An evaluation of IPS in IAPT: The Mental Health Trailblazer Programme

Hazel Wright, Juliane Wiese, Lauren Crouch, Matthew Barnard

THE BEHAVIOURAL INSIGHTS TEAM

Contents

Executive summary

Background

In 2014, the government agreed a series of Growth Deals with local authorities across England to fund various locally led projects that addressed issues in the local economy.
One such project was the Mental Health Trailblazers programme, which responded to linked issues of unemployment and poor mental health by offering an intervention that provided support for both. The programme aimed to support unemployed clients with mental health problems back into work, and was commissioned by three areas: Blackpool, the North East and West London.

This report presents the findings of an evaluation of the programme, which provided a combination of Individual Placement Support (IPS) and Improving Access to Psychological Therapies (IAPT). IPS is an evidence-based model of supported employment designed to help people with mental health conditions enter paid employment. IPS delivery is guided by eight key principles and focuses on individual support, intensive job search guided by the jobseekers preferences, and in-work support upon entry to employment. The intervention integrated IPS support with IAPT, the provision of evidence-based psychological therapies for the treatment of anxiety and depression.

Trial aims

The IPS model has been tested previously and successfully for people with severe mental health conditions; however, it has not yet been evaluated for individuals with common mental health conditions like anxiety and depression. This evaluation aimed to develop the evidence around IPS for a wider population, specifically by testing the effectiveness of a combined model in which IPS and IAPT are offered together. The trial aimed to determine whether the intervention could be effective at improving mental health outcomes and bringing people back into work.

Trial design

This evaluation was designed as an effectiveness trial, using a two-armed randomised controlled trial design (RCT) to determine the effect of the intervention under real-world conditions. When implemented correctly, RCTs are widely considered the gold-standard in evaluations methods, as they allow a comparison of the intervention in question with a counterfactual scenario in which the intervention is *not* rolled out. This is made possible due to the randomised design, in which participants who join the trial are randomised into one of two groups: treatment (in this case IPS + IAPT) or control (IAPT). This RCT included 1,390 participants allocated to treatment, and 425 participants allocated to the control group. By randomising participants into two groups, outcomes observed in the treatment group can be

attributed to the intervention, as this group is assumed to be otherwise similar to the control group.

A number of factors affect the security of findings generated using RCTs, including the extent to which participants adhered to their treatment allocation, the extent to which outcome data was missing, and the number of participants comprising the trial sample, which affects the ability of the trial to detect a meaningful change in outcomes. Issues with sample size, data collection, and potential issues with treatment compliance are discussed later in the report as limitations to the findings.

A process evaluation, comprising interviews with 14 clients, 9 Employment Advisors and 12 referrers, was also undertaken. The evaluation took place between October 2016 and October 2018.

Outcome measures

The outcomes that were selected to evaluate the intervention fall into two categories: health and off-benefit outcomes. Although the aim of the intervention was to improve mental health and bring unemployed people into work, employment data was not readily available, and instead a dataset on benefit spells from the Department for Work and Pensions (DWP) was used, which was considered to be a proxy for employment activity.

Primary health outcomes included clinical recovery from anxiety and clinical recovery from depression. The primary off-benefits outcomes were measured as whether or not a participant had been off benefit for at least one day, measured at 3 months, 6 months, 9 months and 12 months post-randomisation.

Analyses

The outcome variables described above are binary variables; that is, they can take on one of two possible outcomes (e.g.: participant recovered or did not recover; participant was off-benefit for at least 1 day or was not off benefit for at least 1 day). Therefore, logistic regression was used to analyse the effectiveness of the intervention. By accounting for other characteristics that could influence individuals' outcomes such as gender and age, this analysis model efficiently estimates the effects we attribute only to the intervention.

Risks

This trial was delivered as an effectiveness trial to determine the effect of the intervention in a real-world setting. As such, it was not possible to ensure that all participants allocated to treatment or control perfectly complied with their allocation in practice. Given this, the evaluation results reflect both the reality of the delivery of this complex intervention in the field, and its efficacy.
Findings should be interpreted alongside several limitations, including those linked to missing data. DWP provided very complete administrative data on benefit spells; however, merging

participants' baseline data with their mental health outcomes proved to be more challenging. This resulted in very high levels of missing data and bias. Because of this risk, the conclusions drawn around mental health outcomes are less reliable than those around benefits outcomes.

Conclusions

1, When the findings were pooled across the three trials, there were no statistically significant differences between treatment and control groups for any health or off-benefits outcomes (at the $p \leq 0.05$ level). Due to high levels of missing NHS data, there is insufficient evidence to draw conclusions about the programme's impact on health outcomes; however, there was evidence suggesting that there was no overall impact on off-benefit outcomes.

2, There was some evidence of variation between programmes. The Blackpool trailblazer led to better health outcomes but poorer benefits outcomes than the programmes in the Northeast and West London.

3, There was no evidence that the programme had a different impact on participants who had been on benefits longer than average or on those with higher or lower levels of motivation.

4, A client-led approach along with a good relationship between advisors and clients were both crucial to successful provision of the programme and there were cases where this was a challenge.

5, Time pressures meant it was hard for advisors to fully engage with local employers, and the temporary or seasonal nature of some job opportunities undermined efforts to support clients into work.

1. Introduction

As part of its ambition to stimulate long-term economic growth, the Ministry of Housing Communities and Local Government (MHCLG) commissioned the Behavioural Insights Team (BIT) to conduct an evaluation of the Mental Health Trailblazers. This report presents the findings of the evaluation. Chapter one sets out the background to the initiative, describes the principles behind the intervention and discusses how the intervention was implemented within each of the three areas where it took place.

1.1 Background

The Government's 2017 Green Paper, 'Improving Lives', set out proposals for supporting people with disabilities and health conditions back into employment (DWP & DHSC, 2017). Alongside this, through the Growth Deals process, the Government funded Local Enterprise Partnerships (LEPs) for projects that targeted local economies. In doing so, the Government took steps to increase local influence over the design of work and health interventions (Local Government Association, n.d.). Based on these policies and initiatives, Blackpool, West London and the North East worked with the Cabinet Office to design models of integrated mental health and employment support with local institutional coordination. These were known collectively as the 'Mental Health Trailblazers'. In 2015, the Cabinet Office Cities and Growth Unit leading on the Growth Deal agenda was transferred from Cabinet Office to MHCLG.

The interventions drew on a wealth of evidence that demonstrated the relationship between unemployment and poor health. Unemployment has consistently been found to have a negative impact on individuals' health and psychological well-being, with often long-lasting effects (Britt, 1997; Creed, Machin, & Hicks, 1996; Creed, Hicks & Machin, 1998; Creed, et al., 1999; Paul & Moser, 2009). The characteristics of benefit claimants provided additional evidence for need, with 51% of ESA claimants in May 2018 describing their main debilitating condition as a mental or behavioural disorder (Kennedy et al., 2019). Other estimates suggest that as of 2012, almost a quarter of recent Jobseeker's Allowance (JSA) claimants had a common mental health condition, such as anxiety or depression (McManus et al., 2012).

In line with this, the Mental Health Trailblazers were designed for benefit claimants with common mental health conditions, with the aim of supporting them to find work and improve their mental health status. Each intervention aimed to integrate existing mental health support in Improving Access to Psychological Therapies (IAPT) with Individual Placement Support (IPS). All participants were out of work upon joining the trial, and in lieu of employment data, benefit receipt data was used as an outcome.

1.2 The Intervention

IPS is a vocational rehabilitation intervention for individuals who are unemployed, and who have been referred to mental health services with severe and enduring mental health conditions (Becker & Drake, 1994). It offers participants individualised support from employment specialists (ES) who work in partnership with a clinical team in secondary care to help clients manage their mental health. ES and clinical teams are intended to be integrated to enable case conferencing, information sharing and joint working. This integration is designed to allow ES to take into account the clients' health condition and source suitable roles. Together with the ES, clients explore their strengths, skills, interests and work history to find appropriate work. The model applies eight principles, and IPS services have been shown to be more effective with higher fidelity to these principles, outlined in Table 1 below (Bond et al., 2012).

Table 1: The 8 principles of IPS (Bond et al., 2008)

The 8 principles of IPS	
1	The primary aim is to find competitive employment.
2	Employment support is open to all those who want to work.
3	The types of jobs sought are consistent with individuals' preferences and strengths.
4	The job search begins quickly, within the span of four weeks.
5	Employment specialists are embedded into clinical teams.
6	The support is ongoing and aims to accommodate employees' and employers' preferences.[1]
7	Counselling on benefits is included.
8	The programme hinges on building relationships with employers.

The IPS model adopts a different approach to traditional models of employment support by focusing first on securing a role for the client, before providing support to help them sustain employment. This 'place then train' approach has been tested on a small scale with encouraging results. In 2013, RAND Europe was commissioned by the Department of Health (DH) and the Department for Work and Pensions (DWP) to conduct a feasibility pilot of the

[1] While true IPS is expected to provide ongoing support to participants, support was limited to the trial delivery period in each area during the evaluation, due to funding constraints.

integrated IPS/IAPT model. The results suggested that the model could be effective in improving mental health and employment outcomes for those with common mental health conditions; however, the conclusiveness of the results in this pilot is limited by low numbers of referrals and a high rate of attrition (Steadman & Thomas, 2015). Other studies when testing the intervention found that among certain groups, IPS reduced incidence of hospitalisation and emergency service visits, compared to matched control participants (Henry, Lucca, Banks, Simon, & Page, 2004). Finally, a meta-analysis of 30 randomised controlled trials showed IPS to have a positive impact on employment outcome measures (e.g.: obtaining competitive employment, job tenure, job length and income) and mental health-related outcome measures (e.g.: quality of life), when compared to business-as-usual employment support (Frederick & VanderWeele, 2019).

Prior to this evaluation, IPS was yet to be evaluated at scale in conjunction with Improving Access to Psychological Therapies (IAPT) programme. IAPT is an NHS funded service that began in 2008 and offers a range of evidence based psychological therapies for people with anxiety and depression in England. It is a voluntary intervention that offers National Institute for Health and Care Excellence (NICE) approved treatments for mild to moderate mental health conditions, most commonly depression and anxiety. IAPT services are commissioned by Clinical Commissioning Groups (CCGs) and provided by local NHS trusts in primary care to both employed and unemployed people with mental health conditions. Once referred, people are allocated to receive one of a number of treatments available, based on their condition. Treatments include cognitive behavioural therapy (CBT), counselling, couples therapy, interpersonal therapy, and brief psychodynamic therapy. In 2018-2019, more than 1.6 million people were referred to talking therapies, of which 89.4% of referrals began treatment within 6 weeks. 52% of referrals led to clinical recovery (NHS Digital, 2019). The main elements of all IAPT services are:

- the use of evidence-based psychological therapies - CCGs receive guidance on matching adequately trained practitioners to participants, according to the intensity and duration of their condition;
- routine outcome monitoring - this encourages transparency around patients' progress throughout their use of the services; and
- regular and outcomes-oriented supervision - this aims to ensure that a high quality of care is offered to patients.

This evaluation by BIT built on evidence generated by the original feasibility pilot conducted by RAND, with the intention of examining the effect of the integrated IPS/IAPT model when delivered at scale. Traditionally, IPS is offered on a time-unlimited basis; however, in the context of this trial, outcome data was collected only for the duration of the trial (approximately two years). This means that the results of the evaluation do not capture the potential impact that longer exposure to IPS may or may not have on clients.

Participants

The target population for this evaluation consisted of benefit recipients aged 18 and above with common mental health conditions. These were primarily those claiming Employment and

Support Allowance (ESA), as these were individuals who were not employed at the start of the trial due to illness or disability. In addition, Jobseeker's Allowance (JSA) recipients who experienced common mental health conditions were considered for the study.
The identification of a common mental health condition was initially conducted informally by referral teams at either JCP or IAPT. These teams discussed the service and its eligibility criteria with potential participants who were perceived to be suffering from anxiety or depression, before referring them to the randomisation stage of the programme. At this stage, scores for depression and anxiety measures were collected alongside benefit status and other baseline information, in order to confirm programme eligibility. The referral and randomisation process are described for each area in detail later in the report.
Finally, to be included in the study, participants had to give fully informed consent for their data to be used in analysis.

1.3 Local Intervention Models

While the integration of IPS and IAPT is at the core of the Mental Health Trailblazers intervention, there was some variation across the three areas implementing it. The different local models are discussed below.

Blackpool

Blackpool is a single unitary authority, with one CCG and one Jobcentre Plus (JCP) district. Referrals to Blackpool's IPS programme were made by advisors in JCP and the existing IAPT Service. From here, they were directed to the local authority's single point of access (SPA), which served as an administrative centre for the programme. The SPA team was responsible for making an initial appointment with a dedicated IAPT team for potential participants referred from JCP. For all participants, an initial assessment was conducted by the IAPT team over the phone to check their eligibility for the programme, before they were then referred to the IPS Employment Advisor.

The Blackpool IPS advisors and dedicated IAPT staff responsible for providing the intervention were recruited by the local authority as a new team. To facilitate the integration of IPS and IAPT, the services took place in a new site, staffed both by IPS Employment Advisors and IAPT staff, which was a feature unique to the Blackpool area. Within this design, participating staff were hired and worked expressly for the purpose of providing an integrated IPS/IAPT model within the structure of the evaluation. The model offered six months of support from the IPS team to enable the participant to find work, followed by six months of in-work support, with therapy and employment support extended to up to twelve months where possible to the end of the trial period. Those who were referred later in the programme were advised that the intervention might be shorter, and allowed to opt-out based on this information.

The North East

The Mental Health Trailblazer programme was conducted across seven local authorities in

the North East (Newcastle, Northumberland, North Tyneside, Sunderland, South Tyneside, Gateshead and Durham) and multiple CCGs. Referrals were made by JCP Work Coaches. Work coaches received: information sheets explaining the trial and their role in it, a detailed information sheet explaining the eligibility criteria for participants, and advice sheets on the referral process and how to make high quality referrals. These tips set out some of the behaviours that potentially eligible participants could display and provided answers for frequently asked questions.

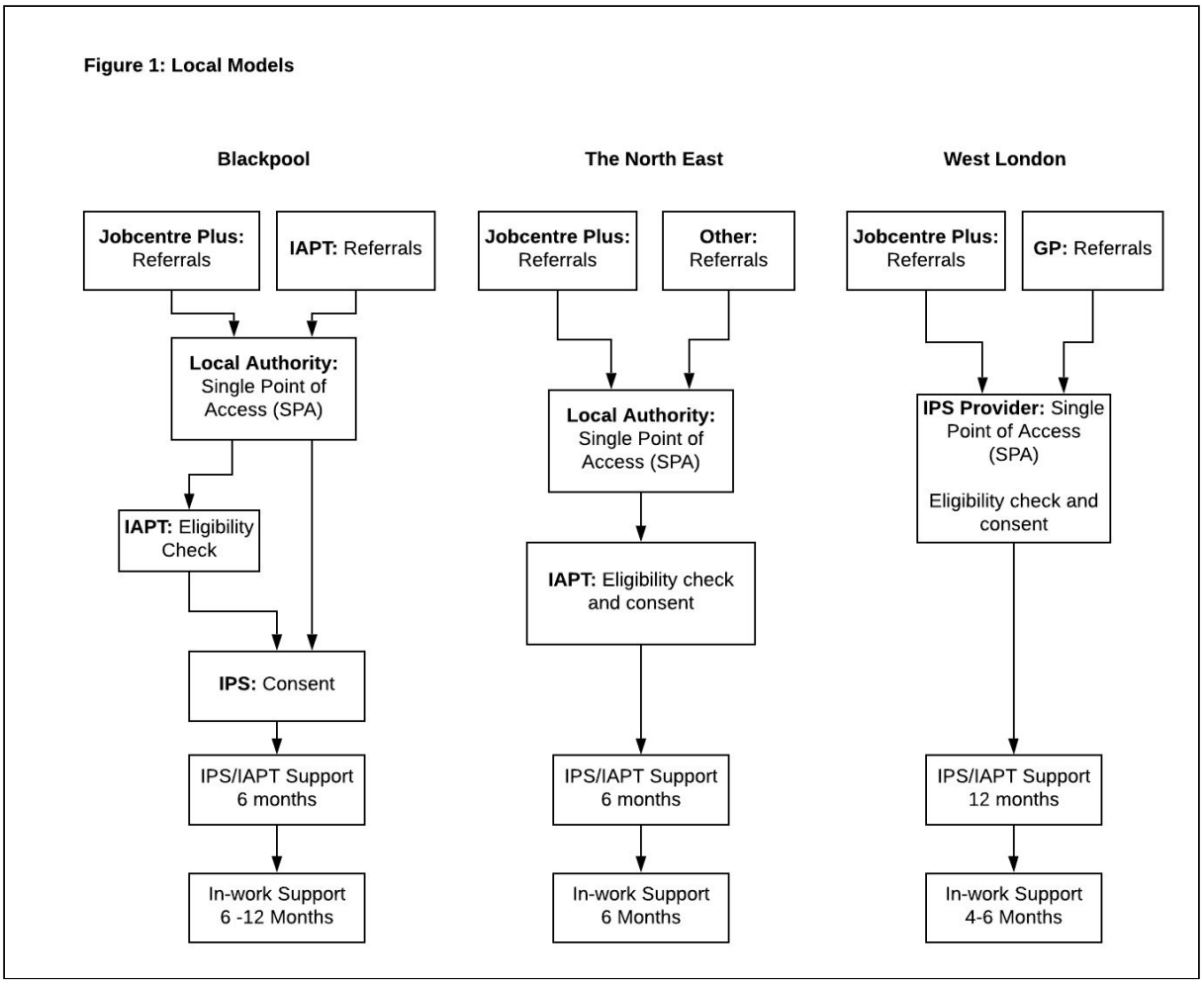
When suitable participants were identified, Work Coaches provided basic information about the programme and asked participants whether they would be happy for the SPA team to contact them to make an initial appointment with the IAPT Provider. The SPA then contacted the participant by phone to arrange an initial appointment at one of the seven participating IAPT provider premises. At the appointment, the participant learned more about the programme. If the participant was still interested in participating, written consent was taken and they were randomised into the trial.

The integrated IPS/IAPT service was provided by a new team of IPS Employment Advisors, recruited by the local authority. The team was trained and co-located with the existing IAPT providers at existing IAPT premises. The model offered six months of the intervention, followed by in-work telephone support for up to six months.

West London

The Mental Health Trailblazer was provided in seven boroughs in West London, comprising the West London Alliance (WLA): Barnet; Harrow; Brent; Ealing; Hounslow; Hillingdon; and Hammersmith & Fulham. Across these boroughs, seven clinical commissioning groups (CCGs) work to plan and commission health care services for the local areas. Referrals were generated by JCP offices across two JCP districts and local GPs, though in some cases participants who had already taken up IAPT services were also referred. All referrals were directed to the SPA, staffed by an existing IPS provider commissioned by the WLA. The integrated IAPT/IPS service was supplied by the IPS Provider and one of three existing IAPT providers. This model offered 12 months of the intervention, followed by a further four to six months in-work support. Figure 1 below summarises the service model in each of the three areas.

Figure 1: Service models across the three areas



1.4 The structure of the report

The structure of the rest of the report is as follows:

- **Chapter 2** outlines the evaluation methodology;
- **Chapter 3** presents the impact evaluation findings;
- **Chapter 4** sets out the findings of the process evaluation;
- **Chapter 5** contains the evaluation conclusions.

