



Supporting employment among individuals with common health problems: Experimental evidence from England[☆]

James Cockett^{a,1}, Richard Dorsett^{b,*,2}, Matthew Gould^{b,2}, Helen Gray^{a,3}

^a Institute for Employment Studies, City Gate, 185 Dyke Road, Brighton BN3 1TL, UK

^b University of Westminster, 35 Marylebone Road, London NW1 5LS, UK

ARTICLE INFO

JEL classification:

C93
I14
I31
J68

Keywords:

Randomised trial
Supported employment
Employment
Earnings
Health
Well-being

ABSTRACT

Individual Placement and Support (IPS) is a vocational rehabilitation intervention with strong evidence of effectiveness at increasing employment among those with severe mental illness in secondary care. This paper assesses its effectiveness for individuals with mild-moderate mental or physical health problems in primary or community care settings. We present the results of three large randomised trials delivered across two sites in England. For those in work when randomised, IPS increased employment 12 months later by 3 percentage points and increased health and well-being by 0.10 and 0.18 standard deviations, respectively. For individuals not working when randomised, employment impacts (4 percentage points) were found in one site and health and well-being impacts (0.10 and 0.12 standard deviations, respectively) in another. The results suggest IPS can work for a broader population than those with severe mental health problems but that its effectiveness may vary according to the nature of implementation and the local context. Projected three-year benefits comfortably exceeded costs where there was a health impact. A positive longer-term return is possible without a health impact but would require employment and earnings impacts to be sustained.

1. Introduction

A substantial proportion of working age adults have a disability or health problem. OECD statistics show disability affecting almost one in five 15–69 year-olds before the COVID-19 pandemic (OECD, 2022). The disability employment gap – the difference in employment rates between those with and without a disability – stood at 27 percentage points in 2019. At the same time, the disability unemployment gap was 9 percentage points, highlighting the greater challenge to finding a job facing those with a disability. In all countries, employment is lower and unemployment is higher for those with a disability than for those without. Transitions from employment into unemployment or economic inactivity are likewise higher for people with a disability.

People with a disability are particularly likely not to participate in the labour market. In the European Union, economic inactivity among 15–64 year-olds was 22% in 2024 for those without a disability but 45% for those with a disability. There is a clear gradient with the severity of individuals' conditions, from an inactivity rate of 35% for those with a less severe disability to 68% for those

[☆] This article is part of a Special issue entitled: 'Field Experiments to Reduce Inequality' published in European Economic Review.

* Corresponding author.

E-mail address: r.dorsett@westminster.ac.uk (R. Dorsett).

¹ Current affiliation: Chartered Institute of Personnel and Development, 46-48 Grosvenor Gardens, London SW1W 0EB.

² Current affiliation: independent researcher.

³ Current affiliation: Learning and Work Institute, Office G12, The Foundry, 17 Oval Way, London SE11 5RR.

<https://doi.org/10.1016/j.eurocorev.2026.105308>

Received 23 December 2024; Received in revised form 23 February 2026; Accepted 27 February 2026

Available online 11 March 2026

0014-2921/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

with a severe disability.⁴ In the US, inactivity in 2024 for 16–64 year-olds was 59% for those with a disability compared to 18% for those without.⁵

Policies and programmes that can encourage or facilitate employment among those with a health problem or disability are of considerable interest. In this paper, we present findings on the effectiveness of Individual Placement and Support (IPS), a model of supported employment that has been found to increase employment among individuals with severe mental health problems. Our central research question focuses on whether IPS has a positive effective on employment, earnings, health and well-being when extended to the broader population of people with common health problems. This was evaluated by conducting large-scale randomised trials in two areas of England.

England, and the UK more generally, provides an interesting case study for such a study. First, economic inactivity on the grounds of long-term sickness is at record levels in the UK. Towards the end of 2024, it stood at nearly 2.8 million and accounted for 30% of all working age inactivity, up from 21% in 1993 (when records started).^{6,7} Second, more recent trends set the UK apart from other G7 countries in that there has been no post-pandemic reduction in economic inactivity. The growth since 2019 in claims for health-related benefits has been much higher in the UK than in other countries for which data are available, most of which instead saw a decline (Latimer et al., 2024). Third, the UK has a long history of developing policies to promote employment among people claiming out-of-work benefits for health reasons. This dates back to 2001, when welfare provision became co-located with the public employment service ('Jobcentre Plus') and performance targets incentivised caseworkers to focus more on getting disabled claimants into work (see, for example, Koenig et al., 2019).

Countries vary in the extent to which people with a disability are able to participate in the labour market. Part of the reason for this lies in institutional differences around employment rights, anti-discrimination protections and welfare systems for disabled people. Strong employment rights may protect disabled workers from dismissal, but can also make employers reluctant to hire them. Anti-discrimination laws can help address this. Welfare systems are likewise important in that those providing generous benefits to individuals with a disability can create a disincentive to work, particularly if taking a job risks losing established income or health coverage.

The three-model typology of welfare regimes proposed by Esping-Andersen (1990) provides a useful framework to understand such institutional differences and to interpret the findings of this paper. The UK is generally classified as following the Anglo-Saxon (liberal) model, characterised by less generous welfare benefits and more flexible labour markets, with anti-discrimination legislation used to safeguard the interests of disabled people in the labour market. The Nordic model, on the other hand, has generous welfare, strong employment protections and robust anti-discrimination laws. There is support for workplace accommodations and integration of disabled people into open employment. The Continental model (e.g. France, Germany, Japan) is conservative-corporatist in the sense that the role of family is promoted, and welfare is earnings-related and administered through the employer. Employers must also meet quotas for hiring disabled workers.

The analysis contained in this paper contributes to the literature in several regards. First, it provides robust evidence on whether IPS has a positive impact on the employment of individuals with low-moderate severity health problems, rather than just those with severe mental health problems. The synthesis report of its findings (Newton et al., 2023) is the only study included in Whitworth et al. (2024) that does not confine eligibility to individuals with mental health problems or other specific attributes. With nearly 10,000 recruits in total, these are by far the largest trials of IPS to ever be conducted. Second, the results cover health and well-being impacts, alongside employment and earnings. IPS is designed to improve employment outcomes but, in line with the view that work can be good for your health, it is relevant to consider these broader outcomes too. Third, the trials extended eligibility to individuals with jobs and thereby provide novel evidence on whether IPS can help with employment retention. As already noted, across all OECD countries, transitions out of work are higher for people with a disability. This represents a major policy challenge for which the current IPS evidence base, focused on those out of work, offers little insight. Fourth, since the trials took place in two sites – Sheffield City Region (SCR)⁸ and West Midlands Combined Authority (WMCA)⁹ – the results add to the evidence on geographic variation in effectiveness.

In terms of findings within the 12 months following randomisation, in SCR, among those in work when randomised, IPS increased sustained employment (defined as working 13 or more weeks in the year following randomisation) by 3 percentage points and self-reported health by 0.10 standard deviations. Both impacts were significant at the 90% confidence level. The impact on well-being was 0.18 standard deviations and significant at the 99% confidence level. Among those out of work in SCR, the only impacts significant at the 90% confidence level were on health and well-being (0.10 and 0.12 standard deviations, respectively). In WMCA, where all recruits were out of work when randomised, sustained employment was increased by 4 percentage points, significant at the 99% confidence level. There was no discernible impact on other outcomes. Outcomes tracked to the 21-month point showed sustained employment impacts but also that a statistically significant earnings impact emerges in the longer-run.

The results show that IPS can work for a broadly-defined population but suggest that its effectiveness may vary according to the nature of implementation and the local context. To have a positive return within the relatively short period of three years requires

⁴ Eurostat, https://doi.org/10.2908/LFSA_IPGAEDDL, accessed 10 July 2025.

⁵ Calculated from Bureau of Labor Statistics 2024 data by dividing the number of people aged 16–64 not in the labour force by disability <https://www.bls.gov/news.release/disabl.t05.htm> by the size of the 16–64 population by disability <https://www.bls.gov/news.release/disabl.t01.htm>.

⁶ <https://www.ons.gov.uk/employmentandlabourmarket/peoplenotinwork/economicinactivity/timeseries/lf69/lms>.

⁷ <https://www.ons.gov.uk/employmentandlabourmarket/peoplenotinwork/economicinactivity/timeseries/lf2m/lms>.

⁸ SCR is made up of 5 clinical commissioning group areas: Barnsley, Bassetlaw, Doncaster, Rotherham and Sheffield.

⁹ WMCA is made up of 4 clinical commissioning group areas: Birmingham, Dudley, Sandwell and Wolverhampton.

that IPS improve health outcomes. Longer-term cost-effectiveness without an impact on health is also possible if employment and earnings effects are sustained.

The rest of this paper proceeds as follows. Section 2 presents the trial design, including an overview of IPS. Section 3 describes the data and Section 4 presents descriptive statistics. Section 5 outlines the estimation approach and Section 6 presents the estimated impacts. The robustness of the estimates to survey non-response are considered in Section 7 and the results as a whole are discussed in Section 8. Section 9 concludes.

2. Experimental design

2.1. Trial registration

The trials were conducted according to protocols and statistical analysis plans registered with ISRCTN (the UK's clinical study registry) prior to outcome data being available. The SCR trial (ISRCTN68347173) is available at <https://www.isrctn.com/ISRCTN68347173>. The WMCA trial (ISRCTN17267942) is available at <https://www.isrctn.com/ISRCTN17267942>.

2.2. The intervention - Individual placement and support

IPS is a voluntary programme of vocational rehabilitation, delivered by specialist employment advisers working alongside health professionals (Becker and Drake, 1994). Originally designed for those living with severe and enduring mental illness, it is a form of supported employment, a tailored model for helping people with disabilities enter and remain in work.¹⁰ In contrast to traditional rehabilitation programmes that seek to address health obstacles prior to entering work – the so-called *train-then-place* approach – supported employment prioritises rapid job entry, which it then seeks to maintain by addressing any threats to remaining in work that may arise. Accordingly, it is regarded as a *place-then-train* model. IPS is based on the belief that, so long as appropriate support is in place, anyone can work (Bond, 2004). The emphasis is on competitive employment, defined as a job that any person can apply for, regardless of disability status. This includes full- and part-time work, either as an employee or self-employed.

For the UK trials, employment specialists worked with participants to develop a tailored action plan and identify the help required to make work possible. They met regularly with clients, initially to assist with finding appropriate employment and later to provide support with managing health at work. In doing so, they facilitated input from the wider team of health professionals. They also actively searched for possible jobs and liaised with employers to identify and encourage further opportunities. Employment specialists were able to advise on financial issues, particularly as these related to the transition into employment. The relationship could also provide participants with a source of moral and emotional support.¹¹

Several surveys of experimental evaluations of IPS attest to its effectiveness in helping people with severe mental illness to find and keep work (Bond et al., 2020; Brinchmann et al., 2020; Frederick and Van der Weele, 2019). However, individuals with such conditions face very specific needs and it is not clear that the same model of IPS would be effective for other populations. De Winter et al. (2022) provide evidence of weaker employment impacts for people with common mental disorders. Where surveys include trials for people with conditions other than severe mental health problems (Bond et al., 2019; Probyn et al., 2021; Fadyl et al., 2020; Whitworth et al., 2024), the underlying studies tend to focus on subpopulations that are distinct in their own way – veterans with PTSD, young people, substance misusers – or to evaluate adapted versions of IPS. The exception is the Californian trial reported in Freedman and Millenky (2022). Eligibility covered a broad population of low-income, out-of-work individuals with a self-reported disability. For both earnings and employment, impacts were statistically significant in the first year post-randomisation but not the second year.

IPS has eight core principles (Bond et al., 1997):

- competitive employment as the primary goal
- eligibility based on patient choice
- integration of vocational and clinical services
- job search guided by individual preferences
- personalised benefits counselling
- rapid job search
- systematic job development
- time-unlimited support.

Adherence to these principles can be measured as the total score across 25 scales assessing distinct dimensions of fidelity.¹² Evidence suggests IPS is most successful when fidelity to its core principles is high (Bond et al., 2012; De Winter et al., 2020; Yamaguchi et al., 2022). However, Burns et al. (2015) tested a variant under which support was time-limited and found this to be

¹⁰ See, for example, <https://www.base-uk.org/home>.

¹¹ In SCR, employment specialists were all employed by South Yorkshire Housing Association, which delivered IPS under the branding *Working Win* <https://southyorkshire-ca.gov.uk/Explore/Working-Win>. In WMCA, there were three providers – Dudley and Walsall Mental Health Trust; Prospects; and Rempjoy – which delivered IPS under the branding *Thrive into Work* <https://www.wmca.org.uk/what-we-do/thrive-into-work/>.

