Department for Work & Pensions

The Work Programme: factors associated with differences in the relative effectiveness of prime providers

August 2016

DWP ad hoc research report no. 26

A report of research carried out by NIESR on behalf of the Department for Work and Pensions.

Summary

The Work Programme is delivered by 18 private, public and voluntary sector organisations, working under contract to DWP. These organisations are known as prime providers, or "primes", and operate within a geographical Contract Package Area (CPA). Each CPA has either two or three primes and individuals entering the Work Programme are randomly assigned to one of these. Comparing the outcomes of individuals assigned to each prime within a CPA provides robust estimates of relative effectiveness. Overall, there are 40 contracts across Britain, allowing 26 within-CPA comparisons.

An interim report confirmed both that randomisation had worked well and that there was statistically significant variation in outcomes between primes. This report extends the interim analysis to relate the variation in outcomes to differences in the services offered by primes. By doing so, it aims to understand what makes some primes more effective than others.

The results provide evidence of certain aspects of service being more associated with better outcomes than others and that much of this is concentrated among particular subgroups. Several relationships have been revealed which could provide some clues as to what is driving the variation in provider performance. However, the analytical approach provides evidence of an association only and does not enable us to establish whether these relationships are causal in nature. In addition, much of the variation in performance is not associated in a statistically significant way with the provider characteristics that we have been able to measure. This may suggest that the key characteristics have not been taken into account.

Contents

List of tables

Acknowledgements

Authors

Richard Dorsett is Principal Research Fellow and Director of Policy Evaluation at the National Institute of Economic and Social Research.

Paolo Lucchino is Senior Research Fellow at the National Institute of Economic and Social Research.

Glossary of Terms

Black box	A term for minimum service prescription, which allows providers to decide which interventions to offer to programme participants in order to achieve sustainable employment.
End-to-end provider	A provider that covers the range of general employment-related services a participant receives throughout their time on a programme.
Herfindahl index	A measure of market concentration
Jobcentre Plus	Jobcentre Plus is part of the Department for Work and Pensions (DWP). It provides services that support people of working age from welfare into work, and helps employers to fill their vacancies.
Job outcome payment	A payment to the prime triggered by a Work Programme participant entering work and remaining employed for a specified period
National Benefits Database	A database recording individuals' benefit payments
Participant	A person on the Work Programme. Referred to as a claimant prior to participation on the Work Programme.
Payment Group	Work Programme participants are divided into nine payment groups based on the benefit they claim and prior circumstances (e.g. prison leavers, young people formerly NEET). Providers are paid at different rates for outcomes achieved by different payment groups.
Provider Referrals and Payments	An IT system which automates the clerical referrals and payments process for providers. This was introduced to replace paper-based systems, as well as to facilitate the smoother exchange of information about participants referred for provision.
Referrals data	Information supplied to DWP by primes on details of their referrals to subcontractors.
Self-delivery	The provision of services to Work Programme participants by the prime rather than a subcontractor.
Specialist provider	A specialist provider typically provides niche services such as provision of support for those wanting to become self-employed or support related to a participant's health or underlying issues, such as drug rehabilitation or debt management.

Subcontractor (or "sub")	An organisation delivering services as part of a primes supply chain.
Supply chain	The organisations providing services to Work Programme participants under contract to a prime.
Sustained job outcome	Employment that lasts for at least 13 or 26 weeks (depending on the customer group).
Sustainment payment	A payment to primes for each successive month of sustained employment following a job outcome payment
The Department	The Department for Work and Pensions

Abbreviations

CPA   Contract Package Area

DWP   Department for Work and Pensions

ESA   Employment and Support Allowance

HMRC   Her Majesty's Revenue and Customs

JSA   Jobseeker's Allowance

MSS   Market Share Shift

NBD   National Benefits Database

PbR   Payment by Results

PRaP   Provider Referrals and Payment

VCSE   Voluntary and Community sector and Social Enterprise

1 Executive summary

1.1 Introduction

The Work Programme is the UK's major welfare-to-work programme. It was launched throughout Great Britain in June 2011, replacing much of employment support previously on offer. It provides personalised support to welfare recipients who have been receiving benefits long-term or who are at risk of doing so.

The Work Programme is delivered by 18 private, public and voluntary sector organisations, working under contract to DWP. These organisations are known as prime providers, or "primes", and operate within a geographical Contract Package Area (CPA). Each CPA has either two or three primes and individuals entering the Work Programme are randomly assigned to one of these. Comparing the outcomes of individuals assigned to each prime within a CPA provides robust estimates of relative effectiveness. Overall, there are 40 contracts across Britain, allowing 26 within-CPA comparisons.

An interim report confirmed both that randomisation had worked well and that there was statistically significant variation in outcomes between primes. This report extends the interim analysis to relate the variation in outcomes to differences in the services offered by primes. By doing so, it aims to understand what makes some primes more effective than others.

1.2 Empirical approach

The analysis used a two-stage regression approach to relate provider impacts to a range of service indicators. An important limitation of the analysis is that it cannot provide causal estimates. That is, the results should not be taken to mean that the differences in measured service type are responsible for the differences in primes' relative effectiveness.

An estimation dataset was constructed from four sources:

1. Administrative data on all 1.65 million individuals referred to the Work Programme between June 2011 and September 2014, with benefit and employment outcomes tracked until December 2014. This allowed the following outcomes to be observed:

   - Whether off benefit 12/18/24 months after referral
   - Whether employed 12/18/24 months after referral

- Whether there has been a Job Outcome payment within 12/18/24 months of referral
- Whether there has been a Sustainment Payment within 12/18/24 months of referral.

2. Information provided by DWP on the discounts offered by primes at the time of bidding. This allowed the following measure to be calculated:

- Total potential revenue from each referral.

3. Information requested of primes by DWP regarding the number of onward referrals to their subcontractors. This allowed the following features of the supply chain to be measured:

- % referrals self-delivered, by year
- % referrals passed to private sector subcontractors, by year
- % referrals passed to public sector subcontractors, by year
- % referrals passed to voluntary sector subcontractors, by year
- % referrals passed to specialist subcontractors, by year
- % referrals passed to end to end subcontractors, by year
- Herfindahl index of supply chain, by year

4. Drawing on information from prime provider websites and their records at Companies House, we identify a dummy variable for primes where the 'Provider is not 100% private sector'.

5. A survey of Work Programme participants, which asks respondents about their experience of the Work Programme. This was used to characterise service with regard to respondents' recall of:
   - Mode of contact
     - % via group meeting
     - % via telephone
     - % via text
   - Frequency of contact
   - Nature of support:
     - % receiving assessment of skills
     - % drawing up an action plan
     - % receiving training in maths, reading, writing
     - % participating in work experience placement or voluntary work
     - % receiving advice for setting up own business
     - % receiving financial advice.
     - % warned of risk of benefit cut.

Regression models were able to capture the relationship between impacts and service type overall and also for subgroups defined across age, sex and payment group.

1.3 Results

Testing confirmed that the nature of service as captured through these measures differed significantly across primes. This is a pre-requisite for examining how primes' effectiveness varied with service type.

The main results are summarised below.

- **Frequency of contact** appears associated with positive outcomes, particularly for employment and Sustainment Payments. More frequent contact may be especially important for females and for Customer Group 1 (JSA 18-24).
- **Mode of contact** may matter for outcomes. Contacting claimants by telephone is associated with better employment outcomes overall. Effects on employment outcomes, Job Outcomes and Sustainment Payments appear to be stronger for younger claimants. The possible effect of other forms of contact varies by claimant type. For example, group meetings may be particularly suited to the 'Other ESA-IB' group, while text messages may be less effective for them.
- The **nature of the support** varies in its association with impacts. Offering work placements is uncorrelated with outcomes while providing help starting a business is strongly associated with a reduction in benefit. Different types of support appear better-suited to particular subgroups. Business start-up support is suited to prime-age JSA claimants. Drawing up an action plan, while not significant overall, appears more effective for older individuals and for non-JSA claimants. Skills assessments appear particularly badly-suited to older people but may work better for non-JSA claimants. English and maths training may be somewhat less effective for older claimants and more effective for the Other ESA-IB Customer Group. Financial advice is associated with higher employment among the JSA 18-24 group. Lastly, our results are suggestive of a negative association between the share of claimants stating they were informed of the risk of losing all or part of their benefit if they fail to comply with Work Programme requirements (i.e. benefit sanctions) and job outcome payments and sustainment payments, but we find no association with employment or benefit receipt outcomes.
- With regards to the **supply chain**, sectoral composition appears unimportant, as does self-delivery and the degree of competition. However, there are some subgroup differences. Prime age, particularly JSA, claimants may do better than other groups when assigned to providers that are not 100% for-profit. For claimants aged 35-44, self-delivery is associated with better outcomes. On the other hand, claimants age 45 and over appear to benefit from referrals to a private sector subcontractors. Individuals in the "Other ESA-IB" group appear to do less well when referred to a public sector subcontractor. The use of specialist as opposed to end-to-end providers is not associated with effectiveness, although again there is subgroup variation. Older claimants appear to do somewhat better than younger claimants under primes that refer more to end-to-end provision. Bigger differences are seen

with specialist provision; this is more effective both for older claimants and the New ESA Customer Group.

- The **total potential revenue** from each individual - as determined by discounts - is not associated with outcomes overall. The absence of associations also largely plays out among the subgroups considered.

1.4 Conclusion

The results provide evidence of certain aspects of service being more associated with better outcomes than others and that much of this is concentrated among particular subgroups. This is perhaps to be expected given the varied nature of Work Programme participants who are likely to differ in the way they respond to differences in provider support. The observed associations may provide some clues as to what might be driving provider effectiveness. However, characteristics used in the analysis may well correlate with other, unobservable or omitted, characteristics, which makes it impossible to disentangle which one may be the true driver of differences in outcomes. As a result, the analysis cannot prove cause and effect – it simply provides some clues as to what might be driving provider effectiveness.

Furthermore, looking across all participants, the conclusion is that much of the variation in provider performance remains unexplained. For example, only one of the observed characteristics is associated with differences in more than three of the twelve performance measures analysed. This is further evidence to suggest that the unobserved characteristics may be the key drivers for performance.

Finally, it is important to view these results in the context of the wider evidence base on what drives performance in employment support. In some cases, other studies specifically designed to test the effectiveness of a given form of support are likely to provide more conclusive evidence on that works.

There is scope for follow-on analysis to probe further the results reported here. As a general comment, randomisation within CPA has provided a low-cost way of being able to assess relative provider effectiveness. This continues the approach used with the Flexible New Deal and one might expect it to be continued with the successor to the Work Programme. Over time, as the number of estimates grows, there will be more opportunity to model the data in a way that can allow more definite statements about the causal relationships in the data.

2 Introduction

The Work Programme is the UK's major welfare-to-work programme. It was launched throughout Great Britain in June 2011, replacing much of the employment support previously on offer, including the New Deals, Employment Zones and Pathways to Work. It provides personalised support to welfare recipients who have been receiving benefits long-term or who are at risk of doing so.

The Work Programme is delivered by 18 private, public and voluntary sector organisations, working under contract to DWP. These organisations are known as prime providers, or "primes".

The aim of this research is to understand what makes some primes more effective than others. Essentially, the idea is to try and open up the "black box" and to explore what might drive any observed differences in outcomes. Examples of the characteristics of primes that might influence outcomes include:

- the type of support provided by the prime
- whether for profit or not for profit
- characteristics of the supply chain
- the degree of competition among subcontractors
- the tendency to 'self-deliver' rather than subcontract.

Furthermore, we can also explore variations across, for example:

- payment groups
- men and women
- age bands.

Importantly, individuals entering the Work Programme are randomly assigned to one of the two or three providers operating within their Contract Package Area (CPA). This allows robust (experimental) estimates of the effectiveness of each prime relative to the other 1 or 2 primes operating locally. The country is divided into 18 CPAs. In 14 of these, there are two primes and in 4 (East London, West London, Birmingham and Manchester) there are three. In total, there are 40 contracts allowing 26 experimental within-CPA comparisons.

