**Health-led Employment Trials (HLET)**

**Statistical analysis plan**

Richard Dorsett, University of Westminster

Helen Gray, Institute of Employment Studies

Supporting document for protocols ISRCTN68347173 and ISRCTN17267942

Note: this statistical analysis plan was registered before any outcome data was available to the authors

|  |  |
| --- | --- |
| **Version history** |  |
| Version 1 | 19/12/2019 |

**Contents**

[1 Description of the trial 3](#_Toc24094085)

[1.1 Principal research objectives to be addressed 3](#_Toc24094086)

[1.2 Trial design 4](#_Toc24094087)

[1.3 Data and timing of data availability 8](#_Toc24094088)

[2 Variables used in the analysis 9](#_Toc24094089)

[2.1 Control variables and baseline variables 9](#_Toc24094090)

[2.2 Outcome variables 9](#_Toc24094091)

[2.3 A comment on primary and secondary outcomes 10](#_Toc24094092)

[Effect modifiers – subgroup analysis 19](#_Toc24094093)

[3 Data analysis plan 19](#_Toc24094094)

[3.1 Recruitment, baseline characteristics and loss to follow-up 19](#_Toc24094095)

[3.2 Non-response and survey weights 19](#_Toc24094096)

[3.3 Missing data 19](#_Toc24094097)

[3.4 Impact estimation approach – primary outcomes 20](#_Toc24094098)

[3.5 Secondary outcomes 21](#_Toc24094099)

[3.6 Subgroup analyses 21](#_Toc24094100)

[3.7 Additional analysis - compliance 21](#_Toc24094101)

[4 Reporting and inference 22](#_Toc24094102)

# Description of the trial

The goal of HLET is to evaluate the effectiveness of the Individual Placement and Support (IPS) model of supported employment. IPS is a well-evidenced voluntary employment programme for supporting people with severe and enduring mental health needs in secondary care settings to find paid employment. This trial assesses its effectiveness among a broader group of people whose health represents an obstacle to employment. The trials were conducted in 5 clinical commissioning group (CCG) areas within the Sheffield City Region (SCR) (Barnsley, Bassetlaw, Doncaster, Rotherham and Sheffield) and 4 CCG areas within the West Midland Combined Authority (WMCA) (Birmingham, Dudley, Sandwell and Wolverhampton).

The intervention is delivered by IPS employment specialists. In SCR, these are all employed by South Yorkshire Housing Association. In WMCA, there are three providers: Dudley and Walsall NHS Trust, Prospects and Remploy. In both areas, the employment specialists are personal caseworkers providing employment support and co-ordinating wider health and support needs.

A full description of the trials is given in the protocol publications.[[1]](#footnote-1)

## Principal research objectives to be addressed

The trial will assess the effectiveness of IPS among people who have either been out of work due to a health condition or disability or who are working but struggling because of such conditions. The model of IPS trialled within SCR differs slightly from that within WMCA. Most notably, while both areas recruited out-of-work individuals to the trial, only in SCR were working individuals also eligible.

The primary hypotheses the trial aims to test is whether IPS encourages employment and improves health and wellbeing outcomes beyond what can be achieved with ‘business as usual’ (BAU or usual care).

Outcomes will be captured through administrative data and surveys. Impacts will be estimated at two points:

* Interim – 4-month impacts on survey outcomes for all recruits
* Final – 12-month impacts on administrative data and survey outcomes for all recruits.

Individuals recruited to the trial are randomised to either a treatment group or a BAU control group. Impacts will be estimated separately for:

* SCR recruits out of work when randomised
* SCR recruits in work when randomised
* SCR recruits in or out of work when randomised
* WMCA recruits out of work when randomised
* SCR and WMCA recruits out of work when randomised.

## Trial design

The evaluation follows an individually randomised control trial design, with equal allocation of recruits to IPS and BAU.

