Department for Work and Pensions

Working Paper No 77

Non-participation in the Employment Retention and Advancement Study: Implications on the experimental first year impact estimates

Barbara Sianesi

A report of research carried out by Institute for Fiscal Studies on behalf of the Department for Work and Pensions

First Published 2010.

ISBN 978 1 84712 704 4

Contents

List of tables

List of figures

List of boxes

Acknowledgements

The analyses in this working paper were done on behalf of the Department for Work and Pensions (DWP).

I would like to acknowledge the many helpful suggestions from Costas Meghir, as well as to thank colleagues from the ERA Evaluation Team at the DWP and consortium colleagues: Electra Small for her hard data work and Ingun Borg, Mike Daly, Richard Dorsett, Lesley Hoggart, Phil Robins and Jim Riccio for detailed comments on previous drafts.

The Author

Dr Barbara Sianesi is a senior research economist in the education, employment and evaluation research sector at the Institute for Fiscal Studies. She joined IFS as a PhD scholar in 1998. Her current research focuses on evaluation methods, applied in particular to labour market programmes and policies as well as to educational investments.

Abbreviations

ASA	Advancement Support Adviser
BIF	Basic Information Form
DWP	Department for Work and Pensions
ERA	Employment Retention and Advancement
HMRC	Her Majesty's Revenue & Customs
IB	Incapacity Benefit
IS	Income Support
JSA	Jobseeker's Allowance
ND25Plus	New Deal 25Plus
NDLP	New Deal for Lone Parents
OLS	Ordinary Least Squares
RA	Random assignment
WTC	Working Tax Credit
WPLS	Work and Pensions Longitudinal Study

Summary

Overall summary

The Employment Retention and Advancement (ERA) programme, a research demonstration project which ran in six UK regions between October 2003 and October 2007, has been evaluated by random assignment. The aim of this report is to explore how the findings from the experimental research relate to the impacts that would have been experienced, on average, by all the people that were eligible for the programme, had they participated in the demonstration.

Overall, our findings largely validate the experimental results of the main ERA evaluation. Specifically, for the New Deal for Lone Parents (NDLP) group the first-year experimental impacts appear to be representative of the impacts that the full eligible NDLP population would have experienced under ERA. For the New Deal 25Plus (ND25Plus) group, the experimental impact findings are found to actually under-estimate the gains that all ND25Plus eligibles would have enjoyed under ERA.

Background

Carefully planned and administered randomised experiments arguably offer the most reliable method for evaluating whether a programme works, on average, for its participants. Since eligible individuals are allocated randomly between a programme group receiving the services and a control group not receiving them, any systematic difference between the two groups in later outcomes can safely be attributed to the programme. Such an experimental approach is currently being used to assess the effectiveness of ERA, a programme which was operational in six Jobcentre Plus districts across the UK between October 2003 and October 2007. Eligible for this new set of support and financial incentives to secure, retain and progress in work were those who were mandated to participate in the

ND25Plus programme and those who had volunteered for the NDLP programme.[1] With over 16,000 individuals being randomly assigned over one year, the ERA study represented at its inception the largest randomised evaluation of a social programme in the UK.

The issue

All individuals flowing into ND25Plus and NDLP in the six evaluation districts during the one-year intake window should automatically have become eligible for the package of support offered by ERA. It has, however, emerged that only parts of the target population actually entered the evaluation sample: some people who were eligible actively refused to be randomly assigned and to take part in the experimental evaluation (the 'formal refusers'), while some eligibles were somehow not offered the possibility to participate in random assignment and hence in ERA (the 'diverted customers'). A sizeable fraction of the eligible population – 23 per cent of ND25Plus and 30 per cent of NDLP – were thus not represented in the experiment.

Research objectives

Overall, this report sets the foundation work for the analysis of non-participation in the ERA study. It aims to:

1 explain the subtle issues that non-participation raises for the ERA demonstration;

2 introduce the different approaches and methodologies to deal with it; and

3 present the intermediate findings (based on 12-month follow-up data) and lessons so far.

Specifically, the report aims to answer the following research questions:

- What kind of impact would the non-participants have experienced, on average, had they been offered ERA services and incentives?
- What would the impact of the ERA study have been on its full intended population?
- How does this estimated impact for all eligibles compare to the experimental impact estimate obtained for the ERA study participants?

1 This analysis focuses on the two main ERA target groups, representing 83 per cent of all ERA study participants. The third group – lone parents working part-time and in receipt of Working Tax Credit (WTC) who had volunteered for ERA – is not considered in this report due to its conceptually different set-up coupled with lack of data.

- The report also sheds light on the issue of whether the non-participants are individuals who even if offered ERA services would not take them up. We do this by asking: What type of involvement would the non-participants have had with ERA and more generally with Jobcentre Plus had they participated in the study?

Issues posed by non-participation for the experimental analysis

The policymaker would arguably be interested in assessing the impact of offering ERA services and incentives for all those eligible to receive such an offer. The experimental evaluation on the other hand provides, under suitable assumptions, unbiased impact estimates only for the ERA study participants – those eligibles who have reached the randomisation stage and have agreed to participate in the experimental evaluation. The concern is that this subgroup may potentially be a selective one. This report, therefore, focuses on the full eligible population in the ERA districts over the study intake window and on the causal effect for all such eligibles of making the ERA package available. This **average effect of the ERA offer for all eligibles in the six districts** is the same type of parameter recovered by the experimental study (the effect of offering ERA in the six districts), but it is averaged over all eligibles, rather than over a potentially adviser-selected and self-selected subgroup of the eligibles.

A related way to appreciate the importance of this group and hence the meaning of this parameter, as well as to envisage more fully how ERA as an official policy could work, is to think of ERA as an integral component of the New Deal programme, specifically as a seamless next stage in which any New Dealer would automatically be enrolled upon having found work. In other words, the customer would make no decisions about ERA per se when enrolling in the New Deal, but would automatically be offered the ERA package once having entered full-time work.[2] A scenario in which all New Deal entrants are automatically 'opted in' for ERA gives direct and high policy relevance to the full New Deal sample, the focus of this report.

The report assesses whether a non-participation rate of 26.6 per cent is likely to have affected the extent to which the experimental results can be generalised to the full eligible population, and hence their representativeness and policy relevance.[3]

---

2 A customer in work could of course always opt out of ERA – both formally if there were such a proviso and de facto as they could not be forced (or sanctioned) into taking up the ERA package.

3 Technically, this relates to the extent to external validity of the experimental findings, or equivalently, to the scope for non-participation bias in the experimental estimate in terms of the impact on all eligibles.

The ERA study offers the rare chance to look at this issue because: (i) the treatment is the offer of ERA support and incentives, (ii) eligible for this offer under an official policy would be the whole population of ND25Plus and NDLP entrants in the six districts, and (iii) such entrants are identified in the available administrative data.

Types of non-experimental analyses

We perform different types of non-experimental analyses seeking to recover the potential impact of ERA on the full eligible population (in the six districts) and compare it to the experimental impact estimate for the ERA study participants. In most cases, identifying and estimating the average impact on all eligibles first requires identifying and estimating the average impact that the non-participants would have experienced had they been offered ERA and agreed to participate in the study.

We separately consider how to deal with non-participation when follow-up information on the outcomes of the non-participants is available (administrative data) or not available (survey data such as earnings). Non-response to the survey and/or to the earnings question among survey respondents can create additional issues when trying to recover the earnings effect of ERA for the full eligible population. An interesting feature of our data is that it allows us to test some conditions under which non-response can be safely ignored.

The analyses are performed under alternative assumptions on the participation process:

- bounding the impacts of interest without making any assumption on the selection process;
- impact estimates under the assumption that we observe all outcome-relevant characteristics that drive selection into the ERA study ('matching and reweighting approach'); and
- impact estimates that allow selection into the ERA study to depend on unobservables ('control function approach').

The specific nature of our set-up and data – randomisation coupled with administrative outcome data for the non-participants – allows us to actually test for the presence of residual selection on some type of unobservable, test to some extent the validity of the instrument needed for the control function model, as well as test two other features of the performance of the control function model.

Based on extensive diagnostic and specification tests, as well as on contrasting and cross-checking the findings and evidence from the different methodological approaches, we have found that the control function approach has produced extremely sensitive, unstable and imprecise estimates, which have to be viewed with extreme care, and as indicative at most. By contrast the most robust findings were those arising from the matching and reweighting estimates. The picture that emerges within the latter framework based on selection on observed characteristics is summarised in the following key findings.

Key findings

- In terms of **selective differences**, we have found that the ND25Plus experimental sample (i.e. those who took part in the ERA study) is composed of individuals with better employment outcomes than the whole population of ND25Plus entrants. By contrast, the experimental NDLP group is made up of somewhat lower performers than the average NDLP entrant. Once we net out the contribution of observable individual characteristics, we find that in the absence of ERA the study participants of both customer groups experience better employment outcomes than non-participants while relying more extensively on benefits. **Non-participants are thus characterised by unobservables that make them more detached from the labour market as well as from the government support system than participants.**
- In terms of **employment outcomes**, the story appears to be quite different for the two customer groups.
  - For the **NDLP group**, the overall experimental impact estimate excluding the non-participants coincides with the average impact ERA would have had on the full population of eligibles. Specifically, no impact was found for the experimental sample either on employment durations or on the probability of being employed during the follow-up year, and the absence of any impact extends to the non-participants, and hence, to all eligibles.
  - By contrast, ignoring the participation decision significantly biases the effect of ERA for all eligibles for the **ND25Plus group**, in the sense that the average effect for all eligibles is statistically different from (and larger than) the experimental estimate. This result is driven by the fact that the effect for the non-participants is considerably larger than the one for the participants. Compared to no effect on employment probabilities for the participants, we estimate that non-participants would have enjoyed an increase of almost six percentage points, resulting in a statistically significant 2.6 percentage points increase for all eligibles. Similarly, compared to an increase in employment durations of 4.5 days for the experimental sample, the non-participants would have enjoyed a ten day increase, yielding a 5.8 day increase for all eligibles. These findings might thus indicate that for the more labour-market detached ND25Plus entrants (i.e. the non-participants) some extra help in the form of advice and financial incentives might be particularly helpful in improving their labour market situation.

- In terms of **benefit outcomes**, the story surprisingly appears to be the same for the two customer groups. First, while ERA has not significantly affected time on benefits for participants, non-participants would have experienced a significant nine days' **increase** had they been offered ERA services and incentives.[4] Second, impact estimates for all eligibles are statistically significantly different from the experimental estimate. Third, these estimates do not, however, tell a **qualitatively** different story, as the impact of ERA on days on benefits for either the experimental sample or the full group of eligibles is not statistically or economically significant. For both the ND25Plus and NDLP eligibles, the point estimates become literally zero.

- The findings on **earnings** impacts have to be taken with extra care, as they rely on the strongest set of assumptions. For both customer groups the impact for the responding study participants appears to be representative of the effect for the full eligible population, in the sense that formal tests fail to uncover any statistically significant difference. However, for the ND25Plus group the qualitative evidence is that the impact on earnings for the responding experimental group (an increase of £393 significant only at the ten per cent level) actually **underestimates** by almost 50 per cent the average ERA impact for the full eligible population (a highly significant increase of £580).

- Finally, we have assessed the conjecture that if ERA became an official policy, non-participants would be mostly uninterested in taking up its support and incentives. We have found no support for this hypothesis for either customer group. In fact, the results show that overall, **the non-participants display observed characteristics that make them quite likely to be involved with ERA and with Jobcentre Plus more generally**. Specifically, had they been randomised into the programme, the non-participants would have been less aware of ERA or less involved with Jobcentre Plus than the programme group only in terms of a couple of measures, and then only marginally. Indeed, had they become eligible to ERA services and incentives, the NDLP non-participants would have been over three percentage points more likely than the programme group to be involved in training or education activities arranged by Jobcentre Plus, as well as **more** likely to be directed to a Jobclub or Programme Centre. Had they been randomised into the control group, NDLP non-participants would have been four percentage points **more** likely than the actual control group to rate advice from Jobcentre Plus staff as very helpful.

[4] The most likely implication of the finding that ERA would have increased both employment durations and time on benefits for the ND25Plus non-participants is that ERA would have reduced the time these customers spend in 'uncompensated' non-employment, i.e. outside the labour market as well as the government support system. Note in any case that time in employment and time on benefits are not mutually exclusive (individuals can be employed at the same time as claiming a benefit such as Income Support (IS)); this is particularly the case with the available administrative data (the Work and Pensions Longitudinal Study (WPLS)), which contains no information on the amount of hours worked.

Conclusions, lessons learnt and further work

How has the presence of the non-participants affected the external validity of the experimental impact estimate?

Overall, we have found that the policy bottom-line 'story' on ERA does not change much – either in statistical or qualitative terms. The tentative conclusion so far (tentative given the uncertainties that are intrinsic to any type of non-experimental analysis) is thus that the external validity of the experimental impact estimate overall is reasonably high. This is the case, especially for the NDLP group, for whom the first-year impact results appear to generalise to the full eligible population. For the ND25Plus group, the external validity of the ERA study is somewhat lower, with the experimental impact findings representing a *lower* bound to the gains all eligibles would have enjoyed had they been offered ERA services and incentives.

- For the **NDLP group**, the story remains unchanged. Specifically, the bottom-line in the first-year follow-up is that ERA has had no effect on employment and benefit outcomes, while it has significantly and substantially increased yearly earnings. This report specifically shows that what the programme has done for the participants, it would have done also for the non-participants and hence for all eligibles. Interestingly, this overall conclusion applies within districts as well.

- For the **ND25Plus group**, the story changes somewhat in the direction of a slightly more effective ERA treatment if the whole eligible population had taken part: positive impacts surface, become larger in size or stronger in statistical significance (while the only negative and large experimental impact – the one on earnings for participants in Wales – decreases both in statistical significance and in magnitude). We thus **do** find evidence of non-participation bias (or of some loss in external validity) in the data for the ND25Plus group. For this group, the employment and earnings impact estimates that rely on experimental data alone **underestimate** the likely impact that ERA would have had on all ND25Plus entrants, both overall and in several districts. Of course, there is always the issue of how different the estimates for the eligibles and for the experimental sample need to be for us to view the issue as a particularly important one. Randomised experiments are however conceptually designed to provide with accuracy the 'true' answer to the evaluation question. Hence, an effect for the eligibles which is 30 or 50 per cent larger (or 15 per cent smaller) than the experimental estimate can be viewed as a finding of substance.

It will be interesting to examine the issue of non-participation in terms of longer term follow-up outcomes; indeed, future work will include updating the findings in this report to outcomes at five years after random assignment. Given the documented differences in characteristics and outcomes, the participants and non-participants might experience ERA impacts that evolve – persist, emerge or fade – differentially. Furthermore, it will be of special importance to account for the issue of survey and/or item non-response for longer-term outcomes.

1 Background, research questions and overview

1.1 Background

Carefully planned and administered randomised social experiments arguably represent the most reliable method for evaluating whether a programme works, on average, for its participants. Since eligible individuals are allocated randomly between a programme group receiving the services and a control group not receiving them, under reasonable assumptions any systematic difference in later outcomes observed between the two groups can be attributed to the programme.

While experimental studies have played an important role in the design of US welfare and training programmes, they have not been widely used in the UK. A recent exception is the Employment Retention and Advancement (ERA) demonstration, which ran in six Jobcentre Plus districts across the UK between October 2003 and October 2007. Eligible for this new set of support and financial incentives to secure, retain and progress in work were those who were mandated to participate in New Deal 25Plus (ND25Plus) and those who had volunteered for New Deal for Lone Parents (NDLP).[5] With over 16,000 individuals being randomly assigned in six districts over one year, the ERA study represented at its inception the largest randomised controlled trial of a social programme in the UK.

5 This analysis focuses on the two main ERA target groups, representing 83 per cent of all ERA study participants. The third group – lone parents working part-time and in receipt of Working Tax Credit (WTC) who have volunteered for ERA – is not considered in this report due to its conceptually different set-up coupled with lack of data.

Since ERA offered of a package of support once in work,[6] all individuals flowing into ND25Plus and NDLP in the six evaluation districts during the one-year intake window should automatically have become eligible to be offered the ERA package. It has, however, emerged that only parts of the target population have entered the evaluation sample: some eligibles actively refused to be randomly assigned and to take part in the experimental evaluation (the 'formal refusers'), while some were somehow not offered the possibility to participate in random assignment and hence, in ERA (the 'diverted customers'). A sizeable fraction of the eligibles – 23 per cent of ND25Plus and 30 per cent of NDLP[7] – were thus not represented in the experiment.

1.2 Research questions

This report sets the foundation work for the analysis of non-participation in the ERA study. It aims at explaining the various and subtle issues that non-participation raises for the ERA demonstration, introducing the different approaches and methodologies to deal with it and presenting the intermediate findings and lessons (i.e. based on 12-month follow-up data; the findings in this report will be updated at the end of the ERA project based on five-year post-random assignment data).

1.2.1 What kind of issues does non-participation pose for the experimental analysis?

The policymaker would arguably be interested in assessing the average impact of offering ERA services and incentives for all those eligible to receive such an offer. The experimental evaluation on the other hand provides, under suitable assumptions, unbiased impact estimates only for the ERA study participants – those customers who reached the randomisation stage and agreed to participate in the demonstration. The concern is that this subgroup may potentially be a selective one, not representative of the full eligible population in the ERA districts who would have been eligible for ERA had it been an official national policy. This report, by contrast, directly focuses on the full population of eligibles and on the causal effect for them of making the ERA package available. This **average effect of the offer of ERA for all eligibles in the six districts** over the study intake

6 Eligible customers have access to in-work emergency payments to overcome short-term barriers to staying in work, those working are further entitled to employment-related assistance from an Advancement Support Adviser (ASA) and qualify for a training bonus, and, provided they work for least 30 hours a week, for a work retention bonus as well.

7 The composition of the non-participants varied markedly between the two customer groups. Formal refusers represent the majority (59 per cent) of ND25Plus non-participants customers, while accounting for barely 13 per cent of the NDLP non-participants On the other hand, diverted customers represent over one-quarter (26.4 per cent) of all NDLP eligibles compared to nine per cent of all ND25Plus eligibles.

window is the same type of parameter recovered by the experimental study (the effect of offering ERA in the six districts), but averaged over all the eligibles, rather than over a potentially adviser-selected and self-selected subgroup of the eligibles.

A related way to appreciate the importance of this group and hence, the meaning of this parameter, as well as to envisage more fully how ERA as a normal policy could work is to think of ERA as an integral, seamless component of the New Deal programme in which any New Dealer would automatically be enrolled upon entering work.[8] A scenario in which all New Deal entrants are automatically 'opted in' for ERA gives direct and high policy relevance to the full New Deal sample, the focus of this report.

The non-participation problem raises the question of the extent to which the conclusions from the experimental study would hold for the whole population of eligibles. Technically, this is the issue of 'external validity' of the experimental impact estimates: how legitimate would it be to generalise these results to the full eligible population?[9]

The beauty of the ERA study is that it offers the rare chance to actually measure the loss in external validity. This is because (1) the treatment is the offer of ERA support and incentives, (2) eligible for this offer under an official policy would be the whole population of ND25Plus and NDLP entrants in the six districts, and (3) such entrants are identified in the available administrative data.

Our previous descriptive report (Goodman and Sianesi, 2007) has explored how representative the group is for whom we can calculate experimental estimates by understanding both how large and how selective the group of non-participants is. Overall, the non-participation problem seems to be a relevant one.

The overall aim of this report is thus to build on these descriptive findings to assess whether a non-participation rate of 26.6 per cent is likely to have affected the extent to which the experimental results can be generalised to the full eligible population, and hence their representativeness and policy relevance.

1.3 Overview

We perform different types of non-experimental analyses seeking to recover the impact of ERA on the full eligible population (in the six districts) and compare it to the experimental impact estimates for the ERA study participants. In most cases, identifying and estimating the average impact on all eligibles requires first

8 A customer in work could of course always opt out of ERA – both formally if there were such a proviso and de facto as they could not be forced (or sanctioned) into taking up the ERA package.

9 Alternatively, non-participation can be viewed as introducing potential bias in the experimental estimate if interest lies in the impact of ERA on the eligibles (in the six districts).

identifying and estimating the average ERA impact that the non-participants would have experienced. These analyses are performed under alternative assumptions on the participation process.

In each case we consider how to deal with non-participation when follow-up information on the outcomes of the non-participants is available (administrative outcome measures) or not available (survey-based outcome measures). Clearly, the latter case will be less informative, and we will have to make more stringent assumptions. Furthermore, non-random non-response to the survey and non-random item non-response among survey respondents potentially create additional issues when trying to recover the effect of ERA on the full eligible population. An interesting feature of our data is that it allows us to test some conditions under which non-response can be safely ignored.

We start by considering analyses that provide bounds for the impact of interest without any assumption on the selection process.

We then move on to providing impact estimates under the assumption that we observe all outcome-relevant characteristics that drive selection into the ERA study. Characteristics that are observed in the data include an individual's demographics as well as information on their current unemployment spell, detailed labour market histories and local factors. This type of analysis is related to matching and reweighting techniques, and when considering survey outcomes we also specifically allow for survey and item non-response. Furthermore, within this framework we estimate the type of involvement that the non-participants would have had with ERA and more generally with Jobcentre Plus had they participated in the evaluation study. This allows us to shed some light on the question of whether the non-participants are indeed individuals who even if offered ERA services would not take them up.

We then consider approaches that allow for selection into the ERA study based on unobservables, i.e. on outcome-relevant characteristics that are not recorded in the available data. In addition to the standard examples of an individual's motivation, ambition, social contacts and health status, the data at our disposal contains no direct information on educational attainment, which is thus among our most important 'unobservables'. These types of analyses follow a so-called control function approach and rely on an exclusion restriction, that is, a variable that affects participation in the ERA study but not outcomes directly. We start with the standard sample selection model, but then extend it in various directions: we relax independence between the observed characteristics and the unobservables; we relax the normality assumption; and we allow for censoring in the outcome variable (both days in employment and earnings are censored at zero).

All of these models build on the standard Heckman selection model. We are however in the rather unique position where for one set of outcomes (the administrative ones), we do observe the outcomes of the selected-out sample – the non-participants. Coupled with randomisation, this feature of the data allows us to test for the presence and extent of residual selection on some type of unobservable. We further exploit it to test the validity of the instrument, as well as two other features of the performance of the model. Specifically, we are in a position to choose between different specifications of the control function model based on two 'metrics': how well the various models capture the presence and direction of the residual selection we have uncovered, and how well the various models predict the (no-treatment) outcome of the non-participants.

The following gives a more detailed overview of how the remainder of the report is organised.

- We start in **Chapter 2** by outlining how non-participation in the ERA evaluation has come about, before focusing on placing the experimental and non-participation analyses into proper context. The section concludes with a summary of the available evidence on non-participants.
- **Chapter 3** briefly describes the data and our working definition of ERA eligibility. It also provides sample breakdowns by customer group and district and describes the rich set of variables we have collated from different sources to capture key characteristics relating to the individuals themselves, their office and their local area.
- Our methodological approaches and the type of analyses we perform are presented in **Chapter 4**. The description is kept as non-technical as its rather technical nature allows us. (A supplementary technical appendix contains an in-depth and formal derivation of all the methods).
  - We start in **Section 4.1** by formally presenting our analytical framework, together with conditions for the experimental impact estimate to coincide with the average impact for the full eligible population. We also briefly overview the type of analyses carried out in the report before turning to the issues raised by survey and item non-response when estimating ERA impacts on survey outcomes such as earnings. In particular, we derive conditions – some of which are testable given the nature of our data – for non-response to be safely ignored.
  - Bounds which make no assumption on the selection process into the ERA study are discussed in **Section 4.2**, first in the case of administrative outcomes, then in the case of survey outcomes. We also sketch some sensitivity analysis to assess how robust the estimate of the average treatment effect for all eligibles is to assumptions about the selection process.
  - **Section 4.3** deals with methods relying on the selection-on-observables assumption. We start by briefly relating the available data to the plausibility this assumption. We then outline our approach to estimate the impact on all eligibles on administrative outcomes and suggest simple sensitivity analyses to assess how sensitive the estimates are to straightforward violations of this crucial assumption. For survey outcomes, we derive estimators that ignore non-response and ones that allow for non-response. This section also outlines an analysis to assess the take-up of services and the contact with Jobcentre Plus staff that non-participants would have had, had they been offered ERA.