We have approached the research in two stages. The first stage was to establish the extent of variation between primes within CPAs. The results of this analysis, Dorsett and Lucchino (2014), showed:

- randomisation worked well; within each CPA, between-prime differences in the characteristics of participants were only ever slight

- provider allocation is important; within each CPA, there were significant differences across providers for all of the outcomes considered.

This report details the result of the Stage 2 analysis. The aim of the research is to understand the extent to which differences in the characteristics of primes and the service they deliver appear to be associated with variations in effectiveness. We emphasise that these associations are not experimental and so cannot be taken to imply causal relationships. Nevertheless, they still provide useful information to policy makers and in the conclusion of the report we consider what it is possible to learn from the results.

3 An overview of the Work Programme

Everyone receiving Jobseeker's Allowance (JSA) or Employment Support Allowance (ESA) can access the Work Programme at a specified point in their claim, depending on their circumstances. Some claimants must join the Work Programme while for others participation is voluntary. Claimants volunteering for the Work Programme are advised by Jobcentre Plus whether that is the most appropriate option for them. Conditions for access to the Work Programme by claimant group are summarised in Table 3-1 and Table 3-2.

Table 3-1 Work Programme Referral points and participation requirements

Mandatory customers Participant	Entry/Access Point	Payment Group
JSA claimants aged 18 to 24.	Required from 9 months	PG 1
JSA claimants aged 25 and over.	Required from 12 months	PG 2
JSA 18 year old NEET participants.	Required from 3 months	PG 3
JSA Repeaters.	Required from 3 months	PG 3
JSA Ex-IB participants.	Required from 3 months	PG 4
JSA Early Access participants.	Optional from 3 months	PG 3
JSA Prison Leaver	Required from Day One of release from prison or the first date of claim if made within 13 weeks	PG 9
Employment and Re-offending Pilot	Required from Day One of release from prison or the first date of claim if made within 13 weeks	PG 9
ESA (IR) WRAG with 3 Month Prognosis.	Required from WCA Outcome	PG 6
ESA (IR) WRAG with 6 Month Prognosis.	Required from WCA Outcome	PG 6
ESA (IR) Ex-IB WRAG with 3 Month Prognosis.	Required from WCA Outcome	PG 7
ESA (IR) Ex-IB WRAG with 6 Month Prognosis.	Required from WCA Outcome	PG 7
Existing ESA (IR) WRAG with 3 or 6 Month Prognosis.	Mandatory from 10/9/12	PG 6
ESA (IR) WRAG with 12 Month Prognosis.	Mandatory from 12 Nov 2012	PG 5
WP ESA Mandatory (IR) WRAG 12mth	Required from WCA Outcome	PG 6
WP ESA Mandatory (IR) WRAG 18mth	Required from WCA Outcome	PG 7
WP ESA Mandatory (IR) WRAG 24mth	Required from WCA Outcome	PG 7

Table 3-2 Work Programme Referral points and participation requirements

Voluntary customers Participant	Entry/Access Point	PG
IS participants.	Optional from benefit entitlement	PG 8
IB participants.	Optional	PG 8
ESA (IR) WRAG (with Youngest Child Under 5 OR Full-Time Carer) with 3 or 6 Month Prognosis.	Optional from WCA outcome	PG 6
ESA (IR) Ex-IB WRAG (with Youngest Child Under 5 OR Full-Time Carer) with 3 or 6 Month Prognosis.	Optional from WCA outcome	PG 7
ESA (IR) WRAG (with Youngest Child Under 5 OR Full-Time Carer) with 12+ Months prognosis.	Optional from WCA outcome	PG 5
ESA (c) WRAG participants.	Optional from WCA outcome	PG 5
ESA (c) WRAG (with Youngest Child Under 5 OR Full-Time Carer).	Optional from WCA outcome	PG 5
ESA (IR) Support Group.	Optional from WCA outcome	PG 6
ESA (IR) Ex-IB Support Group.	Optional from WCA outcome	PG 7
ESA (c) Support Group.	Optional from WCA outcome	PG 6
ESA (c) Ex-IB Support Group.	Optional from WCA outcome	PG 7
ESA Credit Only.	Optional from WCA outcome	PG 5
ESA Voluntary (IR) WRAG 18mth	Optional from WCA outcome	PG 7
Pension Credit Claimants.	Optional from 12 months (claiming benefits)	PG 2
Pension Credit Claimants with Health Conditions.	Optional from benefit entitlement	PG 5
Employment and Re-offending Pilot	Required from Day One of release from prison or the first date of claim if made within 13 weeks – Provider paid more to give extra	PG 9
JSA Benefit Cap	If impacted by benefit cap then can go from 3 months	PG 3

Individuals referred to the Work Programme are assigned to one of the two or three primes operating within their CPA. Assignment is on a random basis, with the number of referrals shared equally by claimant group and payment group (PG) between Prime providers

operating in the CPA. Table 3-3 provides a list of contracts, showing which primes operate in each CPA. It also shows the number of referrals received up until 30 September 2014.[1]

Table 3-3 List of prime provider contracts

CPA	PROVIDER	REFERRALS IN DATASET
East of England	Ingeus UK LTD	59,263
East of England	Seetec	57,056
East Midlands	A4E Ltd	58,478
East Midlands	Ingeus UK LTD	60,329
West London	Ingeus UK LTD	34,971
West London	Maximus Employment UK Ltd	34,002
West London	Reed in Partnership	33,539
East London	A4E Ltd	47,308
East London	Careers Development Group	47,096
East London	Seetec	46,691
North East	Avanta Enterprise Ltd	51,860
North East	Ingeus UK LTD	50,811
Merseyside, Halton, Cumbria and Lancashire	A4E Ltd	56,295
Merseyside, Halton, Cumbria and Lancashire	Ingeus UK LTD	56,372
Manchester, Cheshire and Warrington	Avanta Enterprise Ltd	38,090
Manchester, Cheshire and Warrington	G4S	38,178
Manchester, Cheshire and Warrington	Seetec	37,513
Scotland	Ingeus UK LTD	80,419
Scotland	Working Links	79,474
Thames Valley, Hampshire and Isle of Wight	A4E Ltd	33,000
Thames Valley, Hampshire and Isle of Wight	Maximus Employment UK Ltd	33,246
Surrey, Sussex & Kent	Avanta Enterprise Ltd	40,451
Surrey, Sussex & Kent	G4S	41,040
Devon and Cornwall, Dorset and Somerset	Prospects Services Ltd	23,908
Devon and Cornwall, Dorset and Somerset	Working Links	24,091
Gloucestershire, Wiltshire and Swindon, West of England	JHP Group Ltd	22,104
Gloucestershire, Wiltshire and Swindon, West of England	Rehab Jobfit	22,008

---

[1] DWP announced in the summer 2014 that it would terminate the contract with Newcastle College Group within 12 months. At the time of the latest data used in this analysis, October 2014, the Newcastle College Group had not yet been replaced. We expect the announcement of the termination will have very little effect on the analysis.

CPA	PROVIDER	REFERRALS IN DATASET
Wales	Rehab Jobfit	44,618
Wales	Working Links	45,231
Birmingham and Solihull, the Black Country	EOS-Works Ltd	38,732
Birmingham and Solihull, the Black Country	Newcastle College Group	37,104
Birmingham and Solihull, the Black Country	Pertemps	37,800
Coventry, Warwickshire, Staffordshire and the Marches	ESG	31,344
Coventry, Warwickshire, Staffordshire and the Marches	Serco Ltd	31,608
West Yorkshire	Business Employment Services	39,068
West Yorkshire	Ingeus UK LTD	40,291
South Yorkshire	A4E Ltd	25,026
South Yorkshire	Serco Ltd	25,213
North East Yorkshire and the Humber	G4S	23,711
North East Yorkshire and the Humber	Newcastle College Group	23,098

Following referral to a Work Programme provider, the provider is responsible for contacting that person to discuss the programme and begin planning the steps needed to support them into sustained employment. Once this activity has taken place, the provider registers an attachment to the Work Programme. Individuals remain on the Work Programme for two years and can only be referred and attached once during this period.

A small proportion of claimants sign off benefit prior to being attached. Typically, under 2% do not attach. Among the others, attachment is within 17 days of referral, on average.

Primes are paid according to the results they achieve. Payments are primarily for getting participants into sustained employment; there are no payments for job entry. Only one Job Outcome payment can be paid per participant during the two years they are on the Work Programme. This payment will be made after either three or six months, of continuous or cumulative employment, depending on Payment Group. Following a Job Outcome payment, Sustainment payments are paid for every subsequent four week period in continuous employment, for up to two years. Participants complete the Work Programme after 104 weeks or when the maximum number of Sustainment payments have been made.

The timescales for achieving a Job Outcome payment and Sustainment payment differ according to the payment group to which the participant is assigned by Jobcentre Plus when they are referred to the programme. These timescales and the maximum number of

Sustainment payments which can be paid for each payment group are shown in Table 3-4. The payment of Job Outcomes and Sustainment payments is subject to post-payment validation.

Payment amounts are also shown. These vary by payment group, reflecting differences in the anticipated difficulty in achieving positive outcomes. There are two additional points not obvious from the table. First, attachment fees declined over time and disappeared altogether after year 3. The year 2 attachment fee was 75% of the year 1 amount. The year 3 attachment fee was 50% of the year 1 amount. Second, at the time of bidding, prime providers were invited to offer discounts on the Job Outcome payments. Table 3-4 shows the maximum amount payable but there was variation across providers in the agreed size of these payments.

Table 3-4 Payment structure by payment group

Payment group:	Customer group	Job outcome paid week:	Sustainment payment (4-weekly) from:	Sustainment: maximum amount of 4 weekly payments:	Year 1 attachment fee (£)	Job outcome fee (£ maximum)	Sustainment payment (£)
1	JSA 18-24	26	30	13	400	1200	170
2	JSA 25+	26	30	13	400	1200	215
3	JSA early access	13	17	20	400	1200	250
4	JSA ex-IB	13	17	20	400	1200	250
5	ESA volunteers	13	17	20	400	1000	115
6	New ESA claimants	13	17	20	600	1200	235
7	ESA ex-IB	13	17	26	600	3500	370
8	IB/IS (England only)	13	17	13	400	1000	145
9	JSA prison leavers	26	30	20	400	1200	200

4 Data and measures of outcomes and service delivery

4.1 Outcomes

Administrative data on individuals participating in the Work Programme were provided by DWP. These data detail the circumstances for each of the 1,655,097 individuals referred to the Work Programme between June 2011 and September 2014, and track their outcomes until December 2014. This includes CPA and provider information and dates of referral, attachment, and any payments made to the provider.

This dataset also includes information on the claimant's characteristics, such as the claimant type, payment group, gender, age, health and disabilities, qualification levels, ethnicity and previous occupation. The tables in Annex A present the distribution of these characteristics in the overall Work Programme population as well as broken down by high level Payment Group. As the tables show, data are complete for most characteristics. Qualification levels are an exception to this, as these are either missing or not known for 56% of the sample. Around 6% of individuals in the sample do not report a previous occupation. This is predominantly among individuals on disability benefits. Around 5% of individuals do not report their ethnicity.

A small number of data observations contain possibly contradictory information. Specifically, 2,251 individuals are reported as being in the JSA 18-24 year old payment group, but are aged 25 or over; and 2,038 individuals aged 25 or over are reported as being in the JSA 18-24 payment group. These individuals represent 0.25% of the sample and were dropped.

DWP also provided administrative data on:

- the full benefit history of all individuals who participated in the program over the above-mentioned period, followed forward until December 2014
- employment spells for all participants over the same period.