Referrals come from the local health system – GP practices, Improving Access to Psychological Therapy (IAPT) teams, physiotherapists and pain management teams – as well as via social prescribing and self-referral. Employment specialists conduct the initial meeting with trial recruits. At that meeting, recruits are screened for eligibility, asked to give their consent to be included in the trial (including use of data), complete a baseline survey and are randomly allocated to either the treatment or control condition. Those allocated to the treatment group are enrolled on IPS. Those allocated to the control group are supplied with information about other support available under BAU but are not provided with IPS.

The initial interview is conducted using a web-based randomisation tool. Randomisation itself is carried out using a series of pre-specified allocations accessed by the tool. Recruitment commenced on 8 May 2018 and ended on 31 October 2019. Over this period, 6,117 individuals were randomised in SCR and 3,682 in WMCA. Within SCR, 2,551 recruits were in work at the time of randomisation and 3,66 were out of work.

Recruitment was undertaken for as long a period as possible in order to maximise sample size, subject to the practical constraints around local delivery capacity and budget (the IPS support has a high average unit cost). A large sample size is desirable to maximise the statistical power of the trial, particularly important when considering multiple outcomes and subgroups.

Statistical power calculations were carried out for the case of a binary employment indicator. For the out-of-work (OOW) group, a BAU rate of 15% was assumed, based on official statistics showing that the proportion of new ESA (sickness and disability benefit) claimants who, within a year of referral, had entered work and remained employed for three months averaged 8.4% over the 2015 intake (nationally) for those with 12-month prognosis complaints and 16.1% for those without a 12-month prognosis complaint. The estimated likely average outcome of the treated group was informed by the Working Well pilot in Greater Manchester which was on course to achieve its target of 20%. Hence, for design purposes, it was assumed that IPS would increase labour market outcomes among those out of work at baseline from 15% to 20%. For the in-work (IW) group, there is less evidence available regarding expected BAU outcome levels. Conservatively, we assumed a BAU outcome of 50% and that the assumed 5 percentage point impact of IPS for the out of work groups would hold also for the in-work group.

Initial calculations in the protocol were based on an assumption of 30% of recruits in SCR being in work at baseline. In fact, this proportion has turned out to be closer to 40% and is reflected in the revised power calculations presented below. These calculations are based on simulations performed using the statistical package R. Impacts were estimated using linear regression of the outcome on a treatment group indicator, assuming an R-squared of 0.1. Table 1 presents estimates of power (1 minus the probability of type 2 error) and statistical significance (type 1 error) for the anticipated sample sizes achieved in SCR (the results for the OOW group apply equally to WMCA). We distinguish between outcomes taken from administrative data (which should be available for all recruits) and those taken from survey data (which are available only for survey respondents, assumed to be available for 50% of cases).[[2]](#footnote-2)

The conventional level of power to design trials around is 80%. This is achieved with the OOW group. For the IW group, estimates based on admin data come close to achieving this but those using survey data look underpowered. Earlier simulations suggested the minimum sample size for adequately powered impact estimates in the OOW group to be 1,300-1,350. With this in mind, impacts on outcomes taken from administrative data should be possible for subgroups accounting for as little as (roughly) 45% of the out of work population. For outcomes taken from survey data, the scope for subgroup analysis is more limited. For the IW group, subgroup analysis is likely to be unreliable.

**Table 1: Estimated power to detect 5 percentage point impact in full population (administrative data) and among survey respondent sub-sample (survey data)**

|  |  |  |
| --- | --- | --- |
|  | Out of work at baseline | In work at baseline |
|  | Admin data | Survey data | Admin data | Survey data |
| **Impact** | 0.05 | 0.05 | 0.05 | 0.05 |
| **BAU** | 0.15 | 0.15 | 0.50 | 0.50 |
| power | 0.99 | 0.88 | 0.78 | 0.50 |
| **type 1 error (1 outcome)** | 0.05 | 0.06 | 0.06 | 0.05 |
| N |  3,500  |  1,750  |  2,500  |  1,250  |

The flow chart (Figure 1) shows the process of recruitment and follow-up of participants in the trial.