  - Section 4.4 is devoted to selection on unobservables. We first present tests on whether there are outcome-relevant unobservable differences between ERA study participants and non-participants, as well as tests for the validity of the instrument we plan to use in our control function models. We then move on to consider our different selection models.

- The results of all our empirical analyses are presented and discussed in Chapter 5.
  - Section 5.1 starts by presenting the experimental findings concerning the average impact of ERA for the participants as well as the results of the tests on survey and item non-response.
  - Section 5.2 reports the findings from our bounds and sensitivity analyses.
  - Section 5.3 focuses on those arising from our different models based on the selection-on-observables assumption. This section also includes the results of our analysis of take-up of ERA services and involvement with Jobcentre Plus.
  - Section 5.4 is devoted to presenting and discussing the tests and estimates relating to the different control function models.
- Chapter 6 summarises the key results and briefly concludes.
- The Appendices provide additional material: Appendix A presents the results and summary boxes of the district-level analyses, while Appendices B and C contain intermediate diagnostic and estimation results. A formal and thorough derivation of all estimation methods as well as of the conditions for their validity is contained in a separate Supplementary Technical Appendix.

2 Non-participation in the ERA study: The issues

2.1 How did non-participation come about

In an ideal scenario, all individuals in the six evaluation districts who would take part in Employment Retention and Advancement (ERA) if it were an official policy would have been randomly assigned to either the programme group or the control group. Departures from this ideal situation have arisen from two sources:

1. intake process: not all eligible individuals may have been offered the possibility to participate in random assignment and hence in ERA (the 'diverted customers'); and
2. individual consent: some individuals who were offered the chance to take part in the experimental evaluation actively refused to do so (the 'formal refusers').

Taken together, diverted customers and formal refusers make up the group of the 'ERA non-participants', that is those individuals who while being **eligible** for ERA, for some reason or another have not been included in the experimental sample and have thus not participated in the evaluation.

The 'ERA study participants' are the group of individuals who were eligible for ERA, were offered the chance to participate in the study and agreed to take part in it. These are those making up the evaluation sample, i.e. those who were subsequently randomly assigned either to the programme group, who would receive ERA services and incentives, or to the control group, who would instead receive the baseline New Deal treatment.

2.2 Experimental analysis and non-participation analysis

Many causal parameters of interest are the effect of some 'treatment' averaged over the relevant population. To place the experimental and non-participation analyses into proper context, it is important to clarify both the type of 'treatment' being evaluated as well as the relevant population(s) over which to average its effect.

In order to do this, consider how the decision to participate in a programme can be broken down into a series of steps. The stage at which randomisation is applied determines what can be learnt from an experiment, in other words, the causal parameter it retrieves. In the case of ERA, an individual needs to:

1 satisfy the criteria for ERA eligibility

Starting the New Deal 25Plus (ND25Plus) or New Deal for Lone Parents (NDLP) programmes during the sample intake window in the six evaluation districts would make one eligible for ERA.

2 become aware of ERA and realise own eligibility

In the demonstration, information about ERA and one's eligibility to its services and support came predominantly from Jobcentre Plus staff. However, this information somehow did not reach a non-negligible share of the eligibles (9.4 per cent of ND25Plus eligibles and 26.4 per cent of NDLP eligibles).

3 apply for ERA if application is necessary

As an official policy, one might envisage that:

a) ERA would become an integral component of the New Deal programme in which any New Dealer would automatically be enrolled; in this case, no formal application process would be necessary. Alternatively;

b) New Dealers would need to make their eligibility operational by registering into the ERA programme to be allocated to an advisor and the like.

In the experimental evaluation, there was no formal application process, but study participants had to give their consent to taking part in the research and being randomly assigned.[10]

[10] Consenters would sign that *'I understand that if I sign this form I agree to take part in the study. I understand that I am free to pull out of the study at any time.'* Formal refusers would by contrast sign that *'I do not consent to taking part in this research scheme or to being randomly assigned.'*

4 decide on the take-up of services and entitlement

ERA was a voluntary programme in the sense that it was up to individual customers to decide whether and to what extent to avail themselves of the ERA elements. Specifically, eligible customers became entitled to employment-related assistance from a dedicated adviser and those working at least 30 hours qualified for a work retention bonus as well as for a training bonus should they also undertake training. However, it always remained up to them to decide whether they wanted to avail themselves of such a support package or not. For instance, around 15 per cent of the programme group in either New Deal customer group reported that they had had no contact **at all** with Jobcentre Plus staff during the 12-month period following their randomisation into the treatment group. Furthermore, some programme group members may simply not have been aware of or have forgotten some of the ERA features, as testified by around one-quarter of either New Deal programme group who had not heard of the employment bonus and as many as half or more (49 per cent for NDLP and 57 per cent in the ND25Plus) who were not aware of the training bonus one year into the study.

The experimental estimator of the impact of ERA was applied to stage (3), i.e. unconditional on the take-up of services. In the presence of take-up decisions (stage (4)), it provides an estimate of the mean impact of the **offer** of treatment, not of the mean impact of the treatment itself.

For many purposes, this is the policy-relevant parameter, as it is informative on how the **availability** of ERA services and incentives affect individual outcomes, where it is implicitly acknowledged that non-take-up is a normal feature of any ongoing programme.

Furthermore, the ERA treatment itself represents an **offer** of support and incentives. The experimental estimator is thus perfectly suited to recover the effect of **offering** ERA services and incentives. As mentioned, it is unconditional on the actual take-up of the services or even actual knowledge of the services and incentive structure.

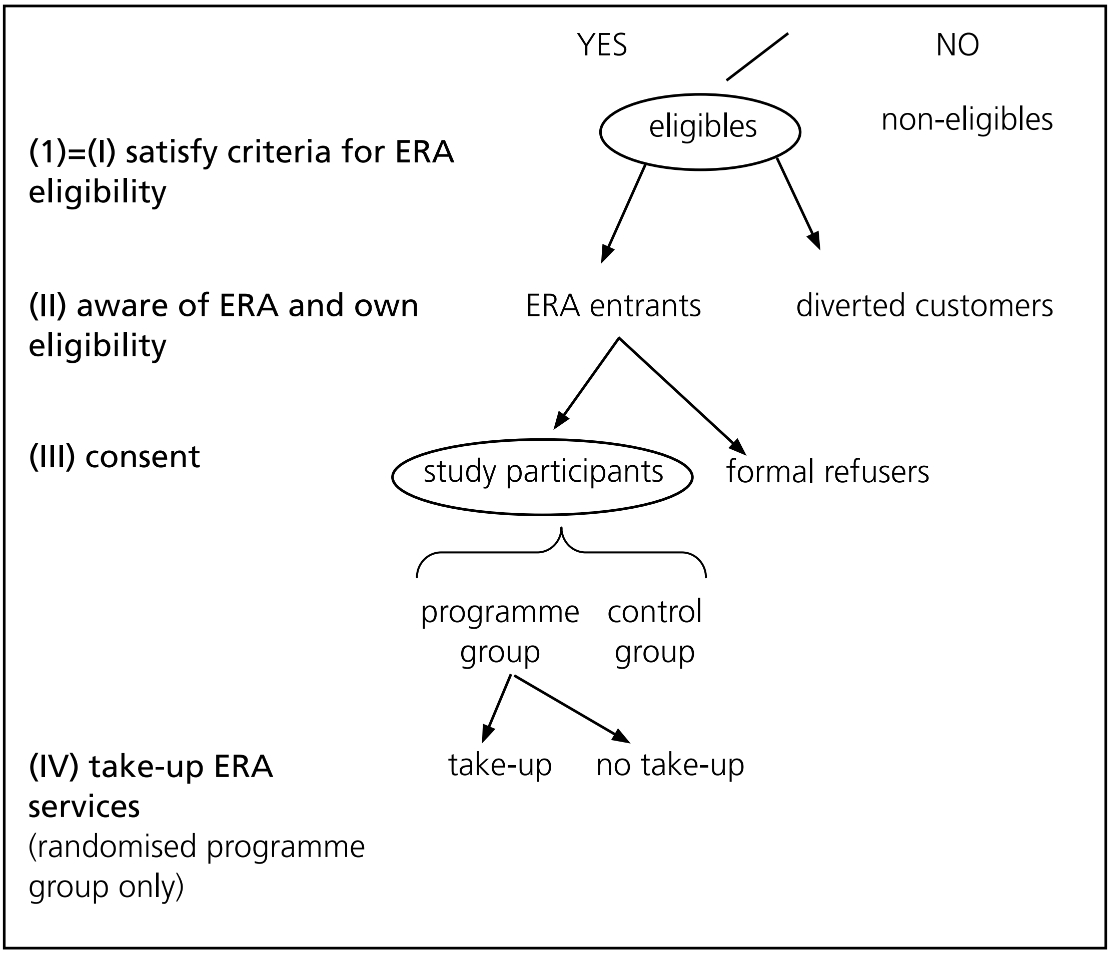
Thus coming back to the first issue we set out to clarify, i.e. the type of 'treatment' being evaluated, both the experimental evaluation and the present non-experimental analysis share the same type of 'treatment': **being offered** the ERA package of support, or, equivalently, **becoming eligible** to the ERA package of support.

Now, let us consider the population of interest over which to average the impact of being offered ERA. One might envisage that the policymaker would be interested in assessing the impact of offering ERA at the eligibility level (stage 1), as well as at the level of application/registration (stage 3).[11]

[11] Yet another interesting – though hard to identify – parameter would be the mean effect of actual receipt of ERA support and services for those who effectively took it up (stage (4)).

To better understand where the experimental parameter fits into this discussion, the following diagram shows the structure of selection into the ERA study group, where the Roman numbeals highlight the fact that the second, third and fourth stages and related populations do not necessarily correspond to those under official-policy running of ERA, where, by 'official-policy running', we mean a situation in which ERA would be available to **all** ND25 Plus and NDLP entrants in the six districts (either as part of a national policy or if the programme had been piloted in the six districts according to a pilot compared to comparison area-based evaluation scheme). In the demonstration, only parts of the target population entered the evaluation sample: some eligibles actively refused to be randomly assigned and to take part in the experimental evaluation (formal refusers), while some were somehow not even offered the possibility to participate in random assignment and hence, in ERA (diverted customers). The experimental estimate is thus conditional on being actually given the chance to participate in the study and on having formally consented to do so and to be randomly assigned, providing an estimate of the impact of ERA eligibility for those who have reached the randomisation stage and have agreed to participate in the study – the ERA **study participants** (stage (III)).

Figure 2.1 From eligibility to service receipt in the ERA demonstration



The problem is that the subgroup of ERA study participants is potentially a selective one, i.e. not necessarily representative of the full eligible population (stage (1)) nor of the subgroup who would apply for ERA if ERA were an official policy which required individuals to actively apply for it (stage (3-b)).

As to the latter point, it seems hard to believe that **all** those who have refused to take part in the experiment and all those who were not even offered such a possibility would not have been interested in registering for ERA had it been an official policy. Quite to the contrary, one could argue that if ERA had been an official policy, a non-negligible share of the current non-participants would have been aware of the programme and consented to taking part in it, i.e. would have applied for ERA if required, as in stage (3-b).

Consider first the diverted customers, eligible customers who were not told about their chance to participate in ERA. As with any government scheme, there is always the issue of how much individuals know about a policy and their eligibility for it. However, under official-policy running of ERA, Jobcentre Plus staff would not be the only source of information. Enhanced eligible individuals' knowledge of ERA would correspondingly reduce advisers' discretion as to how to market, present and sell ERA – including not mentioning it at all.

As to the formal refusers, it is not fully clear how much they actually knew about what they were refusing – according to observations at intake interviews and interviews with customers after those sessions, not much.[12] If ERA were an official policy, there would be no need to severely restrict information on the actual extent of ERA support to prevent disappointment among the control group[13] (nor in fact would there be a need to perform randomisation[14]). It is highly plausible that even under full information some refusers, especially among the ND25Plus group, would still have been reluctant to prolong contact with Jobcentre Plus, all the more likely if they did not intend to be especially pro-active in looking for

---

12 Walker *et al.* (2006) conclude that 'very few customers could be described as understanding ERA, and all of them had already been assigned to the programme group and therefore had been given further details about the services available after random assignment'. More generally, 'there was a consensus among the Technical Advisers who conducted both the observations and the interviews with customers […] that most customers truly did not have a good appreciation of ERA'. (p.43).

13 This was relaxed over time, although Walker *et al.* (2006, p.22) conclude that 'when invited to participate in ERA, customers would generally have known only that some form of extra help was potentially available if they found work and that they had a 50-50 chance of receiving it'.

14 Formally, formal refusers were signing that they did 'not consent in taking part in this research scheme or to being randomly assigned'.

work[15]. With complete information, however, the ERA package would seem very appealing, making it hard to envisage that *all* the formal refusers would have knowingly still refused to become eligible for monetary incentives, training and support once in employment. In conclusion, if ERA had been an official policy of type (3-b), there still would have been some eligibles who would have formally refused to apply for it, but it is reasonable to presume that this group would have been much smaller than the group of formal refusers actually observed in the ERA study.

Based on the above discussion for both the diverted customers and formal refusers, it is thus highly likely that **a large proportion** of the non-participants actually observed in the ERA demonstration would have participated in ERA had it been an official policy of type (3-b), so that the full eligible population might represent a closer proxy than the experimental study group of the population that would participate in ERA were it an official policy that requires eligibles to apply for it. Furthermore, if ERA had been an official policy superimposed by default on the New Deals (i.e. of type (3-a)), the full eligible population would by construction coincide with the group of participants of interest.

One might wonder whether the observed non-participants in the ERA study would actually have not availed themselves of ERA services and incentives (stage 4) even if they had joined the programme. Indeed, under certain assumptions we can assess whether the non-participants are individuals who even if offered ERA services would not take them up. We estimate the type of involvement that the non-participants would have had with ERA and more generally with Jobcentre Plus had they participated in the evaluation study in Section 5.3.3.

The interest of the current report in the full population of eligibles does not, however, hinge on conjectures about what the participation in ERA would have been if ERA had been an official policy requiring individuals to actively apply for it (type 3-b). A policymaker can only make the ERA support package available, but cannot force the eligibles to apply for it or to take up its services. Hence, the causal effect for the eligibles of making such a package available – unconditional on application (if required) and service take-up – is a parameter of paramount policy relevance. Specifically, this report considers the mean effect of ERA offer/availability for all ERA eligibles in the six districts, irrespective of how well informed they are about ERA, of whether they realise their eligibility or not, of whether they apply or not, and of whether they take up its services or not. As mentioned, this is the same type of parameter recovered by the experimental study (the effect of offering ERA), but it is averaged over all the eligibles, rather than over an adviser-selected and self-selected subgroup of the eligibles.

---

15 Although there would be nothing to lose to become formally eligible by registering into the programme (as one can then always decide to refuse to take up its services), the qualitative analysis has highlighted that especially among ND25Plus entrants there is often a tendency to resist any involvement with Jobcentre Plus beyond what is minimally necessary.

A related way to appreciate the importance of this group and hence the meaning of this parameter, as well as to envisage more fully how ERA as an official policy could work is to think of ERA as a type (3-a) policy, that is, as a seamless next stage of the New Deal programmes in which **any** New Dealer would automatically be enrolled. In other words, the customer would make no decisions about ERA per se when enrolling in the New Deal, but would automatically be offered the ERA package once having entered full-time work.[16] A scenario in which all New Deal entrants are automatically 'opted in' for ERA gives direct and high policy relevance to the full New Deal sample, the focus of this report. Indeed, this is how ERA worked for the ERA study participants, who were enrolled into the ERA programme at the time of entering their respective New Deal programme.

The ERA experiment was carefully planned and designed: assignment to the control or to the programme group has taken place after the customer had agreed to participate, and randomisation has been shown to have balanced very well the study participants between a programme group and a control group that are statistically equivalent. The experiment can thus produce highly reliable estimates of the effect of ERA for the ERA study participants (technically, it has high **internal validity** in recovering the effect of ERA for the participants).

The question this report focuses on relates to how the effect for the ERA study participants relates (or generalises) to a wider population. We move beyond the experimental sample to consider what the impact of ERA would have been on its full intended population, how it compares to the impact estimated for the ERA study participants, and what kind of impact the non-participants would have experienced, on average, had they become eligible to ERA. The problems we investigate in this report are thus circumscribed to an issue of **external validity**, or the inference that can be validly drawn from the experimental setup for the population of eligibles (in the six evaluation districts).

An alternative view to consider this issue defines the parameter of interest as the effect for the ERA eligibles (in the six districts) and assesses the scope for bias in the experimental estimate for this parameter. Does a non-participation rate of 26.6 per cent bias the experimental estimate for the treatment effect of interest?

It is important to note that this report is concerned with the **current** experimental evaluation, i.e. it considers the eligibles within the six ERA districts over the study intake window. There is in fact the wider generalisability question that has a national rollout in mind and which relates to how the experimental results obtained in the six evaluation districts would generalise to all other districts in which ERA has not been tested. This complex and very speculative type of analysis would need to address the issue of how the six districts currently offering ERA compare to those not offering ERA in terms of composition of New Deal entrants and of

---

16 A customer in work could of course always opt out of ERA – both formally if there were such a proviso and *de facto* as they could not be forced (or sanctioned) into taking up the ERA package.

local labour market conditions. Ideally, it would also try to take into account entry effects (e.g. more lone parents volunteering for the 'NDLP plus ERA' package, or some long-term unemployed delaying their job entry to become eligible for the 'ND25Plus plus ERA' package), as well as general equilibrium effects.

To summarise the discussion in this section (see also Table 2.1):

- Interest lies in the effect of **offering** ERA services and incentives.
- One can consider the average impact of the ERA offer at various stages of participation, in particular:
  - for all eligibles (the focus of the present report);
  - for the ERA study participants (the focus of the experimental evaluation);
  - for those who would apply for ERA if it were an official policy requiring formal application (neither the experimental study group, nor the full eligible population, though the full eligible population is arguably a closer proxy);
  - for all eligibles if as an official policy ERA were an integral component of the New Deal programmes (the focus of the present report).
- This report is concerned with the external validity of the experimental impact estimate: what one can infer from the ERA study participants for the full eligible population. Alternatively, this report assesses the scope of 'non-participation bias' in the experimental estimate for the average impact on all eligibles.

Table 2.1 Some causal effects of interest

Mean impact of offering ERA for		
• the ERA eligibles	stage 1	current report
• the ERA study participants	stage III	experimental estimate
• those who would apply for ERA if it were an official policy requiring formal application	stage 3	arguably much closer to eligibles than to ERA study participants
• those eligibles for ERA if as an official policy it were an integral component of the New Deal programmes	stage 1	current report

2.3 What do we know about non-participation in the ERA study

Recruitment to ERA greatly differed between the two New Deal customer groups. While lone parents on NDLP were all volunteers to that programme and consequently mostly responded favourably to ERA too, ND25Plus participants, who had just been mandated to start their New Deal programme, were more difficult to recruit and resulted in far higher refusal rates. In-depth qualitative research (Hall *et al.*, 2005, and Walker *et al.*, 2006), which has closely examined the assignment and participation process in ERA at selected sites, has speculated that ERA non-participants are not likely to be random subgroups of the eligible population. Diverted customers seemed to be mostly individuals whom advisers may have had a vested interest in not offering ERA, while formal refusers, especially those among the more problematic ND25Plus group, appeared to have weaker job prospects and poorer attitudes than the average New Deal entrant.[17]

While the insights provided by these in-depth case studies were based on only very few observations and thus could not be safely generalised, Goodman and Sianesi (2007) take the important initial step to thoroughly explore how representative (or policy relevant) the group is for whom we can calculate experimental estimates by understanding both how large and how selective the non-participating groups are. They perform a number of empirical analyses to assess the incidence and determinants of ERA offer and acceptance. This work thus sheds further light on the implementation of random assignment in the ERA study and most important to the current report, on the nature and extent of the non-participation problem. Separately for the ND25Plus and NDLP client groups, they consider the extent to which non-participation was due to diversion and to formal refusal and how the incidence of non-participation has varied across district, Jobcentre Plus office and time. They subsequently formally assess whether eligible individuals who did not participate in the ERA study were different from those who did participate. To this end, they test for significant differences in a wide range of observable, individual, office and local area-level characteristics, as well as for differences in post-inflow labour market outcomes.

The incidence, composition, determinants and selectivity of non-participation were markedly different between the ND25Plus and NDLP client groups, as well as across districts. As to incidence, non-participation overall was lower among the ND25Plus group (23 per cent of all eligibles) than among NDLP clients (over 30 per cent). In terms of composition, the bulk of non-participation in the ND25Plus group was due to formal refusals (59 per cent), though diverted customers, at over nine per cent of all eligibles, remain non-negligible. By contrast, more than 26 per cent of all eligible NDLP entrants in the six districts have been diverted; such eligible customers who did not seem to have been offered ERA account for over 86 per cent of non-participation among this customer group.

There was marked variation in the incidence of non-participation according to ERA district, with some clear outliers in terms of performance. The lowest proportions of non-participants for both client groups were observed in Scotland and in North West England, the highest in the East Midlands and in North East England. In particular, in the East Midlands district **almost half** of all eligible NDLP clients did

[17] Furthermore, the incentive structure arising from Jobcentre Plus job entry targets had an asymmetric influence on New Deal and on ERA Advisers. Specifically, when New Deal advisers undertook the intake interviews, they could benefit if job-ready customers did not participate in ERA and those with bad prospects did participate. Conversely, when ERA advisers were leading the intake process, they could benefit if customers with bad job prospects did not participate, while those with good prospects did.

not take part in ERA, most of them diverted customers. Focusing on the ND25Plus group, the performance of Scotland and North West England is particularly remarkable, with **not one single diverted customer**, while North East England stands out with **over one-quarter** of ND25Plus eligibles formally refusing to give their consent to being randomly assigned. A very strong and interesting role of Jobcentre Plus office affiliation was also uncovered in determining both ERA offer and consenting choice, though, as expected, it was stronger in the former. Over time, a fall in the formal refusal rate was observed for both customer groups, likely to reflect increased adviser experience and confidence in selling ERA, as well as the permission to mention ERA financial incentives.

Non-participants were found to differ from participants in some important respects. Most of the explained variation in ERA offer, acceptance and participation is accounted for by a client's district, office affiliation and inflow month, underscoring the key role played by local practices and constraints. A customer's employment prospects, as well as attitudes towards and past participation in government programmes were, however, also found to matter, leaving only a residual role to demographic characteristics.

In the absence of non-participation bias, the control group and the non-participants should behave similarly, as neither of them has been offered ERA services. However, the analysis of post-inflow labour market outcomes by Goodman and Sianesi (2007) has found non-participants to be somewhat **higher** performers than participants among NDLP customers, but to have significantly **worse** employment outcomes among ND25Plus customers.[18]

To conclude, the non-participation problem seems to be a relevant one, both in terms of its incidence and of the diversity of the excluded groups, the latter being particularly the case in terms of labour market outcomes. Furthermore, the average figures were found to mask at times extreme variation by district, customer group and type of non-participant. Overall, the NDLP ERA study participants are, on average, slightly more likely to depend on government benefits than the average lone parent volunteering for NDLP. By contrast, the study participants in the ND25Plus group are significantly easier to employ than the average ND25Plus entrant; ERA advisers are thus working with a group which is considerably more advantaged than the average population, which potentially raises a creaming question for the experiment.

---

18 ND25Plus non-participants had significantly worse employment outcomes than participants, facing a 21 per cent lower probability of being in employment and spending 19 per cent fewer days in work. By contrast, NDLP non-participants were somewhat higher performers than participants, mainly in terms of benefit outcomes (11 per cent fewer days on benefits).