¹² <https://www.centreformentalhealth.org.uk/ips-resources/>.

equally effective. The support available under the model of IPS considered in this paper was also time-limited. Participants could access 9 months of support before starting or recommencing a job, and 4 months of in-work support after starting a job, subject to an overall maximum of 12 months.

The other regard in which the model of IPS trialled in the UK differed was that eligibility was not universal. Across both sites, individuals were eligible if they were aged 18 or over and registered locally with a GP (medical practitioner). Other criteria varied by site. In SCR, individuals had to want to participate in voluntary support to find and/or stay in paid work, had to have a mild-to-moderate mental health condition and/or a physical health condition which was an obstacle to finding or staying in work, and had not to be receiving mental health support. The SCR criteria did not apply in WMCA, but instead individuals had to be neither in paid employment nor have had a paid job in the four weeks beforehand, had not to currently have an offer of employment and had to have a long-term health condition or disability which made it difficult to find work. In both SCR and WMCA, individuals were ineligible if they were receiving any employment support beyond the standard assistance by Jobcentre Plus.

2.3. Design of the trials

The evaluation was designed as an individual-level randomised control trial. Individuals could refer themselves to the trial or could be referred, on a voluntary basis, by the local health system – GP practices, Improving Access to Psychological Therapy¹³ teams, physiotherapists and pain management teams – as well as via social prescribing. Those joining the trial had an initial meeting with an employment specialist who used a web-based randomisation tool developed for the trial. The tool ensured that all intake interviews followed the same format: screening applicants for eligibility, requesting formal consent to participate in the trial and administering a baseline survey. Eligible applicants who completed the survey were then allocated to either the treatment or control condition. For this, the tool utilised several series of pre-specified allocation outcomes, constructed to achieve equal proportions allocated to treatment and control.¹⁴ In order to prevent predictability, each such series was itself constructed as a series of randomisation sequences within blocks of varying length.

Those allocated to the treatment group were enrolled on IPS. Those allocated to the control group were supplied with information about other ‘business as usual’ (BAU) support available but were not provided with IPS. Recruitment commenced on 8 May 2018 and ended on 31 October 2019. Over this period, 9797 individuals were randomised.

As well as randomisation itself, referral to the trial was also at the individual level. The relevance of this is that trial participants had no obvious connection or interaction with each other. This is in contrast to designs where, for example, employees are nested within workplaces, or pupils within classes, where other individuals within the same workplace or class are randomised to a different trial arm. The advantage of this is that it avoids the possibility of spillover effects that might arise through sharing information between treatment and control group members. Furthermore, the nature of the intervention is such that it is unavailable to members of the control group. For these reasons, while spillover effects are possible, if they do exist they are unlikely to be substantial. Hence, beyond this caveat, we do not consider spillover effects further.

The analysis was designed to allow separate analyses within SCR for those in work at randomisation and those out of work at randomisation. We therefore have three trials:

- SCR in work population (N = 2519)
- SCR out-of-work population (N = 3591)
- WMCA out-of-work population (N = 3675).

These sample sizes are large compared to other trials of IPS (Whitworth et al., 2024). This was in order to increase statistical power, particularly important since the likely size of effect for those with milder health problems is uncertain. We also show results for the pooled sample (N = 9785).¹⁵

For those randomised to the treatment group, IPS began with their adviser carrying out a vocational profile. This was used to create an action plan around which a programme of employment support could be developed. Support included CV development, discussing goals and help with job search and applications. This support was tailored to individual requirements and might include, for example, obtaining references from previous employers, identifying roles that aligned with individuals’ aspirations and providing advice on declaring health conditions and/or disabilities at work. Advisers also had a brokering role, contacting prospective employers about potential vacancies to achieve a job match. In line with IPS principles, the emphasis was on rapid job entry for those out of work. However, importance was also attached to finding the right job and one that could accommodate individuals’ health-related requirements. Mentoring and career counselling was one of the most significant elements of support. With respect to health and well-being, advisers provided ongoing advice and support about work and health issues, challenged the treatment group about what was possible with their condition and supported individuals to engage more effectively with their doctor by, for example, asking them more questions about their condition and letting them know more about their experience of medication. Advisers would also refer individuals to other local health services, such as counselling, where appropriate. Once in work, they provided practical support in the initial stages of employment such as helping to source appropriate clothing, financial help with transport or to bridge

¹³ Now renamed *Talking Therapies*.

¹⁴ The randomisation tool used a separate series for each clinical commissioning group area and, in SCR, also distinguished between whether individuals were in work or out of work at the time of randomisation.

¹⁵ The analysis sample totalled 9785 rather than 9797 (the number randomised) since 12 individuals withdrew from the trial post-randomisation.

the gap between coming off benefits and receiving their first wage, and financial advice to manage the transition from welfare to work. They also advised on whether, when and how to declare health conditions or disabilities to employers, and on reasonable adjustments that could be requested. Ways of maintaining the relationship between the adviser and the individual varied, with some having periodic telephone calls to check-in and others having a more formal schedule of appointments.¹⁶

3. Data

Analysis was based on data from multiple sources. The full trial population comprised 9785 individuals. As part of trial recruitment, baseline data were collected using a questionnaire administered immediately before randomisation. This gathered rich information on personal characteristics, labour market activity, income, health and well-being. This was matched to management information (MI) from providers delivering IPS. The MI recorded details of service delivery such as dates of meetings and their duration. Outcomes were observed using both administrative records and survey data. Individuals were matched to administrative records on employment, earnings and benefits. They were also matched to follow-on surveys conducted 4 and 12 months post-randomisation. The 4-month interview was intended to focus on intermediate outcomes such as job search. The 12-month interview was issued to all respondents who had not opted out from primary data collection (8945) and achieved a total of 4087 interviews, a coverage rate of 46%, or 42% of all trial recruits. To account for this, nonresponse weights were used when estimating impacts on survey outcomes. There was some variation by site in the proportion of trial recruits who responded to the 12-month survey: 46% among the SCR in-work group, 42% among the SCR out-of-work group and 39% among the WMCA out-of-work group. Survey data were collected primarily as web (n = 1768) or telephone (n = 2138) surveys; respondents who requested it were able to take part in a face-to-face interview (n = 181).

The trials have four (pre-registered) primary outcomes:

- Employment – whether employed for 13 weeks or more in the 12 months following randomisation. This outcome was constructed from employment records taken from tax data;
- Earnings – total earnings in the 12 months following randomisation. This outcome was constructed from earnings records taken from tax data;
- Health – self-reported health at the time of the 12-month survey interview. This outcome was constructed using the EuroQol-5D-5L (EQ-5D-5L) instrument that asks about 5 dimensions of health: mobility, self-care, usual activities, pain/discomfort and anxiety/depression¹⁷;
- Well-being – self-reported well-being at the time of the 12-month survey interview. This outcome was constructed using the Short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS) instrument that asks about 7 dimensions of well-being: feeling optimistic, feeling useful, feeling relaxed, dealing with problems well, thinking clearly, feeling close to others and being able to make up one's own mind.¹⁸

Although the trial was designed to provide robust estimates of the impacts of IPS on the primary outcomes, a number of secondary outcomes were also considered. These are more exploratory in nature and act to provide supporting evidence to the primary results. From administrative records, monthly employment and earnings impacts were estimated for the 21 months following randomisation. Impacts on the following administrative data outcomes relating to the twelve months post-randomisation were also estimated:

- Whether employed in month 12
- Number of months employed over 12 months
- Earnings in month 12
- Employed and on out-of-work benefits in month 12
- Whether receiving out-of-work benefits in month 12
- Number of months on out-of-work benefits over 12 months
- Amount of out-of-work benefits in month 12.

Secondary outcomes were also taken from the 12-month survey (and related to the time of interview):

- Whether employed
- Whether working 16+ h per week
- Number of weeks in work
- Number of weeks in work of 16+ h per week
- Number of weeks worked 16+ h per week continuously
- Job search self-efficacy (measured using the 9-item Job Search Self-Efficacy scale)
- Whether has musculoskeletal problems (measured using two items from the musculoskeletal Health Questionnaire (MSK-HQ) to calculate the proportion who have been bothered 'very much' or 'extremely' by joint or muscle symptoms in the last two weeks)

¹⁶ This changed with the onset of the pandemic, as employment specialists focused more on supporting well-being and skill development rather than finding work (Buzzeo et al., 2022).

¹⁷ <https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/>.

¹⁸ <https://www.corc.uk.net/outcome-experience-measures/short-warwick-edinburgh-mental-wellbeing-scale-swemws/>.

Table 1
Characteristics at randomisation and treatment-control balance.

	SCRIW			SCROOW			WMCA			All		
	Control	Treatment	P-val	Control	Treatment	P-val	Control	Treatment	P-val	Control	Treatment	P-val
Age (years)	41.6 (12.40)	41.5 (12.93)	0.85	40.2 (13.61)	40.3 (13.69)	0.91	40.7 (13.34)	41.9 (13.78)	0.01	40.8 (13.21)	41.2 (13.55)	0.11
Female	56.5	57.1	0.75	43.9	43.9	0.99	45	47.6	0.12	47.6	48.7	0.26
Non-white	10.1	10.9	0.53	13.6	14.1	0.67	35	36.7	0.29	20.8	21.8	0.23
Partner	47.3	48.1	0.7	25.5	25	0.7	25.5	24.2	0.35	31.1	30.6	0.58
Dependent children	27.3	26.2	0.57	22.6	23.5	0.53	24	23.7	0.82	24.3	24.3	0.95
Highest qualification			0.57			0.94			0.21			0.7
- No GCSE at A-C grade	17.5	17.9		33.6	32.7		34.1	32.4		29.6	28.8	
- GCSE at A-C grade	23.9	24.5		28.5	29.3		32.7	31.1		28.9	28.7	
- A-level	26.4	23.9		16.9	17.1		15.7	17.4		18.9	19	
- Above A-level	32.1	33.4		21.2	21		17.5	19.1		22.6	23.5	
Health condition			0.62			0.46			0.4			0.54
- Mental health (MH)	5.4	6		5.7	5.7		3.7	4.3		4.9	5.2	
- Musculoskeletal (MSK)	0.6	0.5		0.6	0.9		1.1	1.2		0.8	0.9	
- MH & MSK	1.0	0.7		1	1.8		0.9	0.8		1	1.2	
- MH & other	34.2	33.9		32	31.3		29.5	25.9		31.6	29.9	
- MSK & other	6.9	8.6		9.1	8.6		11.4	11.5		9.4	9.7	
- MH, MSK & other	48.9	48.2		47.4	48.0		47.3	50.4		47.7	48.9	
- Other	2.9	2.1		4.2	3.8		6.1	5.9		4.6	4.2	
Emp at randomisation and over past 2 years			0.24			0.29			0.58			0.82
- Not emp, never emp				50.4	49.8		58.7	58.3		40.5	40.2	
- Not emp, emp < half prior 2 years				23.9	26		22.5	21.8		17.2	17.7	
- Not emp, emp > half prior 2 years				25.7	24.2		18.6	19.8		16.4	16.3	
- Emp, emp < half prior 2 years	9.5	10.9								2.5	2.8	
- Emp, emp > half prior 2 years	90.5	89.1								23.4	23	
Health (EQ-5D-5L)	0.58 (0.25)	0.57 (0.25)	0.36	0.61 (0.25)	0.6 (0.25)	0.73	0.61 (0.26)	0.61 (0.25)	0.85	0.6 (0.25)	0.6 (0.25)	0.59
Well-being (SWEMWBS)	18.87 (3.68)	18.77 (3.90)	0.49	19.85 (4.21)	19.63 (4.22)	0.12	20.44 (4.37)	20.37 (4.18)	0.59	19.82 (4.19)	19.69 (4.17)	0.11
Omnibus test			0.7			0.64			0.14			0.53
Observations	1259	1260		1792	1799		1838	1837		4889	4896	

Notes: standard deviations for continuous variables are shown in parentheses. GCSE is General Certificate of Secondary Education, the qualification mostly gained at the end of compulsory schooling; A-level is Advanced Level qualification (mostly gained at the end of secondary education, at age 18). As described in the text, EQ-5D-5L is a measure of health and SWEMWBS is a measure of subjective well-being. The penultimate row reports the *p*-value from an omnibus test – this regresses treatment status on all the variables and tests for their joint significance.