**Figure 1: Trial flow chart in SCR**

**Out of work**

**N=3,566**

**In work**

**N=2,551**

Incomplete

(N=6)

**Eligible & consenting**

**N=6,223**

**Randomised**

**Randomised**

4-month FU

**Admin N=**

**Survey N = ?,???**

4-month FU

**Admin N=**

**Survey N = ?,???**

12-month FU

**Admin N=**

**Survey N = ?,???**

12-month FU

**Admin N=**

**Survey N = ?,???**

BAU

**N=1,783**

IPS

**N=1,783**

4-month FU

**Admin N=**

**Survey N = ?,???**

4-month FU

**Admin N=**

**Survey N = ?,???**

12-month FU

**Admin N=**

**Survey N = ?,???**

12-month FU

**Admin N=**

**Survey N = ?,???**

BAU

**N=1,274**

IPS

**N=1,277**

**Referred to trial**

**N=6,350**

Ineligible

(N=130)

**Figure 2: Trial flow chart in WMCA**

12-month FU

**admin=**

**Survey N = ?,???**

12-month FU

**admin=**

**survey N = ?,???**

IPS

**N=1,840**

12-month FU

**Admin N=**

**Survey N = ?,???**

12-month FU

**admin=**

**survey N = ?,???**

**Referred to trial**

**N=3,942**

BAU

**N=1,842**

**Randomised**

**N=3,682**

**Eligible & consenting**

**N=3,792**

Incomplete

(N=117)

Ineligible

(N=143)

## Data and timing of data availability

Analysis will be based on a dataset drawn from the following sources:

* baseline and control variables collected by the randomisation tool
* MI data from employment specialists
* administrative data on employment, earnings, benefit receipt and NHS usage
* surveys carried out 4 and 12 months post randomisation.

Data from these sources are sent to the Office of National Statistics, which merges them to give a single dataset which is then pseudonymised before being made available to analysts conducting the evaluation.

At the time of version 1 of this SAP, the timeline for receiving the data is not finalised. However, the earliest possible dates for the interim and final analyses are as follows:

* **Interim analysis** 4-month survey data linked to tool data and MI for interim analysis - Autumn 2020
* **Final analysis** 12-month survey and admin data linked to tool and MI data - 31 July 2021.

# Variables used in the analysis

## Control variables and baseline variables

We distinguish between control variables (background characteristics and trial related fixed effects) and baseline variables (baseline values of the outcome variable, or proxies for that). Both were collected through a survey conducted immediately prior to randomisation and delivered using the randomisation tool.

Control variables

Age

Gender

Ethnicity

Marital status

Dependent children

Qualifications

Employment history

Barriers to employment

Nature of health problem

CCG

Cohort

Baseline variables

Well-being (SWEMBS-7)

Health (EQ-5D-5L)

Life satisfaction (ONS1)

Work search self-efficacy (JSSE)

Health limit (DDA definition)

Looking for work

## Outcome variables

**Primary outcomes**

The protocol identifies three primary outcomes at the interim report stage (employment, health and well-being) and four primary outcomes at the final report stage (employment, earnings, health and well-being). All outcomes at the interim report stage are from the 4-month survey. At the final report stage, the employment and earnings outcomes are from linked tax records while the health and well-being outcomes are from the 12-month survey. We use different measures of employment at the interim and final reports. At the interim stage, we consider whether the individual is employed four months post-randomisation. For the final report, we consider whether the individual has been in work for 13 or more weeks in the 12 months following randomisation.

**Secondary outcomes**

The secondary outcomes are listed below under three domains corresponding to the three primary outcomes:

* Employment
	+ Employed
	+ Number of months employed since randomisation
	+ Earnings
	+ Earnings since randomisation
	+ Receiving out of work benefits
	+ Number of months receiving out of work benefits since randomisation
	+ Amount of benefits received
	+ Employed and receiving benefits
	+ Working (employed or self-employed)
	+ Working 16+ hours per week
	+ Number of weeks working since randomisation
	+ Number of weeks working 16+ hours per week since randomisation
	+ Number of continuous weeks working 16+ hours per week
	+ Job search self-efficacy
* Health
	+ Health
	+ Musculoskeletal Health
	+ Mental health
	+ DDA definition health
* Well-being
	+ Well-being
	+ Life satisfaction
	+ Self-efficacy

We also include one further domain:

* Health services
	+ Days in hospital
	+ Health appointments
	+ Health appointments attended
	+ A & E visits

More detail on outcomes is given in Table 2.