In addition, **technical appendices** provide more detail about the methodology and the approach that was taken to analyse the impact data.

2. Methods

This chapter describes the evaluation objectives and sets out the methodology used for the impact and process evaluations. The impact evaluation examines the success of the programme using quantitative outcome measures, testing whether the intervention has changed the rate of mental health recovery and decreased dependence on benefits. The process evaluation aims to understand the experience of participants using qualitative research methods, including interviews and focus groups with people who have engaged with or provided the intervention. This mixed methods approach allows us to assess the intervention from a variety of perspectives.

2.1 Evaluation Objectives

Impact Evaluation

The impact evaluation sought to answer the following research questions.

1. Does the IPS in IAPT model increase the off-flow from benefits of participants with common mental health conditions, compared to participants only using IAPT?
2. Do participants who cease claiming benefits remain off benefits?
3. Does the IPS in IAPT model improve mental health outcomes for unemployed participants with common mental health conditions, compared to participants only using IAPT?

In addition to these primary research questions, the relative impact of the intervention on participants with different characteristics, including gender, benefit type, initial diagnosis, initial engagement in the programme and prior benefit history, was also of interest.

Process Evaluation

The overall aim of the process evaluation was to explore the factors affecting the implementation and effectiveness of IPS, particularly the facilitators and challenges to the referral process, provision of IPS and impact of IPS on clients. The specific research questions addressed as part of the process evaluation are set out below.

1) How do staff and clients experience IPS?

a) What are the challenges and facilitators to a successful referral process?
b) What are the challenges and facilitators to receiving/providing an integrated service?
c) What are the challenges and facilitators to clients' engagement with IPS?

2) In what ways has IPS impacted clients?

a) What are the challenges and facilitators to IPS having a positive impact on clients?

IPS Fidelity Assessment

IPS Fidelity assessment is recommended by the Centre for Mental Health (CMH) as a means of understanding how closely services adhere to the 8 principles of IPS in practice, and as a way of providing operational recommendations to service teams to improve the quality of delivery.

To deliver these assessments, the evaluation team undertook training with CMH in the use of the adapted Dartmouth Supported Employment Fidelity Scale[2]. This tool translates the 8 key principles of IPS into a 25-item scale focusing on the current behaviour and activities of the IPS team and their partners. The maximum score for the scale is 125, with scores up to 99, 114 and 125 representing fair, good and exemplary provision, respectively. Services scoring 73 and below are not considered to provide true IPS. Annex A summarises this framework.
To assess scores for each item, each provider was asked to provide a range of evidence, including clinical records, vocational profiles, case notes and administrative records of team activity. The assessment included semi-structured interviews and focus groups with staff members, service users and executive team members, both from the employment and clinical side of the service.

Fidelity checks were conducted during a two-day site visit in each area between October 2017 - November 2017. Following the assessments, three members of the evaluation team compiled and discussed their evidence, and rated each item of the scale. Following this, the team held a two-hour meeting with each area lead and IPS Manager to discuss the findings and the scores awarded.

[2] The CMH scale is a slight adaption of The Dartmouth Supported Employment Fidelity Scale (Becker et al., 2008).

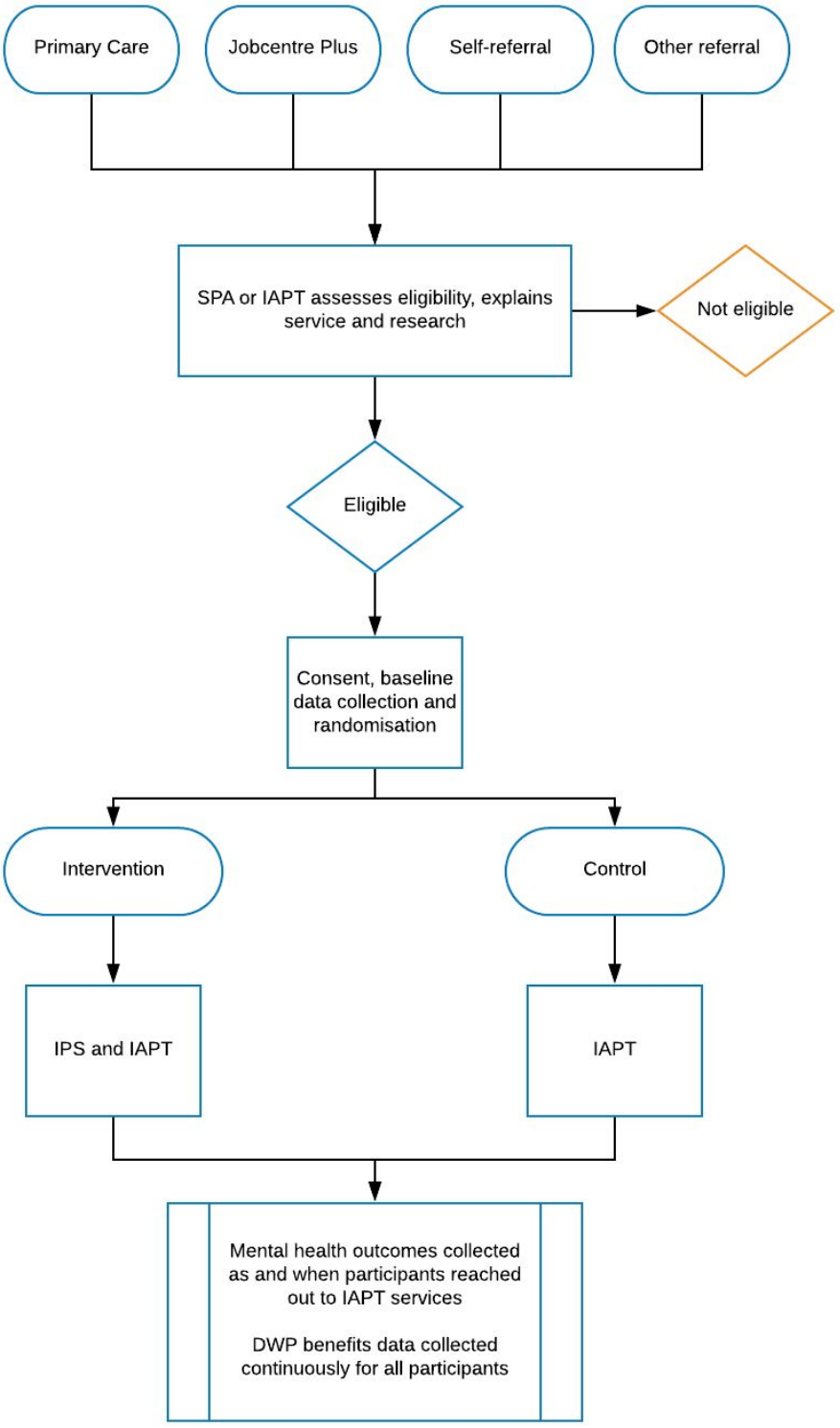
2.2 Impact evaluation

Trial Design

The evaluation in each area was conducted as a two-armed randomised controlled trial with randomisation at the individual level. The design is summarised in Figure 2 (below). Prior to randomisation, a consenting participants' eligibility was determined using a tool created by the evaluation team, which also collected baseline data from participants. This data included self-reported measures of anxiety (Generalised Anxiety Disorder-7, or GAD7 survey) and depression (Patient Health Questionnaire-9, PHQ9), and a statement of the participant's benefit status. Participants' personal contact details (including National Insurance number and NHS number for tracking purposes) were also recorded, along with information on their motivation to find employment, collected through a short survey. Once consent and eligibility were confirmed the individual was randomly allocated to one of the two groups below.

- **Control group:** The control group participants were referred to Improving Access to Psychological Therapies (IAPT) mental health support services.
- **Treatment group:** The treatment group participants were referred to IAPT, linked with employment support provided through the Individual Placement and Support (IPS) model.

Figure 2: Trial design



Participants in each group were notified of their allocation via email, along with information regarding next steps. Randomisation was managed so that there was a relatively smooth flow of participants into each arm to help services plan and provide support across the

duration of the programme. In addition, randomisation was designed so that there were more participants in the treatment groups to enable providers to meet their targets. The precise rate of allocation was dependent on total cohort size, and resulted in 1,390 participants allocated to the treatment group and 425 to the control group across the three trials.

When the trial was launched, it was expected that around 4,000 participants, a much larger volume, would participate in the evaluation. The UK labour market strengthened between 2014 and 2016, which had a substantial impact on volumes of benefit claimants eligible for the trial. Of these, 50% were expected to drop out between referral and randomisation. These estimates suggested that at least 1000 participants would be randomised in Blackpool, with 1866 and 1500 estimated for the North East and West London respectively.

Ultimately, these targets were not achieved in each area. With such a large drop in sample size, it is important to note that the statistical power necessary to detect an effect is reduced. In addition, problems around missing data, discussed in further detail below, compromise the conclusiveness of some of the results. Table 2 below sets out the randomisation process for each trial, and consort diagrams are included in Annex B.

Table 2: Sample sizes by area from randomisation to analysis.

Area	North East		West London		Blackpool	
Total randomised	1,093		435		287	
Treatment / Control	**T:** 838	**C:** 255	**T:** 347	**C:** 88	**T:** 205	**C:** 82
Mental health analysis (missing)	368 (470)	101 (154)	106 (241)	29 (59)	121 (84)	42 (40)
Benefits analysis (missing)	779 (59)	231 (24)	329 (18)	84 (4)	196 (9)	76 (6)

In the North East, the trial was designed in such a way that control participants were put on a wait list, to allay local concerns that it would not be ethical for treatment to be withheld from control group participants. This aspect of the design was signposted by researchers running a feasibility study of the IPS model (Steadman & Thomas, 2015). After nine months, participants in the control/waitlist group were automatically contacted using the contact details provided at registration, to offer them the opportunity to take part in the intervention at that stage. To re-register to take part in the intervention, participants were re-referred to the single point of access, and on providing their details were automatically referred to the IAPT and IPS service.

Outcome Measures

Reflecting the research questions outlined above, the outcome measures for the impact evaluation related to mental health outcomes and benefits outcomes. They were divided into two sets, primary outcomes and secondary outcomes, each of which are discussed below.

Primary outcomes

Primary outcome measures comprise those that most clearly indicate whether the intervention has been a success. For this study, these were the participants' recovery from their diagnosed mental health condition and entry into paid employment. However, given the availability and accuracy of benefits data collected by DWP, benefit off-flow was used as a proxy for employment outcomes. The specific mental health outcomes used in the study were:

- **clinical recovery for anxiety:** a binary outcome indicating whether or not a participant had moved into clinical recovery for anxiety;
- **clinical recovery for depression:** a binary outcome indicating whether or not a participant had moved into clinical recovery for depression.

Clinical recovery occurs if a participant's case is deemed clinical (i.e. 'at caseness,' or scoring beyond a pre-specified threshold[3]) at the start of treatment, and not clinical at the end of their treatment.

The primary outcome around benefit off-flow was measured by whether or not a participant had been off benefit for at least one day, measured at three, six, nine and twelve months post-intervention.

[3] Score of 8 on the GAD7 scale for anxiety and 10 on the PHQ9 scale for depression qualify as being 'at caseness' (Reference the IAPT handbook where these are set out).

Secondary outcomes

In addition to the primary outcomes, the study also included secondary outcomes relating to mental health and benefit status. These outcomes provided more detail about the nature of the impact of the intervention, though achieving them would not in itself be seen as sufficient for the intervention to be considered successful. The secondary mental health outcomes were:

- **reliable change for anxiety:** a binary outcome indicating whether or not a participant's anxiety has reliably changed, which is defined as a minimum decrease of 4 points on the GAD7 scale from start to end of treatment;
- **reliable change for depression:** a binary outcome indicating whether or not a participant's depression has reliably changed, which is defined as a minimum decrease of 6 points on the PHQ9 scale from start to end of treatment;
- **improvement for anxiety:** a binary outcome indicating whether or not a participant's anxiety has improved (measured via any reduction in GAD7 score);
- **improvement for depression:** a binary outcome indicating whether or not a participant's depression has improved (measured via reduction in PHQ9 score).

Please note that the two improvement secondary outcomes are not official metrics, so are reported in Annex E for completion.

There was one secondary outcome for benefits off-flow, which was:

- **remaining off benefits:** in cases in which a participant has come off benefits, a binary outcome indicating whether or not the participant has remained off benefits for 6 months or more.

To calculate the mental health outcomes, we used the GAD7 and PHQ9 scores at the point of randomisation and at time of last recorded outcome in the IAPT system, so long as this record occurred after randomisation. This decision was taken in part because many of the first recorded outcomes in the IAPT system occurred prior to or too close to the date of randomisation for the intervention to have feasibly had an effect. In the analysis, we control for the time between randomisation and mental health outcomes being collected using the covariate "time since randomisation."

Balance checks and outcome data

Balance checks were undertaken after randomisation to ensure that there were equal or almost equal proportions of individuals with relevant characteristics in treatment and control groups, as this reduces the likelihood that results are driven by mere differences in the two groups on observable characteristics. Ensuring that the two groups were equivalent allows us to attribute any differences in outcomes between the two groups to the intervention. In this case, we observed strong balance between treatment and control groups across the variables, shown in Tables 3 and 4. Because the p-values, which compare characteristics between Control and Treatment group, are all greater than 0.01, we are not concerned about

systematic differences between the two groups that could have resulted in the outcomes in our analysis. Although there is a slight imbalance in control participants in Blackpool[4], this difference is not significant at the 0.01 level[5]. In addition, we control for area in our pooled analysis to ensure that any potential area effects are accounted for.

[4] The proportion of Blackpool participants is higher in the control group than the proportion of Blackpool participants in the treatment group.
[5] P-values indicate the probability that any difference between two groups is arisen by chance, rather than a systematic difference. A low p-value suggests that there is a statistically significant difference between the two groups, meaning that the difference between groups is not a result of random chance. Thus, a higher p-value suggests that the difference between two groups on a given characteristic is likely a result of random chance, rather than any meaningful difference.

Table 3: Balance check in mental health outcomes analysis sample

	Control	Treatment	p-value
Age at randomisation *(mean)*	41.39	39.72	0.16
Female *(0=male, 1=female)*	0.45	0.41	0.41
Time since randomisation[6] *(days)*	184.30	178.45	0.70
PHQ9 baseline *(mean)*	17.05	17.57	0.21
GAD7 baseline *(mean)*	14.24	14.69	0.23
Motivation score *(mean)*	26.30	26.46	0.83
Northeast	0.60	0.62	0.61
Blackpool	0.23	0.20	0.39
West London	0.17	0.18	0.80
N	182	616	

[6] Time between randomisation and mental health outcomes being collected

Table 4: Balance check in employment outcomes analysis sample

	Control	Treatment	p-value
Age at randomisation *(mean)*	40.47	39.99	0.55
Female *(0=male, 1=female)*	0.51	0.49	0.43
Motivation score	26.65	26.56	0.87
Northeast	0.59	0.60	0.82
Blackpool	0.19	0.15	0.04
West London	0.22	0.25	0.13
N	391	1304	

Unfortunately, a portion of the sample was also excluded from the analysis due to difficulties obtaining valid outcome measures. Because the NHS and NI numbers were sometimes recorded by IPS and IAPT officers imperfectly, not all participants' baseline information could be matched to outcomes that are recorded on the NHS and DWP systems (where the NHS and NI numbers *are* recorded correctly). A higher proportion of NHS numbers could not be matched due to participants not knowing their NHS number at the point of randomisation, input errors at the same point, and administrative challenges related to matching on alternative data such as name and date of birth.

In the case of NHS data, a number of participants only had outcomes recorded prior to their randomisation date, indicating that they likely received IAPT services in the past, but did not engage with the service after the start of the trial. Other participants did not match with any IAPT outcomes at all, suggesting that they either did not engage with IAPT services, or their records could not be matched. Unfortunately, we do not have data that indicates the proportions for each however, both reasons resulted in participants being dropped from analysis, as we did not receive reliable outcomes for them. In the case of DWP data, a number of participants did not have a recorded benefits spell on record at the time of

randomisation, suggesting that they were likely employed. In both of these cases, participants were excluded from analysis, as they either did not have valid outcome measures or were likely not eligible for the trial. Approximately 58% of participants were excluded from the mental health analysis, and 7% of participants were excluded from off-benefits analysis. Further details can be found in the consort diagrams in Annex B.

Analysis

Analysis was conducted following the standard convention for a multiple regression model. The rationale for using multiple regression is that it is the most efficient way of identifying differences in control and treatment groups that are purely due to the intervention. This is because the regression ensures that any differences due to participant characteristics (known as 'covariates') are accounted for. Participant characteristics in this analysis include things such as age, sex and motivation at baseline, amongst other factors that are likely to influence participants' outcomes. We report our findings in an interpretable way (i.e.: as proportions), and explain in Annex C how these figures can be obtained from the regression coefficients (i.e.: log-odds ratios).

When pooling the data to look at the findings across the three sites, tests for homogeneity (similarity) were carried out, which indicated that most of the findings from each individual trial were sufficiently similar that it was appropriate to pool the data. There was some indication that the results from Blackpool were different from the other two areas, and this is discussed in the findings below.

For every estimation made, there is a chance that a statistically significant result will be found by chance, i.e. a 'false discovery' will be made. The more statistical comparisons that are made, the greater the chance of making such a false discovery. To account for this, the level of confidence that we estimate for a finding can be adjusted based on the number of comparisons made. We use the Hochberg step up procedure, which increases the burden of proof for each additional comparison. Where this has altered the results of the significance test, this has been indicated. Further details of the regression models and output tables can be found in Annex C.

Limitations

The evaluation team faced numerous challenges in gathering relevant and complete outcome data for the analysis sample. These challenges have reduced the size of the analysis sample and compromised the extent to which our analyses are powered to detect a significant result. Broadly, these challenges can be broken into two categories: design-based challenges and logistical challenges.

Design-based challenges

During the design stage of the trial, the DWP estimated that approximately 4,000 individuals would participate in the trial across the three areas between both treatment and control

groups. Over the course of the evaluation, much lower volumes of participants came through the services. The trial was designed to target unemployed participants who were already eligible to engage with IAPT services, to ensure the trial sample would be representative. The drawback to this design was that the evaluation team could not intervene with recruitment to boost the number of participants flowing into the evaluation. The purpose of the design was to evaluate a sample of participants engaging with the services in such a way that is as close to reality as possible. Ultimately, the final sample included 1,815 participants across both treatment arms. The NHS trusts returned records for 1,473 participants; however, only 767 of these had valid PHQ9 or GAD7 outcomes with which we could calculate recovery. In contrast, the DWP returned valid benefit outcomes for 1,695 participants.

Because mental health outcomes were collected only as and when participants had an IAPT appointment, rather than prompted at regular intervals (e.g.: calling each participant 3 months after randomisation), the analysis sample (i.e.: participants for whom outcomes are collected) relied on participants using those services to collect outcomes. This is problematic because it is not unlikely that there could be an association between treatment assignment and reaching out to these services. This could feasibly be the result of the programme working or not working. Participants with improved mental health may not feel the need to reach out to services; conversely, participants whose mental health has not improved may not feel that it is worth contacting services or may lack the motivation to do so. We do not have any data to indicate which, if either, of these explanations is accurate, or what the relative proportion is of each if they are both accurate. Nevertheless, this challenge may provide some insight into why the recovery rates across both treatment and control groups were found to be much lower than the average recovery rate for IAPT; the gaps in data completion, as described above, strongly suggest that the analysis sample is not representative of the target population. However, tests indicated that the attrition due to missing health data did not unbalance treatment and control groups. We ran a regression of missing health outcomes on treatment allocation (including covariates from our primary analysis model) and found no statistically significant relationship between treatment allocation and missing data. The issue of attrition did not affect the benefits outcomes, as DWP provided nearly complete administrative data on benefit status.