All three datasets were merged using a unique and anonymous identifier. Only four pairs of individuals with a duplicate identifier were found in the 1,655,097 Work Programme records, and one arbitrarily selected case from each pair was dropped. A further 367 records showed a unique identifier of 1 and were dropped.

These records were matched then to the National Benefit Database (NBD) data, with a 99.92% success rate. Only 6,143 individuals (0.37% of the sample) were not matched. Possible reasons for this very slight failure to match to NBD data include: a time lag in

logging ex-offenders onto the NBD; having been in prison long enough to not have any benefit record since 1999; lower engagement with Jobcentre Plus; and errors in the data.

Data on Work Programme participants were matched to HMRC data on their employment since referral. 1,497,976 individuals were matched, delivering a match rate of 90.5%. Individuals not matched either had no employment records or had employment spells only before their referral to the Work Programme. It was necessary to carry out a number of data cleaning operations on the employment data. A large number of end dates were missing (indicated by an end date in December 2099). Around 82% of these were followed by spells with a later start date. We interpret this as implying that the end date is missing rather than the spell being ongoing. We set these to the end of the tax year during which the employment spell started. The remaining 18% of spells were not followed by any subsequent spell. Almost half of these had a start date of 2014, suggesting they are likely to be ongoing spells. Their end date was set to a date in the near future (specifically, the end the 2014/15 tax year). The remaining spells had earlier start dates and were assumed to have genuine missing end dates. These were set to the end of the relevant tax year. Other minor cleaning operations include dropping spells with dates in the far future (but not December 2099) or the distant past. These amendments inevitably make assumptions that may not be entirely correct. However, it is important to note that, as individuals are randomly assigned to providers, measurement error will be balanced across these, and will not bias inter-provider comparisons.

Using the above data, we construct the following outcome measures for each individual in our sample:

- Whether off benefit 12/18/24 months after referral (source: NBD)
- Whether employed 12/18/24 months after referral (source: HMRC)
- Whether there has been a job outcome payment within 12/18/24 months of referral (source: Work Programme Administrative Database)
- Whether there has been a sustainment payment within 12/18/24 months of referral (source: Work Programme Administrative Database)

4.2 Characterising aspects of service delivery

Our ability to characterise the nature of the service offered by primes is dependent on the data available. We were able to draw on three sources:

- Details on the financial discounts offered by prime providers at the time of bidding to DWP to win the contracts to provide the service
- Information provided by DWP on the volume of referrals to each subcontractor in the supply chain. This is provided on a monthly basis, separately for each contract.
- A survey of Work Programme participants, which asks respondents about their experience of the Work Programme.

From these sources, we identified numerous measures that could potentially be included in the model. However, as discussed in the Stage 1 report, the fact that there are only a small number of experimental impact estimates means that it is not possible to include all variables of potential interest. Furthermore, some characteristics, while potentially important influences on overall effectiveness, may not vary sufficiently across providers to allow the question of relative effectiveness to be examined. The variables that we consider in the main analysis are described below.

4.3 Payment amounts

Providers were encouraged to offer discounts on the Job Outcome payments. These discounts could vary by year and by payment group. The consequence of the discounts was that providers differed in the amount of revenue each referral could potentially attract. We assess the relevance of discounts by considering a variable for the total potential revenue for each individual. This is calculated as the job outcome fee (where providers can offer discounts) plus the sustainment payment multiplied by the maximum number of sustainment payments possible (which is fixed by payment group).

4.4 The nature of the supply chain

Primes all have supply chains of subcontractors but vary in their use of them. Subcontractors in tier one of the supply chains are responsible for delivering the end-to-end process or a specific element of the service, such as job-brokering. These partners are usually paid by results. Second tier organisations work on a call-off basis, as and when a prime provider judges a participant could benefit from that organisation's help. The voluntary and community sector are represented at all levels of Work Programme delivery i.e. as prime providers and in tiers one and two of prime providers' supply chains.

The referrals data provided by DWP allow us to see how much use has been made of subcontractors in practice. There are limitations to these data. For instance, there is no distinction by payment group, nor do we know how many participants the referrals relate to, since individuals may be referred on by primes more than once. Nevertheless, it is the best available information to characterise how primes tend to distribute work. We break down the total number of referrals into four categories to identify how impacts may vary with sectoral composition of the effective supply chain.

- % referrals self-delivered, by year
- % referrals passed to private sector subcontractors, by year
- % referrals passed to public sector subcontractors, by year
- % referrals passed to voluntary sector subcontractors, by year

We also construct a measure of the sector of the prime. Most primes are privately owned, but a minority are fully or partially controlled by the voluntary or public sector entities. We

therefore construct a dummy variable indicating the primes where the 'Provider is not 100% private sector'.

Providers also vary in the use they made of specialist and end-to-end subcontractors. We capture the extent of this through the following two variables:

- % referrals passed to specialist subcontractors, by year
- % referrals passed to end to end subcontractors, by year

Lastly, we use the referrals data to construct a measure of competition within the suppliers associated with a prime:

- Herfindahl index of supply chain, by year

4.5 Nature of service provided

From the survey of participants, it was possible to characterise different aspects of how individuals experienced the Work Programme. The responses from participants do not distinguish between prime providers and subcontractors. However, since each surveyed individual's prime provider is known, the data can be used to construct contract-level summary measures of individuals' experiences.

We identified the most appropriate variables to consider by applying two criteria. Firstly, to ensure a sufficient sample size, we only considered responses to questions that were asked to all or at least a large majority of the survey respondents. Secondly, we checked that the variables under consideration varied sufficiently across providers to allow the question of relative effectiveness to be examined. We did so by regressing each service variable on a full set of contract dummies and, separately, on CPA dummies. This second regression is a restricted version of the first. We tested whether this restriction was rejected statistically. This amounts to a test of within-CPA provider variation. We selected variables where this test had a p-value below 10% (suggesting statistically significant variation).

The result of this process was that we selected four measures to capture the mode and intensity of the service, as listed below. Survey respondents were asked to identify all the channels of communication used by the provider. From these responses, we were able to construct variables indicating the percentage of customers contacted through a given communication channel:

- Mode of contact: % via group meeting
- Mode of contact: % via telephone
- Mode of contact: % via text

and another variable capturing the frequency of contact:

- number of times provider in contact with customers.

We also selected six measures to capture the nature of support/advice:

- % receiving assessment of skills
- % drawing up an action plan
- % receiving training in maths, reading, writing
- % participating in work experience placement or voluntary work
- % receiving advice for setting up own business
- % receiving financial advice.

Lastly, studies have pointed to the impact of sanctions, sometimes suggesting that they increase benefit exits but not necessarily employment entry. To explore the possible influence of provider warnings of sanctions, we construct a measure of the use of such cautionary messages from the survey of Work Programme participants. Importantly, providers have a legal obligation to inform all claimants of possible sanctions if they fail to comply with Work Programme activities. However, there is considerable variation in the incidence of survey participants at each provider responding affirmatively to the question "Did any of your Work Programme advisers tell you that you had to do something and you could be at risk of losing all or part of your benefit if you did not do it?". This could be due to differential use of mandation by providers or imperfect recall by survey participants. As such, we interpret this as a proxy for the use of warnings by providers or the intensity of such warnings. Our analysis therefore includes:

- % stating they were warned of risk of benefit cut

4.6 Descriptive statistics on service characteristics

Table 4-1 and Table 4-2 present some descriptive statistics on service characteristics. Table 4-1 presents characteristics that vary across contracts, while Table 4-2 presents those characteristics that vary both across contracts and contract years. We present selected percentiles of the distribution of observed values for each service characteristic. The main point to note here is that there is considerable variation in the value of each of the characteristics. In Section 5 we decompose this variation into within- vs. between-CPA variation in service provision.

Table 4-1 Variation in service provision, contract level

Variable	N			Percentile		
	Obs	10th	25th	50th	75th	90th
Mode of contact: % via group meeting	40	0.42	0.45	0.49	0.53	0.56
Mode of contact: % via telephone	40	0.45	0.49	0.52	0.57	0.61
Mode of contact: % via text	40	0.1	0.12	0.2	0.25	0.34
Number of times provider in contact with customer	40	7.71	8.63	9.76	11.36	12.77
Support type: % receiving assessment of your skills	40	0.26	0.29	0.33	0.38	0.44
Support type: % drawing up an action plan	40	0.35	0.39	0.45	0.51	0.56
Support type: % receiving training in maths reading writing	40	0.07	0.1	0.11	0.16	0.18
Support type: % work experience placement or voluntary work	40	0.09	0.11	0.14	0.16	0.18
Support type: % advice for setting up your own business	40	0.05	0.08	0.11	0.15	0.18
Support type: % Financial advice	40	0.09	0.13	0.16	0.21	0.27
% stating they were warned of risk of benefit cut	40	0.62	0.65	0.67	0.73	0.77
Provider not 100% private sector	40	0	0	0	0	1

Table 4-2 Variation in service provision, contract-year level

Variable	Year	N	Percentile				
		Obs	10th	25th	Median	75th	90th
% self-delivered - Annual (DWP)	1	40	0	0.42	0.56	0.7	0.73
% self-delivered - Annual (DWP)	2	39	0	0.41	0.56	0.7	0.82
% self-delivered - Annual (DWP)	3	35	0	0.4	0.57	0.7	0.77
% referrals to private sector subs - Annual (DWP)	1	40	0.05	0.14	0.23	0.36	0.61
% referrals to private sector subs - Annual (DWP)	2	39	0.02	0.12	0.22	0.34	0.61
% referrals to private sector subs - Annual (DWP)	3	35	0.03	0.11	0.24	0.33	0.61
% referrals to public sector subs - Annual (DWP)	1	40	0	0	0.03	0.08	0.15
% referrals to public sector subs - Annual (DWP)	2	39	0	0	0.02	0.07	0.17
% referrals to public sector subs - Annual (DWP)	3	35	0	0	0.03	0.08	0.17
% referrals to voluntary sector subs - Annual (DWP)	1	40	0.03	0.06	0.12	0.24	0.41
% referrals to voluntary sector subs - Annual (DWP)	2	39	0.01	0.05	0.13	0.26	0.42
% referrals to voluntary sector subs - Annual (DWP)	3	35	0.03	0.05	0.11	0.26	0.46
% referrals to specialist subs - Annual (DWP)	1	40	0	0	0	0.01	0.04
% referrals to specialist subs - Annual (DWP)	2	39	0	0	0	0.01	0.05
% referrals to specialist subs - Annual (DWP)	3	35	0	0	0	0.01	0.06
% referrals to end to end subs - Annual (DWP)	1	40	0.2	0.27	0.35	0.52	1
% referrals to end to end subs - Annual (DWP)	2	39	0.14	0.24	0.35	0.54	1
% referrals to end to end subs - Annual (DWP)	3	35	0.15	0.24	0.31	0.46	1
Herfindahl index of supply chain - Annual (DWP)	1	40	0.03	0.07	0.17	0.24	0.39
Herfindahl index of supply chain - Annual (DWP)	2	39	0.03	0.08	0.14	0.21	0.33
Herfindahl index of supply chain - Annual (DWP)	3	35	0.04	0.11	0.15	0.22	0.38
Total revenue achievable from individual (£100s)	1	356	32.85	37.14	55.4	65.04	130.2
Total revenue achievable from individual (£100s)	2	354	32.5	36.14	54.4	64.04	128.7
Total revenue achievable from individual (£100s)	3	349	29.35	33.17	51.6	61.6	108.67

5 Econometric approach

The aim of the project is to understand how provider characteristics or actions can affect performance. The Phase 1 report showed that there are significant differences between providers in their effectiveness. Providers have the freedom to tailor the support they offer to meet the needs of individual participants. In this section, we present the methodological approach we use to explore whether there are particular characteristics of the providers or the support they offer that are associated with higher performance.