## A comment on primary and secondary outcomes

We highlight two potential complications that may affect impact estimates for outcomes drawn from administrative records and survey data, respectively. First, to use outcomes from administrative data requires that it can be linked to trial participants’ data. For the primary outcomes, the relevant administrative data is HMRC tax records since this will be used to identify who is in work. We think it unlikely that this linkage will not be possible, especially since nearly 90% of individuals recruited to the trial have provided their National Insurance number. However, it is still possible that such a situation could arise, in which case we will replace the employment primary outcome with a survey-based measure of how many weeks the respondent has been in work since randomisation.

The second complication would arise in the case where it is felt that nonresponse to the survey has undermined its ability to be used to deliver credible estimates of impact. Early indications are that response rates to the 12-month survey are below 50%. The power calculations presented above suggest that the numbers achieved should still permit adequately-powered tests if the response rate stays at that level. However, with perhaps the majority of the trial population not responding to the survey there is a question over the representativeness of the sample and the impact estimates based on the subset of survey respondents. Furthermore, while it is too early to know what the eventual situation will be, the early indications are of a treatment-control difference in the response rate in SCR. If this persists, it will undermine the credibility of estimates of impact on outcomes taken from the survey.

In such a case, we would demote the survey-based primary outcomes to secondary outcomes and promote earnings (as recorded in HMRC administrative data) to a primary outcome (along sustained employment) resulting in two primary outcomes. This is a substantial departure from the protocol approach and would only be adopted in the case where the survey data is felt to be of too low quality to support an impact analysis:

* **sample size** - if the number of respondents becomes too small, statistical power will suffer. The power calculations presented above suggest that a 50% response rate delivers adequate power for the OOW group. However, should the response rate fall to one third, statistical power would still be close to 70%. Consequently, across the range of likely scenarios, sample size does not appear to be a particular worry.
* **representativeness** – all analysis of survey data will be weighted[[3]](#footnote-3) which may go some way towards restoring representativeness (although it cannot address nonresponse determined by unobserved characteristics). Hence, representativeness does not appear to be a particular worry, although results will have to be interpreted accordingly.
* **treatment-control imbalance** – nonresponse may affect treatment and control groups differently such that it is not possible to separately identify the impact of the treatment from the confounding influence of selection into the respondent sample. We will compare baseline characteristics listed in section 2.1 between treatment and control groups and test for differences. The possibility of survey bias is the main concern surrounding nonresponse and prompts the decision rule below.

*If there is a treatment-control difference in the response rate of more than 5 percentage points* ***and*** *if baseline measures of job search efficacy, employment history, health or well-being differ significantly (p-value <0.05 after adjusting for multiple testing using Westfall-Young (1993)) in weighted regressions on the control variables* ***then*** *regard as primary outcomes sustained employment and earnings taken from the administrative data, and all other outcomes as secondary. [[4]](#footnote-4)*

**Table 2: Outcome measures for which impacts estimated**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Outcome** | **Source** | **Measure** | **Survey waves (I/F -Interim/ Final)** | **Reports (I/F - Interim/ Final)** | **Definition****(shown for survey-based outcomes only since nature of the administrative data is not yet known)** |
|  |  |  |  |  |  |
| **Primary outcomes** |  |  |  |  |  |
|  |  |  |  |  |  |
| Employment | Survey | **Emp**What is your current employment status? If you are currently employed and on sick leave or temporarily reduced hours, please record your normal working pattern prior to taking sick leave or reducing your hours.1. Employed full-time (30 hours a week or more)
2. Employed part-time (16-29 hours a week)
3. Employed part-time (Less than 16 hours a week)
4. Self-employed – working full-time (30 hours a week or more)
5. Self-employed – working part-time (16-29 hours a week)
6. Self-employed – working part-time (Less than 16 hours a week)
7. Unemployed – seeking work
8. Unemployed – not seeking work
9. Retired
10. Full-time education, training scheme/apprenticeship
11. Part-time education, training scheme/apprenticeship
12. Work experience
13. Volunteering
14. Carer (for children)
15. Carer (for adults)