Logistical challenges

The analysis sample also relied on access to and successful merging within 11 different NHS Trusts/IAPT services using a participant's NHS number[7], which led to a number of issues. Firstly, it meant that where participants' NHS or NI numbers were recorded incorrectly at the point of randomisation, they could not be matched with their outcome data. Successful matching across datasets required accurate NHS or NI numbers to be linked to the IAPT and DWP databases, respectively. As a result, participants with unmatched NHS or NI numbers were dropped from the analysis.

[7] Participants consented to this procedure at the point of randomisation, and only anonymised data was returned for analysis

Secondly, because the trial was launched several years prior to analysis, many of the trusts had lost contact with the trial and were hesitant to share data, resulting in major delays in merging the dataset of participants' baseline scores with their mental health outcomes. In many cases the NHS staff originally involved in the trial had moved to other jobs and remaining staff had no relationship to this project; in other cases the staff were concerned about data security and confidentiality. In these cases, new Data Sharing Agreements were put in place where possible to ensure clarity around data anonymity and compliance with GDPR.

Thirdly, once data sharing was agreed, the variability with which data is stored across IAPT services meant that our choice of outcomes was limited and therefore reduced to using the most recent PHQ9 or GAD7 score on record, and comparing this score to the baseline scores that were recorded at point of randomisation. Because the results are analysed as intent-to-treat, it retains participants in the analysis sample based on their randomisation assignment regardless of whether or not they engaged with the treatment. For this reason, we do not exclude from analysis participants for whom only one outcome measure (after randomisation) is recorded, though this may represent the first assessment they received upon entry to IAPT.

2.3 Process Evaluation

In addition to the impact evaluation, a process evaluation was conducted as part of the project to explore the perceptions and experiences of staff and participants using a mixture of focus groups and interviews.

Sampling

A purposive sampling approach was used to select a sample that was representative in terms of range and diversity of relevant characteristics (Ritchie et al., 2013). In terms of participants, the sample included those who had experienced a range of mild to moderate mental health difficulties (including anxiety, depression and stress), and who had been referred via JCP and IAPT services. Referrers selected occupied various job roles, including: cognitive behavioural therapists, psychological wellbeing practitioners, counsellors and team managers from IAPT; Disability Employment Advisors (DEAs), Work Coaches and managers from JCP; and safe care leads from Primary Care Mental Health Services. Participants were selected from across all three locations, to gain further understanding of the differences and similarities in participant's experiences across each location. Details of the sample characteristics are shown in Table 5.

Table 5: Demographic information of the participant sample

	Client	Employment Specialist	Referrer	Female	Male
London	5	2	2	6	3
Blackpool	5	4	6	11	4
Newcastle	4	3	4	6	5

N= 35

Recruitment

Participants were contacted through IPS providers at each location and informed consent was obtained from all staff and clients who agreed to take part in the interviews and focus groups.

Data collection

Semi-structured in-depth interviews were conducted over the phone or face to face, depending on the participant's preference and availability. Focus groups were conducted, and clients were offered a £10 Amazon voucher as a thank you for their participation in the research. Clients were also reimbursed for their travel expenses to the site. The interviews and focus groups were carried out at the IPS provider sites.

Analysis

Interviews were transcribed and uploaded to a qualitative coding programme to simplify the process of organising and managing the data. A coding framework was created and extracts of the transcripts were coded to identify and categorise participants' responses. The process of thematic analysis ensured participants' experiences were organised into themes and sub-themes that represented the differences and similarities in participants experience of IPS. A coding journal was used to map and summarise participants' experiences, and document supporting quotes within the themes. Researchers regularly met to revise the coding framework and thematic structure to produce further insight and interpretation of the qualitative findings. The emergent themes were consistently reviewed and refined in line with the research questions.

2.4 Ethics

The project was approved in October 2016 by an NHS Ethics Committee. As randomisation took place at the individual level, written consent was taken from participants prior to randomisation. A participant information sheet, provided with a consent from, gave

information on the aims of the evaluation and the use of data to allow participants to make an informed decision regarding consent for data sharing. The participant consent and information sheet are available on request.

3. Impact Evaluation

This chapter sets out the findings of the impact evaluation, firstly presenting the results of the analysis of primary outcomes before moving on to presenting the analysis of the secondary outcomes. The sample of participants included in the analysis were on benefits at the time of randomisation, gave consent for their data to be used in analysis, and had accurate NHS and/or NI numbers entered into the randomisation tool, resulting in a successful match with outcome data. All regression tables can be found in the annex.

3.1 Primary analysis

For both health and benefit outcomes, the findings based on data pooled from the three individual trials are first set out, followed by the findings for each individual area.

Health outcomes

Based on the pooled data across all three sites, the evaluation was unable to detect an impact of the intervention on the health outcomes. For anxiety, there was a small but non-significant increase in the percentage recovering when the treatment is compared with the control, with the opposite effect for depression, i.e. there was a small but non-significant decrease in the percentage recovering from depression. This means that there were slight differences in anxiety and depression outcomes between treatment and control groups, but they cannot be attributed to the intervention. In addition, we recognise that the recovery rates in both treatment and control groups are much lower than the historical recovery rate recorded for IAPT (approximately 51%, Clark 2019). This provides further evidence that the missing data, described in the Limitations section above, has resulted in selection bias in our analysis sample.

Table 6.1: Health outcomes across all sites

Recovery from anxiety	763	25.6	28.6	3.0	No
Recovery from depression	759	33.7	31.0	-2.7	No

When we look at the individual sites, there are striking differences between the North East and West London, on the one hand, and Blackpool on the other. We demonstrate this with forest plots, which display in one figure the results of different studies that attempt to answer the same question. This allows us to compare the extent to which results vary or differ across sites. The forest plots show odds ratios[12] for each area and the pooled data; the lozenges mark the odds ratio, with the associated lines indicating the confidence intervals[13]. The size of the squares around the lozenges indicate the sample size of each study and therefore how much the study contributed to the pooled estimate (its weight).

[8] This column denotes the raw control mean of the specification, meaning that it captures purely the percentage of participants in the control group who have recovered from anxiety and depression.
[9] This column denotes the control mean plus the effect size, which is the marginal effect that is calculated from the treatment coefficient in the logit model.
[10] This column captures the marginal effect that is calculated from the treatment coefficient in the logit model. Because the logit model outputs the log of odds ratios, which is difficult to interpret, we calculate the marginal effect of the treatment. More details on how this is done can be found in the Annex.
[11] If a treatment effect is significant at the 5% level, then we are confident that differences between the treatment and control groups are attributable to the treatment. If not, then it is likely that any differences are attributed to unexplained variability in the data.
[12] The odds ratio, in this case, estimates the strength of the correlation between the treatment and the desired outcome (i.e. recovery in anxiety and depression).
[13] The confidence intervals measure the most probable range of values that the odds ratios could take on. The wider the confidence interval, often resulting from small sample sizes, the less certain we can be about the estimate (as the confidence interval covers a wider range of possible values that the real odds ratio could take on).

Figure 3: Anxiety recovery forest plot across the three trial sites

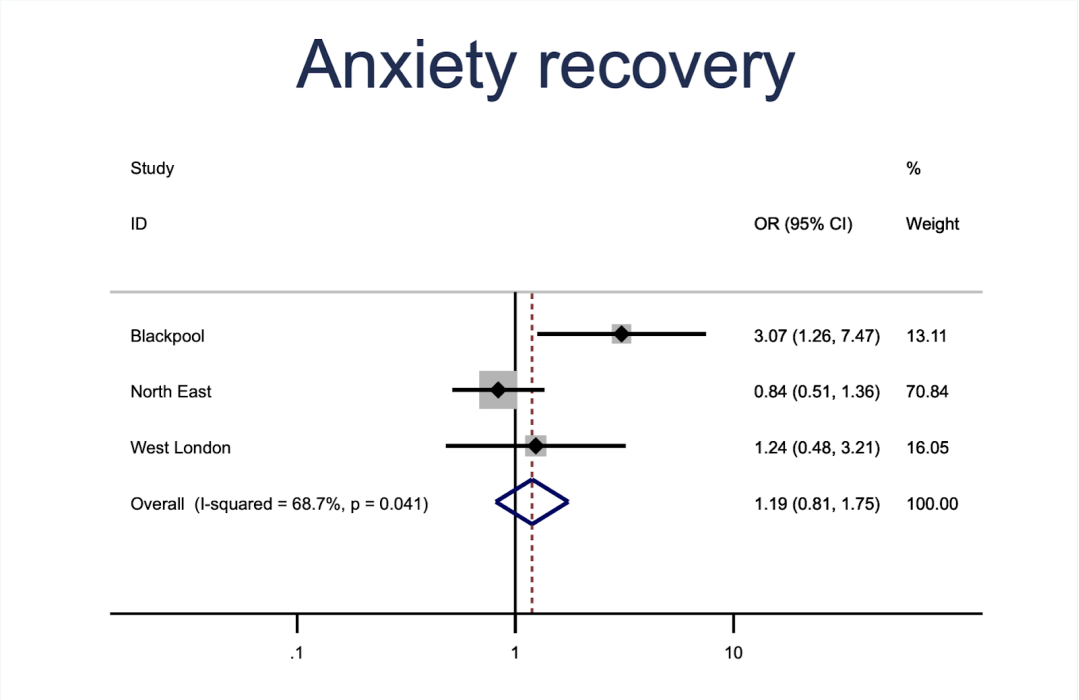
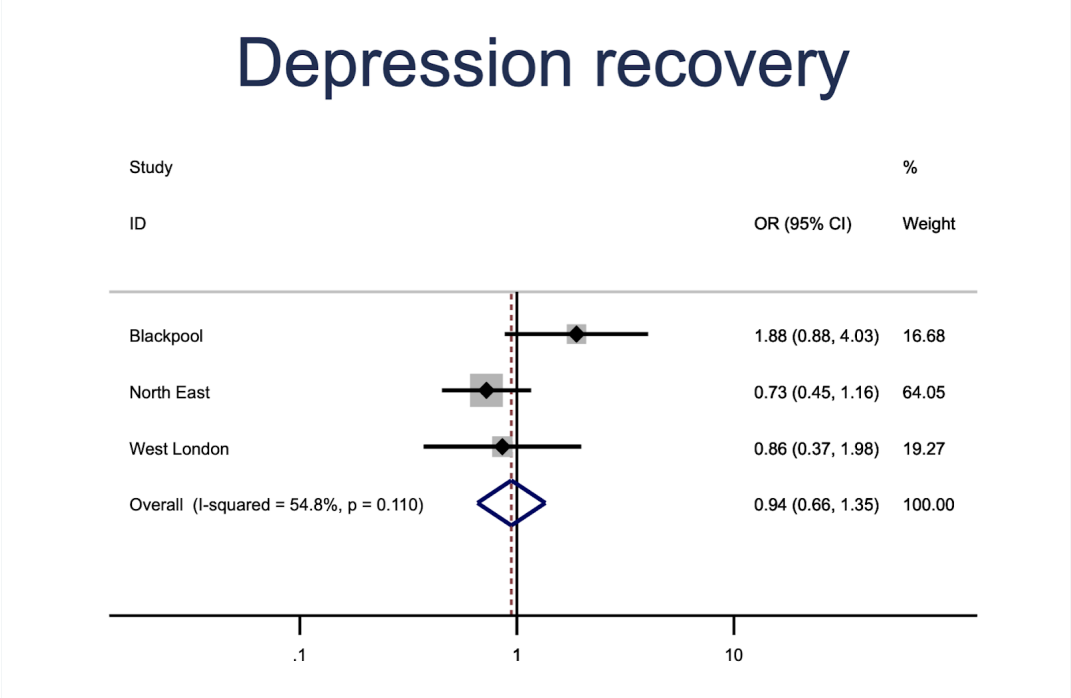


Figure 4: Depression recovery forest plot across the three trial sites



The forest plots illustrate that the North East and West London reflect the overall finding of no significant difference between control and treatment groups for anxiety and depression (with the North East recording small but non-significant decreases in the percentage of people recovering and West London recording small but non-significant increases in the percentage of people recovering). In contrast, for Blackpool there was a large and significant improvement in recovery for anxiety and a large, but non-significant improvement for depression. The health outcomes expressed as percentages for recovery for each site are set out in the tables below.

Table 6.2: Health outcomes for individual sites

North East					
Recovery from anxiety	469	29.7	24.2	-5.5	No
Recovery from depression	469	33.7	23.9	-9.8	No
West London					
Recovery from anxiety	117	24.1	26.3	2.2	No
Recovery from depression	121	41.4	41.4	0.0	No
Blackpool					
Recovery from anxiety	156	16.7	42.4	25.7	Yes
Recovery from depression	140	28.6	47.3	18.7	No[14]

Benefit outcomes

For benefit outcomes, the pooled data across the three sites indicate that there were no statistical differences between the intervention and control groups in terms of the percentage of participants who were off benefit for at least one day. At each time point, the percentage of

[14] Although the original regression output suggests that this difference is significant at the 5% level, after adjusting for multiple comparisons we find that it is not. We discuss this adjustment in the Analysis section under 2.2 Impact Evaluation.

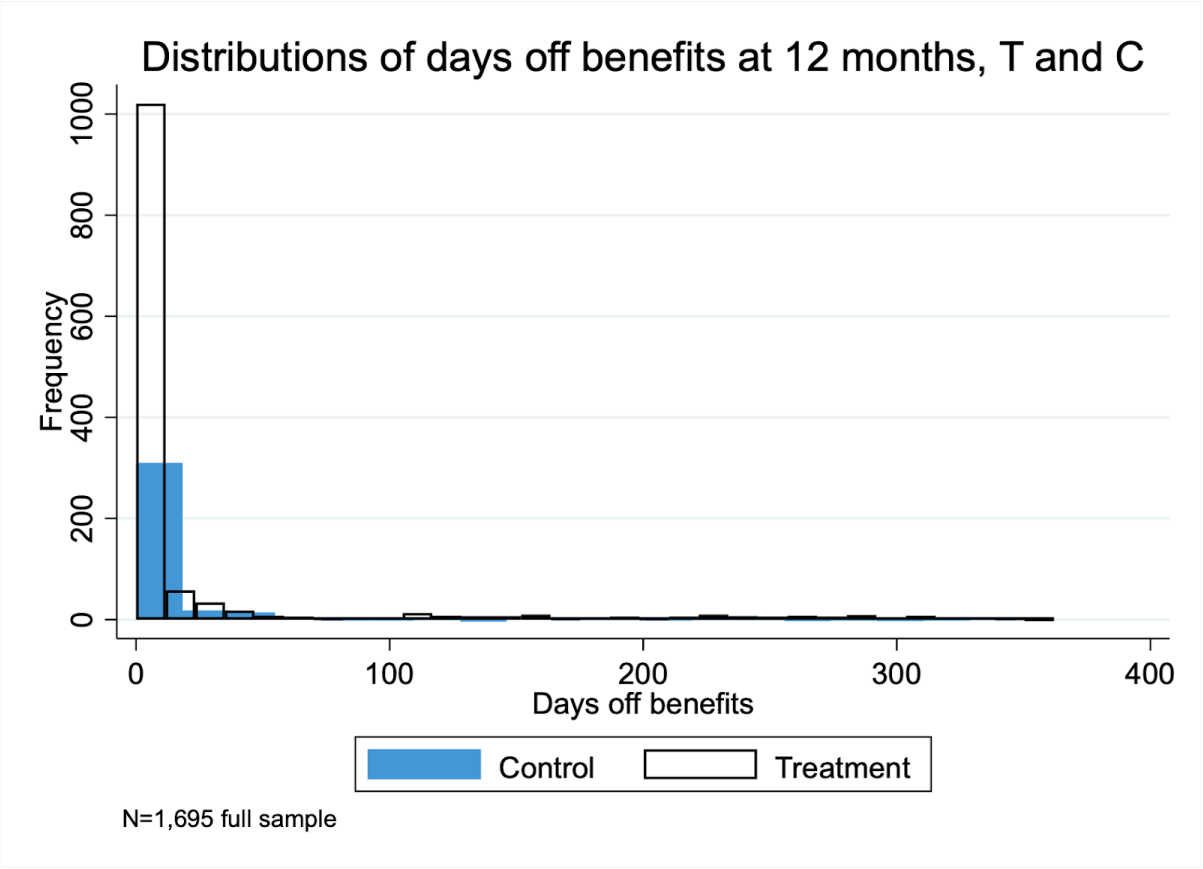
the treatment group who were off benefits for at least one day was slightly less than the control group, though this was not statistically significant.

Table 6.3: Benefit outcomes across all sites (percentage off benefit at least one day in the given time period)

At 3 months	1692	12.3	9.8	-2.5	No
At 6 months	1692	19.4	17.4	-2.0	No
At 9 months	1692	25.6	23.0	-2.6	No
At 12 months	1692	27.4	25.3	-2.1	No

Unlike for the health outcomes, the results for all individual trials were consistent with the overall pooled finding, in that there were no statistically significant differences between treatment and control groups at any time point in any site. In the Northeast, the percentage off benefits was slightly lower than for the control group (but was not statistically significant) and for West London the opposite was true. Blackpool, which looked very different to the other two sites for health outcomes, was more in line with the overall result, with no statistical difference between treatment and control groups. The histogram at Figure 5 sets out the distribution of outcomes overall, and supports the overall findings, showing that the majority of clients randomised into the program did not leave benefit for any length of time.

Figure 5: Distribution of number of days off benefits at the end of the evaluation, for treatment and control groups.



However, over the whole time period of data collection, the treatment group had fewer people off benefits for at least one day, and the gap between treatment and control increased over time without reaching statistical significance. This is somewhat counter-intuitive as it is the opposite result to that seen in the health outcomes, i.e. Blackpool seems to have better mental health outcomes than the other two sites but poorer off-benefit outcomes (though in terms of off-benefit outcomes the difference is not statistically significant).

Table 6.4: Off-benefit outcomes for individual sites (percentage off benefit at least one day in the given time period)

Northeast						
	At 3 months	1008	12.6	10.1	-2.5	No
	At 6 months	1008	19.5	16.8	-2.7	No
	At 9 months	1008	24.2	22.0	-2.2	No
	At 12 months	1008	26.0	24.7	-1.3	No
West London						
	At 3 months	395	8.3	9.3	1.0	No
	At 6 months	412	15.5	20.6	5.1	No
	At 9 months	412	22.6	29.3	6.7	No
	At 12 months	412	23.8	31.2	7.4	No
Blackpool						
	At 3 months	259	15.8	11.3	-4.5	No
	At 6 months	259	23.7	17.8	-5.9	No
	At 9 months	272	32.9	21.4	-11.5	No
	At 12 months	272	35.5	23.5	-12.0	No[5]

[15] The sample size noted in this table represents the sample that is included in the regression. In a logit regression with a binary outcome measure, an observation is dropped if its covariate predicts success or failure perfectly (e.g. all people randomised in a specific month and year were not off benefits for at least one day).The covariate's lack of variation provides no predictive power for the outcome, and the observations with this perfectly predicting covariate are therefore dropped from the regression.

3.2 Secondary analysis

As with the primary outcomes, the findings based on data pooled from the three individual trials are set out first, followed by the findings for each individual area.