5.1 Identification strategy

Formally, we can think of the simple case in which the outcome for individual i in CPA c referred to provider p, is determined by:

Equation 1

$$y_{icp} = c_c'\gamma + l_{cp}'\pi + x_{icp}'\beta + \varphi_c + \upsilon_{cp} + \varepsilon_{icp}$$

Outcomes will depend on the circumstances in each CPA, c_c. Examples of these could be the strength of the local labour market or the sectoral composition of the local economy. Individual characteristics, x_{icp}, such as labour market history or qualification levels will also influence outcomes. Finally, the above equation allows for providers to affect outcomes, through their service characteristics (l_{cp}). Outcomes are also influenced by unobserved factors, which we can decompose into a CPA component (φ_c), a contract component (υ_{cp}), and an individual component (ε_{icp}).[2]

Individuals are randomly assigned to providers within each CPA. Information systems governing the process of randomisations were fully operational from 30th January 2012. Starting from this date, the allocation tool identified, on a daily basis and for each payment group within each CPA, the provider which had the fewest referrals to date, relative to their contracted market share. All referrals for that day would go to that provider.[3] Other than in a limited number of clerical cases and mandatory ESA (IR) WRAG 12 month prognosis cases, opportunities for human intervention are minimal.

Prior to 30 January 2012, the information systems only displayed allocations to providers by Claimant Group (rather than Payment Group) and at the Job Centre Plus district level (rather than the CPA). The same principle as above was then applied to the allocation to providers but was carried out by advisers and was calculated using data at these lower levels. The Phase 1 report confirmed that randomisation delivered balanced customer

---

[2] We refer to "contract" as a provider in a specific CPA.
[3] Up to 5th August 2013, providers had equal market shares within each CPA, meaning that the day's referrals would simply go to the provider with fewer referrals to date.

profiles across providers under both allocation mechanisms (please see Annex B for summary of phase 1 analysis).

Random assignment of individuals allows experimental identification of *relative provider effectiveness* as within-CPA provider differences in outcomes. In a pooled model across all CPAs, all that is required is to introduce CPA fixed effects that capture the influence of both observed (c_c) and unobserved (φ_c) factors which are common at CPA level. In other words, differing customer composition across providers in different CPAs and/or prevailing conditions across CPAs cannot bias our estimates of relative provider effectiveness.

Secondly, randomisation ensures the sets of individuals assigned to each provider exhibit balanced profiles across observable characteristics. Importantly, as we discuss in Annex B, such balanced profiles also lend support to the (untestable) assertion that the groups are balanced over unobservable characteristics. Once CPA fixed effects are netted out, randomisation guarantees that the residual variance in x_{icp} and ε_{icp} is uncorrelated with provider service characteristics l_{cp} . Indeed, controlling for the effect of individual characteristics is not strictly necessary to estimate relative provider effectiveness, though doing so will tend to increase the precision of such estimates. Within-CPA comparisons across providers will therefore also not be biased by compositional differences in customers across these providers. Additionally, they will not be biased should the effect of individual characteristics vary by CPA.

The aim of Stage 2 is to unpack these relative provider differentials and understand which provider characteristics (l_{cp}) are associated with higher impacts. Identifying the *causal* effect of a given service characteristic on outcomes is however much more difficult. On the one hand, by comparing within-CPA differences in outcomes across providers, this setup is robust to endogenous correlations of service provision in response to CPA circumstances or the profile of individuals in each CPA, or indeed the interaction between the two. For example, we need not be concerned if basic Maths and English training is more likely to be administered in CPAs with strong service sectors. However, Equation 1 highlights the fundamental limitation to identifying causal effects of service characteristics. Essentially, this would require having a correctly specified model in which unobserved provider service characteristics (v_{cp}) are not correlated with the observed service characteristics that we include in our estimations. This is the challenge of any observational study. In the specific context under study, omission of relevant service characteristics from the model may bias the coefficient on variables included in the specification. This is an important caveat worth keeping in mind given the limitations in the available data on service provision, as well as the need to keep the model parsimonious. Nevertheless, the estimations can aim to uncover robust correlations across service provision and impacts that are likely to shed important light on optimal forms of intervention.

5.2 Inference

Aside from the issue of bias, the estimation of correct standard errors faces two additional challenges in this context. Firstly, given the complex error structure, careful thought must be given to what the appropriate correction to standard error calculation is. Secondly, and more importantly, despite drawing on a very large number of individual observations, the variation we are most interested in exploiting only occurs across the 40 provider contracts across the 18 CPAs. With a larger number of contracts, this could be addressed by clustering standard errors at the contract, or even CPA, level (Liang & Zeger, 1986). However, the relatively small number of contracts may mean that conventional (asymptotic) calculations of standard errors when estimating the model on the full individual dataset may be biased. When the number of clusters is small, correct standard errors may be obtained using two-step estimators considered by Donald and Lang (2007) and, more recently, by Bryan and Jenkins (2013). To guide the choice of the most appropriate approach, we ran a Monte Carlo simulation using 2,000 random draws from the above error structure and tested this over seven different estimation methods.[4] We identified the Donald and Lang (2007) approach with clustering at CPA level in the second stage as providing an unbiased and efficient estimate of the standard error, and we adopt this as our selected approach across the work in Stage 2.

5.3 Estimation

Drawing on the above discussion, our specific estimation approach is as follows. The equation we wish to estimate is the following:

Equation 2

$$y_{icp} = \alpha_c + l'_{cp}\pi + a_c \cdot x'_{icp}\beta + \varphi_c + v_{cp} + \varepsilon_{icp}$$

There are C CPAs and the model allows for a CPA level effect, α_c. Individual characteristics, x_{icp}, will also influence outcomes, and their effect is allowed to vary across CPAs through an interaction with all CPA dummies. Monte Carlo simulations in Donald and Lang (2007) indicate this flexibility allows for a cautious approach to inference. This minimises the risk of incorrectly identifying correlations as statistically significant.

Following Donald and Lang 2007, Equation 2 is estimated in two stages. The first stage regresses, within each CPA, individual outcomes on individual characteristics and dummies

---

[4] Specifications tested were: OLS without clustering; OLS clustering at contract level; OLS clustering at CPA level; Donald & Lang (2007) approach with 40 contract groups, no clustering in the second stage; Donald & Lang (2007) approach with 40 contract groups, clustering at CPA level in the second stage; random effects at contract level; multilevel model with level 2 defined at the contract level.

for the contracts within the CPA. Together, this set of regressions identifies the parameters of the following equation:

$$y_{icp} = \delta_{cp} + \alpha_c \cdot x'_{icp}\beta + \varepsilon_{icp}$$

From this, we can extract the mean and standard error of the 40 provider contracts dummies. The coefficient on each dummy is the mean outcome for that contract, adjusted for individual characteristics. The second stage involves regressing these estimated coefficients on service level variables as follows:

Equation 3

$$\hat{\delta}_{cp} = \alpha_c + l'_{cp}\pi + \varphi_c + \upsilon_{cp}$$

where l_{cp} are the service level characteristics and the full set of CPA dummies (φ_c) is included to absorb any CPA level effects and errors. Since the φ_c error components are CPA wide, and in line with the results from our Monte Carlo simulation, we cluster the second stage standard errors at CPA level.

5.4 Subgroup analysis

Here we discuss the set-up where we allow for service characteristics to have a heterogeneous effect across population subgroups. The above identification strategy follows through, with one modification: rather than comparing outcomes within-CPA only, we compare outcomes within-CPA-within-subgroup. For example, we would compare the outcomes across providers in CPA c for females only. Because individuals are not randomised to providers conditional on subgroup, the profile of the subgroup members across providers may not be balanced. At the same time, the allocation process is independent of subgroup, meaning that there is no reason to expect systematic imbalance either. Overall, so long as the cell size is sufficiently large, randomisation should hold at the subgroup level too. Controlling explicitly for a number of observed individual characteristics will also improve the estimate of relative provider effectiveness. These comparisons can then be related to service provision in a way that allows for their correlation with impacts to vary across subgroups.

Formally, the model we would like to estimate is the following:

$$y_{iscp} = \alpha_{sc} + S_s \cdot l'_{cp}\pi + \alpha_{sc} \cdot x'_{icp}\beta + \varphi_c + \upsilon_{cp} + \varepsilon_{iscp}$$

We focus on within-CPA-within-subgroup comparisons by allowing for a CPA-subgroup level effect, α_{sc}. As before, individual characteristics (other than the characteristic that defines the subgroup) will also influence outcomes, and their effect is allowed to vary across CPA-subgroup through an interaction with cell dummies, α_{sc}. Service characteristics

(l_{cp}) are interacted with all levels of the subgroup variable (S_s), thereby allowing a distinct correlation to be estimated for each subgroup.

The first stage equation is as follows, with δ_{scp} estimating the covariate-adjusted mean outcome estimated on regressions for each CPA-subgroup:

$$y_{iscp} = \delta_{scp} + a_{sc} \cdot x'_{iscp}\beta + \varepsilon_{iscp}$$

The second stage then takes the form:

Equation 4

$$\hat{\delta}_{scp} = \alpha_{sc} + S_s \cdot l'_{cp}\pi + \varphi_c + \upsilon_{cp}$$

The estimated coefficient dummies from the first stage are regressed on service level characteristics (l_{cp}) and a full set of CPA-subgroup dummies. As above, we cluster the second stage at CPA level, in line with good practice of clustering at the highest relevant level (Cameron & Miller, 2013).

5.5 Time-varying service provision

In some cases, we will be interested in relating outcomes to provider service characteristics that are time-varying. An example of this is the percentage of cases referred to subcontractors each month. The above framework easily generalises to a framework with variation over time. Our overall equation is as follows:

$$y_{icpt} = \alpha_{ct} + l'_{cp}\pi + l'_{cpt}\eta + \alpha_{ct} \cdot x'_{icp}\beta + \varphi_{ct} + \upsilon_{cp} + \theta_{cpt} + \varepsilon_{icpt}$$

The comparison of outcomes is now made within-CPA and by time period t. As above, this is achieved by dummies for such cells, α_{ct}. The effect of individual characteristics is allowed to vary within each cell. The main departure is in the introduction of time-varying service characteristics, l_{cp}, alongside time-invariant ones. Analogously, we decompose the provider level error into a fixed component (υ_{cp}) and a time-varying one (θ_{cpt}).

The first stage is estimated as usual, with cells determined by CPA-time:

$$y_{icpt} = \delta_{cpt} + \alpha_{ct} \cdot x'_{icp}\beta + \varepsilon_{icpt}$$

The second stage is as follows:

Equation 5

$$\hat{\delta}_{cpt} = \alpha_{ct} + l'_{cp}\pi + l'_{cpt}\eta + \varphi_{ct} + \upsilon_{cp} + \theta_{cpt}$$

As with previous cases, CPA-time dummies allow us to identify the overall contribution of provider characteristics. In this case, however, as the effect of time-invariant characteristics is likely to be of less interest, it can be absorbed by contract-level fixed effects. This isolates the estimation of the time-varying service characteristics, l_{cp}. Naturally, any causal interpretation of these effects depends on the time-varying error (θ_{cpt}) being uncorrelated with the variable of interest.

6 Establishing service variation within CPAs

The estimation strategy presented in the previous section seeks to relate within-CPA differences in outcomes across providers to within-CPA differences in service provision across the same providers. For this to be possible, there needs to be sufficient within-CPA variation in both these sets of data. This was partly the focus of the Stage 1 report, which established there were significant differences in the performance of providers within CPAs, as captured by differences in the outcomes of those individuals attached to them. This section establishes that there is enough variation in service provision across providers in the same CPA to ensure our estimations are indeed capable of identifying their correlations with outcomes, should these be present in the data.