Other – please specify |  | I | *emp=1,2,3,4, 5 or 6*Shown at month 4 (interim report) |
| Employment | HMRC | In work for 13 weeks or more in the 12 months following randomisation |  | F | Calculated from HMRC data on the start and end dates of employment spells |
| Earnings since randomisation | HMRC | Total earnings since randomisation  |  | F | Calculated from HMRC data on earningsCovering the 12 months since randomisation |
| Health | Survey | EuroQol-5D-5L (EQ-5D-5L)5-item. Comprises 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. | I/F | I/F | Health state index score calculated from individual health profiles using value set for England[[5]](#footnote-5)Shown at month 4 (interim report) and month 12 (final report) |
| Well-being | Survey | Short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS) 7-item.  | I/F | I/F | Sum of raw scores across 7 items, with the total converted using a [conversion table](https://warwick.ac.uk/fac/sci/med/research/platform/wemwbs/using/howto/swemwbs_raw_score_to_metric_score_conversion_table.pdf) in line with [recommended use](https://warwick.ac.uk/fac/sci/med/research/platform/wemwbs/using/howto/)Shown at month 4 (interim report) and month 12 (final report) |
|  |  |  |  |  |  |
| **Secondary outcomes** |  |  |  |  |  |
|  |  |  |  |  |  |
| ***Employment*** |  |  |  |  |  |
| Employed | HMRC | Employment by month since randomisation |  | F | Calculated from HMRC data on the start and end dates of employment spellsCovering the 12 months since randomisation |
| Number of months employed | HMRC | Number of months employed since randomisation |  | F | Calculated from HMRC data on the start and end dates of employment spellsCovering the 12 months since randomisation |
| Earnings | HMRC | Earnings by month since randomisation  |  | F | Calculated from HMRC data on earningsCovering the 12 months since randomisation |
| Receiving out of work benefits | DWP | Receiving out of work benefits by month since randomisation |  | F | Calculated from DWP data on the start and end dates of benefit spellsCovering the 12 months since randomisation |
| Number of months receiving out of work benefits since randomisation | DWP | Number of months receiving out of work benefits since randomisation |  | F | Calculated from DWP data on the start and end dates of benefit spellsCovering the 12 months since randomisation |
| Amount of benefits received | DWP | Out of work benefits amount by month since randomisation |  | F | Calculated from DWP data on benefit paymentsCovering the 12 months since randomisation |
| Employed and receiving benefits | DWP, HMRC | Employed while receiving out of work benefits by month since randomisation |  | F | Calculated from DWP data on the start and end dates of benefit spells and HMRC data on start and end dates of employment spells.Covering the first 12 months since randomisation |
| Working (employed or self-employed) | Survey | **Emp**What is your current employment status? If you are currently employed and on sick leave or temporarily reduced hours, please record your normal working pattern prior to taking sick leave or reducing your hours.1. Employed full-time (30 hours a week or more)
2. Employed part-time (16-29 hours a week)
3. Employed part-time (Less than 16 hours a week)
4. Self-employed – working full-time (30 hours a week or more)
5. Self-employed – working part-time (16-29 hours a week)
6. Self-employed – working part-time (Less than 16 hours a week)
7. Unemployed – seeking work
8. Unemployed – not seeking work
9. Retired
10. Full-time education, training scheme/apprenticeship
11. Part-time education, training scheme/apprenticeship
12. Work experience
13. Volunteering
14. Carer (for children)
15. Carer (for adults)
16. Other – please specify
 | I/F | F | *emp=1,2,3,4, 5 or 6*Shown at month 12 (final report) |
| Working 16+ hours per week | Survey | Emp, as above | I/F | I/F | *emp=1,2, 4, or 5*Shown at month 4 (interim report) and month 12 (final report) |
| Number of weeks working since randomisation | Survey | {*Ask if currently working, emp=1,2,3,4,5,6 OR ContEmp1=1}***ContEmp2** During the last 4/12 months, approximately how many **weeks** have you been in paid work?  | I/F | I/F | ContEmp2 Covering the 4 months since randomisation in the case of the interim report and 12 months in the case of the final report |
| Number of weeks working 16+ hours per week since randomisation | Survey | {*Ask if ContEmp2>0}***ContEmp3** And out of these [textfill **ContEmp2**] weeks, how many have you worked more than 16 hours a week (on average)? | I/F | I/F | ContEmp3Covering the 4 months since randomisation in the case of the interim report and 12 months in the case of the final report |
| Number of continuous weeks working 16+ hours per week | Survey | {*If ContEmp3>1}***ContEmp4** Have these [textfill **ContEmp3**] weeks been continuous (i.e. without a gap)? | I/F | I/F | ContEmp4Covering the 4 months since randomisation in the case of the interim report and 12 months in the case of the final report |
| Job search self-efficacy | Survey | Job Search Self-Efficacy Scale (JSSE). 9-item. Self-efficacy relating to finding employment |  | I | Scores on each of the individual items will be summed together and divided by the total number of items to derive the mean score on the job search self-efficacy indexShown at month 4 (interim report) and month 12 (final report) |
|  |  |  |  |  |  |
| ***Health & well-being*** |  |  |  |  |  |
| Musculoskeletal Health |  | Musculoskeletal Health Questionnaire (MSK-HQ). 2-item. On joint, back, neck, bone and muscle pain and stiffness, including how much the respondent has been bothered by this in the two weeks prior to interview. | I/F | I/F | **MSKImp = 4 or 5****(**joint or muscle symptoms bothered you very much/ extremely in the last 2 weeks?) Shown at month 4 (interim report) and month 12 (final report) |
| Mental health | Survey | General Anxiety Disorder-7 (GAD-7), 7-item, and Patient Health Questionnaire-8 (PHQ-8), 8-item on anxiety and depression. | F | F | Sum of GAD-7 scoresSum of PHQ-8 scoresShown at month 4 (interim report) and month 12 (final report) |
| DDA definition health | Survey | *{Ask all}*HCondDo you still have one or more health conditions or disabilities?1. Yes
2. No