Health outcomes

The findings for the health secondary outcomes match the findings for the primary outcomes, with no statistical differences between control and treatment groups. Reliable improvement requires a greater degree of change between baseline and follow up in scores than simple improvement, so it is not surprising that absolute percentages and the differences between control and treatment groups are greater for improvement. However, even for improvement, neither change for anxiety or depression is statistically significant and, as previously discussed, results on improvement are not standard IAPT outcome measures.

Table 7.1: Secondary health outcomes across all sites

Anxiety					
Reliable improvement: decrease of 4 or more points in GAD7	763	37.8	37.8	0.0	No
Improvement: decrease of 1 point or more in GAD7	763	69.2	61.8	-7.4	No
Depression					
Reliable improvement: decrease of 6 points or more in PHQ9	763	37.8	36.0	-1.8	No
Improvement: decrease of 1 point or more in PHQ9	763	65.1	68.6	3.5	No

For the individual sites, the pattern reflects that found for the primary outcomes, which is that the Northeast and West London are in line with the pooled findings with no significant differences, but that Blackpool bucks the trend with large differences between control and treatment, with the difference being statistically significant in the case of reliable improvement in anxiety. This reinforces the impression that there is a difference between Blackpool and the other two areas in terms of the mental health outcomes achieved.

Table 7.2: Secondary health outcomes for individual sites

North East					
Anxiety					
Reliable improvement: decrease of 4 points or more in GAD7	469	46.5	35.1	-11.4	No[16]
Improvement: decrease of 1 point or more in GAD7	469	71.3	59.6	-11.7	No[7]
Depression					
Reliable improvement: decrease of 6 points or more in PHQ9	461	38.6	31.7	-6.9	No
Improvement: decrease of 1 point or more in PHQ9	462	68.3	68.3	0.0	No
West London					
Anxiety					
Reliable improvement: decrease of 4 points or more in GAD7	117	24.1	34.0	9.9	No
Improvement: decrease of 1 point or more in GAD7	130	70.0	66.0	-4.0	No
Depression					
Reliable improvement: decrease of 6 points or more in PHQ9	122	37.9	38.6	0.7	No
Improvement: decrease of 1 point or more in PHQ9	121	69.0	77.3	8.3	No
Blackpool					
Anxiety					
Reliable improvement: decrease of 4 points or more in GAD7	152	26.2	57.4	31.2	Yes
Improvement: decrease of 1 point or more in GAD7	146	64.3	73.1	8.8	No
Depression					
Reliable improvement: decrease of 6 points or more in PHQ9	152	35.7	49.5	13.8	No
Improvement: decrease of 1 point or more in PHQ9	144	54.8	76.1	21.3	No[7]

[16] Although the original regression output suggests that this difference is significant at the 5% level, after adjusting for multiple comparisons we find that it is not.

Benefit outcomes

There were very few people off benefit at the end of the trial. For the minority who left benefit, very high proportions in both control and treatment had been off benefit for more than six months. Given this, we do not look to regressions for reliable results; instead, we report the raw proportions of participants who were off benefit at the end of the trial and remained off benefits for 6 months. In the full sample, 92.7% of the control group remained off benefits for 6 months or more if off benefit at the end of the trial, whereas 88.8% of the treatment group had the same result. Given the small sample size, we cannot draw conclusive evidence around the intervention's impact on long-term benefit off flow.

Table 7.3: Secondary employment outcomes - raw means

Full sample	207	92.7	88.8	-3.9
Northeast	89	96.2	84.1	-12.1
West London	72	93.3	93.0	-0.3
Blackpool	46	85.7	90.6	4.9

Subgroup analysis

Sub-group analysis was conducted to explore whether the intervention had different effects for participants with different characteristics. The characteristics included were those participants who had been enrolled on benefits for an above-median amount of time and those who indicate a higher-than-average motivation score at time of randomisation. However, the analysis did not find statistically significant differences between any of those groups and the main sample. Further details of the analysis can be found in Annex D.

[17] Whether or not those participants who were off benefits at the end of the trial remained off benefits for 6 months

4. Process evaluation findings

Reflecting the complexity of supporting individuals back into employment, there was no definitive evidence explaining why the RCT did not find that the programme was effective. However, the process evaluation did identify a range of factors that influenced successful provision and a number of barriers that can help explain why it was difficult to achieve positive outcomes for clients. The factors are grouped into four categories relating to: client characteristics and behaviours; provision challenges; system issues; and the local context. It should be noted that there was no difference in terms of range and diversity between the three sites - i.e. the factors described in this chapter were all relevant to all sites - so the findings of the process evaluation are not presented in a site-specific way.

4.1 Client characteristics and behaviour

One of the primary issues influencing the programme's effectiveness was client's orientation to work. Employment Specialists (ES) said it was helpful when referrers checked whether clients were willing to work and whether they understood that their participation in the programme was voluntary. Some ES felt they had to work with clients who did not feel ready for change or felt that they had to take part in the programme, which meant they were less likely to engage and benefit from IPS.

"You will say, do you feel that you are ready for work? Then [the client] will say, no I don't really want to. They will say that the job centre has said that I have to come on this or they will sanction me if I don't, and the job centre hasn't said that they will sanction them, and we are aware of that. But, that's what they will say to you and they are the inappropriate ones." Employment specialist, female.

Similarly, to be referred to IPS, clients had to be willing to openly discuss their mental health difficulties, which they were sometimes reluctant to do. This hindered referrers' ability to proceed with the referral process and highlights the importance of clients' disclosing their mental health difficulties to receive a referral to IPS.

"That is the hardest thing, getting somebody to turn around and say I have a problem. Sometimes you're constantly hitting a brick wall because the customer hasn't yet admitted it and keeps everything to themselves. It is frustrating at times. You can see it where they can't and you are trying to broach it with them in a very positive way but very delicately without upsetting or making them angry." Referrer, male.

JCP staff also referenced choosing not to refer clients who displayed aggressive behaviour because they did not want to put other services under strain, or clients who were too anxious due to not wanting to make the client's anxiety worse. As a result of these decisions made by

referrers, referrals may have been lower in volume, and clients that were referred were in some cases unlikely to be experiencing extreme or severe symptoms of their mental health difficulties.

4.2 Provision

This section explores three primary aspects of programme delivery: the relationship between advisors and clients; the nature of the client-led support; and, the length of the support that was provided.

Client-employment specialist relationship

At the core of the programme was the relationship between ES advisors and their clients. Some ESs were able to develop positive working relationships with their clients, and others encountered more difficulty. Both ESs and clients felt that specialists who demonstrated active listening, empathy and a non-judgemental approach to support (as opposed to trying too hard to force an understanding of the client's situation straight away), encouraged a more positive, trusting client-ES relationship.

Further qualities and skills that clients valued in their ESs included clients feeling confident that the job searches they were conducting with support from the ES were more productive than if the client had carried these out themselves, due to the ESs employment expertise and knowledge of the current job market. Clients also valued the feeling that their ES had experienced employment or mental health difficulties themselves in the past as this made clients feel that their ESs were relatable.

"It feels like [the employment specialist] has had this kind of experience before. So they're not judgemental. I don't feel judged and I also felt the choice of the [employment specialist] is good because it feels like this person has had mental health issues and is now helping people to get a job and get over [them]." Client, female.

Some, male clients said they had expected to feel more comfortable speaking with a male ES, but on receiving IPS support given by a female, found it surprisingly easy to discuss difficult or emotional topics. This was due to the stigma and norms surrounding men talking about their emotions and mental health.

"Originally I would prefer talking to a man rather than a woman and that's just the way that I am. But when I went in and started talking, you know, I actually ended up preferring it a lot more because I felt that I could open more. Because if you are talking to a man, you can't really open up about your emotional side. But if it's a woman you can open up about that sort of thing." Client, male.

Overall, the extent to which clients felt they were able to connect with their ESs was critical to whether they found their IPS experience a positive or negative one.

In London, JCP work coaches' ability to develop relationships with clients was supported by mental health training that was written by the DWP and provided by a senior member of the provider team. Following the training, DEAs felt that work coaches had gained a better understanding of the local mental health provisions, and could better communicate with clients experiencing mental health difficulties, which DEAs believed increased referrals to IPS from JCP.

> *"We've had mental health training for all of our work coaches... it was a really good approach to talking to people with mental health [difficulties]. When we saw a [referrals] chart at a recent meeting, they had risen after his delivery about mental health training, because it had been explained who the provider is locally - do you know who that is and do you know what they provide, and do you know how to refer."* Referrer, female.

Client-led approach to support

Taking a client-led approach to support was an important element of the IPS approach. Clients reported a positive IPS experience when their discussions (about mental health difficulties, better-off calculations, job searches and goal setting) were client-led, with clients feeling they had been listened to and given some choice. This approach encouraged the development of a trusted relationship between ESs and clients.

> *"For some people it might be that they don't want any input aside from something like interview preparation because they feel quite confident with everything else. They know the industry really well, and it's just about managing the anxiety during the interviews. Then for somebody else it might be that they literally don't have any idea about what to do. They've been out of work for 10 plus years and they don't know what the landscape is like. Then it's more in-depth and starting from the beginning."* Employment specialist, female.

Another client, in contrast, felt her ES was unhelpful when the discussions and support were not client-led, but instead rushed and not listening to the client's needs and preferences.

> *"[My ES] has not really been intently focused on what I've been saying, and he's just been thinking, we've only got this amount of time and my next client is coming along. It's been really rushed, and it hadn't been a smooth experience. The second meeting I had... it was really, really quick again. He just said, I've had a quick scan through your CV and seen you want to be a teaching assistant and you're very musical and that's great. I've had a look, here's a few jobs, and he pulled up like three and said get on and apply with them."* Client, female.

Clients and ESs both felt that the one-to-one nature of support allowed for the personalisation of sessions, because the support could be tailored specifically to the individual's pace and preferences. The one-to-one setting also created a psychologically safe space for clients to discuss their current circumstances and hopes from support.

"I think one-to-one placement support is much, much better. Not everybody feels secure in a group – some people do, and I have done groups and helped people that way but I think one-to-one- the individual and you can work at their pace. In a group, you just have to generalise." Employment specialist, female.

ESs were flexible in the methods they used to provide IPS, depending on clients' needs. These included calls, emails, texts, meeting clients in coffee shops, job centres and IAPT offices. Flexibility in the provision of IPS support allowed clients with busy schedules and extra responsibilities to continue engaging in their IPS journey. For example, one ES met a client who had children in a soft play centre, which meant they did not have to pay for childcare.

"We've got an IAPT centre right near us, and we've got a job centre and JobCentre Plus. Also every so often we go out in the community and just meet in a café. So those three different alternatives are what we do, and we also do telephone meetings because I've been on this course and I've not been available." Client, female.

Length of support

In line with the IPS model staff reported that the length of support was based on the client's needs, though they were adhering to a 26-week maximum support duration. However, the length that support was available for was sometimes dependent on ES's caseload size; if an ES's caseload was smaller than 25, then they felt there was more scope to support clients for longer than the 26-week period, and less pressure to discharge current clients to take on new clients. Clients said that it was important for them to be able to stay in contact with ESs after they had secured employment because they felt apprehensive about dealing with work colleagues. Knowing they could receive feedback and advice from their ES during this time provided important reassurance.

"Well hopefully in the next couple of weeks [I hope] to be in work but I'd love to stay in touch with [employment specialist] I don't really want [support to stop] just because I get a job, that I can't see her anymore because she's been a massive help." Client, female.

4.3 System

Although the programme was primarily based on the relationship between participants and ES advisors, this relationship did not exist in a vacuum. The elements of the system that interviewees particularly highlighted as important were the change in terms of targets; the referral process; support offered within IAPT services; the waiting times for support services; and, the degree to which employment specialists were embedded in the programme. These are discussed in turn below.

Less target-driven support

The fact that ESs experience less pressure to adhere to strict targets (in comparison to other forms of employment support they had provided in previous jobs), allowed them to prioritise the clients' needs. It meant that often clients gained work experience or voluntary placements that were meaningful to them, as opposed to being pressured to find work that wasn't sustainable or appropriate long-term.

"[The client] was attacked once in the early hours of the morning and it plays on her mind. But she wanted to feel like everybody else at work and do shifts. The only way we could do this was to look at Access to Work to see if they will provide you with a taxi to and from work, when finishing late at night or starting early in the morning... It is things like that that you have to look at when they come in with additional needs, so you know that they will sustain. It's not just about getting them in a job." Employment specialist, female.

Referral eligibility

JCP referrers described various factors that made clients unsuitable for IPS support. For example, JCP staff would not refer clients with severe mental health and comorbid difficulties, such as schizophrenia and traumatic brain injury. Although this is in line with the IPS model, employment specialists discussed how clients' whose mental health difficulties increase in severity after accessing IAPT subsequently became ineligible for IPS. Changes in the severity of client's mental health conditions may therefore provide an explanation for lower than expected referral rates.

"I didn't get to see [my client], because she was high-risk. By the time that she had had a couple of therapy sessions, I think the therapist realised that she needed to be stepped up. I did her information session, but I knew that there was something not right." Employment specialist, London, female.

Clients who were not suitable for IPS (such as those not receiving the correct benefits), were sometimes incorrectly referred to IPS. Recent changes to Universal Credit and benefit allowances were widely cited during the interviews, which may have contributed to referrers' lack of understanding regarding the benefits criteria clients' had to meet for an IPS referral.

"A lot of the things that we found is that… the NHS staff are not au fait with the benefit side of things. We would get a lot of people referred, but they weren't on the right benefit or the right criteria." Employment specialist, male.

Referral paperwork and forms

The paperwork and forms that were completed as part of the referral process were seen as too lengthy and it was reported that they caused burden for referrers. IAPT therapists said they

did not have adequate time to fill these out in addition to their assessments, therefore ESs were completing them during their first contact with the client.

"Therapists are even more time-limited because they are trying to deliver the CBT and they are very, very busy… it's easier for the other sites that the employment coach fills it in, that might get more referrals from the therapist. When the therapist has to do it, it's time-consuming for them... A lot of them forget or they haven't got the time." Employment specialist, female.

ES expressed concern that completing the lengthy referral paperwork could put clients off continuing with referral, and therefore prevent them from accessing IPS. Referrers suggested reducing the amount of paperwork required, as simplicity was paramount, and that quick and easily shared online systems would facilitate a more straightforward referral process.

Existing IAPT support

IAPT services had a range of different internal and external support (e.g. bereavement counselling) available to clients. If an IAPT service does not offer the specific support or psychological therapy required, clients are signposted out of IAPT to relevant services. Whilst understanding this is best for clients, an employment specialist found it frustrating when clients suitable for IPS support were being signposted to more relevant services because it resulted in them not remaining an IAPT service-user and therefore not being eligible for a referral to IPS.

"The only problems I've encountered with referrals from IAPT is that there was a stage where a lot of them seemed to be getting referred out. IAPT [in this area] tends not to do a lot of counselling in-house, so if someone needs bereavement counselling, they tend to refer them out [of IAPT to other mental health services] *and I can't work with them."* Employment specialist, female.

Clinical waitlist times

ESs reported that clients encountered waiting lists to access therapy, which varied in length between mental health services across different locations. Clients needed to be referred to IAPT to receive IPS, and ESs felt that clients' motivation could erode if IAPT assessment waitlists were high, which may increase the likelihood of them disengaging from IPS altogether.

Notably, a referrer reported that their provider advised them to prioritise clients already in employment who were at risk of losing their job, over those who were unemployed, when clinical waitlists were high and services were under pressure during a period of staffing issues.

"One thing I'm aware of is that there were times... when waiting lists were quite high. So they asked us to prioritise people that were in work who were struggling and who could

be at risk of losing their job... so I remember that being an occasional focus. Then I think certainly last year or so it's been fine to refer either of those strands [employed/unemployed]." Referrer, male.

Employment specialists' embedment

The IPS model required regular contact between ESs and IAPT/JCP services via weekly team meetings and check-in calls if necessary. While check-ins did occur, referrers reported that inconsistency in the frequency and regularity of ESs attendance at team meetings. This was reinforced by some therapists who found it difficult to recall when ESs had last attended their team meetings. Nonetheless, referrers' that reported less ES attendance at formal meetings, did describe having regular informal check-ins around the office with them, which suggests embedment was not compromised. ESs that did regularly attend service team meetings were able to share learning, and provide and receive current updates on client and service progress and changes, which enhanced communication and integration between services.

"I also work from the IAPT centre on a Monday... and we have team meetings there each week. We will discuss good practice, client caseloads, what's worked well, what hasn't worked well. So I like these meetings because I think there is a great deal of learning and sharing... sometimes in the team meetings someone will just say something that will open up another vision of, 'Oh yeah, I can do that or I can do this.' You have only got so many ideas, you can't come up with every solution." Employment specialist, male.

Successful integration between ESs and IAPT/JCP was also encouraged by ESs efforts to embed themselves within the services. Strategies supporting these efforts including introducing themselves in person, setting up face to face events and avoiding interacting in an overly formal way, including using humour at appropriate times, helped break down barriers between services.

"What I did in [location] and [location] is that they had a speed dating thing. I went to that, and you had all the work coaches come to it, and they knew me also. I said to them ask me the questions that you don't know the answers to, and that was amazing. They were saying, we know that you are the Trailblazer and we know you do this, but what is the actual process – it was those questions." Employment specialist, male.

In contrast, ES and JCP staff turnover hindered successful integration. New ESs had to restart the process of integration and could not rely on existing relationships, and existing ESs had to revisit Jobcentres to remind new JCP staff of the IPS support available.

"I think at one stage yes, they have slipped off but that's because the job centre has been busy with their move. It's been a combination of things because there has been a lot of new staff, so we need to keep going in and reminding them who we are, which is going to happen again to see if we can get the referrals up." Employment specialist, female.

In addition, initial conflicts between existing JCP work coaches and new employment specialists sometimes prevented the development of effective working relationships.

"Where we have had a little bit of conflict in some of the offices between the work coaches and the employment coaches previously, they have moved their coaches around and that has been a refresh that the particular location needed to get the referrals through and get the service working better. It's just personality clashes" Referrer, female.

Locality of client support

ESs met with clients for appointment sessions in JCP and IAPT offices. These were familiar locations for clients who were receiving IAPT therapy or JCP support, which increased clients' perception that they were receiving an integrated service. Attending support sessions in familiar locations also reduced clients' anxiety about travelling to new places. The autonomy ESs had over their diary management allowed them to schedule days in various office locations and in the community in a way that best encouraged integration while accommodating clients needs at the same time.

"Then when we started meeting up it was at the building where I was getting counselling anyway... so it worked hand in hand really speaking about work and how I was feeling at the time. So I had something to look forward to every week, it was a massive help." Client, male.

4.4 Context

The final set of factors influencing the programme were those associated with the local context, including the ability of the programme to engage with local employers, the approach and behaviour of employers and the nature of the local jobs market.

Engaging employers

ESs described engaging less with employers than they had hoped due to time restraints and prioritising their caseload sessions. Employer contact was described as a *"full-time job in itself"* and ESs covering multiple locations, as well as dedicating time to embedding in job centres and IAPT services, said it was difficult to find the time to build relationships with employers.

"I'm covering [multiple] areas, job centres and IAPTs. So my time is limited because I have to have regular visits and we have to do that to the job centres. I also have to go to [multiple] IAPTs… I have to establish relationships with [different] lots of therapists and there's quite a few of them. If I only had the one area I'd probably have more time to deal with employers and talk to them and establish rapport because it takes time." Employment specialist, female.