The fact that ERA was a study and involved random assignment thus seems to have significantly altered how the intake as a whole was handled in the context of Jobcentre Plus, as well as the nature of the adviser/customer interaction in a way that would not have been the case if ERA had been normal policy. The fact that the pool of participants has been both reduced and altered is likely to have led to some loss in external validity or, alternatively, to some non-participation bias in the experimental estimate for the effect on the eligibles. The analyses in the present report aim to formally assess and quantify the loss in external validity, or the amount of non-participation bias.

3 Data and sample definition

3.1 Data

This report uses the same data as the previous descriptive analysis by Goodman and Sianesi (2007).

A number of data files have been put together. The administrative data held by the Department for Work and Pensions (DWP) on New Deal 25Plus (ND25Plus) and New Deal for Lone Parents (NDLP) entrants provide us with the sampling frame. We extracted files for all cases identified as having entered these New Deal programmes in the six districts over the relevant random assignment period, as detailed in Section 3.2. We have further exploited the New Deal extract files for information about past programme participation as well as a number of other relevant individual characteristics.

We have then merged these files with other DWP data on benefit and employment spells – the Work and Pensions Longitudinal Study (WPLS) dataset. This is a relatively recently released, spell-level dataset that contains information from DWP's Master Index about time on benefits (such as Jobseeker's Allowance (JSA), Income Support (IS) or Incapacity Benefits) and from Her Majesty's Revenue & Customs (HMRC) records about time in employment. These administrative records have been used to construct both detailed labour market histories and outcome measures.

We have further combined the administrative data with data collected specifically for the Employment Retention and Advancement (ERA) experimental evaluation in the form of the Basic Information Form (BIF). This file contains all New Deal customers who were approached for recruitment into ERA, including the identifier of those who formally refused to participate. Of this data we mainly use information on customers' decisions as to participation in ERA, as well as the outcome of random assignment (control/programme group) for those who agreed to participate in the study.

We have finally merged in local-area level data (Census, travel-to-work and super-output area data). In Section 3.3 we summarise the extensive variables we have selected and derived from all of these sources.

3.2 Sample

To perform our analyses aiming at estimating the impact of ERA for all ERA eligibles, we obviously need to start by clarifying exactly what we mean by ERA eligibility. This is a conceptual issue which requires us to decide on who should count as eligible. For such a definition to be operational, the criteria that determine ERA eligibility have also to enable us to identify the relevant individuals in the data.[19]

To the purposes of our analysis, which relates to the current experimental evaluation, we thus consider as **eligible** for ERA:[20]

1 those who have become mandatory for ND25Plus during the period when the respective district was conducting random assignment **and** who subsequently also started the Gateway still within the relevant random assignment intake window; and

2 those lone parents who were told about NDLP (had a Work Focused Interview (WFI) and/or expressed an interest in NDLP) during the period when the respective district was conducting a random assignment **and** who subsequently also volunteered for NDLP still within the relevant random assignment intake window.

[19] See Goodman and Sianesi (2007) for a description of how problem cases were handled and what adjustments were performed on the ERA experimental sample.

[20] The rationale underpinning the following definition of ERA eligibility is that Jobcentre Plus staff were instructed that those who became mandatory for ND25Plus or expressed an interest in NDLP during the random assignment window would be eligible for ERA. Those among these customers who effectively came to the office to start their New Deal programme at some point during the random assignment window should thus have been offered the chance to participate in ERA.

The random assignment window (or sample intake window) is actually district- and customer group-specific, since one district started conducting random assignments later than the others and some districts stopped conducting random assignments for some groups earlier. To identify the eligibles, the period when each district was conducting random assignments was defined as follows:

North West England:	3 January 2004	to 31 January 2005
All other districts:	1 November 2003	to 31 October 2004, with the exception of to 21 August 2004 for NDLP in South East Wales.

The report also considers ERA impacts on outcomes (e.g. earnings) collected from the ERA 12-month customer survey. This survey covers the experiences of a sample of the programme group and the control group during the first 12 months following individuals' date of random assignment, with most interviews occurring from December 2004 through November 2005. The intake period for individuals who are eligible to be surveyed is thus 1 December 2003 (3 January 2004 in North West England) to 30 November 2004. When looking at survey outcomes, we thus consider the intersection of the intake window above with this survey sample:

North West England:	3 January 2004	to 30 November 2004
All other districts:	1 December 2003	to 31 October 2004, with the exception of to 21 August 2004 for NDLP in Wales.

There is in fact very good overlap, with only 5.6 per cent of the full eligible sample being lost when imposing consistent intake criteria with those used to select the survey sample.

The following tables provide various sample breakdowns by participation status and survey status, separately for the two customer groups and by district. Section 2.3 has already provided a more detailed discussion of non-participation patterns, here we just summarise the main points for the analysis that will follow.

The incidence of non-participation was substantial: about **one-quarter** (26.6 per cent) of all those eligible to take part in the ERA study did not participate. Non-participation was substantially lower among the ND25Plus group (23 per cent of all eligibles) than among NDLP clients (over 30 per cent).

For both client groups, Scotland saw the lowest proportion of non-participation, with around nine per cent of ND25Plus and five per cent of NDLP eligibles not participating in ERA; North West England is not far behind at around 15 per cent of all ND25Plus eligibles and six per cent of all NDLP eligibles not participating. South East Wales and North East London saw closer to average levels of non-participation, while East Midlands and North East England saw the highest non-participation levels amongst both client groups. In particular, in East Midlands almost half (47 per cent) of all eligible NDLP clients did not take part in ERA. East Midlands in fact accounts for well over one-third of total non-participants, followed by London (27 per cent of all non-participants) and by North East England (20 per cent).

We observe survey outcomes for 31 per cent of ND25Plus and 35 per cent and NDLP study participants. Again this average masks some marked variation between customer groups and districts, ranging from 60 per cent of ND25Plus participants in South East Wales and 60 per cent of NDLP participants in Scotland to only 20 per cent of ND25Plus participants in North West England.

Table 3.1 Sample breakdown by customer group

	ND25			NDLP		
Eligibles	7,796	100.0%		7,261	100.0%	
Study non-participants	1,790	23.0%		2,209	30.4%	
Study participants	6,006	77.0%	100.0%	5,052	69.6%	100.0%
With survey outcome	1,840		30.6%	1,745		34.5%
Without survey outcome	4,166		69.4%	3,307		65.5%

Table 3.2 ND25Plus: Sample breakdown by district

	Scotland			North East England			North West England			South East Wales			East Midlands			London		
	N	%	%	N	%	%	N	%	%	N	%	%	N	%	%	N	%	%
Eligibles	816	100.0		1,080	100.0		1,612	100.0		575	100.0		1,717	100.0		1,996	100.0	
Non-participants	71	8.7		377	34.9		235	14.6		119	20.7		472	27.5		516	25.9	
Participants	745	91.3	100.0	703	65.1	100.0	1,377	85.4	100.0	456	79.3	100.0	1,245	72.5	100.0	1,480	74.1	100.0
With survey outcome	312		41.9	320		45.5	268		19.5	280		61.4	349		28.0	311		21.0
Without survey outcome	433		58.1	383		54.5	1,109		80.5	176		38.6	896		72.0	1,169		79.0

Table 3.3 NDLP: Sample breakdown by district

	Scotland			North East England			North West England			South East Wales			East Midlands			London		
	N	%	%	N	%	%	N	%	%	N	%	%	N	%	%	N	%	%
Eligibles	436	100.0		1,389	100.0		809	100.0		673	100.0		2,140	100.0		1,814	100.0	
Non-participants	23	5.3		406	29.2		50	6.2		159	23.6		1,009	47.1		562	31.0	
Participants	413	94.7	100.0	983	70.8	100.0	759	93.8	100.0	514	76.4	100.0	1,131	52.9	100.0	1,252	69.0	100.0
With survey outcome	253		61.3	308		31.3	288		37.9	268		52.1	306		27.1	322		25.7
Without survey outcome	160		38.7	675		68.7	471		62.1	246		47.9	825		72.9	930		*74.3*

Table 3.4 Non-participation breakdown by district

	Overall		ND25Plus		NDLP	
	N	%	N	%	N	%
Scotland	94	2.4	71	4.0	23	1.0
North East England	783	19.6	377	21.1	406	18.4
North West England	285	7.1	235	13.1	50	2.3
South East Wales	278	7.0	119	6.7	159	7.2
East Midlands	1,481	37.0	472	26.4	1,009	45.7
London	1,078	27.0	516	28.8	562	25.4
Total	3,999	100.0	1,790	100.0	2,209	100.0

3.3 Outcomes and observable characteristics

ERA impacts are assessed during a 12-month follow-up period in terms of two types of outcome measures: administrative and survey outcomes.

As to the former, data on employment and benefits receipt is available from administrative records for the **full** sample of ERA eligibles in the six evaluation districts, i.e. for both for participants and, most importantly for our purposes, for non-participants too. For these administrative outcomes measures we start counting the 12-month follow-up period from the moment individuals flowed in (i.e. from the moment ND25Plus customers started the Gateway, or lone parent customers volunteered for NDLP), and consider the probability of having ever been in employment, the total number of days in employment, and the total number of days on benefits during that period.

Survey outcomes were collected from a first-wave customer survey of a sample of ERA participants during the first 12 months following individuals' date of random assignment. The survey outcomes we consider are total earnings and an indicator for earning above £4,273 (the overall median calculated from those with positive earnings).

We have put together an extensive collection of variables aimed at capturing the widest possible range of individual, office and local area characteristics that are most likely to affect individuals' labour market outcomes, and that might potentially have affected selection into the ERA sample.

Note that all of these variables have to be defined both for the ERA study participants and non-participants, which required us to derive such information from administrative data sources alone.

Table 3.5 groups and summarises the various observable factors we use in our analysis; the table also briefly comments on the variables and lists the omitted category for discrete or categorical variables. Section 4.3 contains a more detailed discussion of the content of the data.

Table 3.5 Summary of observed characteristics

ERA district	
5 District dummies (compared to London)	
Inflow month	
13 dummies for month of 'showing up': 2nd to 13th month (compared to 1st month)	District-specific month from random assignment start when the individual started the ND25Plus Gateway or volunteered for NDLP
Demographics	
Female (compared to male)	
Age at inflow and age squared and missing age	
Ethnic minority customer (compared to white customer)	
Disability indicator Missing disability status (compared to non-disabled customer)	Disability indicator: if client has a disability at in-flow and/or if claiming incapacity benefits at in-flow
Has partner Missing partner information (compared to having no partner)	For ND25Plus
2 children ≥3 children Missing child information (compared to 1 child)	For NDLP
Youngest child <1 yr	
1-5 yrs at inflow Age of youngest child missing (compared to children aged 6-18)	For NDLP
Current spell	
Not on benefits at inflow	For NDLP
Employed at inflow	Indicator of very recent/current employment
Shows up same day Shows up within 30 days (compared to showing up after more than 30 days)	Showing up defined as the time between becoming mandatory for ND25Plus and starting the Gateway (for ND25Plus group), or between being told about NDLP and volunteering for it (for NDLP group)
Early entrant into ND25Plus programme	For ND25Plus Spent <540 days on JSA before entering ND25Plus
	Continued

Table 3.5 Contiinued

Labour market history	
Past participation in basic skills	Indicator of basic skills need
Past participation in voluntary programmes	Number of previous spells on: NDLP, New Deal for Musicians, New Deal Innovation Fund, New Deal Disabled People, WBLA or Outreach
Past participation in ND25Plus programme	For ND25Plus
Spent 0%, more than 0 but less than 50%, more than 50% but less than 100% of the past 3 years on active benefits (compared to having spent 100% of the time)	Summary of active benefit history Active benefits are JSA and compensation from NDYP, ND25Plus, Employment Zones and WBLA and Basic Skills.
Spent 0% more than 0 but less than 50% more than 50% but less than 100% of the past 3 years on inactive benefits (compared to having spent 100% of the time)	Summary of inactive benefit history Inactive benefits are Income Support and Incapacity Benefits
Spent more than 0 but less than 25% more than 25% but less than 50% more than 50% of the past 3 years in employment (compared to never employed in the 3 years before)	Summary of employment history
Local conditions	
Total New Deal caseload at office (100s)	Office indicator
Share of lone parents in New Deal caseload at office	Office indicator
Quintiles of the index of multiple deprivation: bottom, 2nd, 3rd and 4th (compared to top quintile)	Index of local deprivation at the SOA level Note: top quintile is the most disadvantaged
Local unemployment rate	Travel-to-work-level unemployment rate
Postcode missing or incorrect	

4 Methodological approaches

This chapter starts by setting up the framework and basic notation, as well as providing an overview of the types of analyses carried out in the report. We then move on to briefly outline in some more detail the different methodological approaches and their underlying assumptions, mainly to be in a position to highlight some issues which are important for a correct interpretation of the empirical results. Throughout, we try to keep the discussion as informal as its rather technical nature allows us.

The Supplementary Technical Appendix contains an in-depth and formal derivation of all the estimation methods as well as of the conditions for their validity.

4.1 Analysis framework

4.1.1 Set-up and notation

We start by setting up the framework and introducing some basic notation. Figure 4.1 highlights the structure of the problem we need to address, Box 4.1 summarises the notation.

The population of interest are those **eligible** to be offered Employment Retention and Advancement study (ERA) services, i.e. all those becoming unemployed in the six districts over the study intake window. The potential selection into the ERA study is represented by the binary variable Q, where Q=0 denotes individuals who despite being eligible have not been randomly assigned, while Q=1 denotes the ERA study participants, i.e. those eligible individuals who were offered the chance to participate in the ERA study and who gave their consent to be randomly assigned. Participating (Q=1) individuals make up the experimental group which was randomly assigned between a programme group who was offered ERA services (R=1) and a control group who was not (R=0).

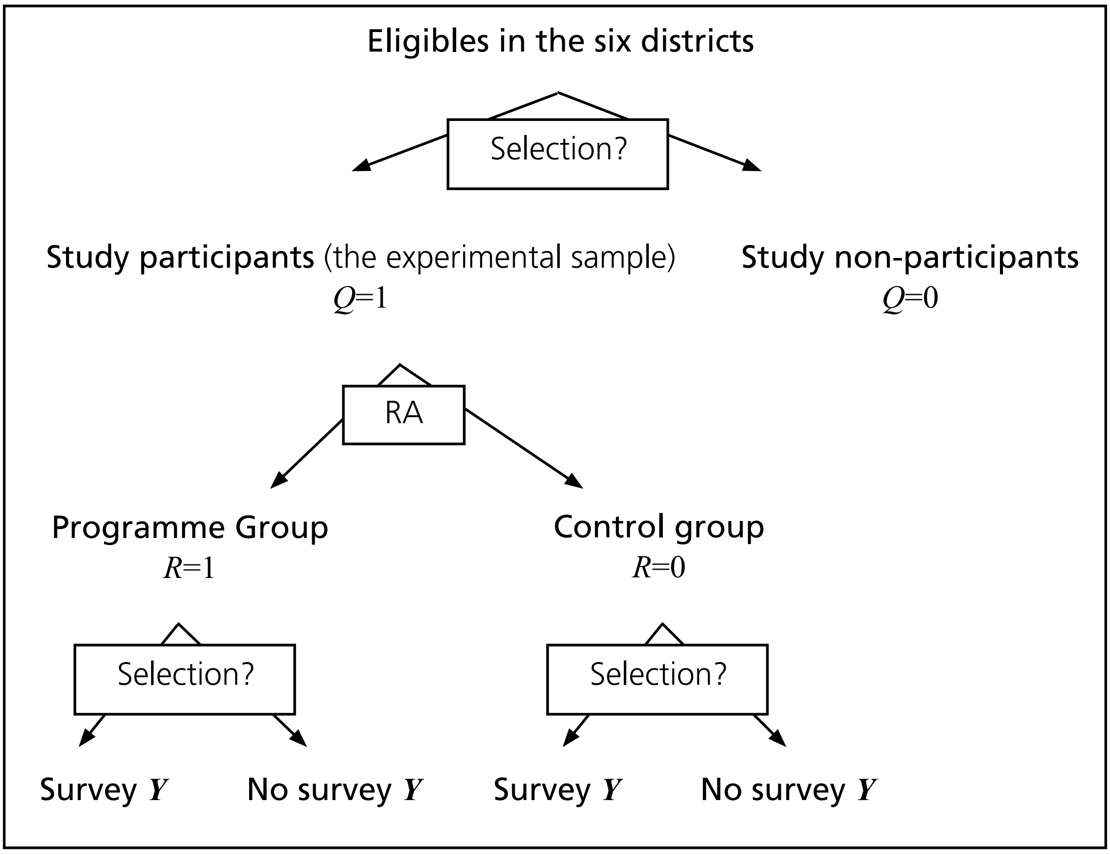
The problem here arises due to changes in participation pattern potentially introduced by the experimental evaluation. In particular, because of diversion and of refusal to be randomly assigned, the population under the experimental evaluation (Q=1) does not correspond to the full eligible population, made up by the (Q=1) **and** (Q=0) groups. If selection has taken place into the participating group, the composition of participants will be different from the composition of the eligible population, and impacts estimated on participants will not necessarily be representative of the impacts that the eligibles would have experienced.

A survey was carried out on participants to obtain specific outcome measures such as earnings. We denote 'respondents' those participants for whom survey outcomes are observed and 'non-respondents' those participants for whom they are not. As Figure 4.1 highlights, it is possible for some selection to have taken place among participants into the responding sample.

Let p be the probability of non-participation among the ERA eligibles. This is directly identified in the data by the proportion of non-participants among the eligibles (see Tables 2.1, 3.1 and 3.2).

Denote the observed outcome by Y and define the two potential outcomes: Y_1 as the outcome if offered ERA services (treatment outcome) and Y_0 as the outcome if not offered ERA services (the no-treatment outcome).

Figure 4.1 Simplified structure of the problem



Box 1: Notation

Q=1	ERA study participants (the experimental sample)
Q=0	non-participants
R=1	individuals randomly assigned to the programme group conditional on Q=1
R=0	individuals randomly assigned to the control group conditional on Q=1
X	observed characteristics
p	probability of non-participation among eligibles
Y_1	potential outcome if offered ERA services
Y_0	potential outcome if not offered ERA services
Y	observed outcome
ATE	average ERA effect on *all* ERA eligibles (parameter of interest)
ATE_1	average ERA effect on ERA study participants (experimental estimate)
ATE_0	average ERA effect on non-participants

The parameter we are interested in is the average effect of ERA on the *full* ERA eligible population in the six districts (the Average Treatment Effect – ATE), defined as the average outcome for the eligibles if they were offered ERA services compared to the average outcome for the eligibles if they were not offered ERA services:

$$ATE \qquad \equiv E(Y_1 - Y_0)$$

What we can however directly identify from the available experimental data is the average effect of ERA for participants in the experiment, $ATE_1 \equiv E(Y_1 - Y_0 \mid Q=1)$. This is because the experiment provides the average effect of the programme for individuals who have been randomly assigned.

Denote the average impact of ERA on the excluded eligibles (i.e. on the non-participants) by:

$$ATE_0 \equiv E(Y_1 - Y_0 \mid Q=0)$$

Using the law of iterated expectations, the parameters ATE and ATE_1 are linked according to:

$$ATE = (1-p)\cdot ATE_1 + p\cdot ATE_0 \qquad (1)$$

Equation (1) simply states that the parameter of interest, i.e. the average impact of ERA on all the eligibles in the six districts, is given by a weighted average of the parameter we can reliably estimate using random assignment, i.e. the impact on the participants ATE_1, and of the impact on the non-participants ATE_0, with weights given by the relative share of participants and non-participants within the eligible pool, p.

Non-participation thus poses a serious problem if it is **both widespread** (the share of non-participants p is sizeable) **and selective** (participants and non-participants are significantly different in terms of (observed and/or unobserved) characteristics that affect potential outcomes and hence programme impacts (i.e. ATE_1 is very different from ATE_0).

In previous work (Goodman and Sianesi, 2007) we have focused on assessing and documenting the size of p, finding that about one-quarter of the target population did not participate. The current report directly aims at estimating ATE_0 and assessing how different it is from ATE_1. Note though that whereas the relative size of non-participants (p) is observed in the data, how different the effect of the programme would have been for them compared to participants remains unobserved, since ATE_0 is not identified in the data. The effect for non-participants and the effect for all eligibles cannot thus be directly identified, unless additional assumptions are made. Before moving on to different non-experimental approaches invoking different assumptions to identify the unobserved parameters, in the next subsection we highlight the conditions under which non-participation, while sizeable, can still be ignored.

4.1.2 Conditions for *ATE* to be equal to *ATE*$_1$

Under what conditions is the average impact for those taken through random assignment the same as the average impact for the full eligible population even in the presence of a non-negligible share of non-participants? In two important cases, the ATE_1 based on experimental data would still provide an unbiased estimate of the ATE of interest:

1. Homogeneous treatment effects

   If the effect of ERA is the same for each individual, then changing the composition of the participants has obviously no effect. The effect for the participants will be trivially the same as for the non-participants, and thus for all eligibles: $ATE_1 = ATE_0 = ATE$.

2. No selection into the ERA study based on individual impacts

   Even in the presence of heterogeneous impacts, if the decisions of eligibles or caseworkers on ERA participation are not affected by the realised individual gain from receiving ERA, the effect for participants will ex post be the same as for non-participants, and thus for all eligibles.

The homogenous-impact assumption is an overly strong one[21], and is only a sufficient but not necessary one.

Programme effects are defined as the difference between the outcome if treated and the outcome if non-treated; in turn, the two potential outcomes depend on observed and unobserved characteristics of the individual and locality he or she lives in. Hence without invoking the questionable homogenous-effect assumption, the issue boils down to whether the participating and non-participating groups systematically differ in terms of observed and unobserved characteristics which affect potential outcomes, and hence programme effects. In particular, if the two groups were not significantly different, or in other words if the experimental sample were just a random sample of the full eligible population, non-participation would only pose an efficiency (precision) issue, but would not bias the impact estimate for the eligibles.

Further analysis is thus needed when effects are allowed to be heterogeneous **and** it cannot be ruled out that selection into the experimental study (at least partially) depends on them (or on variables related to them).

We consider selection on observable characteristics, selection on unobservable characteristics, and bounds that can be obtained without having to make any assumption, as overviewed in the following subsection.

4.1.3 Overview

In this report, we assess how different the average impact on participants is from the average impact on all eligibles based on the following identification strategies:

1. bounds for the ATE that can be obtained without having to make any assumption on the selection process;
2. identification of the ATE under the assumption of selection on observables; and
3. identification of the ATE allowing for selection on unobservables.

For each case, we consider how to deal with non-participants both:

(a) when follow-up information on the outcomes of the non-participants is available – impact estimates based on **administrative** data;

(b) when follow-up information on the outcomes of the non-participants is **not** available – impact estimates based on **survey** data.

---

21 Typically we would not expect all individuals to respond to a programme in exactly the same way. The impacts of a complex programme like ERA may differ across individuals because they differ, e.g. in their responsiveness to the different bonuses, or in the efficiency with which they can exploit the skills conferred by training to raise their productivity.

Clearly, case (a) will be more informative, and we will have to make less stringent assumptions. This can be usefully seen by considering the implications of cases (a) and (b) on equation (1).

In case (a) of administrative data, the observed outcome of the non-participants corresponds to their no-treatment outcome, since they did not take the treatment: $E(Y_0 \mid Q=0) = E(Y \mid Q=0)$. Hence the only unobserved term is the outcome that the non-participants would have experienced, on average, had they been offered ERA services, $\mathrm{E}(Y_1 \mid Q=0)$. Equation (1) thus becomes:

$$ATE = (1\text{–p})\cdot ATE_1 + \mathrm{p}\cdot\{\boldsymbol{E(Y_1 \mid Q=0)} - E(Y \mid Q=0)\} \tag{1a}$$

In case (b) of survey outcomes, both treatment and no-treatment outcomes of the non-participants are unobserved. Furthermore, when relying on survey outcomes there might be an additional reason that prevents the experimental contrast from recovering the parameter of interest (i.e. the effect of ERA on the full eligible population): non-random non-response to the survey, or, among survey respondents, to the questions pertaining to the outcome (in particular, earnings). Thus, in the presence of non-random, non-response, ATE_1 itself will in general remain unobserved:

$$ATE = (1\text{–p})\cdot\boldsymbol{ATE_1} + p\cdot\boldsymbol{E(Y_1 - Y_0 \mid Q=0)} \tag{1b}$$

In the following subsection we discuss non-response in a non-technical way, referring the interested reader to the Supplementary Technical Appendix for more detail.