*{Ask if HCond=2 (does not have a health condition)}*HCChkCan I just check, did you previously have a health condition that you have now recovered from?1. Yes
2. No

*{Ask if HCond=1 (still has a health condition or disability)}*HCLimitTo what extent does your health condition(s) or disability(ies) limit your ability to carry out everyday activities? If you have more than one health condition, or your health condition fluctuates, please consider the overall impact. TEL INTERVIEWER: READ OUT.1. A great deal2. To some extent3. A little4. Not at all | I/F | I/F | HCLimit = 1Shown at month 4 (interim report) and month 12 (final report) |
| Life satisfaction | Survey | Office for National Statistics Personal Well-being Questions (ONS-1).  | I/F | I/F | ONS1Shown at month 4 (interim report) and month 12 (final report) |
| Self-efficacy | Survey | General Self-Efficacy Scale (GSE Scale) 10-item. Designed to assess perceived self-efficacy and ability to cope with daily hassles as well as adaptation after experiencing stressful life events | I/F | I | Sum of raw scoresShown at month 4 (interim report) and month 12 (final report) |
|  |  |  |  |  |  |
| ***Use of health services*** |  |  |  |  |  |
| Days in hospital  | NHS-D | Total number of days in hospital since randomisation |  | F | Calculated from Hospital Episode Statistics on the length of inpatient spellsCovering the 12 months since randomisation |
| Health appointments | NHS-D | Total number of health appointments made since randomisation |  | F | Calculated from Outpatients, Community Services Dataset, Mental Health Services Dataset and Improving Access to Psychological Therapies records on the date of appointmentsCovering the 12 months since randomisation |
| Health appointments attended | NHS-D | Percentage of health appointments attended since randomisation |  | F | Calculated from Outpatients, Community Services Dataset, Mental Health Services Dataset and Improving Access to Psychological Therapies records on the date of appointments and whether the patient attendedCovering the 12 months since randomisation |
| A & E visits | NHS-D | Total number of A & E visits since randomisation |  | F | Calculated from Hospital Episode Statistics A & E data on A & E arrival dateCovering the 12 months since randomisation  |