Methods for recording employer contact were also inconsistent; some used employer logs like Evolutive, and others used YAMO, hangouts and emails interchangeably if they felt that the employer log database was unsuitable because it was used infrequently and therefore was not acting as a shared resource.

> *"We do have [an employer log] but it's a bit rusty which sometimes we go to and sometimes we don't go to. It does exist but it's not well used. The thing is we are all mobile and different, but what we may do is if we find something that is really good and we want to share it, then it happens regularly every week or every fortnight we put it on Yamo… we might do it on an all-round email as well. We network about jobs, but it's just not necessarily on the rusty old database."* Employment specialist, female.

Labour market and employer reliability

Encountering a negative experience by starting a new job that was secured through IPS resulted in clients feeling frustrated about having to start the job search process again. On arrival at a new job, for example, one client described waiting several hours for the manager to arrive only to be told by the receptionist to go home. As a result, the client did not want to pursue the employment they had secured with this organisation. Clients negative initial experiences with employers could be the result of the lack of time ESs had to build contact with them.

> *"It's my first job in 10 years and I was ready to go sort of thing. And then to turn up and walk out. The receptionist said if the boss comes in I'll tell him to give you a ring, this is 10 to 11 in the morning. 2 O'clock in the afternoon I'm sat at home waiting for this phone call… I'm not on call 24 hours a day."* Client, male.

Notably, seasonal jobs (particularly common in seaside locations such as Blackpool) only provided temporary employment for clients and ESs preferred to support clients to find employment in more reliable year-round labour markets.

> *"They have got 'Black Monday' and what that means in this town is at the end of the season all the people re-sign back on. So at the end of October, the beginning of November a lot of the big employers will close down, and that's a lot of people who would work. The balance is finding people work all year round. A lot of people will dip in and out."* Employment specialist, male.

ESs noted that clients who successfully secured employment displayed high levels of motivation and resilience in the face of setbacks. This was demonstrated by the client who was let down on the first day of their employment. Despite the negative experience, the client described feeling more determined to find another job and planned to continue engaging in IPS with a positive attitude.

4.5 Implementation Fidelity Review

Scores on the Supported Employment Fidelity Scale are presented below for each area, followed by a summary of findings particularly relevant to the evaluation, against the three main themes of the fidelity criteria. All areas were found to offer 'good fidelity' IPS, suggesting that the implementation of the IPS service itself did not drive the lack of effect suggested by the quantitative findings. Full scores for each site can be found in Annex A.

Final Scores

West London

The final score awarded to the West London service was 101, which equates to 'good fidelity'. This reflected several strengths, primarily around the role of the ES and their relationship with JCP. While integration with JCP was considered to be strong, integration with IAPT was more limited and it was noted that many participants were not able to access IAPT and their IPS support concurrently. The quality of referrals was also raised as a potential issue.

Blackpool

The Blackpool service was awarded 108. This means the service was well within 'good fidelity'. The service was noted for its strong integration with a dedicated IAPT team, which was co-located with IPS. Following the assessment, recommendations included an increase in activity to promote the service and drive referral numbers, and an increase in ES contact with JCP to improve the quality of referrals. It was noted however that JCP offices relocated much closer to the delivery site during the trial, which increased contact between these teams.

North East

The final score awarded to the North East service was 101. This means the service reached 'good fidelity'. As with the preceding areas, the final score reflects several key strengths of the service, primarily around the role of the ES, who spent additional time at JCP supporting the referral process. Integration with IAPT was also noted to be a challenge at this site, due to existing IAPT service pressures.

Review themes

ES caseloads and the nature of support

This theme focuses on the size of ES caseloads, and the type of work they do with clients. High fidelity IPS involves ES providing only employment services, working with clients to solve barriers to employment and supporting clients through every stage of their journey into work. All sites scored highly for these criteria, as ES in each area were consistently found to be delivering IPS in adherence with its principles.

IPS fidelity guidelines suggest that the optimal caseload size for ES is 20 or fewer clients. While the average caseload size in West London was just under 20 clients, there was some variation across the team, with 16-32 cases per ES, who noted that this was because a proportion of participants on caseloads were in work and only required very light-touch support. Similarly in the North East, while some ES had relatively large caseloads, others had many fewer clients to see. This was considered to be appropriate given some of the specialists had entered their role much later than others. In contrast, caseloads in Blackpool did not exceed 20 participants per person, and it was anticipated caseload numbers would increase as the ES team reduced in size over time.

In line with fidelity guidelines, the vast majority of service activity with clients in each area was focused on employment and job search. Participants confirmed this and highlighted that the range of support offered was both broad and job focused, from help finding appropriate clothes to wear, to interview training and support with right to work issues. Administrative evidence in each area suggested that all stages of employment support were covered, from initial discussions about appropriate roles, to CV work, application support, cover letters and interview preparation. In Blackpool, other activities were also included, such as shopping with clients for interview clothes, supporting them to sign up to professional networking services and joining them to attend assessments and job fairs. In the North East, interviews with ES suggested that where necessary they also provided support to clients for other activities, which generally fell under the umbrella of ‘employment related personal barriers’.

In-work support was found to be offered in all areas, and in each area ES were responsible for all stages of support, from completing the vocational profile, supporting job search, setting up and supporting interviews, sitting in on interviews, and supporting clients when they entered work to make adjustments where needed.

Service integration and engagement

This theme focuses on the level of integration of IPS with IAPT and JCP, including practical arrangements made to support contact between teams, and the level of engagement with the local NHS Trust. All three sites scored relatively well in this area; however, service referrals and collaboration agreements (between IPS and IAPT/JCP) nonetheless proved challenging.

In West London, interviews with ES and JCP staff confirmed that the level of integration between the two was high. JCP had a clear view of the referral process, and Work Coaches had frequent contact with ES. JCP noted that there had been lots of meetings with the IPS provider to discuss the referral process, including a whole team presentation intended to boost referrals. In contrast, the level of integration between IPS and the IAPT team was less strong. While ES were attached to IAPT services and routinely conducted shared management meetings, a very small number of participants were receiving mental health support at the time of the fidelity assessment. This challenge was attributed to local wait times for IAPT, which meant that support from IPS and IAPT could not always be delivered concurrently.

West London teams also noted that the quality of referrals varied, in some cases because those referred had more severe mental health conditions than was appropriate for the model. JCP teams suggested that while potential participants did not always realise they had a common mental health condition, where their condition was more severe, they may have been referred as their needs were more obvious.

Finally, in some instances, West London JCP staff felt that those referred were not bought-in to the intervention, but felt they had to agree to be referred to protect their benefit claims, despite the voluntary nature of the intervention. In contrast, the review suggested that IAPT staff (rather than referred participants) were not convinced by the IPS programme, and more could be done to increase referral numbers by increasing awareness of IPS with the IAPT therapists. In line with this, IAPT staff interviewed could not recall seeing aids used for employment assessments, or seeing good news stories, though leaflets were left with IAPT clinicians to remind them to refer clients to the service. During the review, it was not possible for the review team to meet with a senior member of the Trust to discuss IAPT engagement with the program.

In Blackpool, the relationship with ES and JCP evolved over time. During the review, JCP teams suggested that ES could spend more time at JCP to drive referrals and promote the programme. ES noted that while relationships with JCP were good, at times it could be difficult to get hold of JCP teams to discuss referrals. This improved over time, with the JCP office relocating much closer to the IPS site, which improved collaboration between the teams. In contrast, the new IAPT team, which was built for this programme, was thoroughly integrated with the IPS team. This team sat on the same floor of a shared building with the ES team, and was thought to be far better integrated than the existing IAPT service in Blackpool. Interviews demonstrated that integration with the dedicated team worked well, and clients who accessed the IAPT service at the shared site often had an appointment with their ES in the following hour. Clients reflected that the service felt very joined up, that they received consistent messages and perceived both elements as one service. Both EAs and IAPT therapists were very happy with the regularity and structure of their contact, with regular weekly meetings in place attended by both EAs and IAPT team members. It was noted that weekly meetings tended to be less focused on client case conferencing, and more focused on referrals, given these teams worked so closely together and already held these conversations day to day. At a management level, the IPS Supervisor in Blackpool

also attended regular operational meetings with the mental health team leader, and weekly meetings with the Clinical Team lead, at which progress and project issues were covered.

Referral numbers in Blackpool remained a challenge, and were sometimes driven by concerns that randomisation would have a negative impact on jobseekers subsequently assigned to the control group, who would feel disappointed or distrustful as a result. It was felt that due to this, JCP seemed to take a less active approach in terms of promoting the programme, and tended to highlight the voluntary nature of the programme rather than its potential benefits. Staff in the existing IAPT teams shared these concerns around randomisation, and thus did not distribute posters or leaflets that referenced the trial in detail. Both teams felt that this was a difficult sell, and limited their ability to recommend the trial to potential participants. As a result, the language used to describe randomisation procedures was reviewed, specifically to describe the steps taken to highlight the possibility of randomisation into the treatment group as a 'chance to access an enhanced service' as a means of managing potential participants expectations.

The relationship between IPS and JCP was very effective in the North East, and during interviews with ES, it was noted that ES had invested some time in supporting the referral process. This included ES taking a diary for prospective referrals at JCP, working with clients referred by JCP to warm them up for a more formal meeting where they would provide consent, and scheduled face to face meetings that frequently took place between the teams. Some ES also discussed sending emails to JCP with updates following initial participant appointments, which were considered to be valuable feedback. There were some indications of variation in JCP engagement across the sites in the North East, as during the service user group interview for this area, some participants mentioned that they could not get referred to the service through JCP, and that it was not widely understood across JCP offices. Similarly, the IPS integration with IAPT in the North East was very strong from an administrative standpoint. Non-clinical IPS case notes fed into the IAPT system and ensured that the clinical team had access to notes on employment related support. It was also evident from interviews and diaries that the ES met with clinical teams when possible to discuss cases and potential referrals, and the trust shared documentation indicating clients were asked about employment as a matter of course.

The challenge in this area lay predominantly with the promotion of the IPS service within IAPT. An interview with a senior member of the trust indicated that the trust could not focus on IPS given other service pressures linked to long local waiting lists for entry to the IAPT service, though it was a regular item at board meetings every quarter. The IPS service was not featured in any trust publication or event, and it was reported that the trust was not focused on employment as part of the routine mental health work conducted with clients. In line with this, some service users reflected that they'd had limited contact with IAPT.

Employment and employer engagement

This theme focuses on ES frequency and quality of contact with potential employers, the follow-on support provided to service users, and the quality of employment opportunities secured. High fidelity IPS requires diversity in terms of the types of employers and job opportunities secured, and that employment specialists build strong relationships with local

employers. In a high fidelity service, job search preferences are client-led and frequently reviewed, and the impact of entering employment is reviewed in detail with service users.

Blackpool scored highest in this area, as the ES delivery against fidelity criteria was particularly strong.There were reports of high levels of contact with employers, rapid client-led job searches, and comprehensive work incentives planning to examine the costs and benefits associated with the service user entering employment. In the long term, ES were shown to maintain ongoing relationships with large-scale employers, and routinely connected clients with a diverse offering of employers and roles.

In West London and the North East, the recommended area for development was the level of ES employer contact, which could not be easily evidenced with the available administrative data, as formal records of employer contact were not kept. While this does not necessarily mean the activity didn't take place, for the purpose of the review more evidence of the activity would have been required to score highly on this item.

In West London, in line with previous findings on the level of integration with IAPT teams, there was limited evidence to suggest that assertive employer engagement and outreach was conducted by IAPT staff in partnership with ES. In contrast, outreach work in Blackpool was well evidenced, and outreach attempts were documented in the non-clinical case notes by IAPT staff. These detailed notes evidenced joint working by ES and clinical staff, both in outreach and engagement with clients, as well as during regular meetings with clients.

All sites demonstrated adherence to the principles of IPS when it came to jobsearch activity. On entry to the service at each site, Employment Profiles and About Me records were completed, and detailed action plans kept for each client. Job search was highly individualized, and clients who found work were offered regular follow up meetings (e.g during their probationary period at work).

Service users in the North East were happy that they were getting very tailored support from their ES, and raised that they were very pleased with this element of the service. Vocational profiles demonstrated that previous employment activity and service users' preferences were both taken into consideration, and ES interviewed provided examples of very specific client-led job search activity. From service user feedback and administrative evidence, it was clear that the ES played a major role in advocating for the client, by canvassing for roles through repeated employer contact and requesting interviews on their clients' behalf.

Moving clients into work was a challenge in all areas, for different reasons. Service users and ES in the North East noted that on occasion, voluntary work or work experience would be considered. From a review of action plans and vocational notes in this area, the majority of the work discussed, and for which applications were made, fell under the category of competitive work. Interviews in West London suggested that it was difficult to move clients into work rapidly, but that some positive intermediate outcomes (education, voluntary work) were not recognised as key outcomes for the program. ES at this site reflected that in many

cases, clients needed some time and more personal development before they could be considered ready to move into work.

The majority of the discussions evidenced in clerical records at the Blackpool site were centred around paid work and preparation for interviews for paid work, though as with the other sites, there was some evidence that other options such as work placements were discussed. In this case, the review team took into consideration the labour market characteristics of Blackpool, as the majority of available work was seasonal and transitional, limiting employment options for service users in this area.

All areas were found to offer in-work support, and evidence for this was particularly strong in the North East, where Service users action plans and testimonials confirmed that follow on in-work support was routinely offered, and that this was very much client led. ES described weaning clients from the service at their own pace, and clients recognised that even when they had started work, they were still able to access support from their ES. The medium for support seemed to be very flexible, with follow-up offered in person, on the phone and by text depending on the clients' preferences.

Similarly, all areas scored very highly on the diversity of jobs and employers offered.

5. Conclusions

Overall, these results do not suggest promising significant impacts of the treatment on off-benefit outcomes. Interestingly, in our primary benefit status specifications, the difference in number of days off benefits between treatment and control groups (amongst those who were off benefit for at least 1 day in the given time period) diminishes as the time window is extended. In other words, while the effect is always negative, it becomes smaller over time. The directionality of these results could suggest that the treatment might be helpful for those who are already off benefits at least a little bit, but potentially harmful for those who remain on benefits for the duration of the time.

As previously discussed, a major limitation to the mental health analysis in this trial was the data quality. To calculate the mental health outcomes, we used participants' last recorded GAD7 and PHQ9 scores; however, the timing and availability of these was subject to participants having reached out to NHS services in the first place. A future study could explore mental health outcomes using the earliest valid score as an outcome (comparing it to baseline), as this work suggests that an effect, if found, would likely occur closer to the time of intervention implementation. However, because the data in this study does not allow us to know exactly when implementation occurred, the proposed analysis would run the risk of analysing outcomes of people that haven't yet been treated in this context.

The IPS model is not designed as a clinical solution to mental health problems, but rather as an integrative model in which increased employment is expected to result in improvements in mental health. Given this, another way to consider the findings is to consider whether the IPS in IAPT model helps to negate any potential negative effects on mental health outcomes. It is possible that participants in the treatment group engaging with employment-support activities might have felt more overwhelmed by the sudden increase in demands on their time, resulting in short term negative impacts on mental health. We find no evidence of this effect in the available data.

While the level of missing data in health outcomes reduced the sample size to such an extent that we were unable to conclusively determine whether or not the intervention improved mental health outcomes, the off-benefit data was relatively complete and suggests that the intervention had no effect on off-benefit outcomes. Indeed, the trial did see lower volumes of participants than were expected. However, the histogram in the impact evaluation section shows the distributions of days off benefits at 12 months post-randomisation, and the modal outcome of 0 days off benefits suggests that even a study with a larger sample would not have shown more promising results.
Data were not available to confirm the proportion of those allocated who took up support from IPS. Given IPS attendance was voluntary and the intervention was delivered in the field, the

trial results should be interpreted in light of the fact that they are a function of participant compliance with the intervention, as much as the efficacy of the intervention itself.

Methodological conclusions

A number of lessons were learned from the data collection challenges. These are summarised below:

1. Build the trial design in such a way that all participants who choose to remain in the trial are assessed at regular intervals.
2. Use pessimistic projections for sample size estimates. As the evidence base in this area builds, previous studies can be used to guide assumptions for take-up and attrition.
3. If it is necessary to rely on data held by NHS/IAPT services, reinforce their commitment to the project throughout the duration of the trial.
4. If the analysis sample relies on matching NHS and NI numbers to administrative databases, ensure that invalid NHS/NI numbers cannot be submitted and require double entry.
5. Ensure that the outcomes needed for analysis pre-specified in the Statistical Analysis Plan are available and can be shared. We suggest requiring a data sample prior to the trial launch to ensure that all necessary data will be available at the analysis stage.

6. Annex A: Fidelity Scoring

Items are divided into three sections; staffing, organization, and services. Each set out a number of individual items, alongside scoring criteria for high fidelity, identifying where deviation from recommended practice may take place. Each item is rated up to 5 points, ranging from 1 = no implementation to 5 = full implementation, with higher scores representing progressively greater degrees of implementation. The maximum score for the scale is 125, with scores up to 99, 114 and 125 representing fair, good and exemplary provision respectively. Services scoring 73 and below are not considered to provide true IPS.