4.1.4 Survey outcomes: Survey and item non-response

Survey outcomes, in particular earnings, are only observed for a subsample of participants: those survey respondents who answered the earnings question. Specifically, we define the 'respondents' as those ERA study participants who (1) were randomly selected to be surveyed, (2) could be contacted and accepted to take the survey **and** (3) answered the earnings question.

Correspondingly, the 'non-respondents' include: (1) those participants who were not surveyed, (2) the survey non-respondents and (3) the item non-respondents among surveyed participants. In other words, the non-respondents are those ERA study participants with missing survey outcome information for whatever the reason (not randomly selected for the survey, not contactable, refused to be interviewed, were interviewed but did not fill in the earnings question).

Non-response raises three potential issues for the evaluation of earnings impacts.

First, there is an **internal validity** issue: if the programme and control group experience systematically different non-response, the final programme and control groups are no longer comparable to one other. Should this happen, the benefits of the original random assignment are lost, and a comparison of the responding programme group members and the responding control group members no longer provides unbiased impact estimates.

Second, there is an **external validity** issue: even if the responding programme and control group members have maintained comparability to one another, how do they relate to the original sample? If the responding sample differs substantially from the original one, the results might not generalise to the original target population.

Finally, there is the **precision** loss: as fewer study participants remain, the study's statistical power to detect effects is reduced.

The data directly allows one to calculate the experimental contrast on the responding participants. In order to recover the ATE for the full group of eligibles, we are however interested in the average effect for all study participants, not just those who answered the earnings question. We thus need to consider under what conditions the experimental contrast for respondents recovers the average effect for the full group of participants.

In looking at this issue it is useful to separately consider the following two causality questions relating to the internal and external validity mentioned above.

(a) Internal validity: Under what conditions does the experimental contrast on the respondents recover the average effect for respondents?

Since the average ERA impact for respondents is not identified without additional assumptions, to exploit random assignment one has to assume that randomisation keeps holding within the responding sample, i.e. that the responding programme and control groups are still balanced in terms of any characteristic, observed by the analyst or not, that affects their potential outcomes.

This internal-validity condition cannot be directly tested; supporting evidence can however be obtained by assessing whether randomisation still holds between the two responding subsamples in terms of their **observed** characteristics.

(b) External validity: Under what conditions can the subsample of respondents be assumed to be a representative subsample of the ERA study participants, in the sense that the average effect among respondents is the same as the average effect for the full group of participants?

The average ERA impact is the same for the full sample of participants and for those participants who responded to the survey if participants do not select into responding based on ERA impacts.

Since the impact for respondents is not identified *a priori,* to 'test' this external-validity condition one has first to assume that the internal-validity condition holds. Provided randomisation still holds within the responding sample, the external-validity condition can be tested on administrative data by testing whether the average impacts on administrative outcomes for the full group of study participants are not statistically different from the corresponding average impacts for the subgroup of respondents.

Note that under internal validity, this external-validity condition is implied by the stronger set of conditions that the outcomes of those programme (control) group members who responded to the survey are not statistically different from the outcomes of those programme (control) group members for whom we do not observe the survey outcomes. In other words, conditional on random assignment status, non-response is unrelated to potential outcomes, so that programme and control group members who respond are not selected on outcome-relevant variables. Note that the main driving element behind response was indeed random selection of survey sample members; departure from the intended randomness has, however, arisen due to survey non-response (19 per cent among those selected to take part in the survey) and item non-response (less than 10 per cent among survey respondents). Like the main external-validity assumption, this stronger assumption can be tested on administrative outcomes.

To conclude, the experimental contrast for respondents, which is readily obtained from the data, would recover the average impact for the full group of participants (ATE_1) under the internal-validity condition and either one of the external-validity conditions. In this case, non-response can be ignored in calculating the average effect on earnings for participants.

In Section 5.1, we perform some tests and checks for these conditions.

4.2 Bounds without assumptions on the selection process

For this type of analysis, outcomes need to be bounded. This is obviously the case for discrete events such as being employed or not. To fix ideas, suppose we are evaluating ERA in terms of employment probability, so that the outcome Y is bounded between 0 and 1.

4.2.1 Follow-up data on the non-participants

From equation (1a), the lower bound for the effect on all eligibles is obtained by assuming that none of the non-participants would have gone into employment from the programme, the upper one by assuming that all non-participants would have been in work had they received ERA:

- lower bound: $(1–p)·ATE_1 – p·E(Y | Q=0)$
- upper bound: $(1–p)·ATE_1 + p·(1–E(Y | Q=0))$

The width of the bound is given by p, the proportion of non-participants among the eligibles.

Sensitivity analysis

We can further explore how sensitive the estimate of the effect on all eligibles is to assumptions about the selection process into the group of study participants, as reflected by assumptions on the relative magnitude of the average outcome under ERA for participants and for non-participants.

Specifically, assume that the average ERA outcome that the non-participants would have experienced had they participated in the study is θ times the average ERA outcome of the participants, as identified by the actual outcome of the programme group.

From equation (1a), we can then calculate the effect for all eligibles as a function of θ:

$$ATE_\theta = (1-p)\cdot ATE_1 + p\cdot\{\theta\, E(Y \mid R=1) - E(Y \mid Q=0)\}$$

By varying the values of θ, we can depict different types of selection processes: θ=1 represents the case where decisions to participation in the ERA study are unrelated to treatment outcomes, while θ<1 (θ>1) the case where non-participants would have experienced on average lower (higher) treatment outcomes than what the participants experience.

4.2.2 No follow-up data on the non-participants

In this case, we have to construct bounds for the effect on all eligibles based on equation (1b). It follows that

- the upper bound is $(1-p)\cdot ATE_1 + p$
- the lower bound is $(1-p)\cdot ATE_1 - p$

The width of the bounds is now $2\cdot p$, so that the bounds are twice as large, or twice less informative, as when we did observe the outcomes of the non-participants.

In case non-response cannot be ignored, the bounds will necessarily – and trivially – be the widest possible ones, and unrelated to data content: [–1, 1]

See the Supplementary Technical Appendix for issues concerning the significance of the estimated bounds.

4.3 Impact estimates under selection on observables

This and the next section describe two sets of methods aimed at arriving at a point estimate of the effect for all eligibles. While the two methods differ in terms of the assumptions they make on the selection process into the ERA study (one rules out outcome-relevant unobservable determinants, the other allows for them as well), both rely on the assumption that treatment and no-treatment outcomes among the eligibles are not affected by whether an individual is **offered the chance** to participate in the ERA study or not.

The approaches outlined in this section provide estimates of the average ERA impact for the non-participants (and hence for all eligibles) which can only take into account **observed** differences between non-participants and ERA study participants. To the extent that **unobserved** differences between the two groups are important determinants of subsequent labour market outcomes, these will erroneously show up as part of the ERA impact estimates.

The reliability of such estimates thus crucially depends on the range and quality of characteristics observed. Section 3.3 has summarised the data at our disposal; here we provide a brief discussion of its content in relation to the estimation problem we face.

All our outcomes of interest – employment probabilities and durations, reliance on benefits and earnings – are related to labour market performance. As listed in Table 3.5, we rely on an extensive collection of individual, office and local area characteristics that are most likely to affect individuals' labour market performance, and that might potentially have affected participation into the ERA study.

In addition to a number of individual demographic characteristics contained in the administrative data (gender, age, ethnicity, partner and children, disability and illness), we have summarised information on a customers' current unemployment spell, including, in particular, indicators of a very recent/current employment spell, how long it took them to start the Gateway or volunteer for New Deal for Lone Parents (NDLP) once having become mandatory for it or being told about it, and whether New Deal 25Plus (ND25Plus) entrants volunteered for the Gateway ahead of time.

We have further constructed three years' worth of labour market history, with variables summarising the proportion of time employed and the proportion spent on benefits, separately on active benefits (Jobseeker's Allowance (JSA) and compensation while on a labour market programme) and inactive benefits (Income Support (IS) and Incapacity Benefits (IB)). We have also created variables capturing the extent of past participation in voluntary employment programmes (as a crude indicator of willingness to improve ones circumstances), in the ND25Plus (a mandatory programme) and in Basic Skills (a programme designed to address basic literacy, numeracy and IT skills).

The Census has provided us with information on local labour market conditions (specifically, travel-to-work area unemployment rates), as well as on the deprivation of the area the customer lives in (index of local deprivation). Additionally, we have constructed information at the office level (total New Deal caseload and share of lone parents in such caseload), aimed at capturing office-specific characteristics that might impact on the probability of participating in the ERA study as well as on subsequent labour market outcomes.

Despite offering such rich and detailed information, none of the available administrative data contain reliable information on education – which thus remains an unobservable in our data, together with 'innate ability', discipline or work commitment. The previous literature has, however, indicated the potential for detailed labour market histories (like those we have contructed) to help serve as a proxy for such unobserved traits and thus to eliminate much of the bias due to unobservables (see for example, Dolton *et al.*, 2008, Heckman and Smith, 1999, Heckman *et al.*, 1998, and Heckman *et al.*, 1999).[22]

4.3.1 Follow-up data on the non-participants

In previous work (Goodman and Sianesi, 2007) reviewed in Section 2.3, we have shown the extent to which outcome-relevant observed characteristics X of the participants and non-participants differ.[23] We could build on that work and calculate experimental impacts by some chosen X, in particular by benefit/unemployment history. This would, however, be just an indicative exercise, as it only takes account of a chosen subset of the observables. Also, it would not directly provide the overall average effect for all eligibles.

To estimate the average effect for all eligibles on administrative outcomes, equation (1a) shows that we need to identify the counterfactual ERA outcome of the non-participants, $E(Y_1|Q=0)$.

The methods in this section do so by invoking the 'selection-on-observables' assumption that participants and non-participants with the **same** set of observed characteristics would not differ in terms of the ERA outcome they experience (or would experience) on average:

(A1) $$E(Y_1 \mid Q=0, X) = E(Y_1 \mid Q=1, X)$$

Assumption (A1) thus requires that for the eligibles, selection into the ERA study is not based on unobserved individual characteristics or on unobserved individual ERA impacts.

To give empirical content to assumption (A1), we also need to assume the existence of common support (i.e. overlap in the distribution of observed characteristics X) between participants and non-participants, so that each non-participant has at least a counterpart in the participant group.

22 For their main analysis of the NDLP programme, Dolton *et al.* (2008) rely on the same administrative data we use. When using a subset of their sample for whom detailed additional survey information (including a variety of attitudinal measures) is available, they find that such variables in fact add little to the analysis once the lagged outcomes available in the main administrative data are controlled for. They interpret this finding as indicative of the fact that outcome histories capture these otherwise unobserved factors and supporting of their approach based on the selection-on-observables assumption.

23 Note that we can test whether the two groups significantly differ in terms of observables; we can only speculate about whether such observables are likely to affect impacts.

As for implementation, we match to each non-participant one or more similar programme group member(s) based on the propensity score (the probability that an eligible customer with characteristics X participates in the study). This approach is non-parametric in the sense that it allows the ERA outcome (and the effect) to depend on observable characteristics in an arbitrary way, as well as for eligible individuals to decide to participate in the experiment based on these characteristics.

Sensitivity analysis

As done for the bounding approach, we can explore how sensitive the estimate of the impact for all eligibles is to straightforward violations of assumption (A1) by relaxing it to:

(A1′) $$E(Y_1 \mid Q=0, X) = \theta\, E(Y_1 \mid Q=1, X)$$

and estimating the impacts that arise from different values of θ. Assumption (A1′) implies that the average ERA outcome that non-participants would have experienced are θ times the average ERA outcome experienced by participants with their same observed characteristics. In other words, despite sharing the same observed characteristics, participants and non-participants are allowed to differ in terms of some unobservable, which translates into a proportional difference of θ. For favourable outcomes such as employment probability or days employed, $\theta>1$ implies positive selection into the non-participants sample, while $\theta<1$ negative selection. For unfavourable outcomes such as days on benefits, the opposite holds.

4.3.2 No follow-up data on the non-participants

The task of estimating the impact of ERA on all eligibles when only the impact for the responding participants is available involves making the latter representative – in terms of observed characteristics – of the former. This is accomplished by reweighing the outcomes of the responding participants (i.e. the responding programme and control groups) on the basis of the observed characteristics of the full eligible group (i.e. the non-participants **and** the **full** programme group and the **full** control group).

For this approach to be valid, we need to assume that, once conditioning on observable characteristics X, ERA study participants and non-participants experience the same average outcomes under ERA **and** without ERA:

(A2) (a) $$E(Y_1 \mid Q=1, X) = E(Y_1 \mid Q=0, X)$$

(b) $$E(Y_0 \mid Q=1, X) = E(Y_0 \mid Q=0, X)$$

These ideas can be empirically implemented in several ways; we consider reweighing and matching estimators, both ignoring and allowing for selective survey and/or item non-response (provided in the latter case that selection into the responding sample happens only in terms of observable characteristics). The Supplementary Technical Appendix derives the various weighting schemes and describes the different ways of implementing the matching approach.

4.3.3 Analysis of take-up

This section outlines a simple yet informative analysis which aims at estimating the type of involvement that the non-participants would have had with ERA and more generally with Jobcentre Plus had they participated in the evaluation study – either as part of the programme group or of the control group. Specifically, we aim to answer the following two questions:

1. Are the non-participants individuals who, even if offered ERA services, would not take them up?
2. What kind of involvement would non-participants have had with Jobcentre Plus had they participated in the ERA study and been assigned to the control group?

We can get a handle on these questions by looking at measures of take-up of services and of contact with Jobcentre Plus staff, such as whether the customer has had any type of contact with Jobcentre Plus staff, has received help or advice from Jobcentre Plus staff when not working, has had an education or training course arranged by Jobcentre Plus staff, or, if assigned to the programme group, has heard of the employment and of the training bonuses.

The trick is to simply view such take-up/involvement measures as outcomes, and assess them in essentially the same way as done for employment and earnings outcomes.

To answer question (1), we need to estimate the take-up of ERA services that non-participants would have experienced, on average, had they been offered such services.

To perform this analysis, we again rely on the selection-on-observables assumption (A2.a) requiring that, once conditioning on our rich set of observables X, ERA study participants and non-participants would have taken up the same amount of ERA services on average. In other words, we rule out selection into the ERA study based on unobserved characteristics that also affect take-up of ERA services once in the programme group.

To implement this estimator, we can match to each non-participant one or more 'similar' programme group members and take the latter's reweighted outcomes.

A similar type of analysis can be performed on the non-participants and the control group to answer question (2). It requires that, once conditioning on our observables, ERA study participants and non-participants would on average have had the same involvement with Jobcentre Plus if assigned to the control group.

As a final note, although such take-up/involvement measures are obtained from the 12-month follow-up survey, non-response to these questions is truly negligible (less than one per cent), so that it can be safely ignored when performing both types of exercise.

4.4 Impact estimates under selection on unobservables

This section sketches a class of models which allow selection into the group of ERA study participants to depend on outcome-relevant *unobservables*. For survey outcomes, we rule out selective non-response based on *unobservables*. Though in principle these models could be extended in this direction, the data we have contain no obvious instrument that would make such an extension credible.[24]

All of these models fall within the family of 'control function models' and build on the classical sample selection model introduced by Heckman (1979). We are, however, in the rather unique position where for one set of outcomes (the administrative ones), we do observe the outcomes of the selected-out sample. Together with randomisation, we exploit this feature of the data to:

(a) test the exclusion restriction of the instrument;

(b) test for the presence of residual selection on unobservables related to no-treatment employment or benefit outcomes;

(c) test how well the various control function models capture the presence and direction of the selection on unobservables we have thus uncovered; and

(d) test how well the various control function models predict the no-treatment outcome for the non-participants.

Tests (a) and (b) of course apply irrespective of the actual control function model being considered. By contrast, tests (c) and (d) test some features of the performance of a given model, so that their specific form depends on the actual model under examination. We thus start by presenting tests (a) and (b), then move on to sketch the various models, outlining the idea behind tests (c) and (d) (a formal discussion for each case in contained in the Supplementary Technical Appendix).

4.4.1 Some initial tests

The following two tests exploit the fact that the control group is representative of the participants, but like the non-participants does not receive ERA. Thus for both the controls and the non-participants, the actual outcome coincides with the no-treatment outcome, and in the case of administrative data is observed for both groups.

The general control function approach attempts to control for selection into the ERA study based on unobservables by exploiting some arguably exogenous variation in participation by way of a so-called 'excluded instrument'. Specifically, we need an observable variable Z which affects the decision to participate in the ERA study, but it does not otherwise affect potential outcomes directly. In symbols, Z has to be such that:

---

24 Adviser or surveyor information could be used to model survey non-response and surveyor information item non-response, but we have no such identifier in the data.

(A3) (a) P(Q=1 | X, Z) is a non-trivial function of Z

(b) $E(Y_0 \mid X, Z) = E(Y_0 \mid X)$

(c) $E(Y_1 \mid X, Z) = E(Y_1 \mid X)$

The strength ('power') of the instrument in affecting the choice among eligibles to participate in the ERA study, i.e. condition (A3.a), is, as usual, a testable condition. In our case, however, when modelling administrative outcomes we can test part of the exclusion restriction as well (condition A3.b). Such a test is implemented by pooling the controls and the non-participants, regressing observed (no-treatment) outcomes on the observables and the instrument, and testing the significance of the instrument.

Due to our unique set-up, we are also in a position to test whether there remain differences between participants and non-participants in terms of unobservables related to non-ERA employment or benefit outcomes. We can accomplish this by looking at whether, once controlling for observable characteristics, the outcomes of the non-participants differ on average from those of control group. If in the comparison of the (no-treatment) outcomes of these two groups there remain significant differences conditional on observables, this provides evidence of selection on outcome-relevant unobservables.[25] This test can be performed by running a regression on the pooled sample of controls and non-participants of observed outcomes Y on the group dummy variable G controlling for observables X, and testing the significance of α:

$$Y = \alpha G + \gamma X + \varepsilon$$

A number of alternative methods are also available to minimise all sensitivity to the specification of how the observables should enter the outcome equation or affect differences between the two groups (matching and fully interacted OLS models), as well as to properly take into account the potentially binary or censored nature of the outcome of interest (Probit and Tobit models).

The results of this test are not just informative in themselves, but as we show below, they lend themselves to construct an important specification check for any of the control function models. It is important to note that should the data fail to pass this test, this would not *per se* invalidate the estimates of the effects for non-participants and for eligibles based on the matching and reweighting findings which rely on the selection-on-observables assumption (Section 4.3). This is because this test only concerns the no-treatment outcome and only in the case of administrative data. For administrative outcomes, the matching and

---

25 A crucial assumption underpinning this statement is that there has been no ERA impact on the control group. This is a fundamental assumption for the validity of the experimental impact estimates, which is likely to have been met given that control group members were not allocated a dedicated post-employment advisor nor could they receive the financial incentives.

reweighting estimates do not need to predict the no-treatment outcome. Instead, they need to predict the treatment outcome, as well as the no-treatment outcome based on survey data, but in neither of these cases can such a test be performed.

4.4.2 Standard control function model

The problem of non-participation in the ERA study is akin to the classical sample selection problem: the treatment outcome is only observed for the ERA study participants (indeed, for its representative programme subgroup), but is not observed for the non-participants. In case of survey-based outcomes, it also is the case that the no-treatment outcome is only observed for the participants (via its control subgroup), but is unobserved for the non-participants.

This is a rather formal set-up, requiring technical conditions for identification and at times quite complex estimation methods. In what follows we provide the least detail which is necessary to appreciate the assumptions underlying the estimates and to interpret the output presented in Section 5.4; we refer the interested reader to the Supplementary Technical Appendix for the complete formal derivation of all the steps involved.

For the eligible population, potential treatment (Y_1) and no-treatment (Y_0) outcomes depend on observed (X) and unobserved (u) individual characteristics and on unobserved individual ERA impacts (b) as follows:

$$Y_0 = \beta_0 X + u \qquad u \sim N(0, \sigma_u^2)$$

$$Y_1 = \beta_1 X + u + b \qquad b \sim N(0, \sigma_b^2)$$

As mentioned, treatment outcomes Y_1 are however only observed for study participants (Q=1, as represented by the programme group), not for the non-participants (Q=0). In case of survey outcomes, no-treatment outcomes Y_0 are similarly only observed for study participants (as represented by the control group). Let the observability rule for Y_1 (and Y_0 in case of survey outcomes) be:

$$Q = 1 \quad \text{if } \gamma W + v \geq 0 \qquad v \sim N(0, 1)$$

$$Q = 0 \quad \text{if } \gamma W + v < 0$$

where the observables W are made up of the observed characteristics X as well as by some 'instrument' Z, and where the unobserved determinant of participation in the ERA study, v, is potentially correlated with unobserved individual characteristics (u) and ERA impacts (b):

$$Corr(v, u) = \rho_{uv}$$

$$Corr(v, b) = \rho_{bv}$$

The model thus allows for selection into the ERA study based on both unobserved 'ability' (u) and unobserved individual-specific ERA impacts (b).

The crucial set of assumptions implicit in this model is:

(A3) (a) $P(Q=1 \mid X, Z)$ is a non-trivial function of Z

(b) $E(Y_0 \mid X, Z) = E(Y_0 \mid X)$

(c) $E(Y_1 \mid X, Z) = E(Y_1 \mid X)$

(d) $$\begin{pmatrix} u \\ b \\ v \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \rho_{ub} & \rho_{uv} \\ \rho_{ub} & \sigma_b^2 & \rho_{bv} \\ \rho_{uv} & \rho_{bv} & 1 \end{pmatrix}\right]$$

Apart from the parametric choice of the distribution of the unobservables implied by condition (A3.d) (in particular, joint normality and homoskedasticity), the control function model crucially relies on an exclusion restriction. Specifically, we need an observable variable Z which is contained in W, i.e. which affects the decision to participate in the ERA study (the Q=1 decision – condition (A3.a)), but is not contained in X, i.e. does not affect potential outcomes directly (conditions (A3.b) and (A3.c)).

As discussed in Section 4.4.1, conditions (A3.a) and, for administrative outcomes, (A3.b) can be tested. Also, the parametric assumptions in (A3.d) can be relaxed (and thus tested), as we show in the next subsections.

Under the assumptions of the model, we can derive the exact form of the expected unobserved treatment outcome (and no-treatment outcome for survey outcomes) for each individual non-participant with a given set of characteristics W.

A neat feature of the model is that since it provides estimates of ρ_{uv} and ρ_{bv}, it allows one to separately test for selection into the ERA study based on unobserved 'ability' (u) and based on unobserved individual-specific ERA impacts (b), evidence which can be of interest in its own right.

In the case of administrative outcomes, we can construct two specification tests to assess – and order – the performance of the different control function models.

The Supplementary Technical Appendix describes in detail how to construct a test for how well the control function model captures the actual extent of selection on unobservables between the participants (as represented by the controls) and the non-participants, that is, the parameter α estimated in the test outlined in Section 4.4.1. The idea is to mathematically derive the expression for the control function model which is **equivalent** to α. Maybe unsurprisingly, this expression turns out to be closely related to the selection terms of the model. Given that the different control function models recover potentially different estimates of such selection terms, the difference between α and the selection terms provides a ready metric to 'order' the performance of these models.

The second specification test is based on testing how well a given control function model predicts the average no-treatment outcome for the non-participants. Once estimated, we can use the model to recover the predicted no-treatment outcomes for the non-participants, which we then compare to the average **observed** no-treatment outcome for the non-participants.

We are thus in a position to choose between different specifications of the control function based on these two 'metrics', i.e. how closely a given model matches the difference in adjusted observed outcomes between the control group and the non-participants (reflecting the results from our test of selection on unobservables), as well as the average predicted and observed (no-treatment) outcomes of the non-participants.

4.4.3 Extensions to the standard control function model

We have extended the standard model in two broad directions (see the Supplementary Technical Appendix for all the details and technical derivations of these extensions).