We can consider the level of provider services as being composed of a CPA element and a provider-in-CPA element:

$$l_{cp} = \mu_c + \mu_{cp}$$

The interpretation of the former is that it represents the level at which providers within a given CPA tend to offer a specific form of support. This is a mean across all providers in that CPA. The interpretation of the latter is that it represents the extent to which individual providers choose to offer this type of support at a lower or higher level than this. This latter variation is captured as the residual in a regression of the level of a service characteristic l_{cp} on the full set of CPA dummies. Where service characteristics vary across contract years, we extract the residuals from a regression on a full set of CPA-by-year dummies. The F-test from the regression of l_{cp} on these residuals can be interpreted as the test for whether this residual variance is statistically significant.[5]

Table 6-1 below presents descriptive statistics on the extent of within-CPA variation in service characteristics. For example, relating the within-CPA standard deviation to the overall mean level of the service provision across CPAs (in the leftmost column) allows one to gauge the extent of this variation. Similarly, the range covered by the 10th and 90th percentiles of the distribution of the CPA-adjusted variation in service characteristics gives a sense of the breadth of this variation. Finally, the right-most column presents the p-value from the above-mentioned F-test indicating the statistical significance of this within-CPA variation. We therefore confirm that there is significant within-CPA variation in service characteristics across all the measures considered.

---

[5] Note that the function of this test is identical to that of the approach taken in Section 4.5 to choose measures calculated from the survey of Work Programme participants. For these measures, this test is partly redundant. In this Section, however, we apply this test to all service measures considered, not just those originating from the customer survey.

Table 6-1 Within-CPA variation in service characteristics

Variable	Mean	Residual variance			F test p-value
		Std Dev.	10th pctile	90th pctile	
Number of times provider in contact with customer	10.205	1.097	-1.241	1.691	0
Mode of contact: % via group meeting	0.497	0.051	-0.064	0.064	0
Mode of contact: % via telephone	0.525	0.043	-0.063	0.059	0
Mode of contact: % via text	0.203	0.065	-0.076	0.075	0
Support type: % receiving assessment of your skills	0.349	0.044	-0.057	0.04	0
Support type: % drawing up an action plan	0.457	0.049	-0.059	0.059	0
Support type: % receiving training in maths reading writing	0.131	0.038	-0.05	0.045	0
Support type: % work experience placement or voluntary work	0.136	0.031	-0.048	0.048	0
Support type: % advice for setting up own business	0.112	0.037	-0.039	0.039	0
Support type: % Financial advice	0.171	0.047	-0.059	0.059	0
% stating they were warned of risk of benefit cut	0.683	0.042	-0.048	0.047	0
Provider not 100% private sector	0.197	0.285	-0.333	0.499	0
% self-delivered - Annual (DWP)	0.496	0.202	-0.289	0.276	0
% referrals to private sector subs - Annual (DWP)	0.27	0.139	-0.161	0.183	0
% referrals to public sector subs - Annual (DWP)	0.056	0.052	-0.055	0.052	0
% referrals to voluntary sector subs - Annual (DWP)	0.178	0.106	-0.13	0.133	0
% referrals to specialist subs - Annual (DWP)	0.018	0.041	-0.023	0.028	0
% referrals to end to end subs - Annual (DWP)	0.425	0.208	-0.287	0.289	0
Herfindahl index of supply chain - Annual (DWP)	0.178	0.101	-0.149	0.148	0
Total revenue achievable from individual (£100s)	59.721	9.397	-1.807	3.443	0.04

7 Results

7.1 Format of the results

The results from the estimations come in a number of forms, which we outline in this subsection. At the most aggregate level, we estimate overall correlations between service characteristics and impacts across all participants. These are found in Table 7-1, where we present the estimated coefficient relating each of the service characteristics to the 4 types of outcomes considered. These are estimated using Equation 3 for contract level characteristics and Equation 5 for characteristics that vary by contract year. For each outcome, we present the association at 12, 18 and 24 months from referral. Note however that these are not estimated on the same underlying population, as outcomes at longer horizons are only observed for individuals referred to the Work Programme earlier in the contracts. Any comment on the relative magnitude of coefficients across these points in time would therefore rest on the assumption that the effect of service characteristics does not vary across cohorts of customers. The statistical significance of each coefficient is conveyed through colour-coding. Light shading indicates statistical significance at the 10% level; dark shading, at the 5% level.

The numbers in the table indicate the percentage point change in impacts as the service characteristic in question changes by one unit. In the case of variables that are proportions (such as the proportion of individuals reporting group meetings), this is the equivalent to a change from 0% to 100%. In practice, levels of service characteristics tend to be clustered within narrower ranges. To give a more realistic impression how impacts change, we show the impacts associated with moving from the level of service at the 25th percentile to the level of service at the 75th percentile. Such a representation has an appeal since it conveys how impacts might vary as the level of service changes within a range that is not out of keeping with what is seen in practice. These estimates are found in Table 7-2. The statistical significance of each marginal effect is conveyed through colour-coding as in the previous table.

Table 7-2 also offers a way to gauge how much of the variation in provider performance is statistically accounted for by the correlation with a given service characteristic. To that end, Table 7-2 displays the inter-quartile range in the dependent variable. This measures the difference in performance against a given outcome between providers ranking at the 25th and 75th positions in performance, after having netted out CPA and CPA-year effects. This is shown both for models that consider characteristics that are fixed at the contract level (top panel) and for models that look at characteristics that vary by contract and year (bottom panel). For example, looking at the top panel, the employment rate at 12 months of customers assigned to the provider at the 25th place performance ranking is 0.9 percentage points higher than the employment rate of customers assigned to the provider at the 75th place performance ranking. Comparing the inter-quartile marginal effects of a given service

characteristic with the range in the dependent variable provides a feel for the quantitative relevance of the correlation identified.[6] For example, the figure of 0.004 on 'Mode of contact: % by telephone' at the 12m employment outcome indicates that the employment rate of customers assigned to the provider at the 25th percentile in the extent of use of telephone as a means of contact is 0.4 percentage points higher than the employment rate of customers assigned to the provider at the 75th percentile in the extent of use of telephones. The variation in performance associated with variation in the use of telephone is therefore roughly half of the total variation attributable to provider service characteristics (0.004/0.009).

Next, we estimate how these correlations vary across subgroups by interacting each service characteristic, independently, with sex, age bands and payment groups. This analysis identifies up to 16 coefficients [7] for each service characteristic, making it cumbersome to convey results through a table. Instead, we present these results visually in the extended set of charts included in Annex C. To further simplify the presentation, we consider the heterogeneity in impact across the 5 Customer Groups rather than 9 Payment Groups. These are: (1) JSA 18-24; (2) JSA 24+; (3) Other JSA; (4) New ESA; (5) Other ESA-IB. Finally, we consider only the 18 months outcome, again with the aim of making the presentation of results manageable while still capturing the broad nature of sub-group variation.

Annex C consists of two sets of charts for each service characteristic considered. The first set plots the marginal effect of increasing the level of a given service characteristic for each sex, age and payment group. These estimates answer the question of how much the impact considered would change, for a given subgroup, as the service characteristic increases by an additional unit. In each chart, the marginal effects are enclosed in a 90% confidence interval and a red line is plotted at zero. This gives an immediate representation of whether each marginal effect is statistically different from zero. Each chart also includes the p-value for the test of the hypothesis that the marginal effects over all the levels in a given subgroup type are equal to each other. In other words, a p-value below 0.1 would suggest that there is a 10% statistically significant variation in the marginal effect across the given subgroup type (sex, age or payment groups).

Importantly, this first set of charts cannot reveal the full dynamics of underlying effects. For example, there are a number of cases where the test statistic indicates there is variation in the marginal effects across subgroups, but the visual representation of these does not unambiguously identify which two or more subgroups determine this variation. To provide additional insight into this, we include a second set of charts reporting the coefficient from all estimated pairwise differences in marginal effects within age and Customer Groups that emerge as statistically significant. These tables are read as follows. Cells report the

---

[6] Note that the inter-quartile marginal effect may exceed the inter-quartile range in the dependent variable. This is driven by dynamics outside of the range considered.

[7] 2 sexes, 5 age bands and 9 payment groups.

difference in marginal effects of the row subgroup from the column subgroup. Cells report a value only where this difference is significant at the 10% level.

It is important to note that these tests may appear not to agree with each other. For example, it is possible that, for a certain outcome, marginal effects for younger people could be statistically significant while those for older people are not, yet the hypothesis of similar marginal effect across ages cannot be rejected. Alternatively, all marginal effects across subgroups could be not statistically different from zero, yet the overall test may reject the hypothesis that all marginal effects are equal to each other. The reason for these apparent inconsistencies is that each test hinges on a distinct test statistic. The practical interpretation of the results we will find for each service characteristic will therefore necessarily rely on an all-round consideration of the results for the full battery of tests.

The results from our analysis speak directly to a number of important issues in relation to the effectiveness of support provided via the Work Programme. We discuss each of these in turn, highlighting what we see as the main findings.

Table 7-1 Aggregate correlations with impacts

Service characteristic	In employment, at:			Job Outcome, within:			Sustainment Payments, within:			Off benefits, at:		
	12m	18m	24m	12m	18m	24m	12m	18m	24m	12m	18m	24m
Mode of contact: % via group meeting	0.008	0.01	0.007	-0.006	0.005	0.025	0.002	0.011	0.017	0.02	0.014	0.014
Mode of contact: % via telephone	0.049	0.053	0.043	0.022	0.036	0.051	0.037	0.045	0.047	-0.024	-0.019	-0.021
Mode of contact: % via text	-0.021	-0.007	-0.01	-0.051	-0.049	-0.044	-0.071	-0.059	-0.043	-0.013	-0.013	0.005
Number of times provider in contact with customer	0.002	0.001	0	0.003	0.004	0.003	0.005	0.005	0.003	0	0	-0.001
Support type: % receiving assessment of your skills	-0.025	-0.009	-0.012	-0.045	-0.064	-0.062	-0.011	-0.029	-0.027	-0.03	-0.02	-0.007
Support type: % drawing up an action plan	-0.031	-0.03	-0.009	-0.081	-0.112	-0.098	-0.078	-0.11	-0.086	-0.079	-0.051	-0.052
Support type: % receiving training in maths reading writing	-0.032	-0.012	-0.027	-0.033	-0.029	-0.022	-0.016	-0.011	0	-0.021	-0.009	0.003
Support type: % work experience placement or voluntary work	0.031	0.062	0.048	0.008	-0.034	0.003	-0.031	-0.047	-0.012	-0.002	-0.002	-0.006
Support type: % advice for setting up own business	-0.005	0.017	0.009	0.016	0.008	0.023	0.068	0.073	0.06	0.067	0.058	0.087
Support type: % Financial advice	-0.004	0.017	0.002	-0.059	-0.07	-0.06	-0.037	-0.039	-0.03	0.015	0.019	0.039
% stating they were warned of	-0.034	-0.011	0.018	-0.149	-0.163	-0.15	-0.174	-0.187	-0.14	-0.062	-0.054	-0.038

risk of benefit cut												
Provider not 100% private sector	-0.001	0	-0.002	-0.004	-0.003	-0.002	-0.009	-0.008	-0.006	-0.001	0.001	-0.001
% cases dealt by private providers	-0.004	-0.008	0.057	-0.052	-0.052	-0.12	-0.061	-0.019	0.025	0.007	0.023	-0.047
Total revenue achievable from individual (£100s)	0	0	0	0	0	0	0	0	0	0	0	0
% self-delivered - Annual (DWP)	0.011	-0.018	0.032	-0.025	-0.035	-0.068	0.032	0.009	0.068	-0.027	-0.015	-0.029
% referrals to private sector subs - Annual (DWP)	-0.001	0.022	-0.006	0.029	0.045	0.051	-0.061	-0.009	-0.058	0.049	0.045	0.033
% referrals to public sector subs - Annual (DWP)	-0.109	-0.068	-0.095	0.041	0.017	0.14	0.114	0.061	0.162	0.096	-0.002	-0.043
% referrals to voluntary sector subs - Annual (DWP)	0.01	0.04	-0.057	0.002	0.009	0.057	-0.022	-0.035	-0.166	-0.052	-0.052	0.041
% referrals to specialist subs - Annual (DWP)	-0.099	-0.12	-0.067	-0.035	-0.012	-0.054	-0.011	-0.113	-0.128	-0.122	0.029	0.212
% referrals to end to end subs - Annual (DWP)	-0.057	-0.005	-0.053	0	-0.007	0.022	-0.046	-0.036	-0.053	0.027	0.038	0.013
Herfindahl index of supply chain - Annual (DWP)	0.006	-0.015	0.027	-0.031	-0.043	-0.074	-0.007	-0.02	0.034	-0.02	0.007	-0.026