- Whether has disability (Disability Discrimination Act (DDA) definition of whether respondent has a health condition/ disability which greatly limits the ability to carry out everyday activities)
- Life satisfaction (based on survey question using single item on life satisfaction from the Office for National Statistics Personal Well-being Questions (ONS1))
- Self-efficacy (based on survey question using the 10-item General Self-Efficacy (GSE) scale)
- Mental health (anxiety - measured using the sum of the General Anxiety Disorder (GAD) 7-item score)
- Mental health (depression - measured using the sum of the Patient Health Questionnaire (PHQ) 8-item score)

4. Descriptives statistics

Key characteristics at baseline are summarised in Table 1. Within each site, there is a high level of similarity between treatment and control groups. For each site, the column labelled ‘P-val’ shows the statistical significance of the treatment-control difference for each variable or group of variables. In nearly all cases, the difference is not statistically significant at the conventional level. The exception is age in WMCA. However, with this many comparisons, it is not surprising to find that not all differences are non-significant. The bottom row of Table 1 reports the results of an ‘omnibus’ test of joint significance from a regression of treatment group status on all covariates. The null hypothesis cannot be rejected, providing further support that the randomisation resulted in treatment and control groups that were statistically equivalent. Since the experimental groups are well-balanced in observed characteristics, there is no reason to believe they are likely to be imbalanced in unobserved characteristics.

Looking across sites, some characteristics were relatively stable. For instance, mean age was a little over 40 in all cases, roughly one quarter had dependent children and the combinations of mental and other illness or of mental, musculoskeletal and other illness consistently accounted for the majority of cases. Other characteristics differed by site. Within SCR, those in work were more likely than those out of work to be female, partnered, educated to A-level or higher and to have slightly better health. Considering only those out of work, recruits in SCR were more likely than those in WMCA to be white (but in other respects appeared quite similar).

Panel (a) of Table 2 shows the number of IPS sessions typical among participants. This is restricted to consideration of face-to-face meetings and telephone conversations and ignores other interactions such as email and text messages in order to focus on real-time

Table 2
Number of IPS sessions attended and time to first session.

	SCR (in work)	SCR (out of work)	WMCA (out of work)	All
(a) Number of sessions				
% with no sessions	8.2	11.2	0.9	6.6
Num. sessions	10.3	10.4	14.0	11.7
% with no detailed sessions	8.7	12.0	7.4	9.4
Num. detailed sessions	9.9	9.9	8.3	9.3
Total	1260	1799	1837	4896
(b) Time to first session (weeks since randomisation)				
1	16.8	17.7	51.6	30.2
2	23.0	26.4	29.6	26.7
3	20.1	17.5	9.8	15.3
4+	40.1	38.4	9.0	27.8
Total	1155	1597	1821	4573
(c) Time to first detailed session (weeks since randomisation)				
1	16.6	17.4	26.6	20.6
2	22.5	26.0	28.9	26.2
3	20.2	17.3	20.4	19.2
4+	40.7	39.3	24.1	34.0
Total	1149	1583	1701	4433

personal support in line with the IPS model. In SCR, IPS participants had roughly 10 sessions on average while in WMCA this was higher, at 14. Restricting this to detailed sessions (that is, those lasting more than 15 min) does little to change the average number in SCR, implying that nearly all sessions were for more than 15 min, but the number in WMCA fell to roughly 8 on average. Not attending any detailed sessions was relatively rare in WMCA (7.4%) and in SCR for those in work (8.7%) but slightly more common among those out of work in SCR (12.0%).

Panel (b) of Table 2 shows the week of the first IPS session. There was a marked tendency for this to happen sooner after randomisation in WMCA than in SCR. In WMCA, more than half had their first meeting within a week of randomisation, while in SCR less than one fifth of cases had their initial meeting that quickly. The difference becomes more pronounced when considering the percentage having their first meeting within two weeks: 81% in WMCA compared to 44% for those out of work in SCR and 40% for those in work in SCR. Panel (c) presents similar information for detailed IPS sessions. This narrows the difference but it remains the case that in WMCA there is usually less of a delay before meaningful engagement with IPS than is the case in SCR.¹⁹

5. Estimation approach

5.1. Primary outcomes impact analysis

The impact of assignment to the IPS treatment condition on outcome y was estimated using linear regression with the following specification:

$$y_{it} = \alpha + \beta T_i + \gamma y_{it0} + X'_{it0} \delta + v_{it} \quad (1)$$

where T_i is a dummy variable taking value 1 for an individual in the treatment (IPS) group and 0 for an individual in the control (BAU) group, y_{it0} is a baseline (i.e. at randomisation) measure of y_{it} and X_{it0} is a vector of personal and trial characteristics (also measured at randomisation). The baseline outcome, y_{it0} , was chosen to match as far as possible the outcome y_{it} . For the employment primary outcome, there was no within-trial variation in baseline employment so those regressions instead controlled for employment history in the 2 years prior to randomisation, identified barriers to employment, whether looking for work and work search self-efficacy, all measured pre-randomisation. The coefficient of interest is β and its estimate, $\hat{\beta}$, has an intention to treat (ITT) interpretation. To provide a better intuition of the impact on health and well-being, for those outcomes we also show the estimated impact relative to the standard deviation of the control group outcome, $\Delta = \hat{\beta}/\sigma_0$.

With four primary outcomes, there is an increased risk of finding spuriously significant impacts. To guard against this, as well as conventional p-values, we also report p-values adjusted for multiple testing using the step-down procedure of Westfall and Young (1993).²⁰ This was conducted separately for each of the three trials, with the justification that differences in both eligibility and the nature of treatment mean that each trial is essentially standalone. However, to control the error rate across the three trials, we also present impacts estimated on the pooled sample as a multi-site trial, again using the step-down approach to control for multiple testing.²¹

¹⁹ We return in Section 8 to a consideration of cross-site variation in the delivery of IPS. As a general point, our approach estimates the impact of IPS as delivered, which will vary within and between sites. Such variation is common in programmes involving adviser support. The impacts estimated for each site therefore correspond to an 'average' implementation for that site.

²⁰ This is implemented using the Stata file written by Jones et al. (2019).

²¹ Estimating the impact on the pooled sample was the suggestion of a reviewer and is a deviation from the registered statistical analysis plan.

Table 3
Estimated impacts, primary outcomes.

	$\bar{Y}_0$	$\bar{Y}_1$	$\hat{\beta}$	SE	Δ	P	P_{adj}		N
(a) SCR (in work)									
Employed for 13 or more weeks, %	79.7	82.8	3.2	1.50	–	0.04	0.09	*	2519
Total earnings, £	12,263	12,683	442	361	–	0.22	0.21		2519
Health (EQ5D5L)	0.57	0.60	0.03	0.01	0.10	0.05	0.09	*	1156
Well-being (SWEMWBS)	19.53	20.39	0.80	0.25	0.18	0.00	0.01	***	1153
(b) SCR (out of work)									
Employed for 13 or more weeks, %	26.7	25.0	–1.6	1.37	–	0.25	0.25		3591
Total earnings, £	2382	2120	–233	146	–	0.11	0.18		3591
Health (EQ5D5L)	0.57	0.59	0.03	0.01	0.10	0.02	0.06	*	1474
Well-being (SWEMWBS)	19.39	19.96	0.59	0.24	0.12	0.01	0.05	*	1464
(c) WMCA (out of work)									
Employed for 13 or more weeks, %	17.9	21.8	3.8	1.26	–	0.00	0.01	***	3675
Total earnings, £	1443	1600	150	117	–	0.20	0.36		3675
Health (EQ5D5L)	0.58	0.58	0.02	0.01	0.05	0.22	0.36		1377
Well-being (SWEMWBS)	19.85	20.26	0.45	0.25	0.09	0.07	0.19		1385
(d) All									
Employed for 13 or more weeks, %	37.1	38.6	1.6	0.8	–	0.04	0.07	*	9785
Total earnings, £	4573	4643	86	119	–	0.47	0.47	.	9785
Health (EQ5D5L)	57.2	59.1	2.4	0.8	0.08	0.00	0.01	***	4007
Well-being (SWEMWBS)	19.58	20.18	0.59	0.14	0.12	0.00	0.00	***	4002

Notes: $\bar{Y}_0$ and $\bar{Y}_1$ are the raw (unadjusted) mean outcomes for control and treatment groups, respectively. $\hat{\beta}$ is the ITT impact estimate, with robust standard error, SE. For continuous outcomes, Δ is $\hat{\beta}$ expressed in units of the standard deviation of the outcome. P is the P -value of the estimated impact and P_{adj} is the P -value with a (Westfall and Young, 1993) adjustment for multiple testing (10,000 replications). Asterisks denote statistical significance thresholds: * $P_{adj} < 0.10$, ** $P_{adj} < 0.05$, *** $P_{adj} < 0.01$. Impacts were estimated using linear regression controlling for baseline outcomes as described in Section 5. For employment outcomes, regressions controlled for employment history over the two years prior to randomisation: whether in work for at least half the time (in-work group) or whether in work for at least some of the time (out-of-work groups). Employment regressions also controlled for perceived barriers to employment (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), whether looking for work, and work search self-efficacy. Regressions for all outcomes controlled for the following baseline variables: age, age-squared, female, non-white, partnered, dependent children, highest qualifications (post-A-level; A-level; GCSE A-C); health condition (mental, musculo-skeletal, other), Health (EQ-5D-5L), stratum and randomisation cohort (April–June 2018; July–September 2018; October–December 2018; April–June 2019; July–December 2019).

5.2. Primary outcomes impact heterogeneity

Differential impacts across subgroups (all defined according to observed characteristics/circumstances at randomisation) were estimated using regression specifications that interacted subgroup identifiers with the treatment dummy

$$y_{it} = \alpha + \beta T_i + \sum_{g=2}^G \beta_g T_i S_{ig} + \gamma y_{it0} + X'_{it0} \delta + v_{it} \quad (2)$$

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 Eq. (2)). Statistical significance of impact variation across subgroups was assessed through a joint test of the β_g estimates.

The statistical analysis plan set out the intention to consider impact heterogeneity by gender, age, employment history, type of health condition and self-reported health. We also explored impact variation by time of randomisation. This sub-group was added in response to the outbreak of COVID-19 and the fact that this affected the nature of IPS delivery and is likely to have also affected outcomes.

5.3. Secondary outcomes

As well as the primary outcomes, a number of secondary outcomes were also considered. We note that the trial was designed with a view to providing robust estimates of the impacts of IPS on the primary outcomes. The secondary outcomes are more exploratory in nature and act to provide supporting evidence to the primary results. As with the primary outcomes, the P -values for the secondary outcomes were adjusted for multiple testing.

6. Results

6.1. Impacts on primary outcomes

Table 3 presents estimated impacts for the four primary outcomes. Panel (a) relates to those in SCR who were employed at the time of randomisation, panel (b) to those in SCR who were out of work, panel (c) to those in WMCA who were out of work and panel

(d) to the full randomised population. In each case, mean outcomes for the control and treatment groups are presented, $\bar{Y}_0$ and $\bar{Y}_1$, the estimated impact, $\hat{\beta}$, and its robust standard error (SE). For health and well-being outcomes, Δ , the impact in units of the outcome's standard deviation in the control group, is also shown. Statistical significance for each individual outcome is summarised by the p -value, P , while P_{adj} shows the p -value adjusted to control for the family-wise error rate. Conventional thresholds of statistical significance are highlighted using asterisks corresponding to P_{adj} . Lastly, the number of observations, N , is shown (making apparent the smaller base on which survey outcomes are estimated).

6.1.1. Impacts on employment and earnings

Control group mean outcomes represent the outcome expected in the absence of IPS. Table 3 shows that, within SCR, sustained employment within the 12 months following randomisation was much more common among those in work when randomised than among those out of work at that time. Specifically, four-fifths of the in-work group were employed for 13 or more weeks in the year following randomisation, compared to slightly over one quarter of the out-of-work group. Annual earnings were also much greater; five times higher over the course of the year. Employment and earnings among the WMCA control group, who were all out of work when randomised, were lower still than in SCR. This may reflect differences across the two sites in the composition of trial recruits, the strength of the local economy or the effectiveness of the support available to those out of work.

Despite the relatively high ambient rate of sustained employment in SCR for those in work when randomised, being in the treatment group caused an increase of 3.2 percentage points ($p = 0.04$, $p_{adj} = 0.09$). By contrast, there was no statistically significant effect in SCR for those initially out of work. In WMCA, on the other hand, employment was increased by 3.8 percentage points ($p = 0.00$, $p_{adj} = 0.01$). These employment impacts did not translate into earnings impacts, none of which were statistically significant.