Table A1: Fidelity Check Framework and scores

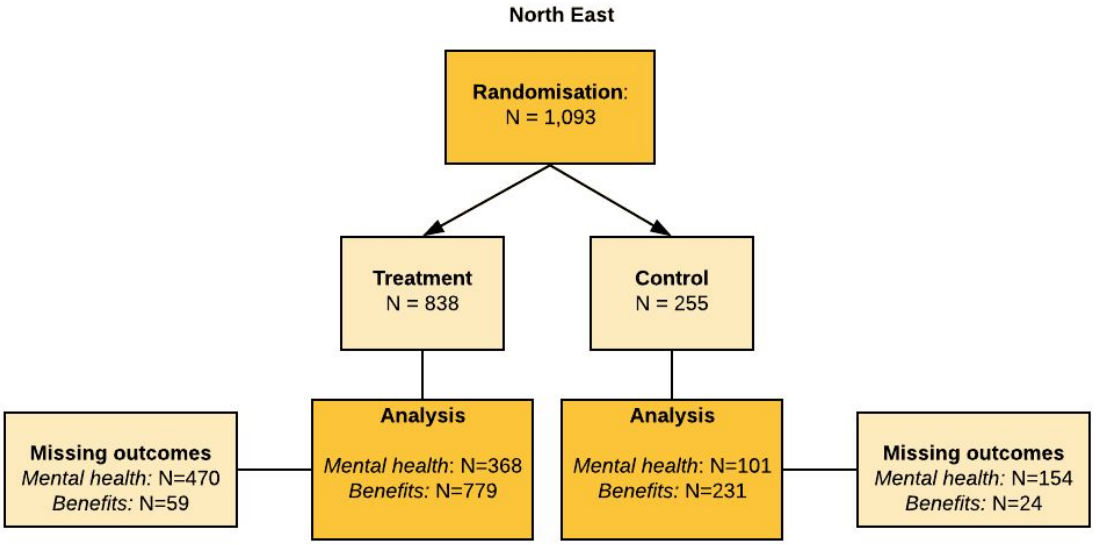
Staffing	1, Caseload size	5	5	5
	2, Staff spend 100% of time on employment tasks	4	5	5
	3, Staff provide all stages of support	5	5	5
Organisation	1, IPS worker is part of the clinical team	5	5	2
	2, Staff attend MDT meetings, use shared notes and are collocated	4	4	4
	3, Regular weekly contact between IPS and government programmes	3	4	5
	4, IPS workers form a team and meet weekly	5	3	5
	5, IPS supervisors provider mentoring and set performance goals	4	3	5
	6, Zero exclusion criteria for referral	3	5	5
	7, NHS organisations promote the service	3	2	3
	8, NHS Director champions IPS at a senior level	3	1	~
Services	1, Work incentives planning	5	5	5

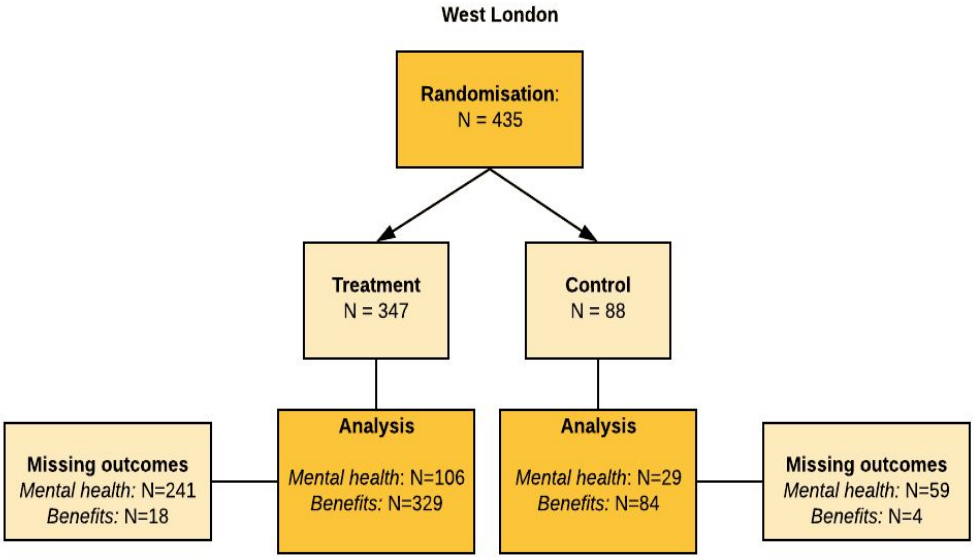
2, Specifics about mental health disclosure to be agreed	4	5	5
3, Live vocational profile tool	5	5	4
4, Job search begins within 30 days of referral	5	3	4
5, Individualised job search based on client preferences	5	5	5
6, Six face to face meetings between ES and employers per week	5	2	2
7, ES makes multiple visits to the same employers	5	4	4
8, Diversity of jobs sourced	4	4	3
9, Diversity of employers sourced	5	5	5
10, Competitive jobs	4	4	5
11, Intensive in-work support	5	5	5
12, On-going in-work support	5	4	3
13, IPS workers out of the office 65% of the time	3	4	5
14, Missed appointments followed up with attempts to re-engage	4	4	2
Total scores	**108**/125	**101**/125	**100**/120

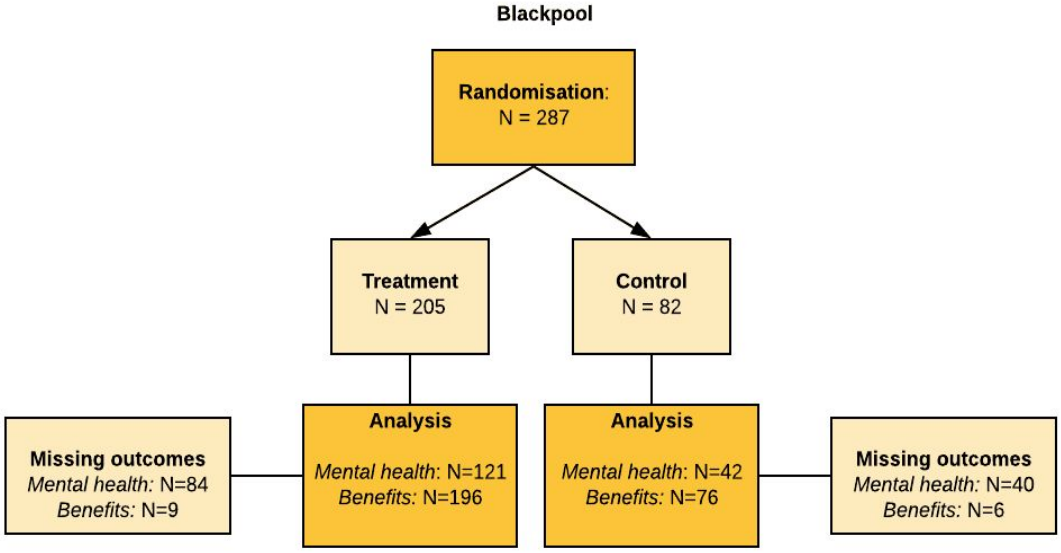
It was not possible to score the West London service for one item in the organisation category, so the total potential score for this area was reduced accordingly. The service was awarded a ‘good’ fidelity rating on the basis of the remaining scores.

7. Annex B: Consort diagrams

Figures B1: Consort diagrams for each area






Blackpool
Randomisation: N = 287
Treatment N = 205
Control N = 82
Missing outcomes
Mental health: N=84
Benefits: N=9
Analysis
Mental health: N=121
Benefits: N=196
Analysis
Mental health: N=42
Benefits: N=76
Missing outcomes
Mental health: N=40
Benefits: N=6

8. Annex C: Technical Appendix

Balance Checks

We observed good balance between treatment and control groups within both analysis samples across the variables in Tables C1 and C2. Although there was a slight imbalance in control participants in Blackpool, this difference is not significant at the 0.01 level and did not significantly impact the analysis.

Table C1: Balance check in mental health outcomes analysis sample

	Control	Treatment	p-value
Age at randomisation *(mean)*	41.39	39.72	0.16
Female *(0=male, 1=female)*	0.45	0.41	0.41
Time since randomisation[18] *(days)*	184.30	178.45	0.70
PHQ9 baseline *(mean)*	17.05	17.57	0.21
GAD7 baseline *(mean)*	14.24	14.69	0.23
Motivation score *(mean)*	26.30	26.46	0.83

[18] This refers to the time between randomisation and the last PHQ9/GAD7 outcome collection on NHS record. We include this because the mental health outcomes were calculated using the baseline score collected at randomisation and the participant's last recorded mental health outcome. We opted against using the first recorded mental health outcome because for many participants these were taken before the intervention was implemented, and we wanted to ensure that participants in the Treatment group had been exposed to the intervention. We control for this in our analysis, as we are aware that time is likely to play an important role in participants' outcomes.

Northeast	0.60	0.62	0.61
Blackpool	0.23	0.20	0.39
West London	0.17	0.18	0.80
N	182	616	

Table C2: Balance check in benefits outcomes analysis sample

	Control	Treatment	p-value
Age at randomisation *(mean)*	40.47	39.99	0.55
Female *(0=male, 1=female)*	0.51	0.49	0.43
Motivation score	26.65	26.56	0.87
Northeast	0.59	0.60	0.82
Blackpool	0.19	0.15	0.04
West London	0.22	0.25	0.13
N	391	1304	

As previously mentioned, we also explore the potential problem of differential attrition for the mental health outcome measures and find that there is not a statistically significant impact of treatment assignment on whether health outcomes are missing. This suggests that in spite of the high level of attrition in health outcomes, we do not expect that participants in the treatment group were necessarily more (or less) likely to attrit from the analysis sample.

Technical explanation of coefficient transformations

For this trial, as the outcome is binary (i.e. participants have recovered or not, or have been off benefits at least one day or not), the most appropriate regression model is a ‘logit model’. The logit model is based on proportions of successes (recovering/ being off benefits for one

day) and failures (not recovering/ not being off benefits for at least one day). Formally the log-odds of the outcome (the left-hand side of the regression equation) is defined as the natural logarithm of the proportion of success (p) divided by the complement of the proportion (known as the odds of success), i.e.

$$\ln(_{1-p})$$

The advantage of using the logit model is that it never predicts impossible values (like that for some trial participants the probability of recovering/being employed is less than zero or greater than one), as some alternative models do (for example linear probability models).

However, the disadvantage of using the logit model is that it is less intuitive to interpret than other forms of regression or a simple comparison of actual proportions. Therefore, BIT takes the following approach to reporting the findings of logistic regression analysis in an understandable way:

Firstly, the proportion of successes that is reported for the control group is the actual proportion in the control sample. (If the regression did not include covariates we would expect this to be very similar to the constant term in the regression equation; however, because there are covariates, the actual control group proportion is very different from the constant term.) There is no statistical uncertainty in this figure because it was actually observed.

Secondly, the proportion of successes reported for the intervention group (pT) is based on the regression equation, which takes into account the other characteristics (the covariates). As this is a randomised controlled trial, the proportion of successes in the treatment group based on the regression equation is very similar to the actual proportion of success recorded for the treatment group. This figure can be interpreted as the expected proportion of successes we would have seen in the control group if we had treated it. Since we didn't treat it, this is a counterfactual quantity and there is some statistical uncertainty as to its true value which is represented by a confidence interval.

To calculate the proportion of successes in the treatment group pT, the following formula is used.

$$T = \frac{\exp\ _1 1^{-}{}_c^c}{1+\exp(\ _1)\ _{1-}{}_c^c}$$

Where:

$_1$ = the regression coefficient for the treatment group

pc = is the actual proportion of successes in the control group

pT = is the reported proportion of success in the treatment group

Finally, to convert the proportion into percentages, pT is multiplied by 100.

In the regression tables in the Annex, the coefficients represent the log-odds of each explanatory variable. At the bottom of the regression tables, the figure in the control mean row is the actual proportion of success in the control group and the figure in the marginal treatment effect row is the difference between the control mean and the proportion of success in the treatment group calculated using the equation above (which as explained above accounts for the covariates).

Primary Analysis Specification

Health - Anxiety and Depression

$$Y_i \sim bernoulli\,(p_i);\ logit(p_i) = \alpha + \beta_1 T_i + \beta_2 X_i + \beta_3 \theta_i$$

where,

Y_i is our binary outcome variable, capturing whether or not the participant has moved into clinical recovery,

α is a constant term. It can be interpreted as the response level of participants in the control condition,

T_i is a binary treatment indicator, set to one if the individual received the IPS service (treatment), or zero otherwise,

X_i is a vector of participant level characteristics, based on data held by the IAPT services. This includes gender, age, area, baseline mental health score, motivation score, and time since randomisation.

θ_i is a vector of time dummies, including month and year at randomisation.

Benefits outcomes

$$Y_i \sim bernoulli\,(p_i);\ logit(p_i) = \alpha + \beta_1 T_i + \beta_2 X_i + \beta_3 \theta_i$$

where,

Y_i is our binary outcome variable, capturing whether or not the participant has spent >0 days off benefits

α is a constant term. It can be interpreted as the response level of participants in the control condition,

T_i is a binary treatment indicator, set to one if the individual received the IPS service (treatment), or zero otherwise,

X_i is a vector of participant level characteristics, based on data held by DWP. This includes gender, age, area, and motivation score.

θ_i is a vector of time dummies, including month and year at randomisation.

Note on the analysis specification

Our primary benefits outcomes were originally the number of days off benefits at 3, 6, 9, and 12 months post-randomisation, separately. We examined these outcomes using a zero-inflated negative binomial regression to estimate the model because:

- This is count data in the form of number of days off benefits in the time period. We chose negative binomial because there is overdispersion in the count data, i.e.: more variation than would be predicted under a simpler model.
- The counts are generated by i) a person becoming coming off benefits (therefore being off benefits >0 days) and ii) a person being off benefits a certain number of days in that timeframe. The zero-inflated part of the model accounts for this structure.
- The area fixed effects account for differences in the outcome mean across areas.

When we ran this analysis, we determined that given the distribution of the data being so heavily skewed to the left (i.e.: many people spending 0 days off benefits), it would be more policy-relevant to turn the outcome into a binary variable, indicating whether or not the person has been off benefits for >0 days in that time period. Distributions of the data can be found in Annex E. We thus decided to run the primary analysis as logits, as previously suggested by DWP.

Secondary analysis

Our secondary analysis takes two major forms – additional outcome measures, and sub-group analysis.

Additional outcome measures

Our model for each of the four additional health outcomes remains constant across all of these outcomes. We run this model for the following outcomes: reliable recovery for anxiety, reliable recovery for depression, improvement in anxiety, and improvement in depression.

$$Y_i \sim bernoulli\,(p_i);\ logit(p_i) = \alpha + \beta_1 T_i + \beta_2 X_i + \beta_3 \theta_i$$

where,

Y_i is our binary outcome variable,

α is a constant term. It can be interpreted as the response level of participants in the control condition,

T_i is a binary treatment indicator, set to one if the individual received the IPS service (treatment), or zero otherwise,

X_i is a vector of participant level characteristics, based on data held by the IAPT services. This includes gender, age, area, baseline mental health score, motivation scores and time since randomisation,

θ_i is a vector of time dummies, including month and year at randomisation.

Sub-group analysis

Our other main form of secondary analysis will be to analyse outcomes according to participants' sub-grouping. To avoid small cell size prohibiting analysis for confidentiality reasons, and to maximise comparability, we opt for an aggregated model rather than several partitioned ones, for each of the subgroup types of interest.

The two main sub-groups of interest are: (1) prior benefit status, and (2) level of motivation/enthusiasm for the project, which will be captured by entry surveys.

Prior Benefit Status:

Benefit types do not significantly vary, but the duration of enrolment varies more substantially across the sample. This enrolment variation will allow us to assign an additional binary characteristic: whether the participant spent below- or above-median length of time on benefits at enrolment into the programme.

We will then estimate a model specified:

$$Y_i \sim bernoulli\,(p_i);\ logit(p_i) = \alpha + \beta_1 T_i + \beta_2 X_i + \beta_3 B_i + \beta_4 \theta_i + \beta_5 D_i + \beta_6 (T_i \cdot D_i)$$

Where,

Y_i is our binary outcome variable, capturing whether or not the participant has come off benefits (provided by DWP),

α is a constant term. It can be interpreted as the response level of participants in the control condition,

T_i is a binary treatment indicator, set to one if the individual received the IPS service (treatment), or zero otherwise,

X_i is a vector of participant level characteristics, based on data held by the DWP. This includes gender, age, baseline motivation score, area

B_i is the initial benefit status of the participant. Specifically, this captures which benefits they are on and how long they have been on them,

θ_i is a vector of time dummies, including month and year at randomisation,

D_i is a binary variable set to 1 if participants are in the top half of benefit claimants by length of time spent on benefits, and 0 else,

$T_i \cdot D_i$ is an interaction between our treatment variable and the median split by length variable.

Engagement with programme:

At the beginning of the programme (prior to randomisation), participants will be asked to indicate how enthusiastic they are about the programme, and in particular whether they think it is likely to be helpful. This was measured using the self report Intrinsic Motivation Inventory (IMI). Similarly to the benefit status data, a median split will be conducted to determine high and low engagement participants, and we will then estimate the models specified:

Health - Anxiety and Depression

$$Y_i - \ bernoulli\ (p_i);\ logit(p_i) = \alpha + \ \beta_1 T_i + \beta_2 X_i + \beta_3 B_i + \beta_4 \theta_i + \beta_5 D_i + \beta_6 (T_i \cdot D_i)$$

Where

Y_i is our binary outcome variable, capturing whether or not the participant has moved into recovery,

α is a constant term. It can be interpreted as the response level of participants in the control condition,

T_i is a binary treatment indicator, set to one if the individual received the IPS service (treatment), or zero otherwise,

X_i is a vector of participant level characteristics, based on data held by the IAPT service. This includes gender, age, baseline mental health score, motivation score, and area,

θ_i is a vector of time dummies, including month and year at randomisation,

E_i is a binary variable set to one if participants are in the top half of participants by IMI score for the project, and 0 else,

$T_i \cdot E_i$ is an interaction between our treatment variable and the median split by IMI variable.

Benefit status

$$Y_i - \ bernoulli\ (p_i);\ logit(p_i) = \alpha + \ \beta_1 T_i + \beta_2 X_i + \beta_3 B_i + \beta_4 \theta_i + \beta_5 D_i + \beta_6 (T_i \cdot D_i)$$

Where

Y_i is our binary outcome variable, capturing whether or not the participant has come off benefits (provided by DWP),

α is a constant term. It can be interpreted as the response level of participants in the control condition,

T_i is a binary treatment indicator, set to one if the individual received the IPS service (treatment), or zero otherwise,

X_i is a vector of participant level characteristics, based on data held by the DWP. This includes gender, age, baseline motivation score, and area,

B_i is the initial benefit status of the participant. Specifically, this captures which benefits they are on and how long they have been on them,

θ is a vector of time dummies, including month and year at randomisation,

E_i is a binary variable set to one if participants are in the top half of participants by IMI for the project, and 0 else,

$T_i \cdot E_i$ is an interaction between our treatment variable and the median split by the IMI variable

9. Annex D: IPE topic guides

Employment specialist

1. Introductions and background

Introduce yourself – stress role as independent research organisation and that we are here to gather all views and explain that all information gathered will be in strict confidence and no-one will be named in any subsequent write-up of this research. Only time confidentiality will be broken is they disclose anything which leads us to believe that they are at risk of harm but we will keep them informed if this was to occur

Explain the aim of the discussions; we are here to talk about their experience of being an employment specialist in the Islington IPS trial.

Stress that you want to understand the world from your (the respondent's) point of view. No answers are right or wrong – and we are not here to judge the decisions made or views held by [the interviewee].

Explain that if at any point they feel uncomfortable or prefer not to answer a specific question they can just say so. The interview can end at any point and any question can be skipped.

Get verbal permission to digitally record and take notes (written permission should already have been obtained).

Explain that recording enables the interview to be transcribed for analysis alongside other interviews. Responses will be anonymised and combined with others', so they should feel free to speak openly.

Once you have consent, start the voice recorder.

Is there anything you would like to ask before we begin?

I'd like firstly to know a little bit about you...

Can you tell me a little bit about yourself?

- What is your current role?
- How long have you been working in this role?

2. First impressions

- Could you explain the IPS service to me, in your own words?
- Can you tell me about your initial thoughts of the IPS service?
    - What appealed to you about it?
    - Did you have any concerns?

3. Clients

- Can you tell me a little about the work you do with your clients?

    - How do you search for jobs?
    - What other areas do you work on?
- What kinds of interactions do you have with employers?
- What is your caseload like?
    - How manageable does this feel?

4. Supervision

- Can you tell me a little about the supervision structure here at IPS?
    - How often do you meet?
    - What do you discuss?
- Do you feel like this is sufficient for you?
    - Is there anything else you would like from supervision?

5. Integration

- Can you tell me about how someone gets referred to your service?
- Can you tell me a little about your relationship with the IAPT teams?
    - How does this work?
    - Do you feel like your services are integrated?
- Are there any other organisations/institutions/agencies that you interact with?

5. Recommendations

- How do you feel about the IPS service overall?
    - Do you think it is effective/achieving its goals?
    - Do you think it is different than other services out there?
- Is there anything you would like to change?
    - *magic wand*

6. Close

Do you have any questions on what we have covered in the interview?

You can round off the interview by summarising the main points you learned from the interview, and ask the respondent if they want to comment.

Thank them for their time and reassure them on the anonymity of the responses, as explained at the beginning of the interview.

Referrer

1. Introductions and background

Introduce yourself – stress role as independent research organisation and that we are here to gather all views and explain that all information gathered will be in strict confidence and no-one will be named in any

subsequent write-up of this research. Only time confidentiality will be broken is they disclose anything which leads us to believe that they are at risk of harm but we will keep them informed if this was to occur

Explain the aim of the discussions; we are here to talk about their experience of referring to the IPS service.