Significant at the 10% level	Significant at the 5% level

Table 7-2 Inter-quartile marginal effects

Service characteristic	In employment, at:			Job Outcome, within:			Sustainment Payments, within:			Off benefits, at:		
	12m	18m	24m	12m	18m	24m	12m	18m	24m	12m	18m	24m
Mode of contact: % via group meeting	0.001	0.001	0.001	-0.001	0	0.002	0	0.001	0.001	0.002	0.001	0.001
Mode of contact: % via telephone	0.004	0.004	0.003	0.002	0.003	0.004	0.003	0.004	0.004	-0.002	-0.001	-0.002
Mode of contact: % via text	-0.003	-0.001	-0.001	-0.007	-0.007	-0.006	-0.01	-0.008	-0.006	-0.002	-0.002	0.001
Number of times provider in contact with customer	0.004	0.003	0.001	0.009	0.01	0.01	0.014	0.012	0.009	0.001	0	-0.002
Support type: % receiving assessment of your skills	-0.002	-0.001	-0.001	-0.004	-0.006	-0.006	-0.001	-0.003	-0.003	-0.003	-0.002	-0.001
Support type: % drawing up an action plan	-0.004	-0.004	-0.001	-0.01	-0.013	-0.012	-0.009	-0.013	-0.01	-0.009	-0.006	-0.006
Support type: % receiving training in maths reading writing	-0.002	-0.001	-0.002	-0.002	-0.002	-0.001	-0.001	-0.001	0	-0.001	-0.001	0
Support type: % work experience placement or voluntary work	0.001	0.003	0.002	0	-0.001	0	-0.001	-0.002	-0.001	0	0	0
Support type: % advice for setting up own business	0	0.001	0.001	0.001	0.001	0.002	0.005	0.005	0.004	0.005	0.004	0.006
Support type: % Financial advice	0	0.001	0	-0.004	-0.005	-0.005	-0.003	-0.003	-0.002	0.001	0.001	0.003
% stating they were warned of risk of benefit cut	-0.003	-0.001	0.001	-0.012	-0.013	-0.012	-0.014	-0.015	-0.011	-0.005	-0.004	-0.003
Interquartile variance in dependent variable	*0.009*	*0.007*	*0.006*	*0.014*	*0.017*	*0.017*	*0.015*	*0.015*	*0.014*	*0.015*	*0.01*	*0.012*

Total revenue achievable from individual (£100s)	0	0	0	0	0.001	0.002	0	0.001	0	0.002	0.001	0.001
% self-delivered - Annual (DWP)	0.003	-0.005	0.009	-0.007	-0.01	-0.019	0.009	0.002	0.019	-0.008	-0.004	-0.008
% referrals to private sector subs - Annual (DWP)	0	0.005	-0.001	0.006	0.01	0.011	-0.013	-0.002	-0.013	0.011	0.01	0.007
% referrals to public sector subs - Annual (DWP)	-0.009	-0.006	-0.008	0.003	0.001	0.012	0.009	0.005	0.013	0.008	0	-0.004
% referrals to voluntary sector subs - Annual (DWP)	0.002	0.008	-0.011	0	0.002	0.011	-0.004	-0.007	-0.032	-0.01	-0.01	0.008
% referrals to specialist subs - Annual (DWP)	-0.001	-0.001	-0.001	0	0	-0.001	0	-0.001	-0.001	-0.001	0	0.002
% referrals to end to end subs - Annual (DWP)	-0.017	-0.001	-0.016	0	-0.002	0.007	-0.014	-0.011	-0.016	0.008	0.011	0.004
Herfindahl index of supply chain - Annual (DWP)	0.001	-0.002	0.004	-0.004	-0.006	-0.01	-0.001	-0.003	0.004	-0.003	0.001	-0.003
Interquartile variance in dependent variable	*0.004*	*0.003*	*0.003*	*0.004*	*0.004*	*0.004*	*0.004*	*0.003*	*0.005*	*0.005*	*0.003*	*0.003*

Significant at the 10% level	Significant at the 5% level

7.2 The nature of support

Under this heading, we consider the frequency of contact, the mode of contact and the types of support offered. The latter are broadly defined. For example, we include among these whether providers inform claimants of the risk of losing all or part of their benefit in the case of non-compliance with the programme.

The frequency of contact appears to have some positive association with outcomes. We estimate generally positive, and in some cases statistically significant, correlation with employment outcomes. Furthermore, despite the small absolute magnitude of the coefficients, we estimate some significant and in some cases substantial inter-quartile marginal effects. For instance, the incidence of Sustainment Payments at the 12 month point among individuals with providers who are at the 75th percentile are 1.4 percentage points higher than among those with providers who are at the 25th percentile. This indicates that the frequency of contact may be a particularly influential determinant for the effectiveness of service provisions, particularly for sustaining outcomes over the time. The subgroup analysis suggests that more frequent contact may be more important for females than males for Sustainment Payments, and for Customer Group 1 (JSA 18-24) when considering Job Outcomes or Sustainment Payments.

We find some evidence that the mode of contact may matter for outcomes. Contacting claimants by telephone is associated with better employment outcomes overall. This is also seen in relatively substantial inter-quartile marginal effects. Effects on employment outcomes, Job Outcomes and Sustainment Payments appear to be stronger for younger claimants.

The possible effect of other forms of contact appears to vary by claimant type. Despite no overall effect, we find that contact via group meetings is correlated with employment and Sustainment Payments for the 'Other ESA-IB' customer group. The plotted marginal effects indicate that contact by text may be less effective for this group, for JSA 18-24 and for older claimants in general. However, the pairwise comparisons do not support this result.

As one would expect, the nature of the support offered is found to be associated with outcomes, though this is not always true. For example, offering work placements is uncorrelated with outcomes, whether overall or for specific subgroups.

Among the types of support considered, we find the strongest evidence of a positive association between help starting a business and being off-benefit, including sizeable inter-quartile marginal effects. This is particularly the case for prime-age JSA claimants, while the association turns negative for Customer Group 5 (Other ESA-IB).

On the other hand, we find some evidence suggesting that help drawing up an action plan is associated with fewer Job Outcomes. While this is only marginally significant, there is more definite subgroup variation. It appears to be more effective for older individuals and for non-JSA claimants.

Similarly, we find some indication that skills assessment is correlated with fewer Job Outcomes, on average. The subgroup analysis, however, reveals a more complex pattern. Again, the negative association is driven by JSA claimants rather than ESA claimants. There are also several statistically significant pairwise comparisons across age groups, but they do not reveal a clear overall pattern.

English and maths training and provision of financial advice are not associated with outcomes overall. However, the subgroup analysis indicates that English and maths training may be somewhat less effective for older claimants and more effective for the Other ESA-IB Customer Group. Financial advice is associated with higher employment among the JSA 18-24 group.

Lastly, we find evidence of negative correlation between claimants stating they were warned that their benefits may be reduced if they fail to comply with Work Programme requirements and some outcomes. Specifically, we find a strong negative correlation with Job Outcomes and Sustainment Payments. The inter-quartile marginal effects are also sizeable. However, we find no association with employment or benefit receipt. It is important to bear in mind that providers are required to issue such warnings as a normal part of the operation of the Work Programme, whenever a claimant is mandated to an activity. This variable does not therefore reflect application of sanctions.

There are a few possible reasons that may explain the apparent inconsistency in the correlation between claimants stating they were warned that their benefits may be reduced if they fail to comply with Work Programme requirements and the different outcomes. Firstly, the Job Outcomes and Sustainment Payment measures capture whether there have been any such events over a specified number of months, while the employment and off-benefit measures capture an individual's circumstances in that same end-month. Additionally, while employment and off-benefit measures are calculated from administrative data, Job Outcomes and Sustainment Payment outcomes are reported by providers. It is therefore possible that customers of providers whose claimants report a higher rate of warnings have similar rates of employment as customers of providers with lower reported warning rates, but are more likely to disengage from providers and thereby reduce providers' ability to evidence job outcome achievement. Looking at the subgroup analysis, the negative associations with Sustainment Payments are particularly strong among JSA claimants. For New ESA claimants, being informed of the consequences of non-compliance is associated with a greater reduction in their probability of being off benefit than for other types of claimant.

7.3 The nature of the supply chain

Sectoral composition

The second main area this analysis can speak to is in relation to the nature of the supply chain. In particular, we can explore its sectoral composition and the level of concentration (a proxy for competition).

Hansmann (1980) argues that the absence of the profit motive implies that non-profit firms will deliver higher standards of quality in markets where this quality is hard to observe. In the context of the Work Programme, this could be interpreted as saying that both non-profits and for-profits will maximise impact on easily verifiable outcomes such as Job Outcome and Sustainment Payments, but non-profits may perform better on outcomes that are less under scrutiny, such as benefit receipt, employment beyond Sustainment Payments milestones, or customer earnings in work. To the extent that we are able, we explore this possible relationship by testing the association of outcomes with the sector of the prime and the percentage of referrals to subcontractors outside the private sector.

Overall, the results provide little evidence that the sectoral composition of the supply chain is systematically associated with negative or positive outcomes. The sector of the prime, the percentage of customers handled by the prime, and the percentage referred to subcontractors in the private, public or voluntary sector are not correlated with impacts overall. Additionally, none of the individual marginal effects for these characteristics is statistically different from zero for any of the subgroups. However, while keeping that in mind, the pairwise tests offer some insights.

We find some evidence that the sector of the prime has different associations with impacts by Customer Group and to a lesser extent by age. The pairwise tests suggest that prime age, particularly JSA, claimants may do better than other groups when assigned to providers that are not 100% for-profit.

While there is no evidence that the proportion of the caseload delivered by the prime is associated with impacts, there does appear to be an inverse-U-shaped variation in the impact of self-delivery over age. Claimants aged between 35 and 44 handled by a prime with a higher rate of self-delivery seem to achieve more Job Outcomes and Sustainment Payments relative to other age groups. On the other hand, claimants aged 45 and over do better than other age groups on the same outcomes when referrals to a private sector Tier 1 provider are higher. Individuals in Customer Group 5 (Other ESA-IB) achieve lower employment outcomes when referrals to a public sector Tier 1 provider are higher.

As a robustness check (not shown), we constructed an additional variable measuring the percentage of cases handled by a private sector provider. This is composed of all self-delivered cases if the prime is private, plus any referrals to private subcontractors. Using

this simpler description of the sectoral composition of the supply chain, we still do not find any clear association with outcomes.

Concentration of the supply chain

We explore the relationship between outcomes and the concentration of the supply chain by using the Herfindahl index of the supply chain over the course of a contract year. We might expect that higher competition would introduce efficiencies and therefore improve outcomes. On the other hand, economies of scale and lower transaction costs may favour having fewer, larger subs. Finally, it could also be that the level of competition in the supply chain simply determines the allocation of rents between primes and subcontractors, but has no effect on outcomes.

The results are most consistent with the latter hypothesis. We find no overall correlation between outcomes and the concentration of the supply chain. We also find very little evidence of variation by subgroup. The most notable exception to this is that we find that, where the supply chain is more concentrated, JSA 25+ claimants may achieve higher proportions off-benefit compared to other groups and that those aged under 24 achieve lower proportions off-benefit compared to other age groups.

