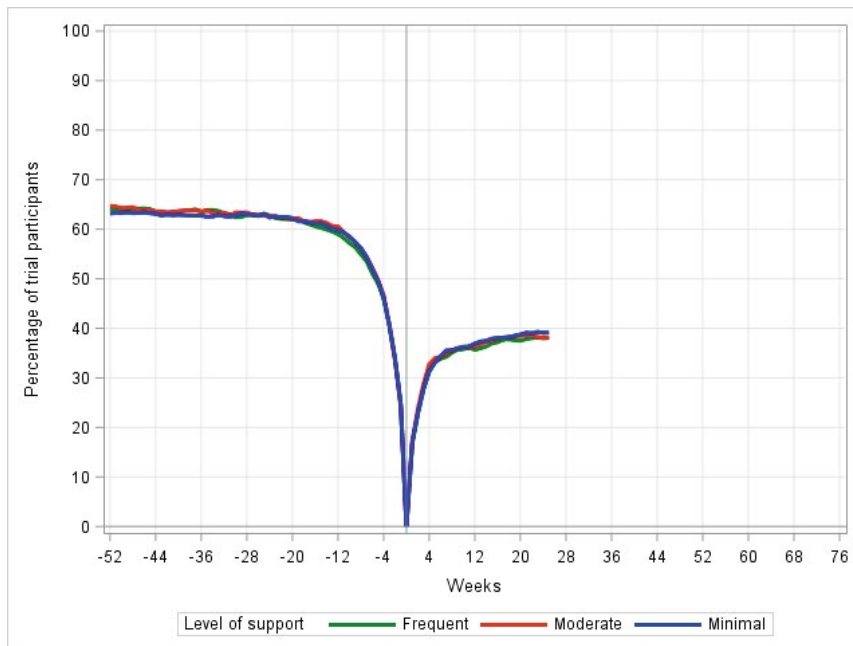


Figure E9: Percentage of trial participants who have decreased earnings by at least 10 per cent since trial start (week 0) to week 26 (sample size of 31,471)

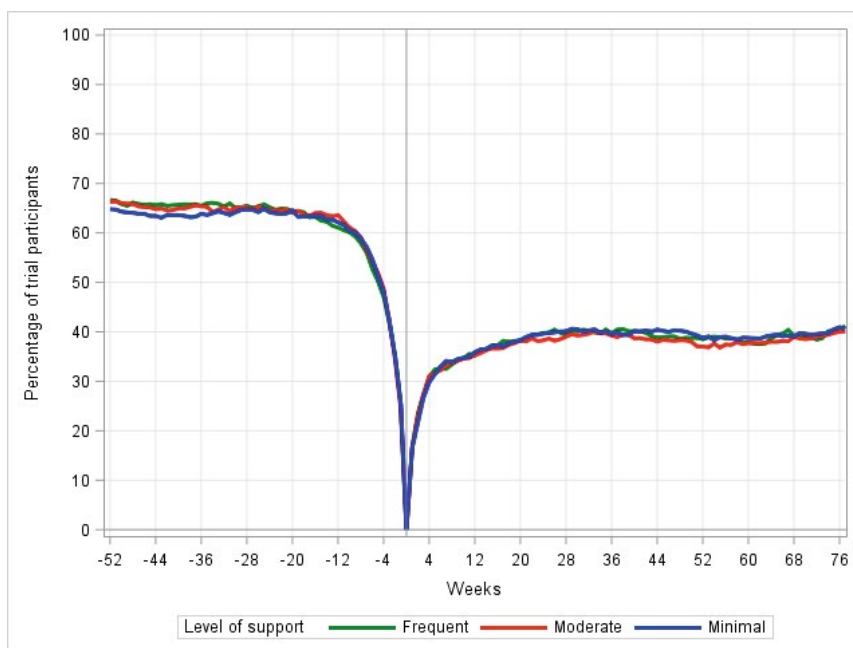


Figure E10: Percentage of trial participants who have decreased earnings by at least 10 per cent since trial start (week 0) to week 78 (sample size of 12,706)

Once again, to assess whether any differences might be significant, it is more helpful to plot the differences between the two treatment group lines and the comparison group, and this is done in the following two charts.