## Effect modifiers – subgroup analysis

All subgroups are based on observed characteristics/circumstances at randomisation. The analysis will include the following subgroups:

* gender (male, female)
* age (<30, 30-39, 40-49, 50+)
* % of time in work over 2 years pre-randomisation (for those out of work at baseline: never, some but less than half, half or more; for those in work at baseline: less than half, half or more)
* Baseline health condition (mental only, musculo-skeletal only, other)
* Baseline health as measured by EQVAS (bottom third, middle third, top third)

# Data analysis plan

## Recruitment, baseline characteristics and loss to follow-up

The CONSORT flow chart (Figure 1) will be produced showing the number of eligible participants, the number randomised and the numbers for which survey and administrative data outcomes are available at 4 and 12 months post-randomisation.

## Non-response and survey weights

Initial analyses will compare control and baseline characteristics of all recruits, 4-month survey respondents and 12-month survey respondents. For survey respondents, we will test the differences in these characteristics. We will report the proportion of trial participants responding to the surveys. In all cases, we will show results for treatment group, control group and treatment and control group combined. All analysis based on survey data will use survey weights.

## Missing data

In principle, the analysis based on administrative data should be unaffected by missing data. In practice, it may be that some trial participants cannot be matched to their administrative records, perhaps because the identifying variables required for linking (NHS number and national insurance number) were wrongly recorded in the randomisation tool. In such a scenario, we would expect there to be no systematic differences across treatment arms and therefore that estimating results on the subsample of individuals who can be matched will not introduce any bias.

The same approach will be used for both administrative and survey data outcomes. Mean imputation will be used in the case of missing control or baseline data, and a dummy variable added to indicate each missing value; these dummies may be dropped if they are not statistically significant. Where outcomes are missing, such cases will be excluded from the regressions used to estimate the impact on that outcome.

## Impact estimation approach – primary outcomes

Impact analyses will follow CONSORT guidelines.[[6]](#footnote-6) It will be conducted in Stata. Intention to treat (ITT) impacts will be estimated using linear regression models of the form

$$y\_{it}=α+βT\_{i}+γy\_{it\_{0}}+δX\_{i}+ε\_{i}$$

where $i$ indexes individuals, $t$ refers to the time of the outcome (4- or 12-months post-randomisation, $t\_{0})$, the outcome variable is $y$, $T$ indicates membership of the treatment group, $X$is a vector of personal and trial characteristics (measured at randomisation) and $ε$ is an error term.

The variables included in $X$ are those listed in section 2.1, parameterised as follows:

Age, age-squared

Female dummy

Non-white dummy

Partner dummy

Dependent children dummy

Highest qualifications dummies: post-A-level; A-level; GCSE A-C

Health condition: mental, musculo-skeletal, other

Health (EQ-5D-5L)

CCG dummies

Cohort dummies: apr-jun 2018; jul-sep 2018; oct-dec 2018; apr-jun 2019; jul- dec 2019

The baseline variable, $y\_{t}\_{0}$, will match as far as possible the outcome variable, $y\_{t}$. For the employment primary outcome, we control for employment history and barriers to employment:

Proportion of time employed in last 2 years before randomisation:

Dummy for at least some of the time (if out of work at baseline)

Dummy for at least half the time (if in work at baseline)

Barriers to employment dummies: Difficulty finding a suitable job; Availability or cost of transport to work; Availability or cost of childcare; Lack of qualifications or experience; Lack of confidence in abilities or skills; Physical health condition; Mental health condition; Caring for a child, or an elderly or disabled family member; Being financially being worse off

Looking for work dummy

Work search self-efficacy (JSSE) index

For the health primary outcome, the baseline measure will be:

Health (EQ-5D-5L)

For the well-being primary outcome, the baseline measure will be:

Well-being (SWEMBS-7)

This approach will be taken for both interim and final impact estimates.