First, we have relaxed the parametric assumptions on the unobservables in terms of both **independence** and **normality** implied by condition (A3.d). Independence in particular was relaxed to allow for heteroskedasticity of the unobservable determinants of treatment and no-treatment outcomes (u and b), as well as for the covariances between the unobservables relating to outcomes (u and b) and the unobservable determinant of participation (v). The latter basically means that the selection process into the ERA study is allowed to be different for customers with different observed characteristics.

The second type of extension takes into account the **censored** nature of the outcome variable. In particular, the outcome is allowed to be censored (at zero in the case of employment duration or earnings) in both the treatment and no-treatment state.

In this extension we utilise all the available information, so that when modelling survey outcomes, participants with missing earnings (because they have not been sampled or because of survey or item non-response) are not dropped from the analysis, but contribute to the estimates in terms of their participation decision. Average ERA impacts can then be separately estimated for responding and for non-responding participants.

As was the case with the other models, in addition to directly testing whether there was selection into the ERA study based on unobserved individual characteristics and/or unobserved gains from ERA, we can perform a number of 'tests' on the performance of the model. Specially, we can construct tests for how well the model captures the actual extent of selection on unobservables and for how well it predicts observed outcomes (i.e. no-treatment outcomes for the non-participants and the control group, and treatment outcomes for the programme group). Furthermore, we use the model to predict the average no-treatment outcome for the programme group and compare it to the observed average outcome of the control group, where as we know the latter provides an unbiased estimate of the former. We also estimate the average effect for the participants using the extended model and compare this estimate to the experimental one.

All these specification tests are summarised as follows, together with the short-cut notation used in the results tables in Section 5.4:

	How well the model…
α – selection terms	… captures the actual extent of selection
Q=0: observed–predicted Y	… predicts (no-treatment) outcomes for non-participants
R=0: observed–predicted Y	… predicts (no-treatment) outcomes for the control group
R=1: observed–predicted Y	… predicts (treatment) outcomes for the programme group
$E(Y\|R=0)–E(Y_0\|R=1)$	… predicts no-treatment outcomes for the programme group
ATE_1	… predicts the average impact for participants

Finally note that our estimate of the average ERA impact for all eligibles uses the full model, taking observed outcomes for the programme group and predicted ERA outcomes for the controls and the non-participants on the one hand, and predicted non-ERA outcomes for the programme group and observed outcomes for the controls and the non-participants on the other.

5 Implications of non-participation for the experimental impact estimates

This section presents our empirical results. The analyses have always been performed separately for the two customer groups, New Deal for Lone Parents (NDLP) and New Deal 25Plus (ND25Plus). For all estimation methods except the control function models, they have been performed both overall and by district. In the following we focus on the overall findings, mentioning district-level ones only if worthy of special note. We refer to Appendix A for the tables with all the district-level results and corresponding summary boxes.

The section starts with the benchmark experimental findings that omit the non-participants.[26]

5.1 Experimental findings

This section presents the experimental findings concerning the average impact of Employment Retention and Advancement (ERA) study for the participants on a series of outcomes. Table 5.1 displays both the raw experimental contrast ('raw') and the impact estimated by linear regression controlling for a number of observed background characteristics ('adjusted').[27] Although randomisation has worked very well so that the ERA programme and control groups are well-

26 As we highlight in the following, these findings do not always correspond to those reported in Dorsett *et al.* (2007). The reasons for any discrepancy are the latter's use of survey-based rather than administrative outcomes, focus on the survey rather than the full sample, adjustment for survey rather than administrative characteristics and use of a different weighting scheme.

27 Background characteristics have all been derived from the administrative data (see Section 3.3).

balanced in terms of such characteristics, controlling for them can increase the precision of the experimental impact estimate by reducing the residual variance of the outcome. This seems to be largely the case in this application, where most standard errors decrease following the regression adjustment. Furthermore, the adjustment allows one to control for differences in observables between the programme and the control group that have occurred by chance. This also seems to matter in our application, as impact estimates are often found to change once conditioning on observables.

No impact could be detected on the probability of being employed in the follow-up year except for those NDLP customers living in North West England (+7 percentage points). Employment durations in the follow-up year have been increased by the ERA intervention for NDLP customers in North West England (by 22 days), but decreased for those in the East Midlands (by 15 days). Overall, the NDLP group has remained unaffected in this dimension. A small positive overall effect of ERA (plus five days) has by contrast been uncovered for the ND25Plus group, driven by the effect in London (plus nine days).[28]

Time spent on benefits appears to have been slightly reduced by the offer of ERA for the NDLP group, an impact driven by the programme effect in North West England; once chance imbalances in the observables are controlled for, though, both effects drop into non-significance. By contrast, evidence of a positive ERA effect on benefit collection is robust to regression-adjustment for ND25Plus customers in the East Midlands and in North West England (14 and 11 fewer days), while benefit dependency for customers in Wales has been significantly increased (plus 29 days). Overall, though, when controlling for observables, neither customer group has been affected in terms of days spent on benefits during the follow-up year.[29]

---

28 Relying on survey-based employment information, Dorsett *et al.* (2007) reach different conclusions. Specifically, for the ND25Plus group no employment impact could be detected, while for the NDLP group a 4.5 percentage point increase in employment probability and a 0.6 month increase in the number of months worked during the first year were uncovered. Note further that the findings in Table 5.1 on employment impacts are still not directly comparable to those based on administrative outcomes in Dorsett *et al.* (2007, Appendix E) due to (a) a slightly different outcome measure (months instead of days, where a respondent is counted as having worked in a month if they worked at least one day); (b) a slightly different sample (survey sample instead of full sample); (c) a different set of control variables (survey – specifically the Basic Information Form – rather than administrative data) and (d) a different weighting scheme.

29 Dorsett *et al.* (2007) also fail to find any impact on the ND25Plus group's reliance on Jobseeker's Allowance (JSA) according to administrative outcome data. However, they do uncover a significant reduction according to survey-based data for this customer group, as well as some decrease in the reliance on Income Support (IS) – both in terms of survey and administrative outcomes – for the NDLP group.

Table 5.1 Experimental findings

	Raw		Adjusted		N
	Effect	Standard errors	Effect	Standard errors	
ND25Plus					
Ever employed	0.014	(0.012)	0.017	(0.011)	6,006
Days employed	4.0	(2.7)	4.6*	(2.4)	6,006
Days on benefits	-3.0	(3.2)	-3.0	(3.0)	6,006
High earnings	0.029	(0.020)	0.026	(0.019)	1,840
Earnings	378.6*	(228.6)	393.2*	(222.7)	1,840
NDLP					
Ever employed	0.003	(0.014)	-0.006	(0.013)	5,052
Days employed	-0.1	(4.0)	-2.2	(3.5)	5,052
Days on benefits	-8.2**	(4.0)	-5.1	(3.7)	5,052
High earnings	0.054**	(0.022)	0.039*	(0.021)	1,745
Earnings	885.2***	(230.3)	730.2***	(225.5)	1,745

Note: adjusted for the observables X constructed from administrative data for the full sample. Robust standard errors for ever employed and for high earnings; *** significant at 1%, ** at 5%, * at 10%.

The impact of ERA on average earnings in the first follow-up year is estimated to be quite substantial and highly statistically significant for the NDLP group (+£730), driven by the impact in Scotland (+£1,443) and Wales (+£1,080), or, when not controlling for observables, by the impact in Scotland and North West England. For the ND25Plus group, the experimental contrast highlights a much smaller impact, which is significant only at the ten per cent level (+£393). This overall effect is driven by the positive and quite substantial impact in the East Midlands (+£869) while being diluted by a large negative impact in Wales (–£1,147).[30] ND25Plus customers were also not affected in their probability of earning above the median, while NDLP customers saw a marginal increase of almost four percentage points.

5.1.1 Testing for survey and item non-response using administrative outcomes

The raw and adjusted experimental contrasts in terms of average earnings in the first follow-up year in Table 5.1 are based on the survey sample with non-missing earnings information. Slightly less than half (49 per cent) of the New Deal ERA study participants were randomly selected to take part in the first-year follow-up survey. Not all the selected customers could however be located, accepted to participate, or could be interviewed. Response rates remained high though: 87 per cent among the NDLP and 75 per cent among the ND25Plus fielded samples. Of these

30 Using the survey sample, observables specifically collected for the ERA experimental evaluation and a different weighting scheme, Dorsett *et al.* (2007) find a comparable positive and significant impact on earnings for the NDLP group, but none for the ND25Plus group.

respondents, ten per cent have, however missing information on yearly earnings. Thus, for only one-third of all ERA study participants do we observe earnings (31 per cent in the ND25Plus and 35 per cent in the NDLP group). It thus follows that earnings information is available for one-quarter of the ERA eligibles (23.6 per cent of the ND25Plus and 24.1 per cent of the NDLP eligibles).

The survey sample was randomly chosen, and while there is good evidence (Dorsett *et al.*, 2007, Appendix G) that the respondents to the survey did not differ dramatically from the non-respondents – both in terms of baseline characteristics and administrative outcomes – no analysis has been performed on item non-response, i.e. on those ten per cent of survey sample members who did not respond to the earnings question. In our definition of non-respondents we have lumped survey and item non-respondents, since impact estimates on earnings can only be obtained for our narrower definition of respondents.

In this context, this section 'tests' a number of conditions (discussed in Section 4.1.4) which help us assess whether comparing the average earnings of those with non-missing earnings information among the programme group with their counterparts among the control group would recover the ERA effect on earnings for the full group of participants (ATE_1).

We start by providing supporting evidence for the assumption that randomisation still holds within the group of respondents (the internal-validity condition). If this is the case, the experimental contrast within the subgroup of respondents will still provide an unbiased estimate of the average effect for respondents. Indeed, the rich set of observables has very little power in predicting whether a respondent is a programme or a control group member; their joint significance is rejected at any level.[31] These findings thus provide very strong evidence that the programme and control respondents subgroups are still balanced in terms of observed characteristics, which spells well for unobservables (and hence, for potential outcomes) to be balanced too.

In the following empirical analyses we thus consider the internal-validity condition to be met, and interpret the experimental contrast taken over the respondents as an estimate of the average effect of ERA for the respondents.

Since employment and benefit outcomes from the administrative data are available for **all** participants (respondents or non-respondents), we can use them to test whether the average impact on such outcomes for the responding participants

31 The pseudo-R squared from a Probit regression of random assignment status on the observables for the respondents' subsample is only two per cent for both the ND25Plus and NDLP groups, with the p-value of the likelihood ratio test of the null that the observables are jointly insignificant in predicting random assignment status being 0.175 for the former and 0.495 for the latter. Across districts, the only exception is the ND25Plus group in North East England, for whom the observables are jointly significant at the seven per cent level.

is the same as the average impact for the full group of participants, i.e. whether the external-validity condition holds. For both customer groups, Table 5.2 shows that differences in impacts for all three administrative outcomes are very small and nowhere near statistical significance, both unconditionally and once controlling for observables. (Considering all the districts as well, only two out of the 42 tests were not passed.)

Given the supporting evidence we have found for the internal-validity assumption and the fact that external-validity condition was found to hold in the administrative data, we can safely ignore non-response in calculating the average effect on earnings for participants; in other words, we can take the experimental contrast for respondents, which is readily obtained from the data, as an unbiased estimate of the ERA impact for the full group of participants.

For completeness, Table 5.3 presents the results of testing the stronger set of external-validity conditions. Again, we rely on the administrative data and test whether (possibly controlling for observables), the administrative outcomes of those programme (control) group members who responded to the survey are statistically different from the outcomes of those programme (control) group members for whom we do not observe the survey outcomes for whatever reason – either because they were not selected for the survey, or because they did not respond to the survey, or because they did not respond to the earnings question. This is an (unnecessarily) stricter test, as this external-validity condition for levels implies the external-validity condition for impacts, but not vice versa, and all we need is external validity in impacts. We do, nonetheless, report these results as they are informative in themselves.

For the ND25Plus group overall, there is evidence that non-responding programme and especially control group members spend significantly fewer days on benefits (13.5 and 8.5) during the follow-up year than do responding programme and control group members. For the programme group, such difference is driven by customers in Wales, while for the control group by customers in Scotland, East Midlands and North East England.

Selective outcome differences between respondents and non-respondents in terms of employment outcomes are restricted to customers in the programme group in London, and to a lesser extent in North East England, where in both cases non-responding programme group members experience significantly better employment outcomes than responding ones.

Controlling for our extensive set of background characteristics does not eliminate such differences; in fact, though the main findings within district remain largely unchanged, selective differences in employment probability arise for the ND25Plus group overall, with non-responding members of both the programme and control groups exhibiting three to four percentage points higher employment probability than their responding counterparts with the same characteristics. Similarly, while outcome differences between responding and non-responding NDLP programme/control group members could only be detected within some districts and affecting in particular the control group, once we condition on observables, selective differences appear quite marked for the overall group. Non-responding NDLP customers experience significantly better employment and benefit outcomes than their responding counterparts, with very similar differences within the programme and the control group.[32]

Table 5.2 Testing equality of impacts for responding and non-responding participants

	Ever employed		Days employed		Days on benefits	
	diff	***p*-value**	***diff***	***p*-value**	***diff***	***p*-value**
Unconditional on X						
ND25Plus	0.022	0.218	6.3	0.131	3.4	0.457
NDLP	-0.015	0.413	-0.4	0.944	2.6	0.636
Conditional on X						
ND25Plus	0.016	0.326	5.0	0.187	4.5	0.310
NDLP	-0.009	0.614	3.0	0.515	1.7	0.749

Notes: *diff* is the difference in the average ERA impact for participants compared to the experimental contrast for responding participants; p-value based on bootstrapped significance (500 reps); *** significant at 1%, ** at 5%, * at 10%.
Sample sizes: 5,724 for ND25 Plus and 4,770 for NDLP.

32 This pattern is consistent with selection into survey/item response within experimental group depending partly on unobservables; in such a situation, conditioning on observed characteristics may accentuate, rather than eliminate, outcome differences between responding and non-responding individuals within the two experimental groups.

Table 5.3 Testing equality of outcomes between non-responding and responding programme (1) and control (0) group members

			Unconditional on X		Conditional on X	
	$P_{S=0\|R=1}$	$P_{S=0\|R=0}$	*diff*(1)	*diff*(0)	*diff*(1)	*diff*(0)
ND25Plus						
Ever employed			0.028	-0.004	0.045**	0.038**
Days employed	0.678	0.680	3.138	-6.125	3.958	1.932
Days on benefits			-8.551*	-13.527***	-6.386	-19.953***
NDLP						
Ever employed			0.009	0.033	0.048**	0.050**
Days employed	0.626	0.642	4.597	5.053	15.271***	7.900
Days on benefits			-6.380	-10.207*	-17.154***	-17.819***

Notes: *p*-values based on heteroskedasticity-robust standard error; *** significant at 1%, ** at 5%, * at 10%.
$P_{S=0|R=1}$ is the proportion of non-respondents among the programme group, $P_{S=0|R=0}$ among the control group.
diff(.) is the difference in average outcomes of non-respondents compared to respondents within the programme group (*diff*(1)) or within the control group (*diff*(0)).
Sample sizes: 5,724 for ND25 Plus and 4,770 for NDLP.

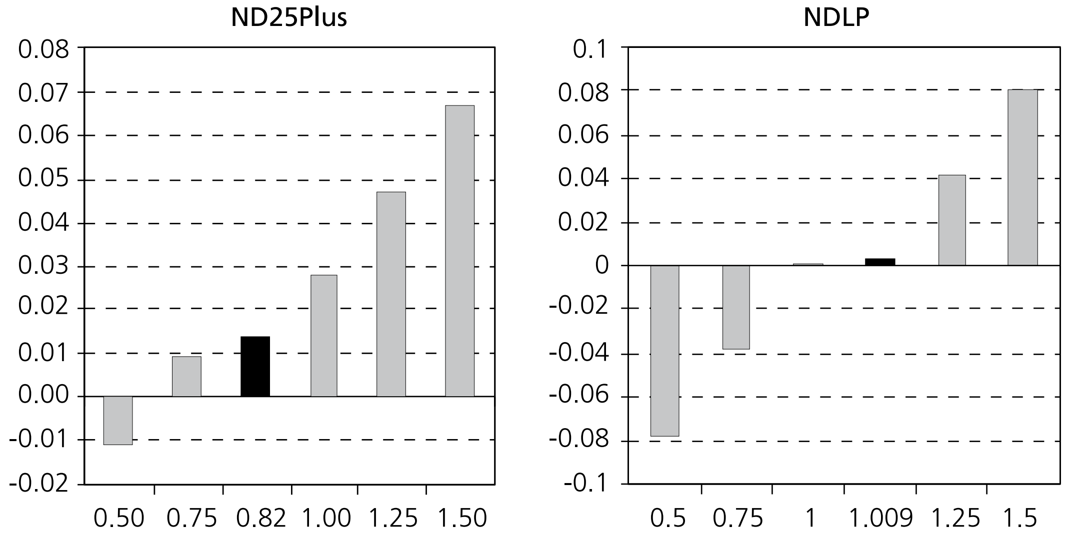
5.2 Bounds

5.2.1 Administrative outcomes

The bounds analysis for administrative outcomes confirms some of the findings to be particularly robust to the non-participation issue, in particular those relating to a combination of low share of non-participants and large experimental estimate, a situation that only applies to within-district results (Scotland for ND25Plus and North West England for NDLP). Conversely, where the share of non-participants is sizeable and the experimental impact negligible, the bounds are very wide. This is indeed the case for the two customer groups overall, for whom the zero impact on employment probability is bounded between -5 and 18 percentage points (ND25Plus) and -16 and 15 percentage points (NDLP).

The sensitivity analysis for administrative outcomes is at times quite informative. Quite in line with the bounds analysis, for some districts it gives some clear indication of the effectiveness of ERA for the whole eligible population. And indeed, for the ND25Plus group overall, the average effect remains positive and small under the most plausible assumptions, in contrast to the NDLP group overall, for whom the ATE could be negative, positive or zero depending on the type of selection mechanism underlying participation in the ERA study (Figure 5.1).

Figure 5.1 Sensitivity analysis: ATE_θ for ever employed, θ from 0.5 to 1.5



In black: experimental impact estimate and corrresponding θ

Another interesting finding from the sensitivity analysis is that the type of assumption (i.e. value of θ) required for the experimental impact to be an unbiased estimate of the average effect for the full eligible population is different for the two customer groups. In particular, in order to ignore non-participation in the NDLP group, one would need to assume a more favourable selection into the ERA study than in the case of ND25Plus.[33]

---

[33] For the ND25Plus customer groups, to take the experimental impact as representative of the impact on the eligibles, one would need to assume that non-participants among the ND25Plus eligibles would have experienced much **lower** employment probabilities had they been offered ERA services than what actual participants receiving ERA are observed to experience (ranging from as low as half in North East England to 93 per cent in North West England and London). Overall, for the experimental estimate to be an unbiased estimate for the *ATE*, the non-participants should have experienced a 20 per cent lower employment probability under ERA as does the actual ERA programme group. For the NDLP customer group by contrast, the experimental estimate would recover the average effect under the assumption that the non-participants did not select into the ERA study based on treatment outcomes; indeed in some districts (North West England and London), non-participants would even need to be those with better ERA outcomes.

5.2.2 Survey outcomes

Based on the evidence of Section 5.1, in calculating bounds for survey outcomes we ignore the issue of potentially non-random non-response, assuming that the experimental contrast calculated on the responding participants consistently recovers the average ERA impact for the full group of participants.

The conclusion that findings relating to a combination of low non-participation and substantial experimental impact are particularly robust no longer necessarily applies to survey outcomes. When analysing survey outcomes in the opposite situation, i.e. one in which the share of non-participants is particularly high and the experimental impact estimate small or negligible, the bounds are by contrast always so wide as to be basically uninformative. This is the case for all estimates except the one for NDLP in North West England and possibly for both NDLP and ND25Plus in Scotland. In all other cases, including for the two overall customer groups, the bounds are totally uninformative.

5.3 Selection on observables

This section reports our impact estimates under the assumption that we observe all outcome-relevant characteristics that drive selection into the ERA study.

5.3.1 Administrative outcomes

Table 5.4 presents the matching results for ND25Plus and NDLP overall, while Appendix A contains all the disaggregated results by district. An overarching comment which applies to the following results is that, provided the selection-on-observables assumption is met, the estimates can be viewed as very reliable, since the matching exercise has performed extremely well in balancing the observable characteristics (see Appendix B).

Starting with the results for the ND25Plus customer group, once we correct for differences in observed characteristics between participants and non-participants in estimating the effect of ERA on non-participants and on the full eligible population, we find that both overall and in all districts except North East England and Wales, non-participants would have experienced a **worse** ERA impact on benefit dependency than participants. In particular, had they been offered ERA services, the group of non-participants taken as a whole would have spent almost nine days longer on benefits (significant at the five per cent level) in the follow-up year than if they had not been offered ERA. By contrast, participants are found to spend a statistically insignificant three days less on benefits thanks to ERA. The ERA impact on eligibles at around 0 is statistically different from the one on the participants. Similarly in the East Midlands and in North West England, the average ERA impact for participants on days on benefits is a statistically significant reduction (14 and 11 days respectively), while for non-participants it would have been a statistically insignificant increase (ten and two days), bringing the ATE for the eligibles to an insignificant reduction (seven and nine days). In Scotland, London and North East England, neither the effect on participants, on non-participants or on the eligibles is significant, although again, in terms of point estimates, the effect on non-participants in the first two districts would have been worse. Wales is the exception, with participants increasing their reliance on benefits due to ERA by a significant 29 days and non-participants' benefit collection not being significantly affected.

In terms of employment outcomes, by contrast, ERA impacts for the non-participants in the ND25Plus group would have been **consistently better** than those experimentally estimated for the subgroup of participants. These findings point to the possibility that the ND25Plus non-participants might in fact be easier to help back into the labour market than the average ND25Plus entrant. To preview the results in Section 5.4.1, the non-participants are more detached from the labour market as well as from the government support system than the control group – experiencing worse employment outcomes as well as fewer days on benefits. The findings in this section might thus indicate that for these more labour-market detached ND25Plus entrants some extra help in the form of advice and financial incentives might be particularly helpful in improving their labour market situation.

The most likely implication of the finding that ERA would have increased both employment durations and time on benefits for the ND25Plus non-participants is that ERA would have reduced the time these customers spend in 'uncompensated' non-employment, i.e. outside the labour market as well as the Government support system.[34]

Moving now to discussing the size of the estimated employment impact, ND25Plus non-participants overall would have enjoyed a highly significant, 5.6 percentage point increase in their follow-up employment probability due to ERA, compared to an insignificant 1.7 increase for participants. The *ATE* for the full group of eligibles would correspondingly have been a significant increase of almost three percentage points. Similarly, non-participants would have enjoyed more than double an increase in days employed (ten) than do participants (4.6), resulting in an overall average impact of six days, all effects being highly statistically significant. This same pattern in terms of both employment probability and employment durations is found for Scotland, North East England, East Midlands and London. In the remaining two districts, North West England and Wales, neither the effects for participants nor for non-participants are statistically significant; interestingly, in North West England the increased size when estimating the *ATE* leads to a

34 Some individuals could still be in work even in the absence of employment records in the available administrative data (the Work and Pensions Longitudinal Study (WPLS)). Note in any case that time in employment and time on benefits are not mutually exclusive (individuals can be employed at the same time as claiming a benefit such as Income Support (IS)); this is particularly the case with the WPLS, which contains no information on the amount of hours worked.

precision gain, with an *ATE* of 7.5 extra days in employment now significant at the ten per cent level.

For the ND25Plus customer group, the experimental impact estimate of ERA thus **underestimates** the contribution that the programme can give to all eligibles in terms of improving their employment outcomes. As we have seen, though, the opposite is true when considering benefit dependency.