Fig. 1 provides more detail on the evolution of employment and earnings impacts and how these continued into the longer term, up to 21 months post randomisation.²² In SCR, early employment impacts for the in-work group lost their statistical significance over time and earnings impacts were not significant in any month. The employment and earnings impacts among the out-of-work group in SCR were also not statistically significant in the longer term. Shorter term, the impact of being in the treatment group was to reduce employment and earnings. This is surprising in the sense that it does not appear consistent with a model of IPS that prioritises rapid employment entry so may raise questions over IPS fidelity in SCR. For WMCA, the monthly results show that the employment impact was sustained over time, remaining statistically significant 21 months post-randomisation. Furthermore, a statistically significant earnings impact emerged. By month 21, monthly earnings were increased by more than £40.

6.1.2. Impacts on health and well-being

Unlike the labour market outcomes described above, control group mean outcomes for health and well-being were very similar across all three trials. In SCR, IPS increased health by 0.1 standard deviations. This held for both the in-work group ($p = 0.05$, $p_{adj} = 0.09$) and the out-of-work group ($p = 0.02$, $p_{adj} = 0.06$). In WMCA, the impact was not statistically significant ($p = 0.22$, $p_{adj} = 0.36$). For well-being, there were also statistically significant impacts in SCR. Among the in-work group, IPS increased well-being by 0.18 standard deviations ($p = 0.00$, $p_{adj} = 0.01$) while for the out-of-work group there was an increase of 0.12 standard deviations ($p = 0.01$, $p_{adj} = 0.05$). In WMCA, evidence of an impact was weaker, an increase of 0.09 standard deviations ($p = 0.07$, $p_{adj} = 0.19$).

In view of the sizeable proportion of individuals for whom survey outcomes are not available and the treatment-control differences in this proportion, Table 4 presents bounds on the estimated impacts (Lee, 2009). With unweighted data, the untightened bounds all span zero, indicating that the impacts for the full sample may not be positive. The tightened bounds control for quartiles of the outcome variable measured at baseline. These are narrower and, in the SCR out-of-work case, exclude zero, indicating a positive impact on both health and well-being. It is relevant to point out that these bounds are themselves estimates. Once this is taken into account, the confidence interval on the bounds widens (and contains zero in all cases). As pointed out by Imbens and Manski (2004), the standard errors on the bounds can be used to derive a confidence interval on the full sample estimated impact that takes account of the uncertainty arising from non-response. Table 4 shows that the Imbens-Manski confidence intervals are consistently narrower in the tightened case than in the untightened case but still span zero.

The impacts for survey-based outcomes were estimated using weights to adjust for non-response. The third set of bounds in Table 4 take these weights into account. Only untightened bounds are shown since tightening resulted in breaches of the monotonicity requirement.²³ For the sample as a whole, the Imbens-Manski confidence interval on the health impact lies almost entirely above zero, while for the well-being impact it is comfortably above zero. For individual sites, the confidence intervals mostly span zero but, relative to the unweighted case, have lower limits that are much less negative.

Overall, the results point to the importance of weighting the data when estimating impacts on survey outcomes and strongly suggest a positive impact on health and well-being for the sample as a whole. It is not possible to rule out null effects on health and well-being in the individual sites. However, it should be noted that this is under the extreme assumptions generating the bounds and the Imbens-Manski confidence intervals; specifically, that the outcome in question is perfectly negatively or positively correlated with survey non-response. Under more realistic scenarios, the results in Table 4 suggest a positive impact on health and well-being in all sites.

²² Appendix A presents event study plots that show the evolution of unconditional treatment-control differences from two years before randomisation.

²³ This is the requirement that, for all quartiles of the outcome variable at baseline, treatment status affects selection in the same direction.

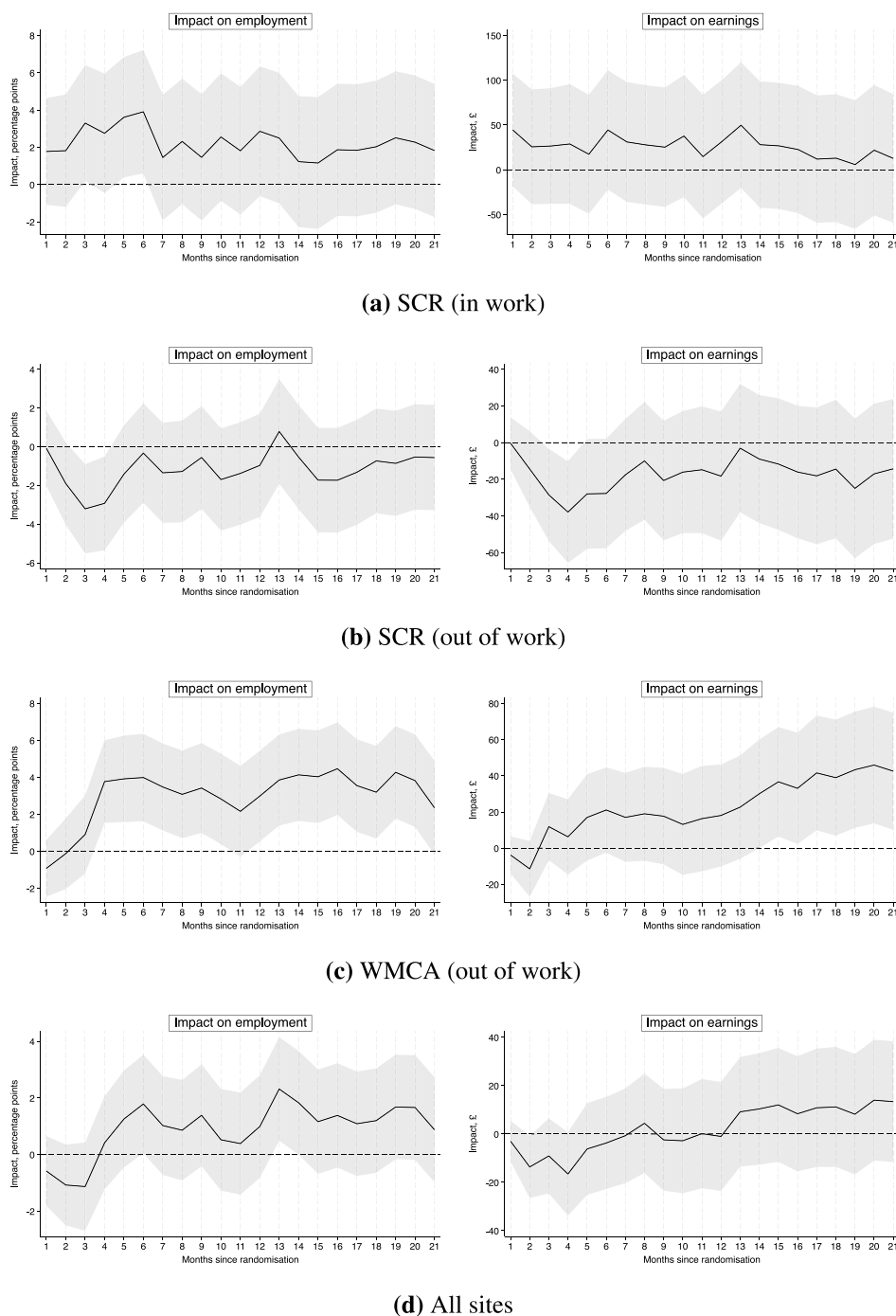


Fig. 1. Employment (left panel) and earnings (right panel) impacts, months 1–21 post-randomisation.

6.1.3. Impacts estimated across all sites

Although not included in the registered analysis plan, we include additional impacts estimated on the pooled sample. The results in panel (d) of Table 3 point to a positive overall effect on employment (1.6 percentage points; $p = 0.04$, $p_{adj} = 0.07$), health (0.08 standard deviations; $p = 0.00$, $p_{adj} = 0.01$) and well-being (0.12 standard deviations; $p = 0.00$, $p_{adj} = 0.00$) but not earnings. Panel (d) of Fig. 1 shows a sustained positive employment effect but provides less evidence of an impact on earnings.

Table 4

Bounds on estimated impacts, for primary health and well-being outcomes.

	Unweighted, untightened			Unweighted, tightened			Weighted, untightened		
	<i>LB</i>	<i>UB</i>	<i>CI</i>	<i>LB</i>	<i>UB</i>	<i>CI</i>	<i>LB</i>	<i>UB</i>	<i>CI</i>
(a) SCR (in work)									
Health (EQ5D5L)	−0.026 (0.022)	0.116 (0.024)	(−0.062, 0.155)	−0.018 (0.023)	0.102 (0.024)	(−0.056, 0.140)	0.016 (0.017)	0.030 (0.017)	(−0.013, 0.059)
Well-being (SWEMWBS)	−0.298 (0.362)	1.939 (0.347)	(−0.893, 2.509)	−0.352 (0.375)	1.836 (0.362)	(−0.969, 2.431)	0.852 (0.279)	0.865 (0.276)	(0.312, 1.401)
(b) SCR (out of work)									
Health (EQ5D5L)	−0.003 (0.024)	0.058 (0.029)	(−0.042, 0.107)	0.005 (0.024)	0.057 (0.028)	(−0.035, 0.103)	0.021 (0.016)	0.029 (0.017)	(−0.007, 0.059)
Well-being (SWEMWBS)	−0.169 (0.439)	0.991 (0.386)	(−0.891, 1.626)	0.024 (0.425)	1.166 (0.413)	(−0.676, 1.845)	0.430 (0.294)	0.702 (0.286)	(−0.066, 1.185)
(c) WMCA (out of work)									
Health (EQ5D5L)	−0.038 (0.023)	0.088 (0.027)	(−0.076, 0.132)	−0.028 (0.025)	0.071 (0.025)	(−0.069, 0.112)	0.006 (0.019)	0.010 (0.017)	(−0.030, 0.042)
Well-being (SWEMWBS)	−0.925 (0.400)	1.446 (0.370)	(−1.583, 2.054)	−0.403 (0.544)	1.200 (0.453)	(−1.297, 1.945)	0.348 (0.293)	0.464 (0.301)	(−0.180, 1.007)
(d) All									
Health (EQ5D5L)	−0.022 (0.013)	0.087 (0.015)	(−0.043, 0.112)	−0.013 (0.014)	0.078 (0.014)	(−0.036, 0.102)	0.017 (0.011)	0.020 (0.010)	(−0.003, 0.038)
Well-being (SWEMWBS)	−0.675 (0.228)	1.582 (0.210)	(−1.051, 1.927)	−0.326 (0.243)	1.382 (0.221)	(−0.726, 1.745)	0.557 (0.170)	0.631 (0.167)	(0.254, 0.929)

Notes: This table presents Lee bounds (Lee, 2009) on the primary health and well-being outcomes using the leebounds Stata package. *LB* is the lower Lee bound on the estimated impact (standard error in parentheses), *UB* is the upper Lee bound and *CI* is the Imbens-Manski confidence interval on the estimated impact. The untightened results do not control for covariates. The tightened results control for quartiles of the outcome variable measured at baseline.

6.2. Primary outcomes sub-group variation

Table 5 shows the extent to which overall impacts on primary outcomes varied across population subgroups. Impact heterogeneity by gender, age, employment history, type of health condition, self-reported health and cohort of randomisation was explored for the pooled sample. Across almost all outcomes and subgroups, the impression is of limited impact heterogeneity, as indicated by the P-values of the joint tests of significance. The exception to this is for well-being, which appears to have been increased less by IPS among those who were in the best health at the point of randomisation and among those with conditions other than mental or musculoskeletal illness.²⁴

The intention to examine impact variation by gender, age, work experience, baseline health condition (mental only, musculoskeletal only, other) and severity of baseline health was pre-registered. Reflecting the outbreak of COVID-19, the subgroup analysis was extended to consider three cohorts of trial recruits. The early cohort comprised those recruited between May 2018 and Feb 2019. For this group, both IPS participation and 12-month outcomes pre-dated the pandemic so impacts should be unaffected. The middle cohort (Mar 2019–Jun 2019) was made up of those whose 12-month outcomes fell after the onset of the pandemic but whose experience of IPS is likely to pre-date it. For them, delivery of IPS was unaffected but outcomes may have been affected by COVID. For the late cohort (Jul 2019–Oct 2019) both the IPS treatment and outcomes will have been affected by the pandemic. Specifically, IPS delivery became remote rather than face-to-face. This restricted the practical support advisers could offer the treatment group. It also raised issues of accessibility, with the consequence that employment specialists often prioritised the wellbeing of their clients rather than employment support (Buzzeo et al., 2022). However, while the pattern of results in Table 5 suggests greater impacts for the early cohort, the variation across cohorts was not statistically significant.