Stress that you want to understand the world from your (the respondent's) point of view. No answers are right or wrong – and we are not here to judge the decisions made or views held by [the interviewee].

Explain that if at any point they feel uncomfortable or prefer not to answer a specific question they can just say so. The interview can end at any point and any question can be skipped.

Get verbal permission to digitally record and take notes (written permission should already have been obtained).

Explain that recording enables the interview to be transcribed for analysis alongside other interviews. Responses will be anonymised and combined with others', so they should feel free to speak openly.

Once you have consent, start the voice recorder.

I'd like firstly to know a little bit about you...

Can you tell me a little bit about yourself?

- What is your current role?
- How long have you been working in this role?

2. Getting involved

- Can you tell me how you first heard about the IPS service?
    - Where did the information come from?
- What were your initial thoughts about a service like this?
    - Did you have any concerns?

3. Making a referral

- Can you tell me about how you decide to make a referral to the IPS?
    - What factors do you consider?
    - Who is most suitable for this service?
    - Is there any type of person who you would think would not be suitable?
- Can you tell me about the process of making a referral?
    - Steps…
    - Do you find this process easy?

4. Integration

- Can you tell me about your interactions with the Employment Specialists?
    - How often you speak with them?
    - Is this scheduled?
    - What kind of things are discussed?
    - Do you find these interactions useful? *(if they have interactions)*

5. Recommendations

- How do you feel about the IPS service overall?
    - Do you think it is effective/achieving its goals?
- Is there anything you would like to change?
    - *magic wand*

6. Close

Do you have any questions on what we have covered in the interview?

You can round off the interview by summarising the main points you learned from the interview, and ask the respondent if they want to comment.

Thank them for their time and reassure them on the anonymity of the responses, as explained at the beginning of the interview.

Service user

1. Introductions and background **5 mins**

Introduce yourself – stress role as independent research organisation and that we are here to gather all views and explain that all information gathered will be in strict confidence and no-one will be named in any subsequent write-up of this research. Only time confidentiality will be broken is they disclose anything which leads us to believe that they are at risk of harm but we will keep them informed if this was to occur

Explain the aim of the discussions; we are here to talk about their experience of the IPS service.

Stress that you want to understand the world from your (the respondent's) point of view. No answers are right or wrong – and we are not here to judge the decisions made or views held by [the interviewee].

Explain that if at any point they feel uncomfortable or prefer not to answer a specific question they can just say so. The interview can end at any point and any question can be skipped.

Get verbal permission to digitally record and take notes (written permission should already have been obtained).

Explain that recording enables the interview to be transcribed for analysis alongside other interviews. Responses will be anonymised and combined with others', so they should feel free to speak openly.

Once you have consent, start the voice recorder.

I'd like firstly to know a little bit about you...

Can you tell me a little bit about yourself?

- Where are you from?

- How long have you lived here?

2. Getting involved **4 mins**

- Can you tell me how you first heard about the IPS service?
    - Where did the information come from?
- What were your initial thoughts about a service like this?
    - Did you have any concerns?
    - Have you ever taken part in a service like this before?
- What made you make the decision to get involved?
    - What was the process?
    - How did you feel about working at that point?

3. Contact with Employment Specialist **4 mins**

- Can you tell me a little bit about your sessions with the Employment Specialist?
    - What kind of things do you work on?
    - Are they different from any other support you are receiving/have ever received?
- Do you find these sessions useful?
- How do you feel about working now?
- Do you feel supported?

4. Integration **4 mins**

- Where do your sessions take place?
    - Do you think this is a positive/negative aspect/doesn't make a difference?
- How do you feel about the IPS integration with mental health services?

5. Experience of employment (If in work) **5 mins**

- Can you tell me about how you got this job?
- Did you feel supported?
    - By ES?
    - By employer?
- Did you choose to disclose your condition/experience of MH difficulties to your employer?
    - Why/Why not?
- How do you feel about your work at the moment?
    - Working environment?

6. Recommendations **5 mins**

- How do you feel about the IPS service overall?
    - Do you think it is effective/achieving its goals?
- Is there anything you would like to change?

- *magic wand*

7. Close **3 mins**

Do you have any questions on what we have covered in the interview?

You can round off the interview by summarising the main points you learned from the interview, and ask the respondent if they want to comment.

Thank them for their time and reassure them on the anonymity of the responses, as explained at the beginning of the interview.

10. Annex E: Regression outputs

Primary Analysis

Anxiety and depression recovery - full sample

	(1) Logit	(2) Logit
	Anxiety recovery	Depression recovery
Treatment	0.157	-0.130
	[0.210]	[0.200]
Baseline	-0.004	-0.007
	[0.018]	[0.016]
Female	0.266	0.146
	[0.194]	[0.185]
Age at randomisation	0.004	0.006
	[0.006]	[0.006]
Motivation Baseline	0.017+	0.021*
	[0.009]	[0.010]
Year and month dummies	Yes	Yes
Blackpool	0.267	0.481*
	[0.225]	[0.218]
West London	0.081	0.530*
	[0.251]	[0.235]
Time since randomisation	-0.002**	-0.002**
	[0.001]	[0.001]
Constant	-1.748*	-1.617*
	[0.631]*	[0.632]

Control mean	0.256	0.337
Marginal treatment effect	0.029	-0.027
	[0.039]	[0.042]
N	763	759

Robust standard errors in parentheses.

Marginal treatment effect is the change in outcome as a function of the change in treatment, all else constant. It is calculated by averaging across each observation's marginal treatment effect.

Standard thresholds for significance as follows: + p < 0.10, * p <0.05, ** p <0.01; however, we increase the threshold in our interpretation using the Hochberg step up procedure.

Anxiety and depression - Northeast only

	(1) Logit	(2) Logit
	Anxiety recovery	Depression recovery
Treatment	-0.287	-0.493^{+}
	[0.261]	[0.259]
Baseline	-0.023	-0.010
	[0.024]	[0.021]
Female	0.375	-0.009
	[0.240]	[0.237]
Age at randomisation	0.007	-0.003
	[0.008]	[0.008]
Motivation baseline	0.017	0.024*
	[0.011]	[0.012]
Year and month dummies	Yes	Yes

Time since randomisation	-0.002*	-0.002*
	[0.001]	[0.001]
Constant	-1.986	-1.687
	[1.273]	[1.343]
Control mean	0.297	0.337
Marginal treatment effect	-0.055	-0.098
	[0.051]	[0.053]
N	469	469

Robust standard errors in parentheses.

Marginal treatment effect is the change in outcome as a function of the change in treatment, all else constant. It is calculated by averaging across each observation's marginal treatment effect.

Standard thresholds for significance as follows: + p < 0.10, * p <0.05, ** p <0.01; however, we increase the threshold in our interpretation using the Hochberg step up procedure.

Anxiety and depression - West London only

	(1) Logit	(2) Logit
	Anxiety recovery	Depression recovery
Treatment	0.137	0.004
	[0.673]	[0.538]
Baseline	0.070	0.047
	[0.053]	[0.045]

Female	-1.244*	-0.016
	[0.506]	[0.462]
Age at randomisation	-0.011	0.025+
	[0.019]	[0.015]
Motivation baseline	-0.012	0.006
	[0.041]	[0.039]
Year and month dummies	Yes	Yes
Time since randomisation	-0.002	-0.002+
	[0.002]	[0.001]
Constant	-1.027	-2.887+
	[1.916]	[1.745]
Control mean	0.241	0.414
Marginal treatment effect	0.022	0.001
	[0.109]	[0.113]
N	117	121

Robust standard errors in parentheses.

Marginal treatment effect is the change in outcome as a function of the change in treatment, all else constant. It is calculated by averaging across each observation's marginal treatment effect.

Standard thresholds for significance as follows: + p < 0.10, * p <0.05, ** p <0.01; however, we increase the threshold in our interpretation using the Hochberg step up procedure.

Anxiety and depression - Blackpool only

	(1) Logit	(2) Logit
	Anxiety recovery	Depression recovery
Treatment	1.609**	0.947+
	[0.581]	[0.494]
Baseline	0.028	-0.074
	[0.049]	[0.048]
Female	1.475*	0.715
	[0.673]	[0.581]
Age at randomisation	0.004	0.006
	[0.017]	[0.015]
Motivation baseline	0.030	0.015
	[0.025]	[0.025]
Year and month dummies	Yes	Yes
Time since randomisation	-0.002+	-0.002+
	[0.001]	[0.001]
Constant	-3.209	-0.039
	[2.085]	[1.865]
Control mean	0.167	0.286
Marginal treatment effect	0.257	0.187
	[0.072]	[0.092]
N	156	140

Robust standard errors in parentheses.

Marginal treatment effect is the change in outcome as a function of the change in treatment, all else constant. It is calculated by averaging across each observation's marginal treatment effect.

Standard thresholds for significance as follows: + p < 0.10, * p <0.05, ** p <0.01; however, we increase the threshold in our interpretation using the Hochberg step up procedure, and apply this threshold to the signs on the marginal treatment effect.

Off benefits - full sample

	(1) Logit	(2) Logit	(3) Logit	(4) Logit
	Binary: Off benefits >1 day in 3 months	Binary: Off benefits >1 day in 6 months	Binary: Off benefits >1 day in 9 months	Binary: Off benefits >1 day in 12 months
Treatment allocation	-0.259	-0.133	-0.145	-0.109
	[0.184]	[0.149]	[0.135]	[0.131]
Female	0.295+	0.177	0.128	0.117
	[0.165]	[0.130]	[0.117]	[0.114]
Age	-0.007	0.001	0.001	0.001
	[0.006]	[0.005]	[0.004]	[0.004]
Motivation score	0.003	0.007	0.010	0.010
	[0.010]	[0.008]	[0.007]	[0.007]
Year and month dummies	Yes	Yes	Yes	Yes
Blackpool	-0.035	0.038	0.062	0.044
	[0.246]	[0.199]	[0.181]	[0.175]

West London	-0.068	0.222	0.344*	0.286*
	[0.215]	[0.163]	[0.147]	[0.143]
Constant	-1.399	-1.345+	-1.327*	-1.409*
	[0.867]	[0.734]	[0.669]	[0.650]
Control mean	0.123	0.194	0.256	0.274
Marginal treatment effect	-0.025	-0.020	-0.026	[-0.021]
	[0.018]	[0.022]	[0.025]	[0.025]
N	1692	1692	1692	1692

Robust standard errors in parentheses.

Marginal treatment effect is the change in outcome as a function of the change in treatment, all else constant. It is calculated by averaging across each observation's marginal treatment effect.

Standard thresholds for significance as follows: + p < 0.10, * p <0.05, ** p <0.01; however, we increase the threshold in our interpretation using the Hochberg step up procedure.

Off benefits - Northeast only

	(1) Logit	(2) Logit	(3) Logit	(4) Logit
	Binary: Off benefits >1 day in 3 months	Binary: Off benefits >1 day in 6 months	Binary: Off benefits >1 day in 9 months	Binary: Off benefits >1 day in 12 months
Treatment allocation	-0.266	-0.191	-0.130	-0.074
	[0.240]	[0.197]	[0.180]	[0.176]
Female	0.353+	0.233	0.195	0.196
	[0.212]	[0.173]	[0.156]	[0.151]

Age	0.005	0.004	0.006	0.004
	[0.007]	[0.006]	[0.006]	[0.005]
Motivation score	0.006	0.011	0.017*	0.016+
	[0.011]	[0.009]	[0.009]	[0.008]
Year and month dummy	Yes	Yes	Yes	Yes
Constant	-1.915**	-1.498**	-1.689**	-1.497**
	[0.677]	[0.570]	[0.531]	[0.516]
Control mean	0.126	0.195	0.242	0.260
Marginal treatment effect	-0.025	-0.027	-0.022	-0.013
	[0.024]	[0.028]	[0.031]	[0.032]
N	1008	1008	1008	1008

Robust standard errors in parentheses.

Marginal treatment effect is the change in outcome as a function of the change in treatment, all else constant. It is calculated by averaging across each observation's marginal treatment effect.

Standard thresholds for significance as follows: + p < 0.10, * p <0.05, ** p <0.01; however, we increase the threshold in our interpretation using the Hochberg step up procedure.

Off benefits - West London only

	(1) Logit	(2) Logit	(3) Logit	(4) Logit
	Binary: Off benefits >1 day in 3 months	Binary: Off benefits >1 day in 6 months	Binary: Off benefits >1 day in 9 months	Binary: Off benefits >1 day in 12 months

Treatment allocation	0.124	0.343	0.357	0.379
	[0.442]	[0.334]	[0.291]	[0.289]
Female	-0.115	-0.102	-0.163	-0.162
	[0.344]	[0.248]	[0.223]	[0.220]
Age	-0.021	-0.003	-0.006	-0.003
	[0.014]	[0.012]	[0.010]	[0.009]
Motivation score	0.005	-0.008	0.008	0.013
	[0.028]	[0.018]	[0.017]	[0.017]
Year and month dummies	Yes	Yes	Yes	Yes
Constant	-1.900	-2.818*	-2.291*	-2.101*
	[1.469]	[1.401]	[1.071]	[0.946]
Control mean	0.083	0.155	0.226	0.238
Marginal treatment effect	0.010	0.051	0.067	0.074
	[0.036]	[0.046]	[0.052]	[0.053]
N	395	412	412	412

Robust standard errors in parentheses.

Marginal treatment effect is the change in outcome as a function of the change in treatment, all else constant. It is calculated by averaging across each observation's marginal treatment effect.

Standard thresholds for significance as follows: + p < 0.10, * p <0.05, ** p <0.01; however, we increase the threshold in our interpretation using the Hochberg step up procedure.

Off benefits - Blackpool only

	(1) Logit	(2) Logit	(3) Logit	(4) Logit
	Binary: Off benefits >1 day in 3 months	Binary: Off benefits >1 day in 6 months	Binary: Off benefits >1 day in 9 months	Binary: Off benefits >1 day in 12 months
Treatment allocation	-0.445	-0.379	-0.651*	-0.648*
	[0.413]	[0.343]	[0.322]	[0.311]
Female	0.427	0.304	0.254	0.163
	[0.430]	[0.347]	[0.316]	[0.309]
Age	-0.035*	0.000	-0.004	-0.004
	[0.017]	[0.013]	[0.012]	[0.012]
Motivation score	-0.006	0.006	-0.020	-0.024
	[0.026]	[0.022]	[0.019]	[0.019]
Year and month dummy	Yes	Yes	Yes	Yes
Constant	-0.836	-1.880	-0.772	-0.714
	[1.494]	[1.313]	[1.264]	[1.237]
Control mean	0.158	0.237	0.329	0.355
Marginal treatment effect	-0.045	-0.059	-0.115	-0.120
	[0.043]	[0.055]	[0.059]	[0.060]
N	259	259	272	272

Robust standard errors in parentheses.

Marginal treatment effect is the change in outcome as a function of the change in treatment, all else constant. It is calculated by averaging across each observation's marginal treatment effect.

Standard thresholds for significance as follows: + p < 0.10, * p <0.05, ** p <0.01; however, we increase the threshold in our interpretation using the Hochberg step up procedure.

Secondary Analysis

Anxiety and depression: (1) reliable change and (2) decrease in score - full sample

	(1) Logit	(2) Logit	(3) Logit	(4) Logit
	Reliable change: anxiety	Reliable change: depression	Decrease in anxiety score	Decrease in depression score
Treatment	-0.001	-0.083	-0.373+	0.176
	[0.196]	[0.192]	[0.197]	[0.195]
Baseline	0.092**	0.075**	0.107**	0.059**
	[0.018]	[0.017]	[0.019]	[0.018]
Female	0.143	0.202	-0.146	0.142
	[0.181]	[0.181]	[0.188]	[0.193]
Age at randomisation	0.007	0.010+	-0.002	0.007
	[0.006]	[0.006]	[0.005]	[0.006]
Motivation baseline	0.014	0.016+	0.016+	0.011
	[0.009]	[0.009]	[0.009]	[0.009]
Year and month dummies	Yes	Yes	Yes	Yes

Blackpool	0.157	0.331	0.372	0.321
	[0.215]	[0.218]	[0.234]	[0.237]
West London	-0.075	0.292	0.290	0.478^{+}
	[0.243]	[0.232]	[0.258]	[0.265]
Time since randomisation	-0.002**	-0.001**	-0.002**	-0.001*
	[0.001]	[0.001]	[0.001]	[0.001]
Constant	-2.289**	-2.882**	-0.766	-1.100^{+}
	[0.598]	[0.600]	[0.594]	[0.613]
Control mean	0.378	0.378	0.692	0.651
Marginal treatment effect	0.000	-0.018	-0.074	0.035
N	763	763	763	763

Robust standard errors in parentheses.

Marginal treatment effect is the change in outcome as a function of the change in treatment, all else constant. It is calculated by averaging across each observation's marginal treatment effect.

Standard thresholds for significance as follows: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$; however, we increase the threshold in our interpretation using the Hochberg step up procedure.

Anxiety and depression: (1) reliable change and (2) decrease in score - Northeast only

	(1) Logit	(2) Logit	(3) Logit	(4) Logit
	Reliable change: anxiety	Reliable change: depression	Decrease in anxiety score	Decrease in depression score
Treatment	-0.506*	-0.325	-0.606*	-0.005
	[0.246]	[0.248]	[0.259]	[0.254]
Baseline	0.099**	0.085**	0.124**	0.064**
	[0.023]	[0.022]	[0.024]	[0.022]
Female	0.079	0.058	-0.352	0.029
	[0.224]	[0.232]	[0.238]	[0.243]
Age at randomisation	0.003	0.007	-0.006	0.007
	[0.007]	[0.008]	[0.007]	[0.008]
Motivation baseline	0.012	0.013	0.019+	0.010
	[0.010]	[0.011]	[0.011]	[0.011]
Year and month dummies	Yes	Yes	Yes	Yes
Time since randomisation	-0.002**	-0.001	-0.003**	-0.001
	[0.001]	[0.001]	[0.001]	[0.001]
Constant	-2.497*	-2.338**	-1.713	-2.098*
	[1.102]	[0.700]	[1.147]	[1.026]

Control mean	0.465	0.386	0.713	0.683
Marginal treatment effect	-0.114	-0.069	-0.117	-0.001
	[0.055]	[0.054]	[0.047]	[0.051]
N	469	461	469	462

Robust standard errors in parentheses.

Marginal treatment effect is the change in outcome as a function of the change in treatment, all else constant. It is calculated by averaging across each observation's marginal treatment effect.

Standard thresholds for significance as follows: + p < 0.10, * p <0.05, ** p <0.01; however, we increase the threshold in our interpretation using the Hochberg step up procedure.