Moving now to the NDLP customer group, the findings are somehow less compelling, as it is more difficult to reach statistical significance. Both overall and within most districts, the employment effect in terms of either employment probability or employment duration would have been the same – and statistically indistinguishable from zero – for the non-participants as for the experimental group. One exception is North West England, where ERA is found to significantly enhance participants' employment outcomes (+7 percentage points employment probability and +22 days in employment); non-participants in this district, however, are estimated to be unaffected by ERA (with negative point estimates), resulting in an *ATE* for all eligibles slightly smaller than the experimental estimate (non-participation in this district is only six per cent). The other district that stands out from the general pattern in terms of employment outcomes is East Midlands, where participants experience significantly fewer days (-14.5) in employment due to ERA, non-participants (who represent almost half of all eligibles) would not have been significantly affected by the programme and all eligibles would thus have experienced no significant adverse ERA impact.

As was mostly the case for the ND25Plus customer group, NDLP non-participants would have experienced a worse ERA impact on benefit dependency than the experimental group. Overall, while participants remained unaffected, non-participants would have seen their time on benefits increase by a significant duration of nine days. The same pattern would have happened in the East Midlands, while in all remaining districts except Scotland the pattern only holds in terms of point estimates, i.e. the impact for non-participants appears to be much worse than the one for participants, though neither can be statistically distinguished from zero. In terms of point estimates, the findings for Scotland are interesting and do not follow the general pattern: non-participants would have experienced a much larger and favourable impact from ERA, in terms of both employment and benefit outcomes. However, the fraction of non-participants is very small (five per cent) and 13 per cent of them have been lost to the common support requirement, which is likely to explain why none of the point estimates reaches statistical significance.

Overall, for the NDLP customer group, the experimental estimate of no ERA impact on employment outcomes is thus largely representative of the average effect for all eligibles; the experimental finding however overestimates the impact ERA would have had on all eligibles in terms of reducing their benefit dependency.

Table 5.4 Matching estimates on administrative outcomes

	p	ATE_1	ATE_0	ATE	$ATE_1 \neq ATE$
ND25Plus					
Ever employed	0.230	0.017	0.056***	0.026**	***
Days employed		4.560**	9.984***	5.805***	*
Days on benefits		-2.966	8.862**	-0.250	***
NDLP					
Ever employed	0.304	-0.006	0.015	0.000	
Days employed		-2.208	-1.957	-2.132	
Days on benefits		-5.078	8.881**	-0.831	***

Notes: Statistical significance based on bootstrapped bias-corrected confidence intervals (1,000 replications); $ATE^1 \neq ATE$: bootstrap-based statistical significance of the difference.
*** significant at 1%, ** at 5%, * at 10%.
Sample sizes: 4,831 for ND25 Plus and 4,768 for NDLP.

Sensitivity analysis

This sensitivity analysis relaxes the selection-on-observables assumption (A1) by allowing participants and non-participants with the same observed characteristics to still differ in terms of some unobserved dimension – summarised by θ – that affects their treatment outcome:

$$\text{(A1')} \qquad E(Y_1 \mid Q{=}0, X) = \theta\, E(Y_1 \mid Q{=}1, X)$$

For favourable outcomes such as employment probability or days employed, $\theta>1$ implies positive selection into the non-participants sample (i.e. non-participants would have enjoyed better employment outcomes under ERA than observably similar participants), $\theta<1$ negative selection. For unfavourable outcomes such as days on benefits, the opposite holds, i.e. for $\theta>1$ ($\theta<1$), non-participants would have spent more (fewer) days on benefits under ERA than do participants with the same observables. For $\theta=1$, we obviously obtain the matching estimates discussed above.

In line with the bounds analysis in Section 5.2, the sensitivity analysis in Table 5.5 is quite informative for the ND25Plus group and paints a rather favourable picture for the impact that ERA would have had on all eligibles. In particular, the employment effect of ERA for the eligibles would have been positive, albeit rather small in size (except than under the most extreme selection scenario of θ much larger than one). Similarly, the impact on benefit outcomes for the eligibles would appear to be quite favourable under most selection scenarios.

The robustness analysis points to particularly encouraging impacts of ERA for all eligibles in Scotland, North West England, East Midlands and, for employment outcomes, London. In Wales, by contrast, the evidence on the ATE is rather negative, with ERA appearing to hurt eligibles' employment and benefit outcomes under all scenarios. Finally, the sensitivity analysis for North East England remains rather inconclusive, as whether ERA would have had a positive or a negative impact on the eligibles in that district greatly depends on what one assumes on the direction of selection on unobservables.

In contrast to what we have found for the ND25Plus group, in terms of both employment and benefit outcomes, relaxing assumption (A1) under a number of plausible values for θ does not allow one to say much for the NDLP group, for whom the average impact for all eligibles would range from substantial and negative to substantial and positive.

Exceptions are the positive employment effects uncovered under all the selection scenarios considered in Scotland and North West England, and the mostly negative impacts in Wales and the East Midlands. In terms of benefit outcomes, the sensitivity analysis is even more inconclusive, with only North West England displaying negative ERA impacts for all eligibles (i.e. a reduction in time spent on benefits) under all the scenarios considered.

Table 5.5 also displays the value of θ for which the experimental estimate coincides with the average impact for the whole eligible population.

As to the value of θ required for the experimental impact on employment outcomes to be an unbiased estimate of the average effect for the full eligible population, for the ND25Plus customer group we again find that it is always well below one, and below the corresponding value for NDLP customers. Thus, in order to take the experimental impact as representative of the impact on the eligibles, one would need to assume that non-participants among the ND25Plus eligibles would have experienced much lower employment probabilities and fewer days in employment had they been offered ERA services than what actual participants receiving ERA are observed to experience. This holds both overall and in all districts except North West England, in which the θ corresponding to the experimental estimate would imply no selection into the ERA study based on unobservables affecting treatment employment outcomes.

In terms of benefit outcomes, though, the θ corresponding to the experimental estimate would imply either a favourable selection (overall, North West England and East Midlands) or no selection (Scotland, North East England, Wales, London) into the non-participation sample.

Given this marked divergence (within group and district) in the direction of selection required for the experimental estimate to recover the average effect for employment as opposed to benefit outcomes, such a set of assumptions would seem questionable.

By contrast, for the NDLP group there seems to be more consistency in the requirements imposed on θ for the two types of outcomes. Overall and in Wales, in order to ignore non-participation one needs to assume no selection into the ERA study in terms of employment outcomes, and a slightly unfavourable selection in terms of benefit outcomes (in particular, had they received ERA, non-participants overall would have spent on benefits 93 per cent of the time that participants spend on benefits, and only 85 per cent in Wales). In London and to possibly a lesser extent in North East England, there should have been no selection. In North West England non-participants should have been favourably selected (experiencing both more time in employment and less on benefits), while the opposite should have been the case in Scotland. Only in the East Midlands do the θ's corresponding to the experimental estimate imply a different direction of selection in terms of employment (unfavourable) and benefit (favourable) outcomes.

Table 5.5 Sensitivity analysis: ATE_θ, θ from 0.5 to 1.5

ND25Plus					
Ever employed		Days employed		Days on benefits	
θ	ATE_θ	θ	ATE_θ	θ	ATE_θ
0.50	-0.011	0.50	-0.783	0.50	-30.424
0.75	0.007	0.75	2.511	0.75	-15.337
0.88	**0.017**	**0.91**	**4.560**	**0.96**	**-2.966**
1.00	0.026	1.00	5.805	1.00	-0.250
1.25	0.044	1.25	9.099	1.25	14.836
1.50	0.062	1.50	12.393	1.50	.
NDLP					
Ever employed		**Days employed**		**Days on benefits**	
θ	ATE_θ	θ	ATE_θ	θ	ATE_θ
0.50	-0.081	0.50	-20.027	0.50	-32.977
0.75	-0.040	0.75	-11.079	0.75	-16.904
0.96	**-0.006**	**0.99**	**-2.208**	**0.93**	**-5.078**
1.00	0.000	1.00	-2.132	1.00	-0.831
1.25	0.041	1.25	6.816	1.25	15.242
1.50	0.082	1.50	15.763	1.50	31.315

In bold: experimental impact estimate and corresponding θ.
Missing ATE_θ denotes an inadmissible θ value.
Sample sizes: 4,831 for ND25 Plus and 4,768 for NDLP.

5.3.2 Survey outcomes

Table 5.6 presents our weighting and matching results for survey-based earnings outcomes, where both methods account for non-response.

Table 5.6 Weighting and matching estimates of the average ERA impact on earnings for all eligibles accounting for non-response

	ND25Plus		NDLP	
	Weighting	Matching	Weighting	Matching
ATE	559.9**	580.2***	644.7**	718.2***
$E(Y_1)$	2,772.3	2,779.6	3,557.9	3,509.2
$E(Y_0)$	2,212.3	2,199.4	2,913.2	2,791.1
Δ	393.2*		730.2***	
N	7,399		6,809	

Notes:
ATE is the average ERA impact for all eligibles;
$E(Y_1)$ are average earnings of all eligibles under ERA treatment; $E(Y_0)$ are average earnings for all eligibles without ERA treatment;
Δ is the experimental estimate ignoring potential non-response bias;
Matching estimator: kernel matching with epanechnikov kernel (bandwidth of 0.06), common support imposed separately for each term.
Statistical significance based on bootstrapped bias-corrected confidence intervals (1,000 replications for the weighting estimator, 500 for the matching estimator): *** significant at 1%, ** at 5%, * at 10%.
In this table, Δ is never statistically significantly different from the *ATE* according to bootstrap-based statistical significance of the difference.

Because of (survey and/or item) non-response, in the following discussion the estimated *ATE* for the full eligible population, $ATE \equiv E(Y_1) - E(Y_0)$, has to be compared to the experimental contrast calculated on the responding participants, Δ.

First of all, the evidence emerging from both the weighting and matching estimators tells a pretty consistent story, despite the former estimator's more pronounced sensitivity and difficulty in achieving statistical significance. The point estimates are also quite close, the only exceptions being North East and North West England for ND25Plus customers, where the point estimates of the two methods are very different, though none statistically significantly different from zero.

Although more formal bootstrap-based tests of the difference between the experimental contrast on respondents and the estimated *ATE* fail to uncover any statistically significant difference, the evidence in terms of both point estimates and their statistical significance tells a consistent story: the ERA impact on earnings estimated on the responding experimental group underestimates the average impact of the programme on the full eligible population for the ND25Plus group while being a representative estimate of the full impact for the NDLP group. Specifically, once non-response and non-participation are taken into account, point estimates increase for ND25Plus and remain largely stable for NDLP customers.

Focusing on the matching estimates, the experimental estimator for respondents of an increase in earnings of £393 (significant only at ten per cent) is contrasted to a highly significant estimated increase for all eligibles of £580 for the ND25Plus group. Behind this overall estimate are the positive experimental impact in Scotland increasing in point estimate and becoming statistically significant, the positive and barely significant experimental impact in the East Midlands increasing in both size and significance, and the highly significantly negative impact in Wales decreasing both in size and significance.

As to the NDLP group, as already mentioned, both the point estimates and their significance remain largely stable. The highly significant overall experimental estimate of £730 is in line with a similarly significant estimate for all eligibles of £718. Looking at results by district, the positive and significant experimental estimates for Scotland (+£1,443) and Wales (+£1,080) compare very well with estimates for all eligibles which are only slightly smaller in size (+£1,343 and +£935) but of the same or higher statistical significance. The point estimates for the other districts actually increase in size once non-response and non-participation are taken into account, though they do not reach statistical significance in either case with the exception of North East England, for which the estimate of the *ATE* attains significance at the ten per cent level.

In terms of the underlying matching quality, which can only be assessed for the matching (as opposed to the weighting) estimator, the indicators are very encouraging, with the possible exception of the overall estimate, driven by London (in which case the different indicators disagree as to the extent of matching quality – see Appendix B).

We have also derived and estimated the matching estimates when non-response can be ignored. For convenience of comparison, in Table 5.7 we report again the matching estimates just discussed which allow for non-response.

In the main, the results for the *ATE* ignoring non-response are much closer to the experimental estimates than those allowing for it (our preferred estimates).

For the ND25Plus group, taking account of non-participation but ignoring non-response still raises the positive impact estimates on earnings estimated on the responding experimental sample, but does so by a smaller magnitude than when allowing for non-response (though this only concerns the point estimates; neither of the estimates of the *ATE* are statistically significantly different from the experimental one at conventional levels). The positive effect bordering significance that emerges for Scotland when allowing for non-response no longer arises when it is ignored; the negative impact in Wales is no longer reduced in size; and instead of increasing the size and significance of the positive and already significant impact in East Midlands, only allowing for non-participation brings the estimate down to insignificance (the experimental and matching estimate in this case being also statistically different from one another).

For the NDLP group, the estimates ignoring non-response line up very closely to the experimental ones. Compared to those allowing for non-response, there is a slightly larger, though still minor, fall in the point estimate for the overall group; no longer does a positive significant impact appear in North East England, though the significant and positive point estimate for Wales increases rather than decreases compared to the experimental benchmark. The impact for London increases in point estimate and is significantly different from the corresponding experimental estimate, though none, taken separately, achieves statistical significance.

In the case where non-response is not taken into account, the two different ways of imposing the common support were found to produce strikingly close point estimates and statistical significance, despite the at times large differences in the proportions of the sample being excluded from the analysis.

Table 5.7 Matching estimates of the average ERA impact on earnings for all eligibles

		ND25Plus	NDLP
Δ		393.2*	730.2***
	allowing for non-response, separate CS	580.2***	718.2***
ATE	ignoring non-response, separate CS	442.8*	662.8***
	ignoring non-response, joint CS	443.5*	660.4**
% lost to joint CS		0.8	1.0
N		7,399	6,809

Notes: Statistical significance based on bootstrapped bias-corrected confidence intervals (500 repetitions): *** significant at 1%, ** at 5%, * at 10%.
In this table, Δ (the experimental estimate ignoring potential non-response bias) is never statistically significantly different from the ATE according to bootstrap-based statistical significance of the difference.
Kernel matching with epanechnikov kernel (bandwidth of 0.06).
Separate CS: common support imposed on the non-participants separately for each term; Joint CS: estimates pertain to those non-participants satisfying both support conditions.
When ignoring non-response, Δ is assumed to be equal to ATE_1.

5.3.3 Analysis of take-up

Although as argued in Chapter 2, an analysis of the effect of ERA **eligibility** would need to include the non-participants irrespective of their potential take-up of the programme, it is still very interesting to know the type of involvement they would have had with ERA – and more generally with Jobcentre Plus – had they participated in the evaluation study, either as part of the programme group or of the control group.

Table 5.8 presents the results of these analyses in terms of a number of measures of take-up of services and of contact with Jobcentre Plus staff.

Specifically, we consider

- measures of presence, type and intensity of contact with Jobcentre Plus staff (any contact, customer has initiated face-to-face visits, very intense contact in the form of ten or more face-to-face meetings);
- measures of help or advice received from Jobcentre Plus staff when the customer was not working (staff offered any help/advice, performed a better-off calculation, suggested customer attend a Jobclub/Programme Centre, arranged an education or training course, offered advice without being requested);

- measures of the customer's assessment of the advice received; and
- for the programme group analysis only, measures directly linked to knowledge of ERA features (whether the customer has heard of the employment and of the training bonuses).

Recall from Section 4.3 that all results hinge on the assumption that there is no selection into the ERA study based on **unobserved** characteristics that also affect take-up of ERA services or involvement with Jobcentre Plus if participating in the study. Subject to this proviso, the findings provide interesting evidence on the two sets of questions we consider.

First we focus on the take-up that the non-participants would have exhibited had they been assigned to the programme group. Are the non-participants individuals who even if offered ERA services would not take them up? And could this be the underlying reason for Jobcentre Plus caseworkers not offering them the chance to participate in the randomisation in the first place, or, for those who were offered such a chance, the reason driving their own refusal to participate in the demonstration? If this is the case, one might argue that even if ERA became an official policy, they would not be interested in effectively taking up the support and incentives it offers.[35]

For the ND25Plus group, there are statistically significant differences between the non-participants and the programme group in two measures of involvement with Jobcentre Plus staff and in terms of awareness of the ERA bonuses, but such differences are not striking. Specifically, while 85 per cent of the programme group has received help or advice from Jobcentre Plus staff while not working, our model predicts that 82.5 per cent of the non-participants would have received such help had they been assigned to the programme group. Similarly, the non-participants would have a two percentage point lower likelihood than the programme group of being offered help by staff without being requested. Non-participants would also have been less aware of the bonuses than the actual programme group is (72.9 per cent rather than 75.4 per cent for the employment bonus and 40.1 per cent rather than 43 per cent for the training bonus).

Overall, had they been randomised into the programme, the ND25Plus non-participants would have been quite heavily involved with ERA and Jobcentre Plus. And although we find that they would have been statistically significantly less aware of ERA features and would have experienced slightly less contact than the actual programme group, such differences are arguably small from a substantive point of view.

---

[35] Again note that if some eligibles are not fully informed about ERA or do not otherwise avail themselves of its services, they will dilute the effect of ERA eligibility on the eligibles.

The conjecture that if the programme became official, non-participants would be mostly uninterested in taking up its support and incentives finds no strong support for the NDLP group either. In fact, had they become eligible to ERA services and incentives, the non-participants would have been over three percentage points more likely than the programme group to be involved in training and education activities arranged by Jobcentre Plus, as well as more likely to be directed to a Jobclub or Programme Centre. The two groups are not found to differ significantly in any other measure of awareness and involvement, with the notable exception of the likelihood to receive help or advice from Jobcentre Plus when not working. As was the case for ND25Plus customers, it is again the programme group who is 2.4 percentage points more likely to receive such help than the non-participants. As many as 75 per cent of the latter are however still predicted to receive such support when out of work.

The second question we have looked at concerns the kind of involvement that non-participants would have had with Jobcentre Plus had they participated in the ERA study and been assigned to the control group. Among the reasons that the qualitative research has highlighted for ND25Plus customers to formally refuse to participate, there was a feeling of being close to getting a job in the near future and not wanting to stay in touch with Jobcentre Plus, or a strong antipathy to government and systems of support and governance. The question thus arises of whether the ND25Plus non-participant group is made up of individuals who would shun involvement with Jobcentre Plus at all cost. This supposition is not borne out in the data: had they been assigned to the control group, the involvement that the ND25Plus non-participants would have had with Jobcentre Plus would not have been statistically different from the one displayed by the actual control group in any of the dimensions considered.

As opposed to ND25Plus customers, NDLP customers were easy to recruit to the ERA study once having been offered the chance to participate in it. In fact, most (87 per cent) of the non-participants among the NDLP group were diverted customers. One might thus conjecture that had they been offered the chance to participate, the NDLP non-participants would in fact have been quite involved with Jobcentre Plus even if assigned to the control group. According to the results in Table 5.8, this seems to be the case. Indeed, it is estimated that compared to the control group, NDLP non-participants would have had the same type and intensity of involvement with Jobcentre Plus staff, while being four percentage points more likely to rate their advice as very helpful.

Overall, the share of the eligible population that has been excluded (i.e. the diverted customers) or has formally refused to take part in the ERA study displays observed characteristics that make them quite likely to be involved with Jobcentre Plus generally, both with and without ERA.

Table 5.8 Take-up and involvement with Jobcentre Plus predicted for the non-participants both under ERA and without ERA

	ND25Plus				NDLP			
	ERA outcome		Non-ERA outcome		ERA outcome		Non-ERA outcome	
	Programme group	Non-participants	Control group	Non-participants	Programme group	Non-participants	Control group	Non-participants
Has had contact with Jobcentre Plus staff	84.8	83.7	78.2	78.2	85.3	86.4	71.9	74.6
Has ever initiated face-to-face visits	55.4	54.5	50.4	49.7	62.0	61.3	55.5	56.5
Had face-to-face contact with Jobcentre Plus staff ≥10 times	43.0	43.5	41.0	42.1	14.2	15.5	9.8	9.1
Received help/advice from Jobcentre Plus staff when not working	85.0	82.5***	84.9	85.8	77.2	74.8*	73.7	71.2
Jobcentre Plus staff did better-off calculation when not working	41.6	41.0	38.6	39.4	63.8	63.2	64.2	64.7
Jobcentre Plus staff suggested attend a Jobclub/Programme Centre	32.7	34.3	32.9	35.2	5.3	6.6*	6.2	7.1
Jobcentre Plus staff arranged education/training	30.4	31.3	31.5	31.4	14.6	17.8***	12.3	14.0
Jobcentre Plus staff offered help/advice without being requested	18.4	16.2**	7.8	7.9	26.3	27.6	9.4	9.9
Found advice from Jobcentre Plus staff overall very helpful	33.1	31.2	23.6	22.8	42.6	43.2	31.1	35.1**
Found advice from Jobcentre Plus staff overall not at all helpful	4.7	5.0	5.8	5.2	3.4	2.5	4.1	3.7
Has heard of employment bonus	75.4	72.9**	–	–	72.8	71.0	–	–
Has heard of training bonus	43.0	40.1**	–	–	50.8	52.9	–	–
N	1,014	1,675	996	1,675	1,014	2,039	994	2,039

Note: Programme group and control group columns report the observed rates; non-participants columns report the predicted rate for participants under ERA and without ERA.

Statistical significance of the difference in rates between non-participants and programme (or control group) is based on bootstrapped bias-corrected confidence intervals (500 replications): *** significant at 1%, ** at 5%, * at 10%.

Note: the sample sizes shown for the programme and control groups refer to those with non-missing information on 'has had contact with Jobcentre Plus staff'.

5.4 Selection on unobservables

Before presenting and discussing in Section 5.4.3 the findings from the different control function models we have estimated, this section reports the results from testing for selection in terms of no-treatment unobservables (Section 5.4.1) and presents the chosen instrument together with evidence on its power and validity (Section 5.4.2).

5.4.1 Testing for selection on specific unobservables

In Section 4.4.1 we have suggested a simple way to test for the presence of residual selection into the ERA study based on unobservables related to no-treatment employment and benefit outcomes. Specifically, this involves assessing whether, once controlling for observable characteristics, the non-ERA outcomes of the participants differ on average from those of the non-participants. Table 5.9 reports the results of this test.

The evidence is exceptionally robust to the specification of the regression function, with simple Ordinary Least Squares (OLS), fully interacted regression, non-parametric matching and Tobit or Probit to take account of censoring or of the binary nature of the outcome variable painting the same picture in terms of selection into the ERA study based on unobservables.

Interestingly, non-participants are subject to the same type of selection on unobservables in both customer groups. Specifically, non-participants overall have unobservables leading them to experience worse (non-ERA) employment outcomes but fewer days on benefits than observationally similar participants.

For both customer groups, this overall result is driven by East Midlands. For both New Deal groups, non-participants in North East England have unobservables that cause them to experience fewer days on benefits but the same employment outcomes as participants, while non-participants in London have not been subject to any residual selection on unobservables. No selection on unobservables has taken place for ND25Plus non-participants in Scotland and Wales as well.

Overall, when selection on unobservables has been uncovered, the picture that emerges is one of worse employment outcomes and less dependence on benefits for non-participants.[36] Non-participants thus seem to be less attached to the labour market as well as to the government support system than participants – experiencing shorter employment durations and a smaller incidence of employment as well as fewer days on benefits. The unobservables characterising the non-participants might thus relate to those more on the fringe of both the labour market and the benefits system.

---

36 The only exceptions are NDLP non-participants in North West England (better employment outcomes) and in Scotland (more days on benefits). In these two districts, though, non-participants are very few in absolute terms and as percent of the eligibles (see Table 3.3).

Table 5.9 Differences in outcomes for participants (control group) compared to non-participants with the same observed characteristics

Outcome	Method	α	Given observables, participants
ND25			
Days employed	OLS	7.9***	spend more days employed
	FILM	9.8***	
	Matching	9.5***	
	Tobit	30.4***	
Ever employed	OLS	0.044***	are more likely to be employed
	FILM	0.057***	
	Matching	0.056***	
	Probit	0.057***	
Days on benefits	OLS	10.1***	spend more days on benefits
	FILM	9.7**	
	Matching	9.2**	
	Tobit	14.7**	
NDLP			
Days employed	OLS	10.3***	spend more days employed
	FILM	11.5***	
	Matching	11.4**	
	Tobit	21.7***	
Ever employed	OLS	0.042***	are more likely to be employed
	FILM	0.045***	
	Matching	0.041**	
	Probit	0.055***	
Days on benefits	OLS	8.2**	spend more days on benefits
	FILM	9.3**	
	Matching	9.6*	
	Tobit	10.3	(not really according to Tobit)

Significance based on robust standard errors for OLS and FILM, and on approximate standard errors for kernel matching. ***: significant at 1%,**: at 5%, *: at 10%.
Sample sizes: 4,755 for ND25 Plus and 4,702 for NDLP.