6.3. Impacts on secondary outcomes

Table 6 reports the estimated impacts, across all sites, on secondary outcomes. This provides additional context to the primary outcome impact estimates. Impacts on employment 12 months post-randomisation or at the time of the final survey interview (which should have taken place at a similar point) suggest a positive impact when considered alone but controlling for multiple testing renders these estimates non-significant. In fact, none of the employment, earnings or benefit secondary outcomes show a statistically significant impact. There was though a positive impact on job search self-efficacy (0.13 standard deviations; $p = 0.00$, $p_{adj} = 0.00$). This provides an insight into possible causal mechanisms. It suggests IPS improves individuals' confidence in their ability to find work, which in turn raises the prospect that employment impacts may strengthen in the longer-term.

²⁴ An alternative approach to exploring subgroup variation is to conduct a separate analysis for each subgroup. This has the advantage of not constraining the relationship between outcomes and control variables to being fixed across the sample but has the disadvantage of not allowing the significance of subgroup variation to be directly tested. The results of this by-subgroup analysis are shown in Appendix B, Table B.1.

Table 5
Variation in impacts on primary outcomes, pooled across all sites.

		Employment	Earnings	Health		Well-being	
Gender							
	- Male (base)	1.73 (1.12)	-12.32 (179.90)	0.03 (0.01)	***	0.55 (0.21)	***
	- Female	-0.23 (1.66)	-199.89 (253.88)	-0.01 (0.02)		0.11 (0.29)	
	P (joint significance)	0.89	0.43	0.57		0.71	
Age							
	- <30	-1.20 (2.10)	9.63 (295.22)	0.02 (0.02)		-0.13 (0.35)	
	- 30–39	-3.13 (2.02)	-493.68 (317.58)	0.01 (0.02)		0.07 (0.35)	
	- 40–49	-1.93 (1.82)	-166.37 (304.07)	-0.01 (0.02)		-0.03 (0.32)	
	- ≥50 (base)	2.96 (1.31)	215.24 (212.27)	0.02 (0.01)	**	0.62 (0.22)	***
	P (joint significance)	0.46	0.32	0.44		0.95	
Work in 2 years pre-randomisation							
	- ≥ 50%	2.02 (1.80)	324.57 (304.20)	-0.01 (0.02)		0.08 (0.32)	
	- < 50% (base (in work))	0.37 (2.34)	-272.08 (235.51)	-0.02 (0.02)		-0.08 (0.42)	
	- Never (base (out of work))		10.36 (107.43)	0.03 (0.01)	***	0.57 (0.24)	**
	P (joint significance)	0.53	0.23	0.64		0.92	
Health							
	- bottom third	1.77 (2.03)	140.22 (320.23)	0.03 (0.02)		-0.28 (0.34)	
	- middle third (base)	-0.01 (1.35)	-99.36 (208.83)	0.02 (0.01)	**	1.03 (0.22)	***
	- top third	3.85 (2.00)	469.04 (302.81)	-0.01 (0.02)		-0.87 (0.35)	**
	P (joint significance)	0.16	0.29	0.15		0.04	††
Health condition							
	- not MH or MSK (base)	1.15 (3.66)	-166.81 (510.74)	-0.01 (0.03)		-1.34 (0.79)	*
	- MH	0.85 (3.93)	326.00 (558.01)	0.02 (0.03)		1.96 (0.83)	**
	- MSK	4.21 (4.46)	537.97 (637.82)	0.06 (0.04)		2.21 (0.94)	**
	- HM & MSK	-0.55 (3.84)	161.33 (540.25)	0.04 (0.03)		2.04 (0.81)	**
	P (joint significance)	0.39	0.76	0.27		0.09	†
Cohort							
	- May 2018 - Feb 2019 (base)	1.81 (1.33)	195.72 (199.17)	0.02 (0.01)	**	0.59 (0.24)	**
	- Mar 2019 - Jun 2019	1.29 (2.17)	-98.99 (328.18)	0.02 (0.02)		-0.05 (0.37)	
	- Jul 2019 - Oct 2019	-1.31 (1.89)	-237.65 (290.76)	0.01 (0.02)		0.05 (0.33)	
	P (joint significance)	0.48	0.72	0.60		0.96	

Notes: For each primary outcome, 6 regressions are reported. Each allows subgroup variation across a different dimension. The base category for each regression is indicated and other rows show how impacts vary with subgroups: * $P < 0.10$, ** $P < 0.05$, *** $P < 0.01$. The P value of a test of joint significance for each regression is also shown: † $P < 0.10$, †† $P < 0.05$, ††† $P < 0.01$. This does not adjust for multiple testing. For control variables, see notes to Table 3.

Table 6

Estimated impacts, secondary outcomes.

	$\bar{Y}_0$	$\bar{Y}_1$	$\hat{\beta}$	SE	Δ	P	P_{adj}	N
<i>Outcomes from administrative data</i>								
Employed in month 12, %	34.1	35.5	1.5	0.82		0.08	0.45	9785
Number of months employed	3.9	4.0	0.1	0.07	0.03	0.10	0.52	9785
Earnings in month 12, £	420	427	9	1256	0.01	0.48	0.95	9785
Employed and on OOW benefits in month 12, %	2.5	2.9	0.5	0.33		0.16	0.69	9785
Receiving OOW benefits in month 12, %	49.6	49.5	0.0	0.89		0.98	1.00	9785
Number of months on OOW benefits	6.2	6.2	0.0	0.09	0.00	0.97	1.00	9785
Amount of OOW benefits in month 12, £	283	281	−2	753	−0.01	0.76	1.00	9785
<i>Outcomes from survey data</i>								
Employed, %	34.8	38.8	2.7	1.30		0.04	0.29	4087
Working 16+ hours %	29.2	31.5	1.1	1.22		0.38	0.92	4087
No. weeks in work	14.3	15.1	0.1	0.48	0.01	0.80	1.00	4026
No. weeks in work 16+ hrs	11.2	11.7	0.1	0.46	0.00	0.86	1.00	4037
Worked 16+ hours continuously %	27.0	29.4	1.2	1.30		0.35	0.92	4036
Job search self-efficacy	3.1	3.3	0.1	0.03	0.13	0.00	0.00	*** 3882
Musculo-skeletal problems %	31.4	28.2	−4.5	1.28		0.00	0.00	*** 4070
Disability %	28.9	26.6	−2.1	1.33		0.11	0.11	4038
Life satisfaction	5.1	5.4	0.3	0.08	0.10	0.00	0.00	*** 4038
Self-efficacy	27.1	27.8	0.7	0.22	0.10	0.00	0.00	*** 3736
Anxiety	9.6	8.9	−0.6	0.19	−0.09	0.00	0.00	*** 3947
Depression	11.0	10.4	−0.6	0.20	−0.09	0.00	0.00	*** 3804

Notes: $\bar{Y}_0$ and $\bar{Y}_1$ are the raw (unadjusted) mean outcomes for control and treatment groups, respectively. $\hat{\beta}$ is the ITT impact estimate, with robust standard error, SE. For continuous outcomes, Δ is $\hat{\beta}$ expressed in units of the standard deviation of the outcome. P is the P -value of the estimated impact and P_{adj} is the P -value with a (Westfall and Young, 1993) adjustment for multiple testing (10,000 replications). Asterisks denote statistical significance thresholds: * $P_{adj} < 0.10$, ** $P_{adj} < 0.05$, *** $P_{adj} < 0.01$. Impacts were estimated using linear regression controlling for baseline outcomes as described in Section 5. For control variables, see notes to Table 3. OOW — out of work; DDA — Disability Discrimination Act.

Secondary health and well-being outcomes show more signs of improvement due to IPS, even after controlling for multiple testing. IPS reduces the incidence of musculoskeletal problems by 4.5 percentage points ($p = 0.00$, $p_{adj} = 0.00$) and increases life satisfaction and self-efficacy by 0.1 standard deviations ($p = 0.00$, $p_{adj} = 0.00$ in both cases). Two standard measures of mental health were constructed from survey respondents' answers. For both of these, a higher score indicates poorer mental health. The generalised anxiety disorder (GAD) measure showed a reduction in anxiety of 0.09 standard deviations ($p = 0.00$, $p_{adj} = 0.00$). The patient health questionnaire (PHQ) is designed to measure depression and similarly showed a reduction of 0.09 standard deviations ($p = 0.00$, $p_{adj} = 0.00$). Hence, it appears that the positive impacts on health and well-being are quite general in the sense that they apply to the two largest categories of health problem – musculoskeletal and mental – and also improve subjective well-being. The fact that these impacts appear stronger than the labour market impacts suggests that they are likely to be direct effects of IPS rather than mediated through increased employment.

6.4. Cost–benefit analysis

Information on the costs of delivering IPS was taken from the expenses claims providers submitted to the Department for Work and Pensions. In addition, providers were asked to supply details on the composition and activities of staff. This was supplemented by a staff survey, administered in October 2020. Associated estimates of central government costs were also made available.

All costs associated with IPS occurred during the period support was available. By contrast, benefits were estimated for the three years following randomisation. While this period was arbitrary, it arguably allows sufficient time for impacts to be realised while limiting the need for assumptions regarding longer-term impacts. Nevertheless, this three-year period is longer than the period for which outcome data were available and so it was necessary to make some assumptions about how impacts would evolve over this timeframe. To explore how the ratio of costs to benefits varied depending on the assumptions underlying the analysis, costs and benefits were estimated under three different scenarios. These included a central estimate, which was based on assumptions commonly used in other cost-benefit analyses, and more optimistic and pessimistic scenarios. Table 7 provides a summary of the assumptions underlying each of the three scenarios.

The central estimate of benefits from employment effects assumes that the impacts in the second year post-randomisation (Fig. 1) persist into year 3. Estimates of health impacts were only available for the year following randomisation. These were likewise assumed to persist into years 2 and 3 and were monetised using quality-adjusted life years (QALYs).

Table 8 shows the cost per recruit and the returns for every pound spent. Monetary amounts are expressed in 2018/19 prices, applying a 3.5% discount rate to employment-related impacts and a 1.5% discount rate to health impacts.²⁵ As noted in Table 7, the

²⁵ This follows official guidance, see <https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-government/the-green-book-2020>.

Table 7

Assumptions for the reference, optimistic and pessimistic cost-benefit analysis scenarios.

Input	Central estimate	Optimistic scenario	Pessimistic scenario
Employment-related discount rate ^a	3.5%	3.0%	4.0%
QALY discount rate ^b	1.5%	1.5%	3.5%
2021/22 CPI inflation forecast ^c	3.3%	3.8%	2.8%
QALY value ^d	Fixed	GDP deflator applied	Fixed
Substitution effects ^e	0.0%	0.0%	20.0%
Social cost of exchequer finance ^f	0.0%	20.0%	0.0%
Employment impact trajectory	Year 2 impact = 22-month impact scaled up to 24-months linearly, Year 3 impact = Year 2 impact	Year 2 impact = 22-month impact scaled up to 24-months linearly, Year 3 impact = upward trend from Year 1 to Year 2 continues	Year 2 impact = 22-month impact scaled up to 24-months linearly, Year 3 impact = average of Year 1 and Year 2 impacts
Health impact trajectory	Year 2 and Year 3 impacts = Year 1 impact	Year 2 impact = +50% Year 1 impact, Year 3 impact = +50% Year 2 impact	Year 2 impact = -50% Year 1 impact, Year 3 impact = -50% Year 2 impact

Notes: The scenarios also made different assumptions about the impact of participation in IPS on the costs of healthcare, travel and childcare costs. Full details are provided in [Edney et al. \(2022\)](#).

Table notes:

^a Central estimate taken from HM Treasury Green Book Social Time Preference Rate.

^b Central estimate based on HM Treasury Green Book recommendation for discounting health-related quality of life. Pessimistic scenario based on National Institute for Health and Care Excellence (NICE) recommendations for sensitivity analysis ([NICE, 2012](#)).

^c Use to forecast forward prices used to monetise benefits. Central estimate based on the Office for National Statistics estimate of Consumer Price Inflation (CPI) and the Office for Budget Responsibility forecast of CPI for the 2021/22 tax year, from the OBR Economic and Fiscal Outlook for October 2021.

^d Central estimated based on a valuation of £60,000 per QALY per year, in line with guidance in the HM Treasury Green Book. Raw scores from the primary health outcome (EQ5D) were converted to a measure of health-related quality of life using an approach described by [Devlin et al. \(2018\)](#). The change in quality of life was multiplied by the duration in years of the change.