The role of specialist providers

We also consider the relative use of specialist vs. end-to-end providers. Like with the sectoral composition of the supply chain, we find no association with outcomes overall, and none of the individual subgroup marginal effects is statistically different from zero. However, the pairwise tests reveal some patterns. There is some evidence that older claimants achieve higher Job Outcome and Sustainment payments if assigned to a prime that makes greater use of end-to-end providers. On the other hand, a large number of pairwise test results indicate that the prime's use of specialist providers is associated with levels of Job Outcomes and Sustainment Payment that increase with age. The use of specialist providers is also associated with lower employment outcomes among the JSA 18-24 Customer Group and higher proportions off-benefit among the New ESA Customer Group.

The role of payment levels

Finally, our analysis speaks to the very relevant question of whether higher payments to providers are associated with higher outcomes. The importance of this is obvious, given pressure on programmes to deliver value for money. At the same time, all else equal, higher discounts may cause providers to reduce the intensity of the service to maintain profit margins, thereby likely delivering lower outcomes. Alternatively, it could be that providers vary in their productivity and, at the contracting stage, see the quantity to be provided (for example, number of Job Outcomes) as fixed or at least largely outside of their control, and prefer to compete on prices. Under this scenario, the contracting mechanism would allow DWP to price-discriminate and extract all or most of the producer surplus.

Under this hypothesis, outcomes across providers would not be correlated with total revenue payable for each individual.

Our results would appear to be consistent with this latter hypothesis. We find no association of outcomes with total potential revenue from each individual, and we find no evidence of subgroup marginal effects that are statistically different from zero. The pairwise tests also confirm that there is little evidence of the associations with outcomes differing by subgroup. This is notable considering that it is precisely the Job Outcome fee that providers were invited to discount. The interpretation of these results may be that rather than payments acting as an incentive to deliver results, Job Outcome discounts are more likely to be a contracting mechanism to ensure the contract is assigned to the most efficient provider.

8 Conclusion

This report has attempted to provide an indication of what drives higher performance among Work Programme providers. The random assignment of individuals to one of the providers operating within their CPA offers the ideal circumstances to estimate the relative effectiveness of providers. Furthermore, the large number of individuals randomised means that the precision of these estimates is maximised.

Notwithstanding the large number of individuals randomised, when estimating variation across providers in their effectiveness, only 40 observations are available (one observation for each contract, or CPA-provider combination). Failure to take this point into consideration is likely to lead to erroneous results, dramatically overstating the strength of observed correlations. We have taken care to avoid such misleading inference and adopted a 2-stage approach that, we argue, represents best practice.

Following this approach makes the structure of the estimation problem much more evident, and the constraint of small sample size readily apparent. As discussed in Brewer et al. (2013), with samples of this size there may be an issue of statistical power; in other words, smaller correlations may go undetected.

Other than the issue of small sample size, the essential point to bear in mind is that relating impacts to service characteristics is a non-experimental exercise; randomisation provides the experimental provider impact estimates but does not help beyond that. In view of this, the estimated results do not support a causal interpretation but should instead be viewed as correlations. Observing a significant correlation alerts us to the possibility that there may be a causal relationship at play. It does not formally confirm it, but may add to the weight of evidence in support of a particular type of service. The corollary is that a non-significant correlation suggests that there may be no significant causal relationship. Again, this is not formally a sufficient condition but it may be felt to be strongly indicative.

With these caveats in mind, the analysis has provided some interesting findings, mostly in the sense that they suggest overall provider effectiveness is largely unconnected with the kinds of variations in service delivery observed in practice. That is not to imply that providers do not matter. The results from the Stage 1 report confirmed they do. However, the drivers of these differentials may be outside of what is currently observable and measureable. They may also relate to intangible characteristics of the provider, such as staff know-how, which are clearly difficult to include in quantitative analysis.

However, the analysis identifies some potentially interesting results. The intensity of the service, as captured by the number of meetings, appears particularly important in practical terms. The mode of interaction may also play a role, with telephone contact associated with better employment outcomes, and group meetings appear to be most suitable for ESA claimants. This should be seen in the context of existing related evidence. In particular,

Middlemas (2005) provides experimental evidence that 'signing on' by telephone rather than in person reduces benefit off-flow rates among JSA claimants. While this may appear at odds with the findings in this report, we note that this earlier finding relates only to JSA claimants during the first 13 weeks of their claim rather than individuals who have been claiming for much longer and may have been claiming benefits other than JSA.

One type of support that does appear promising is providing advice for setting up in business; this occurs against the backdrop of a slow but steady rise in self-employment in the UK since the 2008 recession. Other than that, there is no strong indication that the actual content or nature of the services provided are consistently associated with stronger impacts. There are, though, numerous differences across subgroups. Again, this can help direct certain types of support to those more likely to benefit from them.

The share of claimants stating they were informed of the consequences of non-compliance on the amount of benefit received is associated with considerably lower rates of Job Outcome and Sustainment Payments, but not employment or benefit rates. However, it is important to be cautious about over-interpreting this result. As we have already stressed, the results are not necessarily causal.

At the more conceptual level, there is little evidence to support the view that the sectoral composition of the supply chain or the degree of competition among subcontractors has a bearing on effectiveness. This is an important result, and one that has received very little attention in the empirical welfare-to-work literature. The results also suggest that the discounts offered by providers at the time of bidding to DWP are not associated with reduced performance.

We put forward a possible explanation for these results that relates to the structure of the market and contracting process, and has implications on the extent to which rents are appropriated by DWP, primes and subcontractors. These hypotheses are sufficiently fundamental as to warrant further detailed exploration. However, this is beyond the scope of this report.

As a final comment, while the introduction of the Work Programme did not offer any ready way of evaluating its success, the within-CPA randomisation was a low-cost design feature that worked extremely well to provide estimates of relative provider effectiveness. This continues the approach used with the Flexible New Deal and there is every reason to repeat this with the successor to the Work Programme. In this way, we can begin to amass sufficient observations on the correlates of programme effectiveness to alleviate concerns about possible low power and also to entertain the application of more sophisticated modelling techniques in an attempt to address the endogeneity of service characteristics that prevents us from being more definite about the causal relationships that exist in the data. It would be especially helpful from this perspective to increase the number of primes within each CPA. In the current case, two primes within a CPA give a single experimental estimate of relative effectiveness, while three primes give three such estimates. Adding

additional primes within a CPA would increase the number of estimates by the total number of primes, less one. For instance, having four primes rather than three would increase the number of estimates from three to six. Having five primes would increase it to 10. However, evaluation objectives are not the only consideration; there may be practical and efficiency arguments for keeping the number of primes per CPA relatively low.

9 References

Brewer, Mike, Thomas F. Crossley, and Robert Joyce. 2013. Inference with Difference-in-Differences Revisited. IZA Discussion Paper 7742. Institute for the Study of Labor (IZA). https://ideas.repec.org/p/iza/izadps/dp7742.html.

Bryan, M. and Jenkins, S. (2013) Regression analysis of country effects using multilevel data: a cautionary tale IZA discussion paper 7583

Cameron, A. Colin, and Douglas L. Miller. 2013. A Practitioner's Guide to Cluster-Robust Inference.

Card, D., Chetty, R. and Weber, A. (2007) The Spike at Benefit Exhaustion: Leaving the Unemployment System or Starting a New Job? American Economic Review 97: 113-118.

Dorsett, R. and Robins, P. (2013) A Multi-Level Analysis of the Impacts of Services Provided By the UK Employment Retention and Advancement Demonstration. Evaluation Review 37(2): 63-108.

Donald, S. and Lang, K. (2007) Inference with difference-in-differences and other panel data Review of Economics and Statistics 89(2): 221-233.

Hansmann, Henry. 1980. "The Role of Nonprofit Enterprise." The Yale Law Journal 89 (5): 835. doi:10.2307/796089.

Liang, Kung-Yee, and Scott L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." Biometrika 73 (1): 13–22. doi:10.1093/biomet/73.1.13.

Middlemas, J. (2005) "Jobseeker's Allowance intervention pilots quantitative evaluation" DWP Research Report 382.

Annex A - Descriptive Statistics

Customer group	JSA 18-24	JSA 25+	Other JSA	New ESA	Other ESA-IB	Total
Frequency	287,720	701,485	369,029	107,551	91,711	1,557,496
Percent	18.47	45.04	23.69	6.91	5.89	100

Age of claimant	JSA 18-24	JSA 25+	Other JSA	New ESA	Other ESA-IB	Total
	%	%	%	%	%	%
<=24	100	0	35.3	11.7	5.6	28
25/34	0	33.6	20.8	19.8	15.9	22.4
35/44	0	30.1	18.1	27.7	26.3	21.3
45/54	0	26	17.9	28.7	34.6	19.9
55+	0	10.3	7.9	12.2	17.6	8.4
Total	100	100	100	100	100	100

Gender of claimant	JSA 18-24	JSA 25+	Other JSA	New ESA	Other ESA-IB	Total
	%	%	%	%	%	%
Female	32.7	35.9	25.4	47.9	47.5	34.3
Male	67.3	64.1	74.6	52.1	52.5	65.7
Total	100	100	100	100	100	100

Primary health condition - Summary	JSA 18-24	JSA 25+	Other JSA	New ESA	Other ESA-IB	Total
	%	%	%	%	%	%
Claimants without any diagnosis code on the system	100	99.9	99.8	1.3	2.5	87.4
Mental and Behavioural Disorders	0	0	0.1	49.1	53.4	6.6
Diseases of the Nervous System	0	0	0	3.1	4.5	0.5
Diseases of the Circulatory or Respiratory System	0	0	0	4.8	4.1	0.6
Diseases of the Musculoskeletal system and Connective Tissue	0	0	0	13.5	14.8	1.8
Injury, Poisoning and certain other consequences of external causes	0	0	0	8.3	4.3	0.8
Other	0	0	0	20	16.4	2.4
Total	100	100	100	100	100	100

Disadvantaged due to ex-off, drugs or alcohol	JSA 18-24	JSA 25+	Other JSA	New ESA	Other ESA-IB	Total
	%	%	%	%	%	%
None (default)	97.3	97	84.5	95.6	96.7	94
Ex-Offender	2.3	1.8	13.2	1	0.5	4.5
Misuser of Drugs	0.1	0.4	0.4	0.7	0.6	0.4
Ex-Offender and Misuser of Drugs	0.1	0.2	0.8	0.5	0.2	0.4
Misuser of Alcohol	0.1	0.4	0.4	1.3	1.2	0.4
Ex-Offender and Misuser of Alcohol	0	0.1	0.3	0.2	0.2	0.1
Misuser of Drugs and Misuser of Alcohol	0	0.1	0.1	0.4	0.4	0.1
Ex-Offender, Misuser of Drugs and Misuser of Alcohol	0	0.1	0.3	0.3	0.2	0.1
Total	100	100	100	100	100	100

Disability Indicator	JSA 18-24	JSA 25+	Other JSA	New ESA	Other ESA-IB	Total
	%	%	%	%	%	%
Not Disabled	85.7	74.4	65.6	19.5	13.5	67
Disabled	14.3	25.6	34.4	80.5	86.5	33
Total	100	100	100	100	100	100

Ethnicity of the claimant (high level)	JSA 18-24	JSA 25+	Other JSA	New ESA	Other ESA-IB	Total
	%	%	%	%	%	%
White	79.6	75.1	81.9	83	80.9	78.4
Mixed	2.6	2	2	1.4	1.3	2
Asian	5.9	6.2	4.5	4.7	4.9	5.6
Black	5.5	9.2	5.7	3.8	4.6	7
Other	1.8	3.2	1.4	2.1	2.7	2.4
Missing	4.7	4.4	4.5	5	5.5	4.6
Total	100	100	100	100	100	100