5.4.2 Instrument

This section motivates our choice of instrument and presents evidence on its power and validity.

Conditional on the observables, the instrument should affect the probability to participate in the ERA study but should not directly affect ERA outcomes, nor, in case of survey outcomes, non-ERA outcomes.

Motivated by the idea that for both customer groups the observed fall over time in the non-participation rates is likely to reflect increased adviser experience and confidence in selling ERA, as well as the permission to mention ERA financial incentives (see Goodman and Sianesi, 2007), we have chosen to use elapsed days since random assignment started in an individuals' district and for that individual's customer group. This measure is thus relative to random assignment start in each district and for each New Deal customer group, and is conditional on controlling for calendar time using five-month dummies.[37]

This instrument based on the increased persuasiveness of the advisers and the greater promotion of the ERA bonuses indeed looks like a very promising one, both in terms of its relevance and validity (see Appendix C). Specifically, for both customer groups and for both the full sample and for the survey-eligible sample, it displays a very powerful first stage (that is, it greatly contributes in explaining whether an eligible individual participates in the ERA study or not) and it passes our exclusion restriction test at any significance level.

5.4.3 Control function models

In the following we present the results from our extensive search of an appropriate control function model to take account of the residual selection uncovered in Section 5.4.1. To preview our conclusions on this part of our analysis, unfortunately none of the various models we have implemented pass our strict specification tests. The unsatisfactory performance of this class of models in our application calls for careful interpretation of these results, which can only be viewed as indicative.

Administrative outcomes

Tables 5.10 and 5.11 present our findings from the four types of control function models we have implemented: the standard model, a model where independence of the observables and the error terms is relaxed, a model where normality of the unobservables is relaxed, and a control function model embedded in a tobit model to explicitly take account of censoring in the outcome variable. For the first three models we have performed different estimations including non-linear and interaction terms in the first-stage probit ('interactions') or not ('no inter').

For both customer groups, the first three models, interacted or not, yield estimates of the average ERA impact for non-participants as well as for all eligibles that are not statistically different from the average impact for participants. Specifically, for ND25Plus customers, they predict a statistically insignificant average impact on employment durations for the non-participants (with one exception of a large positive impact significant at the ten per cent level), compared to a small, barely significant impact of four days for the participants. Overall, the impact of the eligibles

[37] Originally we had explored the possibility of using a series of individual office dummies, within district and controlling for important local and office characteristics such as travel-to-work-level unemployment rate, local index of multiple deprivation, total New Deal caseload at that office and share of lone parents in New Deal caseload at that office. Interestingly, except than for some districts (e.g. Scotland and London), this instrument does not pass the exclusion restriction test.

in all but one case is again undistinguishable from zero. As mentioned, however, we cannot statistically distinguish between the impact for the participants, for the non-participants and for the eligibles. The NDLP case is even more clear-cut: for this customer group, no effect (for participants, non-participants or all eligibles) is ever statistically different from zero. In conclusion, for both customer groups these first three models imply that the experimental impact is **representative** of the impact that the full eligible population would have experienced, on average, had they been offered ERA services and incentives.

Almost all of these models perform well in terms of our two criteria and in terms of recovering an average impact for participants which basically coincides with the one estimated on the experiment groups using linear regression.[38] These models however suffer from two shortcomings. First, we obtain very noisy estimates, both of the selection terms (never statistically significant) and of the predicted impacts (again never significantly different from zero despite at times large point estimates). Low precision of the estimates is in fact a problem often encountered in the estimation of control function models (see e.g. Blundell *et al.*, 2005). Second, we have seen that participants spend eight (or 17 per cent) more days in employment compared to observationally-equivalent non-participants in the ND25Plus group, and ten (or nine per cent) more days in the NDLP group, these differences being highly statistically significant. The terms for selection on unobserved characteristics should thus be positive and significant for both customer groups, but none of the three models manage to detect this statistically significant amount of selection.

In this dimension, the model allowing for censoring performs very well: for both the ND25Plus and NDLP groups, it detects highly statistically significant, positive selection on unobserved characteristics. This model however performs extremely poorly in terms of our two criteria, and when applied to the ND25Plus group, also in terms of matching the observed outcomes of the programme groups as well as capturing the average effect for participants.[39] Furthermore, the estimated impacts appear implausibly large: an ERA impact on days in employment of 268 days for ND25Plus non-participants and of 153 days for NDLP non-participants. The average effect for all eligibles is correspondingly large and significant (100 for ND25Plus and 45 for NDLP). These large effects are of course statistically different from the small (ND25Plus) or zero (NDLP) effect for participants. We can only conclude that the effect for non-participants and all eligibles would have been larger than what the experimental estimate reveals for the participants. Given how inadequately the censored model performs though on our two criteria, such findings should be interpreted with extreme care.

38 The only exception is the model relaxing normality estimated for the NDLP group. For both interacted and non-interacted versions, neither of our two criteria is passed.

39 The average impact for participants of 49.4 days – to be compared to the experimental estimate of 4.4 days – is calculated in the control function model by subtracting the average non-ERA outcome predicted for all participants from the average ERA outcome again predicted for all participants.

Table 5.10 Control function models: ERA impacts on days in employment accounting for selection on unobservables – ND25Plus

Selection: α	(OLS)	8.0***	(Tobit)	8.5***
ATE_1	(OLS)	4.4*	(Tobit)	3.7
Non-participants' observed Y		47.3		
N	7,796			

Results from the control function models

	Standard		No independence		No normality		Censoring
	No interactions	Interactions	No interactions	Interactions	No interactions	Interactions	No interactions
Selection on unobserved characteristics			None				yes: ρ_{uv} = 0.947***
Selection on unobserved ERA gains			None				yes: negative*
α – selection terms	-30.6	-12.2	159.4	45.1	-8.0	-46.5	295.1***
Q=0: observed–predicted Y	-53.4	-52.0	-168.0	-49.7	-0.8	41.4	-287.4**
Independence: Y_0			Yes	No*			
Independence: Y_1			Yes	Yes			
Normality: Y_0					Yes	Yes	
Normality: Y_1					Yes	Yes	
R=0: observed–predicted Y							9.6
R=1: observed–predicted Y							-82.2***
$E(Y\|R=0)–E(Y_0\|R=1)$							10.1
ATE_1	4.4*	4.4*	4.5*	4.5*	4.4*	4.4*	49.6***
ATE_0	-30.9	-11.4	-103.2	-18.8	242.7	157.0*	268.5**
ATE	-3.7	0.8	-20.2	-0.8	59.1	39.5*	99.8***
$ATE_1–ATE$		Never statistically significantly different from one another					-50.3**
$ATE_1–ATE_0$		Never statistically significantly different from one another					-218.9**

Note: Statistical significance based on bootstrapped bias-corrected confidence intervals (500 replications): *** significant at 1%, ** at 5%, * at 10%.

Table 5.11 Control function models: ERA impacts on days in employment accounting for selection on unobservables – NDLP

Selection: α	(OLS)	10.4***	(Tobit)	12.3***
ATE_1	(OLS)	-2.3	(Tobit)	-2.7
Non-participants' observed Y		119.8		
N	7,261			

Results from the control function models

	Standard		No independence		No normality		Censoring
	No interactions	Interactions	No interactions	Interactions	No interactions	Interactions	No interactions
Selection on unobserved characteristics			None				Yes: ρ_{uv} = 0.880***
Selection on unobserved ERA gains			None				No
α – selection terms	2.7	14.5	178.5	66.1*	301.6**	137.5***	188.4***
Q=0: observed–predicted Y	29.7	-8.8	-189.8	-63.2*	-327.5*	-145.7**	-168.3***
Independence: Y_0			Yes	Yes			
Independence: Y_1			Yes	Yes			
Normality: Y_0					No*	No*	
Normality: Y_1					Yes	Yes	
R=0: observed–predicted Y							5.4
R=1: observed–predicted Y							5.3
$E(Y\|R=0)–E(Y_0\|R=1)$							2.6
ATE_1	-2.3	-2.3	-2.0	-1.9	-2.3	-2.3	-2.7
ATE_0	-9.6	3.7	-46.2	5.1	-80.0	43.4	153.3***
ATE	-4.5	-0.5	-15.4	0.2	-26.0	11.6	44.9***
$ATE_1–ATE$	Never statistically significantly different from one another						-47.6***
$ATE_1–ATE_0$	Never statistically significantly different from one another						-156.0***

Note: Statistical significance based on bootstrapped bias-corrected confidence intervals (500 replications): *** significant at 1%, ** at 5%, * at 10%.

Survey outcomes

The data and identification requirements that modelling survey outcomes such as earnings impose are quite demanding. First, both treatment and no-treatment outcomes need to be predicted for the non-participants. Second, the observed characteristics themselves turned out to have at times very low predicting power in modelling earnings. In particular, OLS regressions of earnings on the administrative variables separately for the programme group and for the control group, show that hardly any observable, or group of observables, is significant. Indeed, for most districts, one cannot reject that all the regressors are jointly insignificant.[40] An important lesson in this context is thus that the available administrative data, however rich in detailed employment histories and other background information, does not appear adequate enough to explain earnings outcomes for the two New Deal groups examined.[41]

As was the case for employment outcomes, neither the standard model, nor the ones relaxing normality or heteroskedasticity show any significant selection on unobservables, and this, once again, despite a strong first stage of the instrument. We cannot directly test whether there was selection into the ERA study based on unobservables related to **earnings**. Indirect evidence is, however, provided by our tests concerning unobservables related to employment and benefit outcomes: given our findings of relatively strong selection in terms of such unobservables, the presence of selection on earnings-related unobservables would seem plausible. In any case, these control function models reject such selection and would thus lead us to use simple regression.

In addition, in most cases these models predict **average** treatment and/or no-treatment earnings for the non-participants that are negative. While the aforementioned low predictive power of the observables is likely to be a contributing factor, negative average predictions also point to some mis-specification problem and the corresponding need to properly account for the censored nature of the outcome variable.

We thus do not report results for these models in terms of earnings, and in Table 5.12 just focus on the model taking censoring into account.

Interestingly, the model uncovers positive selection on unobserved characteristics for the ND25Plus group (i.e. participants have higher earnings-related unobservables than do non-participants), but no such selection for the NDLP group. Furthermore, the model finds that there has been selection into the ERA study based on

---

40 The same conclusion applies if using a Tobit specification to account for the spikes at zero.

41 Nor does survey information from the Basic Information Form (BIF) seem to help in modelling earnings. By contrast, employment outcomes are well modelled using administrative data, in fact even better than when using survey (BIF) data.

unobserved ERA impacts: ND25Plus participants appear to have lower unobserved ERA components than do non-participants, while the (more plausible) reverse is the case among the NDLP customers.

Since we do not observe the earnings outcomes for the non-participants, we cannot check the performance of this model against our two criteria. However, we can assess how well this complex model manages to predict the observed earnings of the programme and of the control group, as well as the average impact for (responding) participants. Formal tests show that the model cannot adequately predict observed outcomes, this being especially the case for the programme group. It can however satisfactorily match the average observed earnings of the control group as an estimate of earnings that the programme group would have experience if it had not been offered ERA services and incentives. Similarly, its estimates of the effect for responding participants (£488 and £847) are not too far off from the experimental ones using OLS (£405 and £746).[42]

The model allows us to recover estimates of the average effect for the non-responding participants. Though substantially higher for the ND25Plus group, the effects for non-responding and responding participants are not statistically different from one another for both customer groups. For the whole group of participants, the model thus estimates an average increase in earnings due to ERA of around £670 for the ND25Plus group and £820 for the NDLP group.

Finally, the model provides us with an estimate of the average impact on earnings that the non-participants would have experienced, on average, had they been offered ERA services and incentives. The estimates are implausibly large: a **decrease** in earnings of about £5,500 for the ND25Plus group and an increase of £6,170 for the NDLP group. The separate estimate of the ERA and non-ERA earnings for the non-participants shed some light on these implausible results: ND25Plus non-participants are estimated to earn on average over £11,700 per year without ERA, while NDLP non-participants are predicted to earn on average over £10,400 had they been offered ERA. These predicted outcomes are of course implausible and completely out of line with the observed ones for the participants (around £2,600 for ND25Plus and £3,300 for NDLP participants), and reiterate the difficulty we have found in satisfactorily modelling the earnings of the New Deal groups using the available administrative (or even survey) data. More generally, control function models rely on a complex structure, and often suffer from imprecisely estimated and unstable parameters. Blundell, Dearden and Sianesi (2005) suggest a practical way to exploit the control function approach where suitable instruments can be found. Specifically, by contrasting the relative magnitude of the estimates that result from applying different approaches to a common dataset, one can try to infer what kind of selection and outcome models underlie the data. In this

[42] It might, however, be more appropriate to assess the censored (Tobit) control function model against the experimental estimates using a Tobit model. In this case, the control function estimates do not line up very well with the experimental Tobit estimates of an insignificant £241 and £597.

context, the role of the control function model is thus to provide us with the basis for assessing the validity of assumptions on selection (i.e. residual selection on unobserved individual characteristics and on unobserved ERA impacts), as well as with an indication of the direction of bias when ignoring such selection.

Following this logic, the only overall conclusion one can take away – again with great care – from this model is that the experimental impact might be overestimating the impact for the non-participants for the ND25Plus group, while underestimating it for the NDLP group.

Table 5.12 Censored control function model: ERA impacts on earnings accounting for selection on unobservables

Effect for responding participants (ATE_1_S)

	ND25	NDLP
OLS	404.6*	745.8***
Tobit	241.2	597.2***
N	7,796	7,261

Censored control function model results

		ND25	NDLP
Selection on unobserved characteristics		Yes: ρ_{uv} = 0.936***	No
Selection on unobserved ERA gains		Yes: negative***	Yes: positive***
R=0: observed–predicted Y		-70*	-77**
R=1: observed–predicted Y		-200*	-229**
$E(Y\|R=0)–E(Y_0\|R=1)$		-88	-136
ATE_1_S	Effect for responding participants	488**	857***
ATE_1_NS	Effect for non-responding participants	753***	800**
$ATE_1\_S – ATE_1\_NS$	Difference in effect for responding and non-responding participants	-265	57
ATE_1	Effect for (all) participants	668***	821**
ATE_0 =	Effect for non-participants =	-5,511**	6,173***
$E(Y_1\|Q=1)$ –	Average predicted earnings for non-participants under ERA	6,222	10,422
$E(Y_0\|Q=0)$	Average predicted earnings for non-participants without ERA	11,733	4,249
$ATE_1–ATE_0$	Difference in effect for all participants and for non-participants	6,179**	-5,352***
ATE	Effect for all eligibles	-731	2,424***
$ATE_1–ATE$	Difference in effect for all participants and for all eligibles	1,399**	-1603***

Note: The control variables X we use in these models are marginally different from the full set, which accounts for the slight differences in the benchmark OLS estimates. Statistical significance based on bootstrapped bias-corrected confidence intervals (500 replications): *** significant at 1%, ** at 5%, * at 10%.

6 Summary and conclusions

6.1 Drawing the findings together

We start this concluding section by drawing together the findings from the different types of analyses we have performed, the results of which are summarised in Table 6.1 for the two customer groups overall (district-level summaries are contained in Appendix A).

- We have found that the New Deal 25Plus (ND25Plus) experimental sample is composed of adviser- and self-selected individuals with better employment outcomes than the population of ND25Plus entrants. By contrast, the experimental New Deal for Lone Parents (NDLP) group is made up of somewhat lower performers than the average NDLP entrant. Once we net out the contribution of observable individual characteristics such as extensive labour market histories, we find that in the absence of Employment Retention and Advancement (ERA) the study participants of both customer groups experience better employment outcomes than non-participants while relying more extensively on benefits. Non-participants are thus characterised by unobservables that make them more detached from the labour market as well as from the government support system than participants.

- Given the extent of non-participation (with almost one-quarter of the eligibles not participating in the ERA study) as well as such important selective differences between study participants and non-participants, we have extensively explored whether and how much the experimental impact estimates are representative of the potential impact of offering ERA services and incentives to the population of New Deal entrants, that is, to the full group of ERA eligibles, in the six evaluation districts.

- This has necessarily involved invoking a number of suitable assumptions and using a range of techniques to estimate the likely impact that the non-participants would have experienced, on average, had they participated in ERA. Based on extensive diagnostic and specification tests, as well as on contrasting and cross-checking the findings and evidence from the different methodological approaches, we have come to the following conclusions.

- The control function approach ('selection on unobservables analysis') has produced extremely sensitive, unstable and imprecise estimates. Based also on the specification tests we could devise thanks to our unique set-up and data, such results have to be viewed with extreme care, and as indicative at most. Particularly suspect are those related to earnings outcomes: the available individual data was found to be totally inadequate in modelling this outcome for the two New Deal groups of interest.[43]
- By contrast the most robust findings – based on our diagnostic checks and confirmed by our sensitivity analyses – are those arising from the matching and weighting estimates ('selection on observables analysis').[44] The picture that emerges within this framework is as follows:
  - For employment outcomes, the overall experimental impact estimate excluding the non-participants coincides with the average impact ERA would have had on the full population of eligibles for the NDLP group. Specifically, no impact was found for the experimental sample either on employment durations or on the probability of being employed during the follow-up year, and the absence of any impact would have extended to the non-participants, and hence, to all eligibles. For the NDLP group, there is thus no heterogeneity in impacts between participants and non-participants: for both subgroups, the point estimates are close to zero (or literally zero) and certainly never statistically significantly different from zero.

---

43 Indeed, the inability of administrative (or survey) individual data of explaining yearly earnings for the ND25Plus and NDLP groups is among the interesting peripheral findings that have emerged as we performed our main analyses. Another of such findings is the invalidity as an exclusion restriction of office affiliation, often used for such purposes.

44 Of course, finding no selection into the ERA study based on unobservables related to employment and benefit outcomes in the absence of ERA would have greatly increased our confidence in the results based on the selection-on-observables assumption. We reiterate, however, that such an assumption is not *per se* invalidated by the actual finding that non-ERA employment (benefit) outcomes are worse (better) in unobserved dimensions for the non-participants than for the participants. This is because matching and reweighting only need to rely on observables to recover estimates of participants' employment and benefit outcomes **under** ERA.

– By contrast, ignoring the participation decision significantly biases the effect of ERA for all eligibles for the ND25Plus group, in the sense that the average effect for all eligibles is statistically different from – and larger than – the experimental estimate. This result is driven by the fact that the effect for the non-participants would have been considerably larger than the one for the participants. This might indicate that for the more labour-market detached ND25Plus entrants (i.e. the non-participants) some extra help in the form of advice and financial incentives might be particularly helpful in improving their labour market situation. Specifically, compared to no effect on employment probabilities for the participants, non-participants would have enjoyed an almost six percentage points increase. Similarly, compared to an increase in employment durations of 4.5 days for the experimental sample, the non-participants would have enjoyed a ten day increase. However, even though the ERA effects for all eligibles and for participants are statistically different, some might not regard them as qualitatively that different: compared to no impact for the participants, the eligibles would have experienced a rather small increase in employment probability of 2.6 percentage points, and compared to the 4.5 days' increase in durations for participants, the eligibles would have experienced a 29 per cent higher increase of 5.8 days.

– In terms of benefit outcomes, the story is the same for the two customer groups. First, compared to a non-significant reduction for participants of three and five days in time spent on benefits, non-participants would have experienced a significant nine days' increase had they been offered ERA services and incentives. Second, impact estimates for all eligibles are statistically significantly different from the experimental estimate. Third, these estimates do not however tell a **qualitatively** different story, as the impact of ERA on days on benefits for either the experimental sample or the full group of eligibles is not statistically or economically significant. For both the ND25Plus and NDLP eligibles, the point estimate becomes literally zero.

– As far as earnings are concerned, we find that for both customer groups the impact for the responding study participants is representative of the effect on the full eligible population, in the sense that formal bootstrap-based tests of the difference between the experimental contrast on respondents and the estimated impact for all eligibles fail to uncover any statistically significant difference. However, for the ND25Plus group the qualitative evidence is that the impact on earnings for the responding experimental group (an increase of £393 significant only at the ten per cent level) actually **underestimates** by almost 50 per cent the average ERA impact for the full eligible population (a highly significant increase of £580). Note however that all the earnings estimates have in any case to be taken with extra care, as both ERA and non-ERA earnings had to be predicted for the non-participants, and on the basis of observed characteristics alone.

- Finally, we have assessed the conjecture that if the non-participants had been offered ERA, they would have been mostly uninterested in effectively taking up its support and incentives. We have found no support for this hypothesis for either customer group. In fact, the results show that overall, the eligibles who have been excluded or have formally refused to take part in the ERA study display observed characteristics that make them quite likely to be involved with ERA and with Jobcentre Plus more generally. Specifically, had they been randomised into the programme, the non-participants would have been less aware of ERA or less involved with staff than the programme group only in terms of a couple of measures, and then only marginally. Indeed, had they become eligible to ERA services and incentives, the NDLP non-participants would have been over three percentage points more likely than the programme group to be involved in training or education activities arranged by Jobcentre Plus, as well as **more** likely to be directed to a Jobclub or Programme Centre. Had they been randomised into the control group, NDLP non-participants would have been four percentage points **more** likely than the actual control group to rate advice from Jobcentre Plus staff as very helpful.

6.2 External validity of the experimental findings

How has the presence of the non-participants thus affected the external validity of the experimental impact estimate?

In our descriptive examination of the non-participation problem (Goodman and Sianesi, 2007), we speculated that it would be hard for the non-participants to give rise to an estimate for all eligibles that tells a different 'story' from the one arising from the experimental estimate (where the 'story' could be one among: ERA is harmful, it has basically no effect, it has a 'relatively small' effect or it has a 'relatively large' effect – whatever one may mean with 'relatively large' or 'relatively small').

Indeed, we have found that **the story does not change much** – in statistical as well as qualitative terms.

Indeed, for the NDLP group **the story remains unchanged**. Specifically, the bottom-line in the first-year follow-up is that ERA has, so far, had no effect on employment and benefit outcomes, while it has significantly and substantially increased their yearly earnings. Furthermore, what the programme has done for the participants, it would have done also for the non-participants and hence for the whole eligible population. Interestingly, this overall conclusion applies within district as well.

For the ND25Plus group, **the story changes somewhat** in the direction of a slightly more effective ERA treatment: positive impacts surface, become larger in size or stronger in statistical significance. Specifically, while no significant impact has been detected on the employment probability of participants both overall and in any of the districts, statistically significant effects for all eligibles emerge overall and in Scotland, East Midlands and London. As mentioned above, the treatment effect on employment durations for all eligibles is 29 per cent higher than the one obtained using the experimental sample, while the ERA effect on earnings would have been 48 per cent higher for all eligibles than it is for the study participants. No significant earnings effect could be detected in Scotland for the participants, while eligibles are estimated to enjoy a £620 increase in their yearly earnings under ERA that attains the ten per cent confidence level. Similarly, the positive, but barely significant, earnings impact in East Midlands becomes highly significant and increases by five per cent in magnitude. Another interesting finding in this 'estimates-get-better' direction is that the negative and large impact on earnings for participants in Wales decreases in statistical significance and by 15 per cent in magnitude from a £1,147 reduction significant at the five per cent level to a £967 reduction significant only at the ten per cent level. The only exception to estimates being either unchanged or improved concerns impacts on benefit receipt in North West England and East Midlands, where the significant reduction for participants drops into non-significance when considering all eligibles.

We thus do find evidence of non-participation bias (or of some loss in external validity) in the data for the ND25Plus group. When we adequately account for non-participation, we find that the employment and earnings impact estimates that rely on experimental data alone underestimate the true impact of ERA on all ND25Plus entrants, both overall and in several districts. Of course, there is always the issue of how different the estimates for the eligibles and for the experimental sample need to be for us to view the issue as a particularly important one. Randomised experiments are however conceptually designed to provide with accuracy the 'true' answer to the evaluation question. Finding an effect for the eligibles which is 30 or 50 per cent larger (or 15 per cent smaller) than the experimental estimate can be viewed as a finding of substance.