Anxiety and depression: (1) reliable change and (2) decrease in score - West London only

	(1) Logit	(2) Logit	(3) Logit	(4) Logit
	Reliable change: anxiety	Reliable change: depression	Decrease in anxiety score	Decrease in depression score
Treatment	0.568	0.035	-0.208	0.457
	[0.610]	[0.551]	[0.552]	[0.557]
Baseline	0.166**	0.122*	0.127**	0.030
	[0.051]	[0.048]	[0.045]	[0.050]
Female	-0.419	0.197	-0.581	0.250
	[0.475]	[0.452]	[0.449]	[0.481]
Age at randomisation	0.002	0.005	-0.004	0.004
	[0.013]	[0.014]	[0.012]	[0.015]
Motivation baseline	0.065	0.055	-0.015	-0.001

	[0.044]	[0.043]	[0.033]	[0.032]
Year and month dummies	Yes	Yes	Yes	Yes
Time since randomisation	-0.002+	-0.002+	-0.001	-0.002
	[0.001]	[0.001]	[0.001]	[0.001]
Constant	-5.680**	-5.266**	-0.199	-0.507
	[2.069]	[1.937]	[1.769]	[1.827]
Control mean	0.241	0.379	0.690	0.690
Marginal treatment effect	0.099	0.007	-0.040	0.083
	[0.103]	[0.109]	[0.103]	[0.105]
N	117	122	130	121

Robust standard errors in parentheses.

Marginal treatment effect is the change in outcome as a function of the change in treatment, all else constant. It is calculated by averaging across each observation's marginal treatment effect.

Standard thresholds for significance as follows: + p < 0.10, * p <0.05, ** p <0.01; however, we increase the threshold in our interpretation using the Hochberg step up procedure.

Anxiety and depression: (1) reliable change and (2) decrease in score - Blackpool only

	(1) Logit	(2) Logit	(3) Logit	(4) Logit
	Reliable change: anxiety	Reliable change: depression	Decrease in anxiety score	Decrease in depression score
Treatment	1.621**	0.653	0.495	1.100*

	[0.545]	[0.442]	[0.511]	[0.482]
Baseline	0.053	-0.004	0.097*	0.060
	[0.046]	[0.045]	[0.048]	[0.048]
Female	0.772	0.491	0.958	-0.098
	[0.629]	[0.564]	[0.653]	[0.577]
Not known	1.361*	1.129*	0.926+	0.585
	[0.542]	[0.515]	[0.542]	[0.538]
Age at randomisation	0.014	0.011	0.021	-0.002
	[0.018]	[0.015]	[0.017]	[0.014]
Motivation baseline	0.020	0.003	0.027	0.017
	[0.023]	[0.023]	[0.028]	[0.026]
Time since randomisation	-0.001	-0.001	-0.003*	-0.001
	[0.001]	[0.001]	[0.001]	[0.001]
Constant	-4.721*	-2.289	-4.546*	-3.153+
	[1.856]	[2.271]	[1.952]	[1.752]
Control mean	0.262	0.357	0.643	0.548
Marginal treatment mean	0.312	0.138	0.088	0.213
	[0.085]	[0.090]	[0.091]	[0.095]
N	152	152	146	144

Robust standard errors in parentheses.

Marginal treatment effect is the change in outcome as a function of the change in treatment, all else constant. It is calculated by averaging across each observation's marginal treatment effect.

Standard thresholds for significance as follows: + p < 0.10, * p <0.05, ** p <0.01; however, we increase the

threshold in our interpretation using the Hochberg step up procedure.

Subgroup Analysis

Above median time on benefits

	(1) Full sample	(2) North East only	(3) West London only	(4) Blackpool only
	Has come off benefits	Has come off benefits	Has come off benefits	Has come off benefits
Treatment	-0.320	-0.316	-0.223	-0.869+
	[0.206]	[0.315]	[0.420]	0.470]
atabovemedian_onbenefit=1	-0.847+	-0.660	-0.370	-1.596
	[0.511]	[0.763]	[1.148]	[0.997]
Control # atabovemedian_onbenefit=0	0.000	0.000	0.000	0.000
	[.]	[.]	[.]	[.]
Control # atabovemedian_onbenefit=1	0.000	0.000	0.000	0.000
	[.]	[.]	[.]	[.]
Treatment # atabovemedian_onbenefit=0	0.000	0.000	0.000	0.000
	[.]	[.]	[.]	[.]
Treatment # atabovemedian_onbenefit=1	0.728	0.577	0.370	1.207
	[0.504]	[0.776]	[0.970]	[0.989]

Female	-0.043	0.198	-0.317	-0.629
	[0.162]	[0.247]	[0.300]	[0.402]
Age	0.004	0.008	0.004	-0.002
	[0.006]	[0.009]	[0.012]	[0.017]
Motivation score	0.042**	0.037*	0.032	0.092**
	[0.012]	[0.017]	[0.024]	[0.029]
Year and month dummies	Yes	Yes	Yes	Yes
Blackpool	0.670**	-	-	-
	[0.240]			
West London	0.506*	-	-	-
	[0.210]			
ICA	1.672*	2.852**	0.000	0.000
	[0.676]	[0.821]	[.]	[.]
IS	0.300	0.835	0.000	0.000
	[0.622]	[0.684]	[.]	[.]
JSA	0.434+	0.267	0.707	0.125
	[0.243]	[0.387]	[0.442]	[0.570]
UC	-0.616*	-0.925*	-0.081	-1.302+
	[0.250]	[0.402]	[0.444]	[0.666]
priorbenefitdays	-0.000**	-0.000**	-0.000+	-0.000
	[0.000]	[0.000]	[0.000]	[0.000]
Constant	-2.667**	-2.497+	-1.466	-1.909
	[0.592]	[1.284]	[1.090]	[1.390]
Marginal effect: Control # below	0.156	0.110	0.212	0.333

median on benefit	[0.023]	[0.023]	[0.059]	[0.093]
Marginal effect: Treatment # below median on benefit	0.122	0.086	0.181	0.194
	[0.012]	[0.013]	[0.028]	[0.050]
Marginal effect: Control # at/above median on benefit	0.080	0.065	0.163	0.113
	[0.030]	[0.037]	[0.114]	[0.067]
Marginal effect: Treatment # at/above median on benefit	0.111	0.081	0.181	0.147
	[0.025]	[0.027]	[0.073]	[0.070]
N	1674	998	383	236

The marginal effect is the change in outcome as a function of the change in treatment, all else constant. It is calculated by averaging across each observation's marginal treatment effect.

Standard thresholds for significance as follows: + p < 0.10, * p <0.05, ** p <0.01; however, we increase the threshold in our interpretation using the Hochberg step up procedure.

Above median motivation score

Anxiety

	(1) Full sample	(2) North East only	(3) West London only	(4) Blackpool only
	Anxiety recovery	Anxiety recovery	Anxiety recovery	Anxiety recovery
Treatment	0.313	0.153	-0.211	1.037
	[0.319]	[0.384]	[1.350]	[0.751]
atabovemedian_motivation =0	0.000	0.000	0.000	0.000
	[.]	[.]	[.]	[.]
atabovemedian_motivation =1	0.179	0.737	-0.516	-1.068
	[0.445]	[0.551]	[1.693]	[1.129]

Control # atabovemedian_motivation =0	0.000	0.000	0.000	0.000
	[.]	[.]	[.]	[.]
Control # atabovemedian_motivation =1	0.000	0.000	0.000	0.000
	[.]	[.]	[.]	[.]
Treatment # atabovemedian_motivation =0	0.000	0.000	0.000	0.000
	[.]	[.]	[.]	[.]
Treatment # atabovemedian_motivation =1	-0.283	-0.890+	0.476	1.287
	[0.422]	[0.531]	[1.595]	[1.115]
GAD7 Baseline	-0.004	-0.022	0.067	0.033
	[0.018]	[0.024]	[0.057]	[0.048]
Female	0.266	0.372	-1.225*	1.581*
	[0.194]	[0.243]	[0.521]	[0.698]
Age at randomisation	0.004	0.006	-0.011	0.003
	[0.006]	[0.008]	[0.019]	[0.017]
Motivation baseline	0.018	0.015	-0.003	0.031
	[0.015]	[0.018]	[0.078]	[0.039]
Year and month dummies	Yes	Yes	Yes	Yes
Blackpool	0.266	-	-	-
	[0.226]			
West London	0.077	-	-	-
	[0.252]			
Time since randomisation	-0.002**	-0.002*	-0.002	-0.002+
	[0.001]	[0.001]	[0.002]	[0.001]
Constant	-1.898**	-2.254+	-0.964	-2.592

	[0.698]	[1.288]	[2.494]	[2.278]
Marginal effect: Control # below median motivation	0.241	0.243	0.353	0.210
	[0.552]	[0.064]	[0.235]	[0.099]
Marginal effect: Treatment # below median motivation	0.299	0.271	0.317	0.385
	[0.036]	[0.042]	[0.133]	[0.080]
Marginal effect: Control # at/above median motivation	0.273	0.391	0.268	0.095
	[0.051]	[0.079]	[0.116]	[0.061]
Marginal effect: Treatment # at/above median motivation	0.279	0.243	0.310	0.428
	[0.031]	[0.040]	[0.070]	[0.079]
N	763	469	117	156

The marginal effect is the change in outcome as a function of the change in treatment, all else constant. It is calculated by averaging across each observation's marginal treatment effect.

Standard thresholds for significance as follows: + p < 0.10, * p <0.05, ** p <0.01; however, we increase the threshold in our interpretation using the Hochberg step up procedure.

Depression

	(1) Full sample	(2) North East only	(3) West London only	(4) Blackpool only
	Depression recovery	Depression recovery	Depression recovery	Depression recovery
Treatment	-0.378	-0.436	-1.447	0.518
	[0.288]	[0.362]	[1.137]	[0.622]
atabovemedian_motivation=0	0.000	0.000	0.000	0.000

	[.]	[.]	[.]	[.]
atabovemedian_motivation=1	-0.045	0.815	-2.163+	-0.589
	[0.425]	[0.557]	[1.199]	[1.119
Control # atabovemedian_motivation=0	0.000	0.000	0.000	0.000
	[.]	[.]	[.]	[.]
Control # atabovemedian_motivation=1	0.000	0.000	0.000	0.000
	[.]	[.]	[.]	[.]
Treatment # atabovemedian_motivation=0	0.000	0.000	0.000	0.000
	[.]	[.]	[.]	[.]
Treatment # atabovemedian_motivation=1	0.451	-0.116	1.887	0.977
	[0.396]	[0.511]	[1.289]	[1.008]
PHQ9 Baseline	-0.007	-0.014	0.049	-0.069
	[0.016]	[0.021]	[0.048]	[0.049]
Female	0.144	-0.018	-0.001	0.729
	[0.185]	[0.238]	[0.464]	[0.609]
Age at randomisation	0.005	-0.004	0.026	0.005
	[0.006]	[0.008]	[0.016]	[0.015]
Motivation baseline	0.008	-0.007	0.052	0.007
	[0.015]	[0.019]	[0.071]	[0.042]
Year and month dummies	Yes	Yes	Yes	Yes
Blackpool	0.492*	-	-	-
	[0.219]			
West London	0.537*	-	-	-

	[0.236]			
Time since randomisation	-0.002**	-0.002*	-0.002	-0.002+
	[0.001]	[0.001]	[0.001]	[0.001]
Constant	-1.223+	-1.110	-2.968	0.671
	[0.682]	[1.386]	[2.196]	[2.065]
Marginal effect: Control # below median motivation	0.353	0.280	0.739	0.357
	[0.060]	[0.066]	[0.158]	[0.116]
Marginal effect: Treatment # below median motivation	0.277	0.204	0.459	0.460
	[0.035]	[0.035]	[0.145]	[0.087]
Marginal effect: Control # at/above median motivation	0.343	0.454	0.320	0.251
	[0.057]	[0.084]	[0.096]	[0.113]
Marginal effect: Treatment # at/above median motivation	0.359	0.332	0.404	0.539
	[0.035]	[0.049]	[0.070]	[0.088]
N	759	469	121	140

The marginal effect is the change in outcome as a function of the change in treatment, all else constant. It is calculated by averaging across each observation's marginal treatment effect.

Standard thresholds for significance as follows: + p < 0.10, * p <0.05, ** p <0.01; however, we increase the threshold in our interpretation using the Hochberg step up procedure.

Benefit status

(1) Full sample	(2) North East only	(3) West London only	(4) Blackpool only

	Has come off benefits	Has come off benefits	Has come off benefits	Has come off benefits
Treatment	-0.124	-0.260	0.891	-0.586
	[0.298]	[0.446]	[0.833]	[0.607]
atabovemedian_motivation=0	0.000	0.000	0.000	0.000
	[.]	[.]	[.]	[.]
atabovemedian_motivation=1	0.283	0.258	1.774+	-1.108
	[0.392]	[0.560]	[0.972]	[1.092]
Control # atabovemedian_motivation=0	0.000	0.000	0.000	0.000
	[.]	[.]	[.]	[.]
Control # atabovemedian_motivation=1	0.000	0.000	0.000	0.000
	[.]	[.]	[.]	[.]
Treatment # atabovemedian_motivation=0	0.000	0.000	0.000	0.000
	[.]	[.]	[.]	[.]
Treatment # atabovemedian_motivation=1	-0.073	0.083	-1.416	0.073
	[0.377]	[0.586]	[0.932]	[0.827]
Female	-0.058	0.188	-0.308	-0.628
	[0.162]	[0.248]	[0.301]	[0.416]
Age	0.004	0.008	0.003	-0.001
	[0.006]	[0.008]	[0.012]	[0.017]
Motivation score	0.028	0.021	-0.005	0.153*
	[0.019]	[0.025]	[0.038]	[0.066]
Year and month dummies	Yes	Yes	Yes	Yes
ESA	0.000	0.000	0.000	0.000

	[.]	[.]	[.]	[.]
IB	0.000	0.000		
	[.]	[.]		
ICA	1.641*	2.806**	0.000	0.000
	[0.677]	[0.810]	[.]	[.]
IS	0.310	0.794	0.000	0.000
	[0.624]	[0.696]	[.]	[.]
JSA	0.431+	0.246	0.820+	0.084
	[0.243]	[0.384]	[0.447]	[0.591]
PIB	0.000	0.000		
	[.]	[.]		
RP	0.000	0.000		
	[.]	[.]		
SDA	0.000	0.000	0.000	0.000
	[.]	[.]	[.]	[.]
UC	-0.627*	-0.955*	0.060	-1.305+
	[0.250]	[0.399]	[0.453]	[0.674]
Number of days on benefits prior to trial	-0.000**	-0.000**	-0.000**	-0.000*
	[0.000]	[0.000]	[0.000]	[0.000]
Blackpool	0.663**	-	-	-
	[0.236]			
West London	0.500*	-	-	-
	[0.210]			
Constant	-2.515**	-2.258+	-1.814	-3.367+

	[0.654]	[1.322]	[1.397]	[1.923]
Marginal effect: Control # below median motivation	0.119	0.089	0.075	0.352
	[0.027]	[0.028]	[0.054]	[0.114]
Marginal effect: Treatment # below median motivation	0.108	0.073	0.154	0.263
	[0.017]	[0.017]	[0.041]	[0.083]
Marginal effect: Control # at/above median motivation	0.148	0.109	0.278	0.194
	[0.024]	[0.028]	[0.067]	[0.060]
Marginal effect: Treatment # at/above median motivation	0.128	0.095	0.199	0.139
	[0.015]	[0.018]	[0.037]	[0.035]
N	1674	998	383	236

The marginal effect is the change in outcome as a function of the change in treatment, all else constant. It is calculated by averaging across each observation's marginal treatment effect.

Standard thresholds for significance as follows: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$; however, we increase the threshold in our interpretation using the Hochberg step up procedure.

References

Becker, D., Drake, R. (1994). Individual placement and support: a community mental health center approach to vocational rehabilitation. *Community Mental Health Journal*, 30(2), 193-206.

Becker, D. R., Swanson, S. J., Bond, G. R., Carlson, L., Flint, L., Smith, G., & Lynde, D. (2008). Supported Employment Fidelity Scale, Unpublished scale (http://dms.dartmouth.edu/prc/employment/). Lebanon, NH: Dartmouth Psychiatric Research Center.

Bond, G., Drake, R., Becker, D. (2008). An update on randomized controlled trials of evidence-based supported employment outside the US. *Psychiatric Rehabilitation Journal*, 31(4), 280-90.

Bond, G., Drake, R., Becker, D. (2012). Generalizability of the Individual Placement and Support (IPS) model of supported employment outside the US. *World Psychiatry*, 11(1), 32-39.

Britt, C. (1997). Reconsidering the unemployment and crime relationship: variation by age group and historical period. *Journal of Quantitative Criminology*, 13(4), 405-428.

Clark, David. (2019). IAPT at 10: Achievements and challenges. Retrieved from https://www.england.nhs.uk/blog/iapt-at-10-achievements-and-challenges/.

Conservative Party. (2015). Strong leadership: a clear economic plan: a brighter, more secure future. Retrieved from https://cedrec.com/cedrec_images/upload/other_docs/ConservativeManifesto15.pdf.

Creed, P.A., Hicks, R., & Machin, M.A. (1998). Behavioural plasticity and mental health outcomes for long-term unemployed attending occupational training programs. *Journal of Occupational and Organizational Psychology*, 71(2), 171-191.

Creed, P.A., Machin, M.A., & Hicks, R. (1996). Neuroticism and mental health outcomes for long-term unemployed youth attending occupational skills training programs. *Personality and Individual Differences,* 21(4), 537-544.

Creed, P. A., Machin, M. A., & Hicks, R. (1999). Improving mental health status and coping abilities for long-term unemployed youth using cognitive-behaviour therapy based training interventions. *Journal of Organizational Behavior, 20*(6), 963-978.

Department for Work and Pensions & Department of Health & Social Care. (2017). Work, health and disability green paper: improving lives. Retrieved from https://www.gov.uk/government/consultations/work-health-and-disability-improving-lives.

Henry, A., Lucca, A., Banks, S., Simon, L., Page, S. (2004). Inpatient hospitalizations and emergency service visits among participants in an individual placement and support (IPS) model program. *Mental Health Services Research*, 6(4), 227-237.

Frederick, D., VanderWeele, T. J. (2019). Supported employment: meta-analysis and review of randomized controlled trials of individual placement and support. *PLoS ONE,* 14(2).

Kennedy, S., McInnes, R., Mackley, A., Bellis, A., Brown, S., Steele, S. (2019). Mental health and the benefits assessment process. *House of Commons Library*, CDP-0005.

Knapp, M. , Patel, A., Curran, C., Latimer, E., Catty, J., Becker, T., Drake, R., Fioritti, A., Kilian, R., Lauber, C., Rössler, W., Tomov, T., VB., Jooske, C., Adelina, W., Sarah, W.,
(2013) Supported employment: cost-effectiveness across six European sites. *World Psychiatry,* 12 (1), 60-68.

Local Government Association. (n.d.). Devolution deals. Retrieved from https://www.local.gov.uk/topics/devolution/devolution-deals.

McManus, S., Mowlam, A., Dorsett, R., Stansfeld, S., Clark, C., Brown, V., Wollny, I., Rahim, N., Morrell, G., Graham, J., Whalley, R., Lee, L., Meltzer, H. (2012). Mental health in context: the national study of work-search and wellbeing. *Department for Work and Pensions*, Research Report 810.

NHS Digital. (2019). Annual report on the use of IAPT services, England 2018-19. Retrieved from https://files.digital.nhs.uk/1C/538E29/psych-ther-2018-19-ann-rep.pdf.

Paul, K. & Moser, K. (2009). Unemployment impairs mental health: meta-analyses. *Journal of Vocational Behavior*, 74(3), 264-282.

Ritchie, J., Lewis, J., Nicholls, C.M. and Ormston, R. eds. (2013). *Qualitative research practice: A guide for social science students and researchers*. Sage.

Steadman, K., Thomas, R. (2015). An Evaluation of the 'IPS in IAPT' psychological wellbeing and work feasibility pilot. *The Work Foundation, Department of Health, Department for Work and Pensions.*