^e Central estimate assumes that IPS participants who enter employment do not displace other workers. The pessimistic scenario is based on an assumption of 20% displacement, as set out in [Fujiwara \(2010\)](#).

^f This captures the welfare loss resulting from the disincentive effects of the taxation required to fund IPS. As the costs to the Exchequer of IPS outweigh the benefits, the Social Cost of Exchequer Finance is assumed to be positive under the optimistic scenario.

Table 8

IPS costs and returns.

	SCR (in work)	SCR (out of work)	WMCA (out of work)	All
Costs per recruit to treatment group	£2116	£2416	£3893	£2893
Central estimates:				
- Financial return for every £1 spent	£0.02	£0.00	£0.01	£0.01
- Economic return for every £1 spent	£2.32	£2.02	£0.00	£1.34
Optimistic scenario				
- Financial return for every £1 spent	£0.02	£0.00	£0.02	£0.01
- Economic return for every £1 spent	£4.18	£3.77	£0.20	£2.54
Pessimistic scenario				
- Financial return for every £1 spent	£0.01	£0.00	£0.01	£0.01
- Economic return for every £1 spent	£1.26	£1.03	£0.00	£0.70

employment-related discount rates were varied under the optimistic and pessimistic scenarios (3.0% and 4.0% respectively), while the central estimate used a discount rate of 3.5%.

[Table 8](#) shows that the cost per treatment group member is roughly similar for the two SCR groups but substantially higher in WMCA. As set-up and ongoing delivery of IPS was funded through a series of grants, rather than varying with the number of participants, the differences in costs for treatment group members in SCR and WMCA are largely due to the smaller number of individuals recruited to the trial in WMCA.²⁶

The financial return refers to the costs and benefits to the Exchequer and is essentially zero across all groups and scenarios, indicating that IPS does not appear to be advantageous from this perspective. The economic return has a broader scope, taking into account the wider costs and benefits to society, as well as the individual participant. For example, economic returns may include gains to society such as increased productivity, reduced costs to the state in providing healthcare and administering benefits, as well as private gains from improvements in health-related quality of life. There is a positive economic return in both SCR sites, even under the pessimistic assumptions, but not in WMCA. This is driven by the health impacts found in SCR. It should be noted

²⁶ The figures for total costs in each pilot site included those required to set-up the IPS service and the costs of ongoing delivery. In SCR the cost of ongoing delivery was £1670 for the out-of-work group and £1371 for the in-work group, whereas in WMCA the delivery cost was £2989 per participant.

Table 9
Results of regressing survey response on treatment and other characteristics.

	SCRIW	SCROOW	WMCA	All
<i>(a) Regressing response on treatment and baseline characteristics</i>				
Treatment group	0.0761*** (3.87)	0.0195 (1.17)	0.0459** (2.85)	0.0455*** (4.56)
<i>(b) Regressing response on treatment, baseline characteristics and 12-month employment and earnings</i>				
Treatment group	0.0751*** (0.0194)	0.0183 (0.0163)	0.0420*** (0.0159)	0.0428*** (0.00982)
Employed 12 month(s) after randomisation	0.0377 (0.0303)	0.0722** (0.0298)	0.0416 (0.0332)	0.0349** (0.0169)
Earnings (£000s) 12 month(s) after randomisation	−0.0002 (0.0150)	−0.0631*** (0.0226)	−0.0455 (0.0289)	−0.0157 (0.0111)
Observations	2519	3591	3675	9785

Notes: In panel (a), the result from a single regression of survey response on treatment and baseline characteristics is reported for each site and for all sites combined. Only the treatment group coefficient is shown. In panel (b), the result from a single regression of survey response on treatment, baseline characteristics, and 12-month employment and earnings is reported for each site and for all sites combined. Coefficients for the treatment group are shown, as well as coefficients for 12-month employment and earnings. For control variables, see notes to Table 3. Standard errors are in parentheses. Asterisks denote statistical significance thresholds: ** $P < 0.10$, *** $P < 0.05$, **** $P < 0.01$.

that there is considerable uncertainty around these estimates. This is apparent both from the variation in the returns in each of the three scenarios and from Monte Carlo probabilistic sensitivity analysis. This involved randomly selecting values from the probability distributions of the impact estimates, based on their standard errors and using these distributions to simulate distributions for the economic return under each of the three scenarios and for each pilot site.²⁷ In summary the sensitivity analysis showed that even for the SCR sites, it was possible that the positive economic returns found in this study would not materialise if it were repeated.

Nevertheless, the analysis does suggest that for IPS to provide a social return within three years, it needs to improve participants' health. Longer-term, it is possible that a positive return may be seen even without an impact on health. This follows from the fact that there are no associated costs beyond those incurred during the support period. Consequently, should employment and earnings impacts be sustained longer-term they will deliver financial benefits without any offsetting costs. However, it is important to bear in mind that as the study was only able to observe outcomes for a limited period following randomisation, it is uncertain whether any effects seen initially would be sustained over time.

7. Assessing robustness of the health and well-being results

As noted in Section 3, survey outcomes were observed for 42% of trial recruits. To restore representativeness of the results, impacts on outcomes taken from survey data were estimated using survey weights to adjust for non-response. Particularly worrying in the context of a trial is the possibility that nonresponse may differ for treatment and control groups. In this case, the causal basis of the impact estimates may be undermined.

Table 9 shows the key results from two sets of regressions. In the first set, survey response is regressed on a treatment group indicator and baseline characteristics. For each site and for all sites combined, only the treatment group coefficient is presented. Except for those in SCR who were out of work when randomised, this treatment coefficient is positive and statistically significant, indicating a greater tendency to respond among the treatment group after controlling for other characteristics.

This higher tendency among the treatment group to respond to the survey may bias impact estimates if it is correlated with outcomes. While this cannot be examined for outcomes only observed in the survey, some insight is possible by incorporating outcomes from the administrative data that are observable regardless of survey response. The second set of regressions augments the first set by including employment and earnings 12 months post-randomisation (i.e. roughly at the time of survey interview). Earnings and employment at 12 months are not statistically significant for the SCR in-work group, nor the WMCA group. This provides some reassurance that the differential response rates among experimental groups in these sites are not related to post-randomisation outcomes. In the case of the SCR out-of-work group, earnings and employment at 12 months are both statistically significant. As before, the treatment group coefficient is not statistically significant and so internal validity is not threatened. However, looking across all sites, both the treatment group and employment status 12 months post-randomisation are statistically significant. This raises a more direct concern that, for the trial population as a whole, the differential response rates may bias the estimated impacts on survey-based outcomes. This underscores the importance of the estimated bounds reported in Table 4 that suggest a positive impact on health and well-being in the full sample.

²⁷ In each case the random sampling was repeated 10,000 times.

8. Discussion

Taking all sites together, the trial suggests that IPS had a small positive effect on sustained employment but not earnings, and increased health and well-being by roughly one-tenth of a standard deviation. However, these results vary by site. Perhaps most novel are those relating to individuals who were in work when randomised. For them, IPS increased employment by 3 percentage points. For those in the same area (SCR) who were out of work when randomised, there was no statistically significant employment effect. The subgroup analysis gives little reason to believe that this disparity is due to compositional differences between the in-work and out-of-work groups. A more likely explanation is that there was less alternative (BAU) support available to the in-work group than there was for the out-of-work group. Among those out of work when randomised, the control group spent an average of more than seven months receiving benefits in the following year and would have had access to standard employment support and incentives during this time. For the in-work group, who were all experiencing health problems that made employment difficult for them, available support is less formalised and a common progression may be simply to go on sick leave. In this sense, there may be more of a support gap for IPS to fill than in the case of the out-of-work group.

Given this difference in the support available to the in-work and out-of-work control groups in SCR, it is no surprise to find a difference in impacts, despite being in the same local labour market. It is important to note the nature of the intervention's aim in the two cases. For the in-work group, IPS is intended to support employment retention. For the out-of-work group, on the other hand, it is intended to support (sustained) employment entry. Comparing control group outcomes reinforces the extent of this distinction; the in-work group have a sustained employment rate of 80% compared to 27% for the out-of-work group (Table 3).

The bigger puzzle is why employment impacts for the out-of-work groups in SCR and WMCA differ so much. In WMCA, IPS had a positive impact on employment and also on earnings in the longer-run. Why should this differ from SCR, where no employment or earnings impacts were evident? There are three possible differences between the sites that could potentially offer an explanation: compositional differences, differences in the delivery of IPS and differences in the context within which IPS was delivered.

Table 10 summarises the results of three analyses that probe the first of these possibilities. Panel (a) presents key coefficients from a regression of the primary employment outcome estimated on a pooled SCR and WMCA out-of-work sample. The regression controls for those covariates specified in the statistical analysis plan and also lets the impact vary with each covariate. If the SCR-WMCA difference in impact were due to those characteristics associated with a greater impact being concentrated within WMCA, we would expect the interaction of the treatment dummy with the WMCA dummy to be non-significant. Instead, this interaction is statistically significant ($p = 0.02$), suggesting that compositional differences do not explain the stronger employment impact in WMCA. Panel (b) shows results from a regression on the covariates estimated using the pooled control groups. The WMCA coefficient is found to be statistically significant ($p < 0.01$), indicating that, even without the influence of IPS, the rate of sustained employment is lower in WMCA than SCR after controlling for observed characteristics. Another perspective on this control groups result is provided by the Blinder-Oaxaca decomposition presented in Panel (c) (Blinder, 1973; Oaxaca, 1973). Across both formulations of the decomposition, the clear suggestion is that the difference in the employment outcome arises predominantly from differences in the coefficients rather than composition.

The second possibility is that the two sites may have differed in how IPS was delivered.²⁸ Some evidence to this effect has already been noted. For example, IPS sessions in WMCA began sooner after randomisation than in SCR and there were more of them (Table 2). The spend per head was also higher in WMCA than in SCR (Table 8). Qualitative evidence summarised in Newton et al. (2023) highlights other differences. In SCR, the job of employment advisers was more demanding since they had to cater for both in-work and out-of-work groups. By contrast, WMCA advisers could specialise solely on those out of work when randomised. This, plus their greater resourcing, allowed WMCA advisers to focus on employer engagement more than was possible in SCR. Employer engagement is a key element of IPS but one that staff found time-consuming and often challenging. The staff survey conducted to inform the cost-benefit analysis suggested that WMCA advisers spent roughly 30% more time liaising with employers than SCR advisers.

The possibility that the greater effectiveness of IPS in WMCA for those out of work when randomised is at least partly due to it being a more intensive, employment-focused treatment than in SCR makes intuitive sense. Despite this, individuals in the out-of-work treatment group in SCR had a more favourable impression of the IPS support than their counterparts in WMCA. They had a higher opinion of advisers' ability to provide helpful support for both employment and health needs than in WMCA, and were more likely to report that IPS had increased their motivation and work-readiness (Crowley et al., 2023). Among the out-of-work control groups, on the other hand, individuals in WMCA made more use of BAU support than those in SCR and also tended to have a higher opinion of that support. Taken together, these findings imply that the contrast between the perceived usefulness of support available under IPS and that available under BAU was greater in SCR than in WMCA. This would lead us to expect greater effects in SCR than in WMCA. The fact that the opposite is the case is intriguing but does not invalidate the argument that treatment intensiveness is an important influence on effectiveness. The impressions of programme participants are subjective and not necessarily accurate measures of effectiveness. This is particularly the case when important aspects of the advisers' role – such as engagement with employers, as discussed above – are not always visible to the individual participant.

The third possible explanation for the differences in estimated employment impacts for the out-of-work groups in SCR and WMCA is that the effectiveness of the programme may depend on the context, particularly the labour market, within which it operates. Fig. 2 shows monthly employment rates and earnings for the treatment and control groups in each trial site. It is striking that the

²⁸ Unlike when comparing results for the in-work and out-of-work groups in SCR, when comparing SCR and WMCA out-of-work groups, the BAU support available is the Jobcentre Plus provision in both cases.

Table 10

Three analyses exploring the difference in primary employment outcome impacts between SCR and WMCA out of work groups: (a) Regression of primary employment outcome for out-of-work groups in both areas; (b) Regression for out-of-work control groups in both areas; (c) Blinder-Oaxaca decomposition of difference in employment impacts between SCR and WMCA.