##

## Secondary outcomes

Impacts on secondary outcomes will be estimated using the same approach as for primary outcomes. The baseline variable, $y\_{t}\_{0}$, will match as far as possible the outcome variable.

## Subgroup analyses

Differential impacts on subgroups will be estimated using an augmented form of the regression models, interacting the subgroup dummy variable with the treatment dummy

$$y\_{it}=α+βT\_{i}+\sum\_{g=2}^{G}β\_{g}T\_{i}S\_{ig}+γy\_{it\_{0}}+δX\_{i}+ε\_{i}$$

where $S\_{ig}$ indicates membership of the subgroup $g$ so that $β\_{g}$ captures the differential impact of treatment for that subgroup. Each $β\_{g}$ coefficient captures the extent to which the mean impact differs for members of that subgroup relative to the reference subgroup (implicitly group 1 in the equation). Statistical significance of impact variation across subgroups will be assessed through a joint test of the $β\_{g}$ estimates.

## Additional analysis - compliance

We will use management information from IPS providers to report on the extent of nonparticipation among those in the treatment group. Members of the treatment group will be considered to have not complied with IPS if they have not attended any IPS interviews following random assignment, as recorded on the Management Information systems of providers. The randomisation tool includes checks to ensure that individuals cannot be assigned to the treatment group if they have previously been assigned to the control group. Therefore, the analysis will be based on an assumption that it is not possible for members of the control group to be non-compliers by receiving the treatment. This means that any non-compliance will be one-sided i.e. it can only arise if those assigned to the treatment group do not take-up IPS.

We will conduct an instrumental variable (IV) regression to estimate the local average treatment effect (complier average causal effect), using the randomisation outcome as the instrument. In the first stage regression we will estimate the probability of attending at least one IPS session for those assigned to the treatment group. In the second stage regression we will use the predicted take-up rate from the first stage to estimate the impact of treatment for compliers.

# Reporting and inference

The intention to treat impact estimates will be presented as the difference in regression-adjusted means (continuous outcomes) or proportions (binary outcomes) between the treatment and control groups. For binary outcomes, results tables will report the control group proportion and the estimated impact expressed as a percentage point difference. For continuous outcomes, results tables will report the control group mean and the estimated impact expressed as a difference in means. For continuous outcomes, the impacts will also be expressed in units of control group standard deviation (Glass’ Delta):

$$Δ=\frac{μ^{T}-μ^{C}}{σ^{C}}$$

where $Δ$ is Glass’ Delta, $μ^{T}and μ^{C}$ are the (regression-adjusted) mean outcomes for the treatment and comparison group respectively, and $σ^{C}$ is the standard deviation of the outcome in the control group.

Robust standard errors will be reported along with p-values that are adjusted for multiple testing using the step-down procedure of Westfall and Young (1993). This adjustment for multiple testing will be carried out for the primary outcomes, for each domain of secondary outcomes and for sub-groups within each primary outcome.

1. See https://www.isrctn.com/ISRCTN68347173 and https://www.isrctn.com/ISRCTN17267942 [↑](#footnote-ref-1)
2. Should the response rate be lower than 50%, statistical power of impacts estimated for outcomes taken from survey data will also be lower than those presented in Table 1. [↑](#footnote-ref-2)
3. These weights will be supplied by the fieldwork organisation, NatCen [↑](#footnote-ref-3)
4. Westfall, P. and Young, S. (1993) Resampling-based multiple testing: Examples and methods for p-value adjustment. [↑](#footnote-ref-4)
5. Devlin, N. J., Shah, K.K., Feng, Y., Mulhern, B., van Hout, B. (2018) ‘Valuing health-related quality of life: An EQ-5D-5L value set for England,. Health Economics 27(1): 7-22. [↑](#footnote-ref-5)
6. Schulz, K., Altman, D. and Moher D. (2010) CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC medicine 2010;8:18 [↑](#footnote-ref-6)