Highest qualification level	JSA 18-24	JSA 25+	Other JSA	New ESA	Other ESA-IB	Total
	%	%	%	%	%	%
No Qualifications	6.7	10.8	13.5	7.8	5.1	10.2
Entry Level	3	2.8	4.3	1.8	1	3
Level 1	10.7	6.6	10.4	4.5	2.2	7.9
Level 2	16.4	11	13.4	7.5	3.7	11.9
Level 3	6.8	4.5	4.4	2.9	1.4	4.6
Level 4	1.2	1.8	1.2	0.8	0.5	1.4
Level 5	0.5	1	0.6	0.4	0.3	0.8
Level 6	0.7	1.3	0.5	0.4	0.3	0.9
Level 7	0.1	0.4	0.1	0.1	0.1	0.2
Level 8	0	0.2	0.1	0	0	0.1
Not Known	2.4	2.5	4.3	1.4	0.7	2.7
Missing	51.4	57.1	47	72.3	84.9	56.4
Total	100	100	100	100	100	100

1 digit SOC	JSA 18-24	JSA 25+	Other JSA	New ESA	Other ESA-IB	Total
	%	%	%	%	%	%
Managers, directors and senior officials	0.5	3.2	1.3	1.7	1.3	2
Professional occupations	0.9	2.5	1.1	1.4	1.3	1.7
Associate professional and technical occupations	4.3	5.2	3.3	3.3	2.9	4.3
Administrative and secretarial occupations	9.1	9.6	6.8	6.3	5.6	8.4
Skilled trades occupations	6.9	9.2	9.9	7	5.1	8.6
Caring, leisure and other service occupations	8.6	7.3	5.9	6.9	5.1	7.1
Sales and customer service occupations	31.9	16.5	19.8	10.8	7.6	19.2
Process, plant and machine operatives	4.2	11	10.1	7.2	4.8	8.9
Elementary occupations	33.4	35.4	40.4	20	14.7	33.9
Missing	0.1	0.1	1.3	35.5	51.5	5.8
Total	100	100	100	100	100	100

Annex B - Summary of Phase 1 results

The aim of this research project is to understand how provider characteristics or practices are associated with levels of performance. As a precondition to carrying out this work, Stage 1 of the project focussed on ascertaining that:

a) The randomisation process assigning customers to providers within each CPA delivered a balanced profile in observable characteristics across providers.
b) There is a statistically significant difference in provider effectiveness over and above that which one might expect by pure chance.

We conduct these statistical tests by estimating a series of regression equations. Formally, we can modify Equation 2 in Section 5.3 to combine the effects of all provider service characteristics, $l'_{cp}\pi$, and of unobserved provider service characteristics, v_{cp}, into a single provider indicator, P_{icp}:

$$y_{icp} = \alpha_c + \delta_{cp}P_{icp} + a_c \cdot x'_{icp}\beta + \varphi_c + \varepsilon_{icp}$$

To avoid perfect multi-collinearity between the set of provider indicators in each CPA, we first randomly define one provider in each CPA to be the 'reference' provider. The other provider, or providers, in each CPA are each identified as 'treatment' providers. This assignment is purely arbitrary, does not affect the results and involves no loss of generality. The set of indicators, P_{icp}, is therefore only included for 'treatment' providers.[8]

Importantly, because individuals within a CPA are assigned at random to providers, P_{icp} is independent of ε_{icp} and so δ_{cp} has a causal interpretation as the mean effect on individuals in CPA c of being assigned to provider p rather than the reference provider. Each two-prime CPA provides one estimate of relative effectiveness. Each three-prime CPA provides two estimates of relative effectiveness.[9] In total, this gives us 22 distinct experimental estimates of relative effectiveness for each outcome.

With a background characteristic as y_{icp}, the joint significance of the δ_{cp} terms across all treatment providers and CPAs provides a test of whether that characteristic is balanced in the sense of being similar across providers within a CPA. Failure to reject the null hypothesis of no significant coefficients would indicate that background characteristics are statistically similar across providers in the same CPA, while a rejection of the null hypothesis would indicate differences in the customer profiles across providers in the same CPA. Analogously, with y_{icp} as an outcome, the null hypothesis is of no significant variation in effectiveness. Now, rejection of the null hypothesis

---

[8] Following this operation, $\alpha_c + \varphi_c$ captures both the CPA level effect and the effect of the reference provider. The two cannot be separately identified.

[9] In 3-provider areas, 3 pair-wise comparisons can be made. However, the third estimate can be derived from the other two: A-C= (A-B) - (C-B). Therefore, only two are independent.

indicates that there is significant variation in within-CPA provider effectiveness, meaning that the question of which provider an individual is assigned to has an important bearing on their outcomes

The regressions are estimated using a linear probability model (OLS) with robust standard errors. The dependent variable is dummy taking the value 1 if an individual exhibits a given characteristic or achieves a given outcome and 0 otherwise. When testing for variation in outcomes, the model allows for the effect of individual background characteristics, X_{icp}. These are omitted when testing for balance. This implies that our tests for variation in outcomes are more stringent.

We run the tests over a set of 45 customer characteristics. These are: 6 age bands; gender; disability; being an ex-offender or with a history of alcohol or substance abuse; being a lone parent; being on benefit 12 months before referral; 7 indicators of mental or physical health conditions; 5 Work Programme Customer Groups; 6 ethnic groups; 7 indicators for educational qualification levels; and 9 indicators of previous occupational classification. We do this for the whole sample as well as by high level-Payment Group aggregates (JSA 18-24, JSA 25+, Other JSA, New ESA, Other ESA/IB). This delivers a large number of estimates. We therefore do not present these in full detail, but only discuss the overall patterns across the results.

The first thing to note is that we find more evidence of imbalance over observable characteristics than we should expect purely by chance, if all tests were statistically independent. Given our lowest threshold of statistical significance is 10%, we might expect 4 or 5 of the F-tests on the 45 balancing variables to come up as statistically significant by chance. Instead, 14 of the tests emerge as statistically significant in the pooled sample; 8 for the JSA25+ group; 9 for New ESA customers; and 13 for Other ESA/IB customers. Only 1 and 4 tests come up as significant for JSA18-24 and Other JSA respectively. This suggests that there might be differences in observable customer characteristics across providers that, from a statistical point of view, depart from zero. However, it is also relevant to repeat that the expected number of significant results is on the assumption of independence; it may be that the higher number found reflects a correlation between outcomes.

However, in socio-economic terms the magnitude of the coefficients is very small. For example, the typical difference in the share of lone parents in the whole sample of customers across any two providers is 0.003 or, equivalently, 0.3 percentage points. Therefore, while statistically significant, it is nevertheless materially very small. The largest typical difference estimated related to the share of customers with No Qualifications and is 0.6 percentage points.

The fundamental point at stake is whether these balancing test results should alter our view on the ability of our approach to provide causal estimates of the impact of providers on outcomes. This would be the case if we suspected that differences in outcomes across providers could be driven by these small differences in the profile of customers at different providers. Our view is that this cannot be the case as, while there is evidence of statistically significant imbalance with regard to (say) sex, the estimated impacts presented in the next section are sometimes 5 or even 10 times larger than the observed imbalances, suggesting that any such differences would have to be implausibly influential on outcomes in order to suggest that the impact is not attributable to the

provider. We therefore conclude that there is evidence of sufficient balancing to proceed with our proposed approach.

The F-tests for provider variation in outcomes generally reject the hypothesis that the provider dummies are jointly irrelevant to the outcome levels achieved. Furthermore, the mean absolute deviation between pair-wise provider comparisons is several times larger than was typically the case for background characteristics. The mean absolute values of the differences in the outcomes across providers are found to be in the order of 1-2 percentage points. Providers therefore appear to have a measurable effect on outcomes.

To overcome the difficulty interpreting such a large number of estimates, we make use of visualisations, which can perhaps give a more immediate impression of what the data are telling us. Specifically, in Figure 1 we display the 26 point estimates and 95% confidence intervals for a selection of characteristics (top) and outcomes (bottom). Importantly, all charts use the same x-axis scale, allowing an immediate comparison of the relative magnitudes.

Such visual exploration reveals two important messages. Firstly, despite the statistical significance of some of the coefficients on pair-wise comparisons across background characteristics, their magnitude almost always essentially zero or negligible. On the other hand, the variation in provider effectiveness is much more pronounced, and in the majority of cases statistically significant. On this basis we interpret the results as confirming that background characteristics are generally balanced and that variation in provider effectiveness is present, statistically significant and of material magnitude.
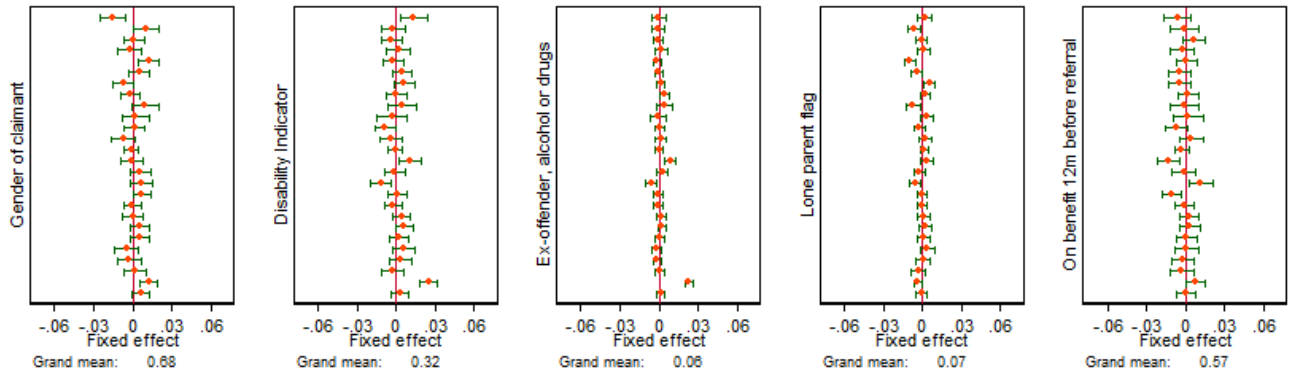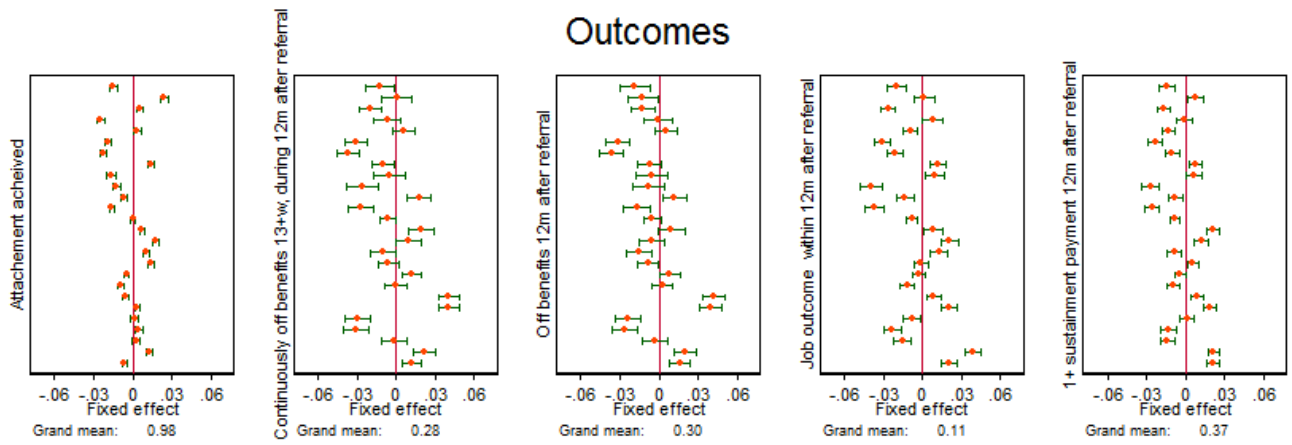
Figure 1 - Visualisation of balancing tests (top row) and impacts (bottom row)