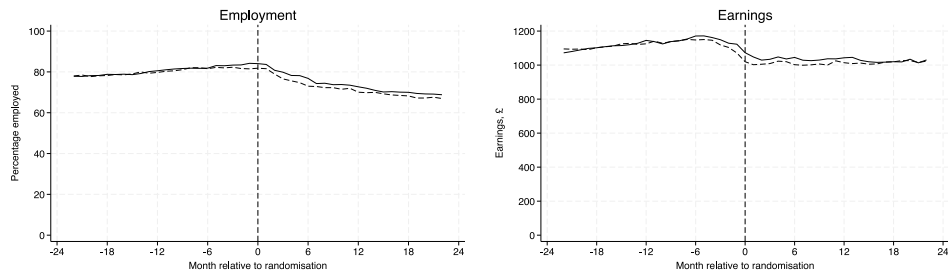
<i>(a) Regressing employment on treatment, WMCA dummy and treatment-WMCA interaction</i>	
WMCA	−0.07*** (0.01)
Treatment	0.16 (0.16)
Treatment × WMCA	0.06*** (0.02)
N	7266
<i>(b) Regressing employment on WMCA dummy, control groups only</i>	
WMCA	−0.07*** (0.00)
N	3630
<i>(c) Blinder-Oaxaca decomposition</i>	
	(i) $(X^{WMCA} - X^{SCR})\delta^{WMCA} + X^{SCR}(\delta^{WMCA} - \delta^{SCR})$
Variation due to differences in:	
- Composition	−0.01 [13.9%]
- Coefficients	−0.08 [86.1%]
	(ii) $(X^{WMCA} - X^{SCR})\delta^{SCR} + X^{WMCA}(\delta^{WMCA} - \delta^{SCR})$
Variation due to differences in:	
- Composition	−0.01 [16.8%]
- Coefficients	−0.07 [83.2%]

Notes: for control variables included in regressions (a) and (b), see notes to Table 3. For both (a) and (b), standard errors are shown in parentheses and asterisks denote statistical significance thresholds: * $P < 0.10$, ** $P < 0.05$, *** $P < 0.01$. Panel (c) presents the two-component Blinder-Oaxaca decomposition (Blinder, 1973; Oaxaca, 1973), estimated using the nldecompose Stata package (Sinning et al., 2008). Variants (i) and (ii) differ in the choice of reference group used to compute the composition and coefficient effects. The percentage contributions of each component to the overall difference in impacts between SCR and WMCA are shown in square brackets.

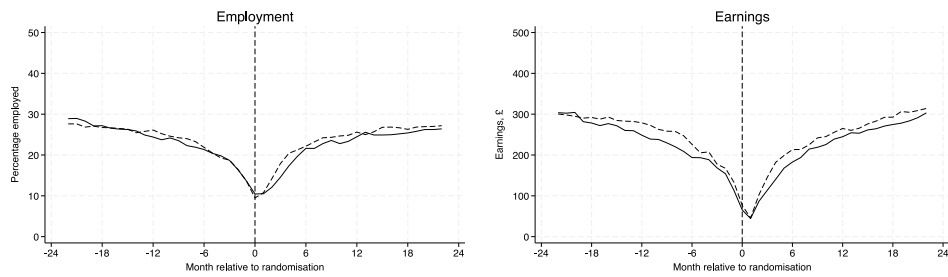
employment rate in the SCR out-of-work control group reaches 20% within four months of randomisation whereas, in WMCA, only 13% were employed by this point and it took 21 months before the 20% level was reached. In fact, the employment rate in the SCR out-of-work control group is close to symmetric around the point of randomisation, indicating that the group as a whole entered work at roughly the same rate that it had been leaving work prior to randomisation. In WMCA, by contrast, pre-randomisation employment rates were never achieved post-randomisation, remaining lower by 4–5 percentage points. The impression is of a labour market in SCR that is more dynamic than that in WMCA and offers more opportunities for those who lose their job to find new employment. In this context, there is less of a role for IPS since the local labour market is functioning well enough to allow individuals to find work without the need for additional support. In WMCA, by contrast, the labour market is such that those losing their job are less likely to find new work without additional support. In this sense, the greater impact of IPS on employment in WMCA may reflect the greater need for assistance there. This is consistent with Metcalfe et al. (2018), who found that the impact of IPS on employment for those with a severe mental health problem was greater in areas with higher unemployment (although not significantly so).

In summary, while it is not possible to make definitive statements about the moderators of IPS effectiveness in this trial, we have suggestive evidence that both the nature of delivery and the strength of the local labour market may play a role. This has implications when considering how effective IPS might be if operated in countries other than the UK. That the nature of delivery appears important highlights the implementation point that, even with a tightly-defined programme, there remains scope for programme heterogeneity in practice. A strength of IPS is that having a set of clearly-articulated principles allows fidelity to the IPS model to be measured on a consistent basis. Because of this Whitworth et al. (2024), for example, were able to compare fidelity across the 18 studies they consider, revealing considerable variation. This study has shown that even when a single agency is funding the intervention, differences in delivery can emerge. The fact that delivery in WMCA, with its greater emphasis on employer engagement, was more in line with the IPS model concurs with findings from other studies that effectiveness is greater when fidelity is higher. IPS offers a template which, if followed, increases the effectiveness of the programme.

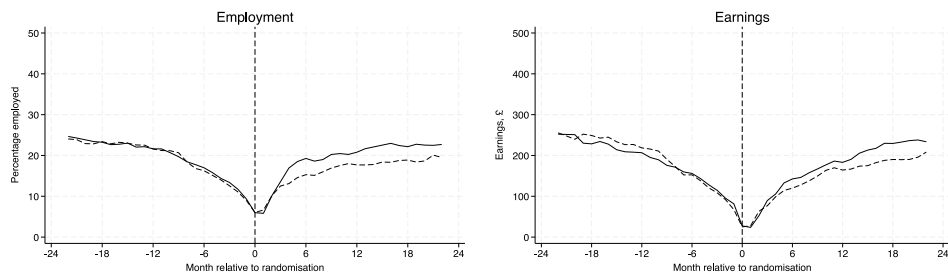
The suggestion that the effectiveness of IPS varies with the strength of the local labour market is also relevant when considering the potential role of IPS in Europe and elsewhere. A programme that appears to work better in less dynamic labour markets offers the potential to reduce regional imbalances in unemployment, for example. In this sense, it is a programme that indirectly targets areas of need. While not a dimension that could be considered in this paper, institutional context is also important. De Winter et al. (2022) argue that this is an important moderator of effectiveness and speculate that smaller IPS impacts in Europe than elsewhere are attributable to a benefits trap in those higher welfare countries. This interpretation is consistent with Metcalfe et al. (2018) who find IPS employment effects are greater when disability benefits are less generous, employment protection is weaker and less effort



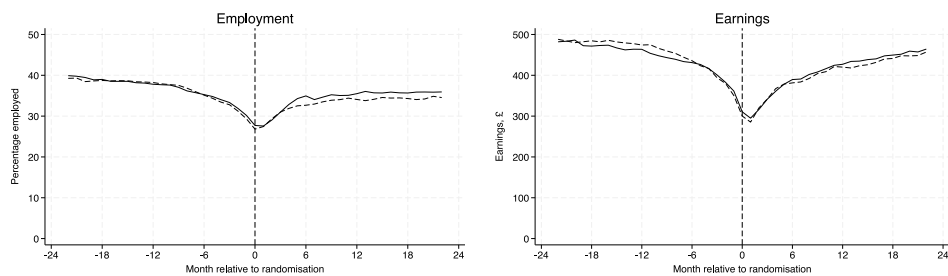
(a) SCR (in work)



(b) SCR (out of work)



(c) WMCA (out of work)



(d) All sites

Fig. 2. Mean employment (left panel) and earnings (right panel), pre- and post-randomisation.
 Notes: Mean employment or earnings for treatment group (solid line) and control group (dashed line).

is made to integrate those with disabilities into the workforce. As such, we might expect that the results of this study are likely to be most relevant to countries that, like the UK, have a liberal welfare regime.

9. Conclusion

Health-related employment and participation gaps are a policy priority in all advanced economies. Through successive small trials, IPS has repeatedly been shown to be effective in encouraging employment among individuals with severe mental health problems. The question of whether it can be effective for a broader population is less clear. The past decade has seen policy makers experiment with extending IPS to a range of population subgroups. The results of this study go further, broadening eligibility to the general population of individuals with a health problem. As the largest trial of IPS to date by far, it provides robust evidence on how this model of employment support can operate at scale when catering to the diverse needs associated with a wide range of health conditions.

The results demonstrate that IPS can work in this more general application. The trials variously improved labour market, health and well-being outcomes. The results for those already working when randomised are particularly novel given the near-total lack of existing evidence on the effect of IPS on employment retention. For those out of work when randomised, the results highlight how effectiveness may vary according to the nature of implementation and the local context. While the impact of IPS for those with severe mental health problems has been consistently demonstrated across multiple trials to be positive, the results from this trial suggest the effectiveness of IPS may vary more when applied to the broader population. This carries a lesson for policy makers considering IPS in their own locality.

Overall, the trials have demonstrated the promise of IPS as a programme suited to a population with a wide range of health conditions. The balance between employment and health objectives is a key consideration. Only where IPS achieved an improvement in health did the benefits of the programme outweigh the costs within three years. Longer-term, if employment and earnings effects can be sustained, these too will have a positive return in their own right.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This paper is based on research commissioned by the Work and Health Unit, a UK cross-government unit, jointly sponsored by the Department for Work and Pensions and the Department of Health and Social Care. We are grateful to Chris Gunning, David Johnson, Huw Meredith, Hayley Moore-Purvis and Anna Silk for comments and support. NHS England part-funded the trials and we thank Elizabeth Dyer, Johanna Ejbye, Shabana Janjua, Belinda McIntosh, Kevin McKenna and Elizabeth Wade. The commissioned research was conducted by a consortium of research organisations, led by the Institute for Employment Studies. Stephen Bevan was principal investigator. Within the evaluation consortium, Becci Newton, Rosie Gloster and Joanna Hofman provided important insights. Ethical approval was given by the Health Research Authority Research Ethics Committee: <https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/sheffield-city-region-health-led-trial> and <https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/west-midlands-combined-authority-health-led-trial>. We thank Adam Whitworth, two anonymous reviewers, the guest editor and the editor for helpful comments that have improved the paper. The usual disclaimer applies.

Appendix A. Event study plots

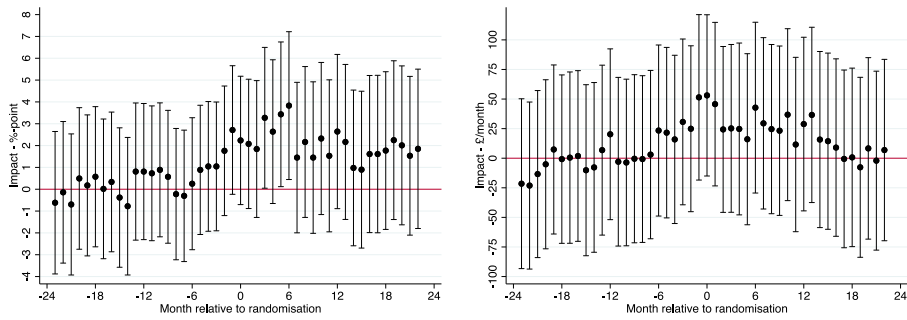
Fig. A.1 presents event study plots for the employment and earnings outcomes. These show unconditional treatment-control differences in employment rates (left panel) and mean earnings (right panel) by month relative to randomisation. For months prior to randomisation, the treatment-control differences are never significantly different from zero in any site. For the months after randomisation, the treatment-control differences show similar patterns to those in Fig. 1. Since treatment is randomised, these differences are unbiased estimates of the impact of IPS. The impacts depicted in Fig. 1 control for baseline characteristics and so provide more precise estimates.

Appendix B. By-subgroup analysis

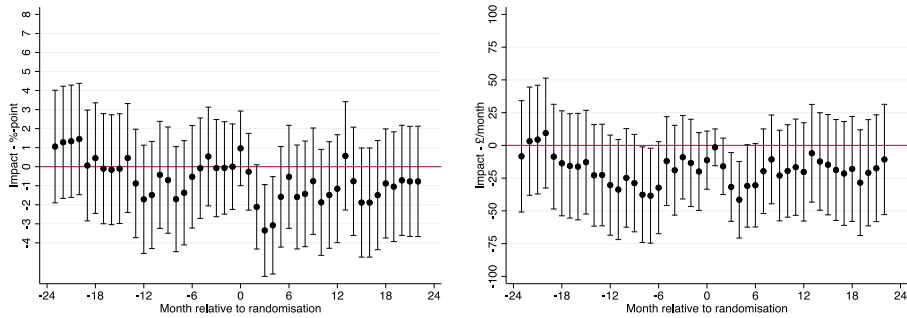
See Table B.1.