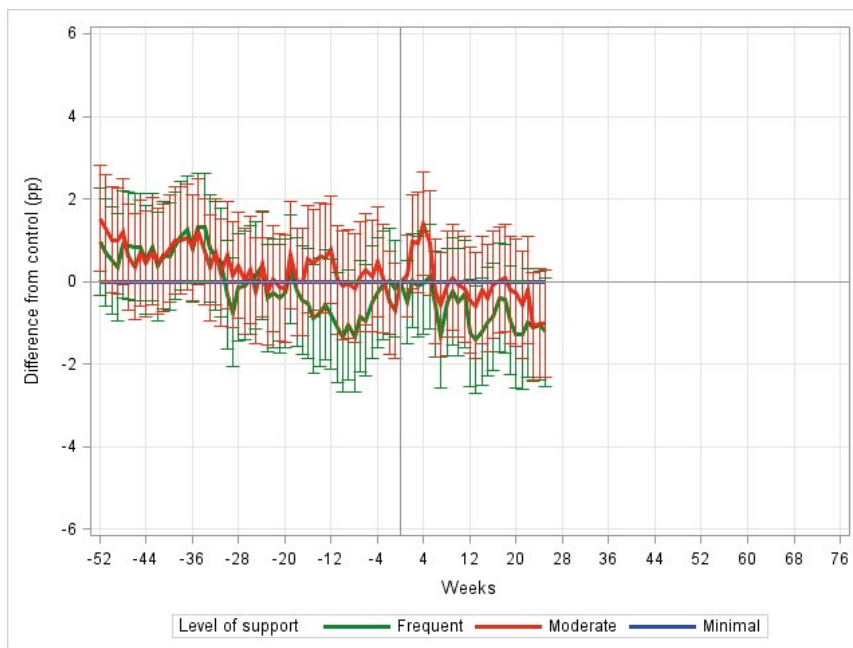


Figure E11: The percentage point difference between IWP treatment groups (Frequent and Moderate) and the comparison group (Minimal) for participants that have decreased their earnings by at least 10 per cent since trial start (week 0) to week 26

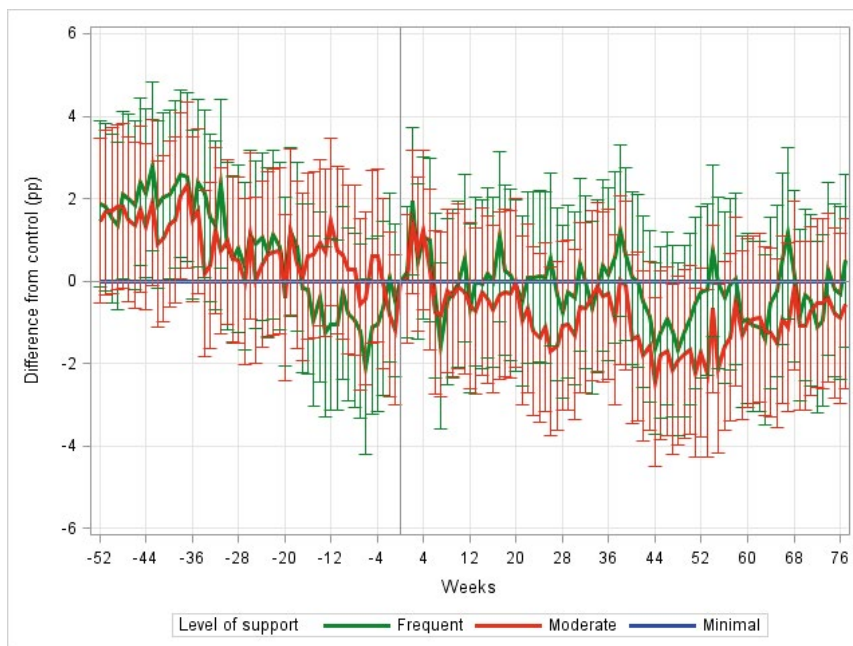


Figure E12: The percentage point difference between IWP treatment groups (Frequent and Moderate) and the comparison group (Minimal) for participants that have decreased their earnings by at least 10 per cent since trial start (week 0) to week 78

At 26 and 78 weeks (Figure E11 and Figure E12 respectively), no statistically significant impacts were observed for the percentage of participants who had *decreased* earnings by at least 10 per cent. Again, note the wider confidence intervals in the 78-week chart (Figure E12). This is due to the smaller sample size of 12,706 compared to the sample size of 31,471 used in the 26-week analysis.

References

1. Bloom HS, Michalopoulos C, (2010), When is the story in the subgroups? Strategies for interpreting and reporting intervention effects for subgroups, MDRC.
2. DWP (2018), Official statistics, benefit sanctions statistics to January 2018, Department for Work and Pensions.
3. DWP (2017), Official Statistics, In-Work Progression trial update: April 2015 to October 2016, Department for Work and Pensions.
4. Bruhn M, McKenzie D (2009), In pursuit of balance: randomization in practice in development field experiments, *American Economic Journal: Applied Economics*, 1:4, 200-232.
5. Cornish R (2007), *Statistics: Analysing repeated measures data*, Mathematics Learning Support Centre.
6. Anders J, Dorsett R, Stokes L (2018), The relative effectiveness of blended versus face-to-face adult English and Maths learning, Department for Education, NIESR.
7. Matthews JNS, Altman DG, Campbell MJ, Royston P (1990), Analysis of serial measurements in medical research, *British Medical Journal*, Vol. 300: 230-235.
8. Hesterberg T, Monaghan S, Moore DS, Clipson A, Epstein R (2003), *Bootstrap methods and permutation tests, companion chapter 18 to the practice of business statistics*, WH Freeman and Company.
9. Hesterberg T (2014), What teachers should know about the bootstrap: resampling in the undergraduate statistics curriculum.
10. Efron B. and Tibshirani RJ (1993), *An Introduction to the Bootstrap*, Chapman and Hall.