Appendix C. Supplementary data

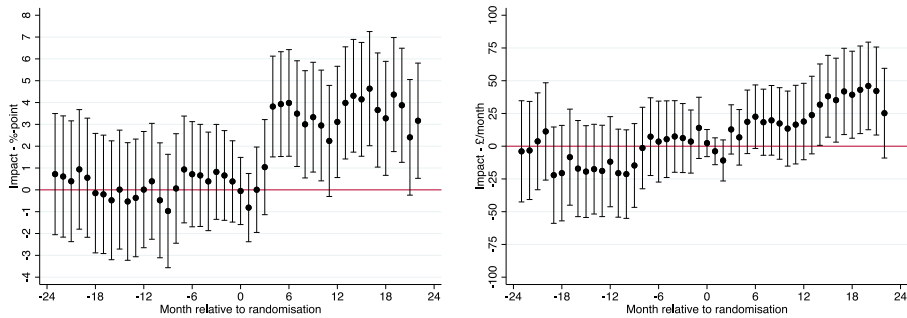
Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.eurocorev.2026.105308>.



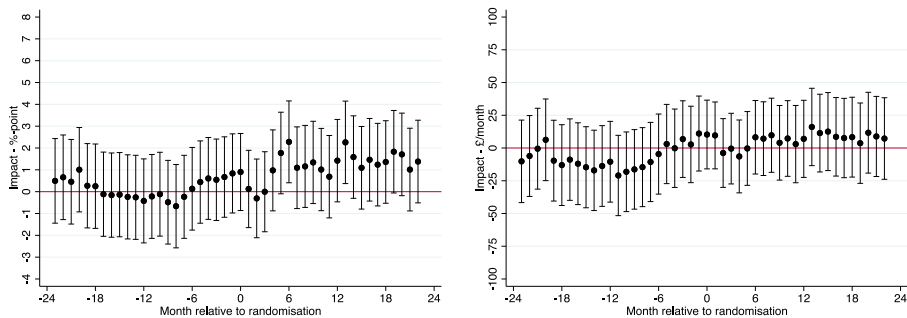
(a) SCR (in work)



(b) SCR (out of work)



(c) WMCA (out of work)



(d) All sites

Fig. A.1. Event study plots for employment (left panel) and earnings (right panel) impacts.

Table B.1
Estimated impacts on primary outcomes by subgroup, pooled across all sites.

			Employment	Earnings	Health	Well-being		
Gender								
	- Male	1.84	*	4.20	0.03	***	0.58	***
		(1.09)		(169.09)	(0.01)		(0.21)	
	- Female	1.49		177.78	0.02	**	0.63	***
		(1.17)		(168.38)	(0.01)		(0.20)	
Age								
	- <30	1.21		374.88	**	0.04	***	0.52
		(1.65)		(189.84)	(0.01)		(0.28)	*
	- 30–39	1.07		–449.05	0.02		0.71	**
		(1.86)		(294.59)	(0.02)		(0.35)	
	- 40–49	0.99		142.80	0.03		0.64	*
		(1.71)		(287.49)	(0.02)		(0.33)	
	- ≥50	2.23	*	140.62	0.02	**	0.58	**
		(1.31)		(212.08)	(0.01)		(0.24)	
Work in 2 years pre-randomisation								
	- ≥ 50%	1.58		–184.29	0.01		0.45	
		(2.05)		(195.86)	(0.02)		(0.34)	
	- < 50%	3.12	**	379.79	0.02	**	0.63	***
		(1.34)		(262.87)	(0.01)		(0.21)	
	- Never			–6.87	0.03	***	0.61	**
				(88.97)	(0.01)		(0.24)	
Health								
	- bottom third	1.67		55.36	0.05	**	0.73	***
		(1.45)		(230.41)	(0.02)		(0.26)	
	- middle third	0.15		–48.77	0.02	**	0.98	***
		(1.30)		(195.45)	(0.01)		(0.22)	
	- top third	3.50	**	334.30	0.00		0.10	
		(1.44)		(208.03)	(0.01)		(0.27)	
Health condition								
	- not MH or MSK	1.51		–146.60	–0.01		–1.59	*
		(3.72)		(486.78)	(0.03)		(0.84)	
	- MH	2.03		158.70	0.01		0.57	**
		(1.39)		(213.33)	(0.01)		(0.25)	
	- MSK	4.28	*	113.90	0.04	**	0.69	
		(2.55)		(370.50)	(0.02)		(0.51)	
	- HM & MSK	0.90		54.16	0.04	***	0.73	***
		(1.10)		(164.92)	(0.01)		(0.19)	
Cohort								
	- May 2018 - Feb 2019	1.85		207.53	0.02	**	0.57	**
		(1.30)		(189.12)	(0.01)		(0.24)	
	- Mar 2019 - Jun 2019	3.24	*	131.23	0.04	**	0.55	**
		(1.67)		(249.05)	(0.02)		(0.28)	
	- Jul 2019 - Oct 2019	0.46		–58.19	0.03	***	0.63	***
		(1.26)		(197.74)	(0.01)		(0.23)	

Notes: For each characteristic, the results shown are for a regression carried out on the subgroup with a particular value of that characteristic. For control variables, see notes to Table 3. Asterisks denote statistical significance thresholds: * $P < 0.10$, ** $P < 0.05$, *** $P < 0.01$.

Data availability

The data that has been used is confidential.

References

- Becker, D.R., Drake, R.E., 1994. Individual placement and support: A community mental health center approach to vocational rehabilitation. *Community Ment. Health J.* 30 (2), 193–206.
- Blinder, A.S., 1973. Wage discrimination: reduced form and structural estimates. *J. Hum. Resour.* 436–455.
- Bond, G.R., 2004. Supported employment: Evidence for an Evidence-Based practice. *Psychiatr. Rehabil. J.* 27, 345–359.
- Bond, G.R., Becker, D.R., Drake, R.E., Vogler, K.M., 1997. A fidelity scale for the individual placement and support model of supported employment. *Rehabil. Couns. Bull.* 40, 265–284.
- Bond, G.R., Drake, R.E., Becker, D.R., 2020. An update on individual placement and support. *World Psychiatry* 19 (3), 390–391.

- Bond, G.R., Drake, R.E., Pogue, J.A., 2019. Expanding individual placement and support to populations with conditions and disorders other than serious mental illness. *Psychiatr. Serv.* 70 (6), 488–498.
- Bond, G.R., Peterson, A.E., Becker, D.R., Drake, R.E., 2012. Validation of the revised individual placement and support fidelity scale (IPS-25). *Psychiatr. Serv.* 63 (8), 758–763.
- Brinchmann, B., Widding-Havneraas, T., Modini, M., Rinaldi, M., Moe, C.F., McDaid, D., Park, A.-L., Killackey, E., Harvey, S.B., Mykletun, A., 2020. A meta-regression of the impact of policy on the efficacy of individual placement and support. *Acta Psychiatr. Scand.* 141 (3), 206–220.
- Burns, T., Yeeles, K., Langford, O., Montes, M.V., Burgess, J., Anderson, C., 2015. A randomised controlled trial of time-limited individual placement and support: IPS-LITE trial. *Br. J. Psychiatry* 207 (4), 351–356.
- Buzzee, J., Campbell, B., Cetera, R., Gloster, R., Shukla, R., Wilson, T., Newton, B., 2022. Health-led Employment Trial Evaluation: The Effects of COVID-19 on IPS Delivery and the Trial Site Labour Markets. Department for Work and Pensions Research Report No. 1038.
- Crowley, J., Martin, I., Weston-Stanley, P., Lau, R.L., Khambhaita, P., Newton, B., Gloster, R., 2023. Survey findings: Health-led Employment Trial Evaluation. Department for Work and Pensions Research Report No. 1035.
- De Winter, L., Couwenbergh, C., van Weeghel, J., Bergmans, C., Bond, G.R., 2020. Fidelity and IPS: does quality of implementation predict vocational outcomes over time for organizations treating persons with severe mental illness in the Netherlands? *Soc. Psychiatry Psychiatr. Epidemiol.* 55, 1607–1617.
- De Winter, L., Couwenbergh, C., Weeghel, J.v., Sanches, S., Michon, H., Bond, G.R., 2022. Who benefits from individual placement and support? A meta-analysis. *Epidemiol. Psychiatr. Sci.* 31.
- 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 Econ.* 27 (1), 7–22.
- Edney, C., Gray, H., Muir, D., Newton, B., Gloster, R., 2022. Economic evaluation 12 month outcomes report: Health-led Employment Trial. Department for Work and Pensions Research Report No. 1036.
- Esping-Andersen, G., 1990. The three worlds of welfare capitalism. *Polity*.
- Fadyl, J.K., Anstiss, D., Reed, K., Khoronzhevych, M., Levack, W.M., 2020. Effectiveness of vocational interventions for gaining paid work for people living with mild to moderate mental health conditions: systematic review and meta-analysis. *BMJ Open* 10 (10), e039699.
- Frederick, D.E., Van der Weele, T.J., 2019. Supported employment: Meta-analysis and review of randomized controlled trials of individual placement and support. *PLoS One* 14 (2).
- Freedman, L., Millenky, M., 2022. Two-Year Findings from the Evaluation of Breaking Barriers: An Individual Placement and Support (IPS) Program (OPRE Report 2022-35). Office of Planning, Research and Evaluation, Administration for Children and Families, US Department of Health and Human Services, Washington, DC.
- Fujiwara, D., 2010. The Department for Work and Pensions Social Cost-Benefit Analysis Framework. Department for Work and Pensions Working Paper 86.
- Imbens, G.W., Manski, C.F., 2004. Confidence intervals for partially identified parameters. *Econometrica* 72 (6), 1845–1857.
- Jones, D., Molitor, D., Reif, J., 2019. What do workplace wellness programs do? Evidence from the Illinois workplace wellness study. *Q. J. Econ.* 134 (4), 1747–1791.
- Koenig, F., Petrongolo, B., Van Reenen, J., Bagaria, N., 2019. Can helping the sick hurt the able? Incentives, information and disruption in a welfare reform. *Econ. J.* 129 (624), 3189–3218.
- Latimer, E., Pflanz, F., Waters, T., 2024. Health-related benefit claims post-pandemic: UK trends and global context. *IFS Report* R333.
- Lee, D.S., 2009. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Rev. Econ. Stud.* 76 (3), 1071–1102. <http://dx.doi.org/10.1111/j.1467-937X.2009.00536.x>.
- Metcalfe, J.D., Drake, R.E., Bond, G.R., 2018. Economic, labor, and regulatory moderators of the effect of individual placement and support among people with severe mental illness: A systematic review and Meta-analysis. *Schizophr. Bull.* 44 (1), 22–31.
- Newton, B., Gloster, R., Hofman, J., 2023. Health-led Trials: Evaluation Synthesis Report. Department for Work and Pensions research report no. 1025.
- NICE, 2012. Methods for the Development of NICE Public Health Guidance. National Institute for Health and Care Excellence.
- Oaxaca, R., 1973. Male-female wage differentials in urban labor markets. *Internat. Econom. Rev.* 693–709.
- OECD, 2022. Disability, Work and Inclusion: Mainstreaming in All Policies and Practices. OECD Publishing, Paris.
- Probyn, K., Engedahl, M.S., Rajendran, D., Pincus, T., Naeem, K., Mistry, D., Underwood, M., Froud, R., 2021. The effects of supported employment interventions in populations of people with conditions other than severe mental health: a systematic review. *Prim. Health Care Res. Dev.* 22, e79.
- Sinning, M., Hahn, M., Bauer, T.K., 2008. The blinder-oaxaca decomposition for nonlinear regression models. *Stata J.* 8 (4), 480–492.
- Westfall, P.H., Young, S.S., 1993. Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment. John Wiley & Sons.
- Whitworth, A., Baxter, S., Cullingworth, J., Clowes, M., 2024. Individual placement and support (IPS) beyond severe mental health: An overview review and meta-analysis of evidence around vocational outcomes. *Prev. Med. Rep.* 102786.
- Yamaguchi, S., Sato, S., Shiozawa, T., Matsunaga, A., Ojio, Y., Fujii, C., 2022. Predictive association of low-and high-fidelity supported employment programs with multiple outcomes in a real-world setting: A prospective longitudinal multi-site study. *Adm. Policy Ment. Health Ment. Health Serv. Res.* 49 (2), 255–266.