6.3 Conclusions

The tentative conclusion so far (tentative given the uncertainties that are intrinsic to any type of non-experimental analysis) is that the external validity of the experimental impact estimate overall is reasonably high. This is the case especially for the NDLP group, for whom the first-year impact results appear to generalise to the full eligible population. For the ND25Plus group, the external validity of the ERA study is somewhat lower, with the experimental impact findings likely to represent a lower bound to the gains all eligibles would have enjoyed had they been offered ERA services and incentives.

It will be interesting to examine the issue of non-participation in terms of longer term follow-up outcomes. First, given their different characteristics and outcomes, the participants and non-participants might experience ERA impacts that evolve – persist, emerge or fade – differentially. For instance, our take-up analysis has found that had the NDLP non-participants become eligible to ERA services and incentives, they would have been over three percentage points more likely than the programme group to be involved in training and education activities arranged by Jobcentre Plus. If such activities yield longer term returns, it might well be that the effect of ERA for the NDLP non-participants would have grown more over time than the one for the participants. Furthermore, it will be of special importance to examine and account for the issue of survey and/or item non-response for longer-term outcomes, a problem which has worsened in the second-year follow-up survey.

In this report we have not only extensively assessed the external validity of the intermediate ERA findings (based on 12-month follow-up data), but we have set the foundation work and developed a sound and thorough methodological framework for the continued analysis of non-participation in the ERA study, which will include an update of the findings in this report to outcomes experienced five years after random assignment. Given that in many evaluation settings the problem of non-participation is an empirically relevant one (see e.g. Kamionka and Lacroix, 2005), the framework we have developed can be applied to assessing this issue in any study which can exploit the three critical features of (1) being interested in assessing the impact of offering a new treatment, (2) eligible for this offer under an official policy would be a well-defined population, (3) for whom background (and ideally, outcome) information is recorded in the available data.

Table 6.1 Summary of overall findings by analysis type and customer group

	ND25Plus	NDLP
Employment		
Experimental impact for participants	None on probability, positive on duration (+4.6 days).	None on probability nor duration.
Bounds for the effect for all eligibles	Sensitive.	Non-informative.
Sensitivity analysis	The effect on employment probability for all eligibles is positive but small under most scenarios.	Sensitive.
Selection on observables analysis	Better effect for non-participants (+6pp vs none, +10 compared to +5 days); Experimental impact on employment underestimates the effect for all eligibles (+2.6pp, +5.8 days).	Same – zero – impact for non-participants; Experimental impact on employment is representative of the effect for all eligibles (i.e. none).
Selection on observables: sensitivity	Impact for all eligibles is positive but small overall.	Non-informative.
Selection on unobservables	Non-participants experience worse employment outcomes than participants with the same observables.	
Selection on unobservables analysis	Experimental impact on employment is *representative* or *underestimates* the average impact for all eligibles.	
Benefits		
Experimental impact for participants	None on duration.	None on duration.
Selection on observables	Worse effect for non-participants (+9 days vs none); Experimental impact on benefits **overestimates** the effect for all eligibles, though same story of no effect.	
Selection on observables: sensitivity	Impact for all eligibles is mostly negative overall.	Non-informative.
Selection on unobservables	Non-participants experience fewer days on benefits than participants with the same observables.	
Earnings		
Experimental impact for participants	None on incidence of high earnings; Positive on amount (+£393).	Positive on incidence of high earnings (+4pp); Large and positive on amount (+£730).
Bounds for the effect on all eligibles	Non-informative.	Non-informative.
Selection on observables	Impact on earnings for the responding participants is **representative** of (possibly **underestimates** when allowing for non-response) the effect for all eligibles.	Impact on earnings for the responding participants is **representative** of the impact for the full eligible population.
Selection on unobservables analysis	Impact on earnings for the responding participants *overestimates* the effect for all eligibles.	Impact on earnings for the responding participants *underestimates* the effect for all eligibles.

Note: in italics the more tentative findings.

Appendix A
District-level results

A.1 Experimental findings

Table A.1 Experimental findings for the ND25Plus group – administrative outcomes

	Raw		Adjusted		N
	Effect	Standard errors	Effect	Standard errors	
Ever employed					
All	0.014	(0.012)	0.017	(0.011)	6,006
Scotland	0.060*	(0.036)	0.048	(0.033)	745
North East England	-0.034	(0.037)	-0.036	(0.035)	703
North West England	0.026	(0.025)	0.034	(0.022)	1,377
Wales	-0.053	(0.044)	-0.036	(0.042)	456
East Midlands	0.030	(0.028)	0.029	(0.026)	1,245
London	0.007	(0.022)	0.023	(0.019)	1,480
Days employed					
All	4.0	(2.7)	4.6*	(2.4)	6,006
Scotland	11.1	(7.6)	9.2	(7.2)	745
North East England	-7.9	(8.7)	-10.2	(8.2)	703
North West England	6.2	(5.6)	7.7	(5.0)	1,377
Wales	-15.3	(9.9)	-14.0	(8.8)	456
East Midlands	8.8	(6.1)	7.6	(5.6)	1,245
London	6.1	(5.1)	9.1**	(4.3)	1,480
Days on benefits					
All	-3.0	(3.2)	-3.0	(3.0)	6,006
Scotland	-10.2	(7.8)	-7.2	(7.6)	745
North East England	8.9	(10.0)	9.1	(9.8)	703
North West England	-10.8	(6.6)	-11.2*	(6.3)	1,377
Wales	29.3**	(11.8)	29.3**	(11.6)	456
East Midlands	-13.5**	(6.9)	-13.6**	(6.6)	1,245
London	0.6	(6.3)	0.9	(6.1)	1,480

Note: adjusted for the observables constructed from administrative data for the full sample.

Robust standard errors for ever employed.

Table A.2 Experimental findings for the NDLP group – administrative outcomes

	Raw		Adjusted		N
	Effect	Standard errors	Effect	Standard errors	
Ever employed					
All	0.003	(0.014)	-0.006	(0.013)	5,052
Scotland	0.013	(0.048)	0.037	(0.045)	413
North East England	-0.022	(0.032)	-0.019	(0.030)	983
North West England	0.097***	(0.036)	0.067**	(0.034)	759
Wales	-0.023	(0.043)	-0.044	(0.042)	514
East Midlands	-0.022	(0.030)	-0.033	(0.027)	1,131
London	-0.012	(0.028)	0.000	(0.025)	1,252
Days employed					
All	-0.1	(4.0)	-2.2	(3.5)	5,052
Scotland	3.2	(14.5)	7.9	(14.1)	413
North East England	-2.9	(9.0)	0.5	(8.2)	983
North West England	32.8***	(10.5)	22.1**	(9.3)	759
Wales	-6.6	(12.7)	-14.4	(12.1)	514
East Midlands	-12.3	(8.1)	-14.6**	(7.0)	1,131
London	-6.9	(7.8)	-3.1	(6.6)	1,252
Days on benefits					
All	-8.2**	(4.0)	-5.1	(3.7)	5,052
Scotland	-5.3	(13.8)	2.8	(14.4)	413
North East England	-13.7	(9.2)	-12.1	(8.5)	983
North West England	-22.5**	(10.3)	-16.0	(9.9)	759
Wales	-12.8	(12.8)	-13.2	(12.6)	514
East Midlands	-2.0	(8.4)	0.8	(7.7)	1,131
London	2.5	(7.7)	1.6	(7.3)	1,252

Note: adjusted for the observables constructed from administrative data for the full sample.

Robust standard errors for ever employed.

Table A.3 Experimental findings – survey outcomes

	Raw		Adjusted		N
	Effect	Standard errors	Effect	Standard errors	
ND25Plus					
High earnings					
All	0.029	(0.020)	0.026	(0.019)	1,840
Scotland	0.060	(0.047)	0.041	(0.046)	312
North East England	0.038	(0.049)	0.067	(0.052)	320
North West England	0.082*	(0.049)	0.057	(0.051)	268
Wales	-0.079	(0.052)	-0.102*	(0.052)	280
East Midlands	0.045	(0.048)	0.032	(0.049)	349
London	0.026	(0.043)	0.037	(0.044)	311
Earnings					
All	378.6*	(228.6)	393.2*	(222.7)	1,840
Scotland	739.4	(451.3)	497.9	(421.9)	312
North East England	155.0	(550.5)	429.7	(559.3)	320
North West England	611.6	(437.2)	467.0	(449.1)	268
Wales	-1,202.3**	(609.2)	-1,146.5*	(627.3)	280
East Midlands	919.1**	(455.4)	869.4*	(451.4)	349
London	918.1	(770.1)	1,044.2	(826.7)	311
NDLP					
High earnings					
All	0.054**	(0.022)	0.039*	(0.021)	1,745
Scotland	0.088	(0.061)	0.067	(0.065)	253
North East England	0.056	(0.049)	0.032	(0.048)	308
North West England	0.118**	(0.054)	0.073	(0.058)	288
Wales	0.049	(0.056)	0.082	(0.057)	268
East Midlands	0.050	(0.049)	0.046	(0.048)	306
London	-0.019	(0.050)	-0.053	(0.052)	322
Earnings					
All	885.2***	(230.3)	730.2***	(225.5)	1,745
Scotland	1,613.4***	(534.7)	1,443.3**	(588.7)	253
North East England	915.3	(576.1)	561.7	(555.5)	308
North West England	1,165.4**	(525.6)	680.0	(554.3)	288
Wales	820.0	(499.5)	1,080.0**	(524.6)	268
East Midlands	485.8	(605.3)	491.4	(613.5)	306
London	513.2	(596.2)	310.0	(609.5)	322

Note: adjusted for the observables constructed from administrative data for the full sample.

Robust standard errors for high earnings.

Summary Box A.1 Experimental findings for the ATE_1 (adjusted for administrative observables)

	ND25Plus	NDLP
Employment	none on probability positive on duration (+4.6 days) driven by London (+9 days)	none on probability except in NW England (+7pp) none on duration except in NW England (+22 days) East Midlands (-15 days)
Benefits	none on duration except in NW England (-11 days) East Midlands (-14 days) Wales (+29 days)	none on duration
Earnings	none on incidence of high earnings except in Wales (-10pp) positive on amount (+£393) driven by East Midlands (+£869) diluted by Wales (-£1,147)	positive on incidence (+4pp) large and positive (+£730) driven by Scotland (+£,1443) Wales (+£1,080)

Testing for survey and item non-response using administrative outcomes

Table A.4 Balancing of observed characteristics between program and control group members in the responding sample – Internal-validity condition (I-V)

	ND25		NDLP	
	Pseudo R^2	p>chi2	Pseudo R^2	p>chi2
All	0.022	0.170	0.022	0.495
Scotland	0.080	0.748	0.129	0.458
North East England	0.121	0.071	0.079	0.861
North West England	0.078	0.882	0.127	0.191
Wales	0.108	0.484	0.138	0.205
East Midlands	0.062	0.899	0.064	0.989
London	0.094	0.425	0.069	0.959

Note: Pseudo-R squared from a Probit of random assignment status on X on the respondents' subsample and p-value of the likelihood ratio test of the null that the X's are jointly insignificant in predicting random assignment status.

Sample sizes: see Table A.3.

Table A.5 Sample sizes for tests on non-response (Tables A.6-A.9)

	ND25Plus	NDLP
All	5,724	4,770
Scotland	718	386
North East England	686	946
North West England	1,268	728
Wales	447	460
East Midlands	1,196	1,073
London	1,409	1,177

Table A.6 Testing equality of impacts for responding and non-responding participants – external-validity condition (E-V)

	Ever employed		Days employed		Days on benefits	
	diff	*p*-value	*diff*	*p*-value	*diff*	*p*-value
ND25Plus						
All	0.022	0.218	6.3	0.131	3.4	0.457
Scotland	-0.033	0.429	-5.9	0.516	21.8**	0.015
North East England	0.044	0.299	8.6	0.386	11.0	0.307
North West England	0.018	0.711	2.9	0.804	3.1	0.815
Wales	-0.009	0.790	5.1	0.511	-9.1	0.338
East Midlands	-0.015	0.741	-1.7	0.865	17.1	0.100
London	0.073*	0.059	13.4	0.132	-2.4	0.832
NDLP						
All	-0.015	0.413	-0.4	0.944	2.6	0.636
Scotland	-0.031	0.384	-9.5	0.378	-1.5	0.885
North East England	-0.003	0.947	3.6	0.781	-3.2	0.810
North West England	-0.036	0.426	-2.6	0.844	16.6	0.207
Wales	-0.048	0.230	-13.5	0.270	5.3	0.652
East Midlands	0.017	0.703	0.1	0.991	-7.6	0.562
London	-0.001	0.981	16.7	0.201	5.3	0.678

Notes: *diff* is the difference in the average ERA impact for participants compared to the experimental contrast for responding participants; *p*-value based on bootstrapped significance (500 reps).

Sample sizes: see Table A.5.

Table A.7 Testing equality of impacts for responding and non-responding participants controlling for observables – external-validity condition (E-V) given X

	Ever employed		Days employed		Days on benefits	
	diff	***p*-value**	***diff***	***p*-value**	***diff***	***p*-value**
ND25Plus						
All	0.016	0.326	5.0	0.187	4.5	0.310
Scotland	0.004	0.917	0.0	0.997	21.6**	0.034
North East England	0.022	0.623	10.5	0.348	13.7	0.296
North West England	0.004	0.937	-3.0	0.788	-9.8	0.507
Wales	-0.055	0.164	-0.9	0.915	-5.1	0.650
East Midlands	0.011	0.817	4.1	0.688	10.1	0.357
London	0.035	0.364	0.7	0.941	11.2	0.374
NDLP						
All	-0.009	0.614	3.0	0.515	1.7	0.749
Scotland	-0.035	0.395	-6.0	0.650	-0.1	0.994
North East England	0.000	0.994	8.3	0.549	-2.9	0.829
North West England	0.028	0.593	18.0	0.188	-0.3	0.984
Wales	-0.079*	0.095	-18.0	0.205	21.4	0.138
East Midlands	0.012	0.810	1.4	0.910	-10.5	0.453
London	0.032	0.489	27.3	0.030	-3.8	0.782

Notes: *diff* is the difference in the average ERA impact for participants compared to the experimental contrast for responding participants; p-value based on bootstrapped significance (500 reps).

Sample sizes: see Table A.5.

Table A.8 Testing equality of outcomes between non-responding and responding programme (1) and control (0) group members – external-validity condition (E-V'): ND25Plus

			Unconditional on X		Conditional on X	
	$P_{S=0\|R=1}$	$P_{S=0\|R=0}$	*diff*(1)	*diff*(0)	*diff*(1)	*diff*(0)
All						
Ever employed			0.028	-0.004	0.045**	0.038**
Days employed	0.678	0.680	3.138	-6.125	3.958	1.932
Days on benefits			-8.551*	-13.527***	-6.386	-19.953***
Scotland						
Ever employed			-0.040	0.019	0.018	0.018
Days employed	0.554	0.577	-16.114	-5.206	-8.100	-4.409
Days on benefits			9.027	-29.067***	3.153	-37.095***
North East England						
Ever employed			0.094*	0.014	0.094*	0.080
Days employed	0.543	0.524	4.594	-11.638	9.732	-1.310
Days on benefits			-4.594	-25.694*	-4.418	-34.470**
North West England						
Ever employed			0.063	0.040	0.047	0.041
Days employed	0.789	0.789	5.959	2.240	0.734	0.323
Days on benefits			-5.804	-9.743	-11.045	-2.054
Wales						
Ever employed			0.044	0.067	0.059	0.146**
Days employed	0.369	0.378	18.191	4.230	18.043	29.220*
Days on benefits			-42.043**	-17.111	-35.882*	-33.640*
East Midlands						
Ever employed			0.001	0.022	-0.004	0.012
Days employed	0.715	0.701	-1.246	1.166	-2.805	-3.158
Days on benefits			3.550	-20.736**	8.172	-16.282
London						
Ever employed			0.097**	0.003	0.061*	0.026
Days employed	0.773	0.785	16.633*	-0.677	6.910	5.526
Days on benefits			-12.325	-9.136	-7.701	-15.469

Notes: p-values based on heteroskedasticity-robust standard error. (S-1): not controlling for observables; (S-1.X): controlling for observables.
$P_{S=0|R=1}$ is the proportion of non-respondents among the programme group, $P_{S=0|R=0}$ among the control group.
diff(.) is the difference in average outcomes of non-respondents compared to respondents within the programme group (*diff*(1)) or within the control group (*diff*(0)).
Sample sizes: see Table A.5.

Table A.9 Testing equality of outcomes between non-responding and responding programme (1) and control (0) group members – external-validity condition (E-V′): NDLP

			Unconditional on X		Conditional on X	
	$P_{S=0\|R=1}$	$P_{S=0\|R=0}$	*diff*(1)	*diff*(0)	*diff*(1)	*diff*(0)
All						
Ever employed			0.009	0.033	0.048**	0.050**
Days employed	0.626	0.642	4.597	5.053	15.271***	7.900
Days on benefits			-6.380	-10.207*	-17.154***	-17.819***
Scotland						
Ever employed			0.050	0.142**	0.038	0.203***
Days employed	0.349	0.340	11.317	39.395*	12.340	64.799***
Days on benefits			-38.801*	-35.531	-25.717	-62.479**
North East England						
Ever employed			0.111**	0.112**	0.110**	0.124***
Days employed	0.665	0.684	27.569**	21.563	26.507**	22.341*
Days on benefits			-24.786*	-19.381	-24.089*	-20.123
North West England						
Ever employed			0.059	0.118**	0.095*	0.080
Days employed	0.602	0.607	16.622	20.814	30.518**	7.728
Days on benefits			-6.735	-34.055**	-24.427	-29.960*
Wales						
Ever employed			0.049	0.159**	0.014	0.139**
Days employed	0.407	0.429	12.515	43.396**	17.652	42.256**
Days on benefits			-26.376	-37.466*	-16.471	-19.100
East Midlands						
Ever employed			-0.015	-0.038	-0.005	-0.040
Days employed	0.705	0.725	-0.926	-1.100	1.250	1.272
Days on benefits			-15.062	-4.159	-13.381	-0.265
London						
Ever employed			0.018	0.019	0.035	0.005
Days employed	0.719	0.734	14.107	-8.999	18.307*	-12.350
Days on benefits			-1.006	-8.204	-2.514	-8.212

Notes: p-values based on heteroskedasticity-robust standard error.
$P_{S=0|R=1}$ is the proportion of non-respondents among the programme group, $P_{S=0|R=0}$ among the control group.
diff(.) is the difference in average outcomes of non-respondents compared to respondents within the programme group (*diff*(1)) or within the control group (*diff*(0)).
Sample sizes: see Table A.5.

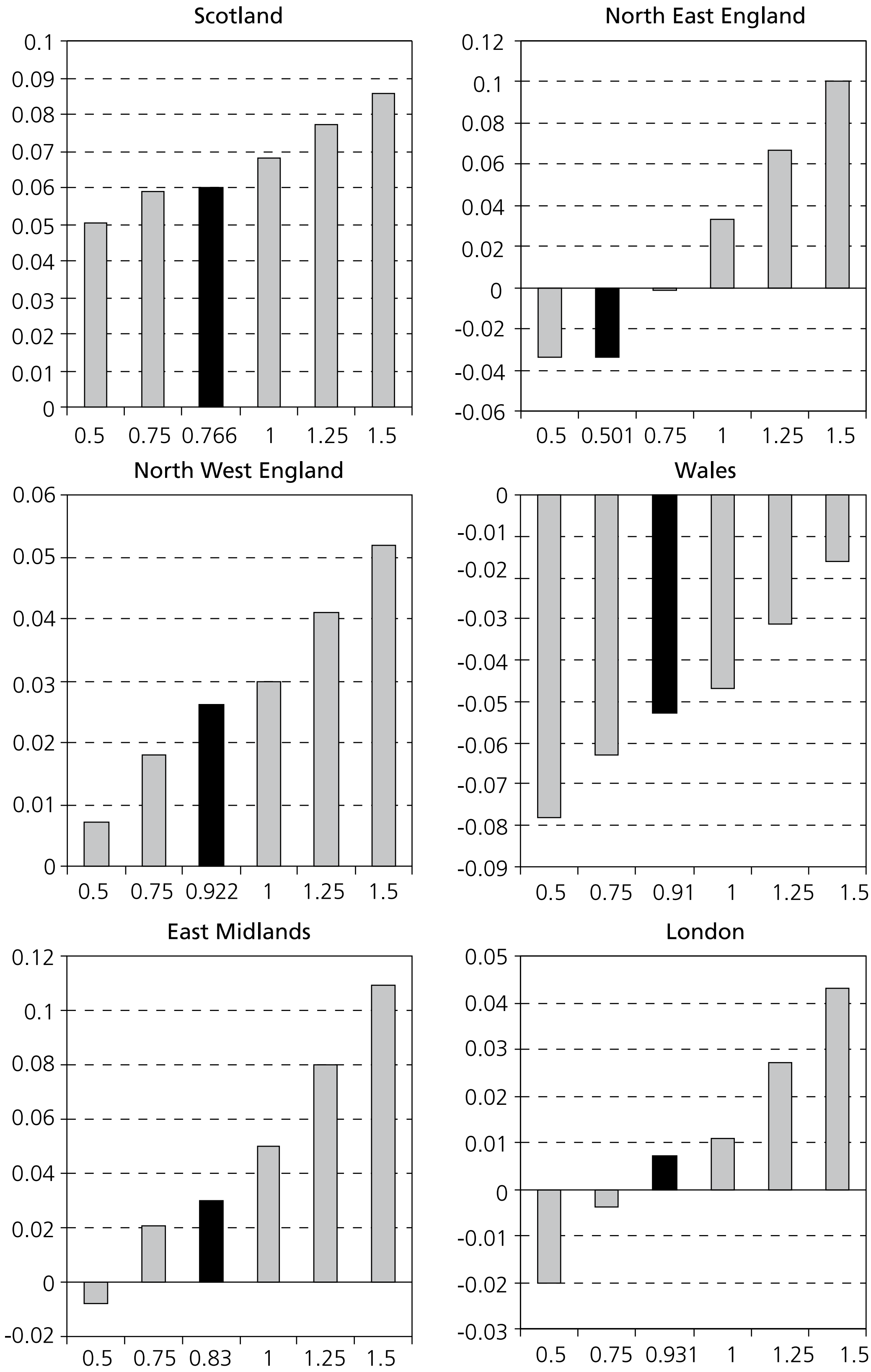
A.2 Bounds

Table A.10 Non-parametric bounds for the *ATE* – Outcome: ever employed

	p	ATE_1	$\underline{ATE}$	$\overline{ATE}$	95% CI lower	95% CI upper	N
ND25Plus							
All	0.230	0.014	-0.050	0.180	-0.069	0.199	7,796
Scotland	0.087	0.060*	0.033	0.120	-0.033	0.186	816
North East England	0.349	-0.034	-0.101	0.249	-0.155	0.303	1,080
North West England	0.146	0.026	-0.016	0.130	-0.056	0.171	1,612
Wales	0.207	-0.053	-0.110	0.097	-0.180	0.168	575
East Midlands	0.275	0.030	-0.067	0.208	-0.108	0.249	1,717
London	0.259	0.007	-0.051	0.207	-0.086	0.242	1,996
NDLP							
All	0.304	0.003	-0.157	0.147	-0.177	0.167	7,261
Scotland	0.053	0.013	-0.013	0.040	-0.102	0.128	436
North East England	0.292	-0.022	-0.168	0.124	-0.216	0.171	1,389
North West England	0.062	0.097***	0.049	0.110	-0.023	0.182	809
Wales	0.236	-0.023	-0.176	0.060	-0.244	0.128	673
East Midlands	0.471	-0.022	-0.254	0.218	-0.289	0.253	2,140
London	0.310	-0.012	-0.157	0.153	-0.200	0.196	1,814

Note: Confidence intervals covering the identification region with 95 per cent probability have been derived from 1,000 bootstrap replications following Horowitz and Manski (2000).

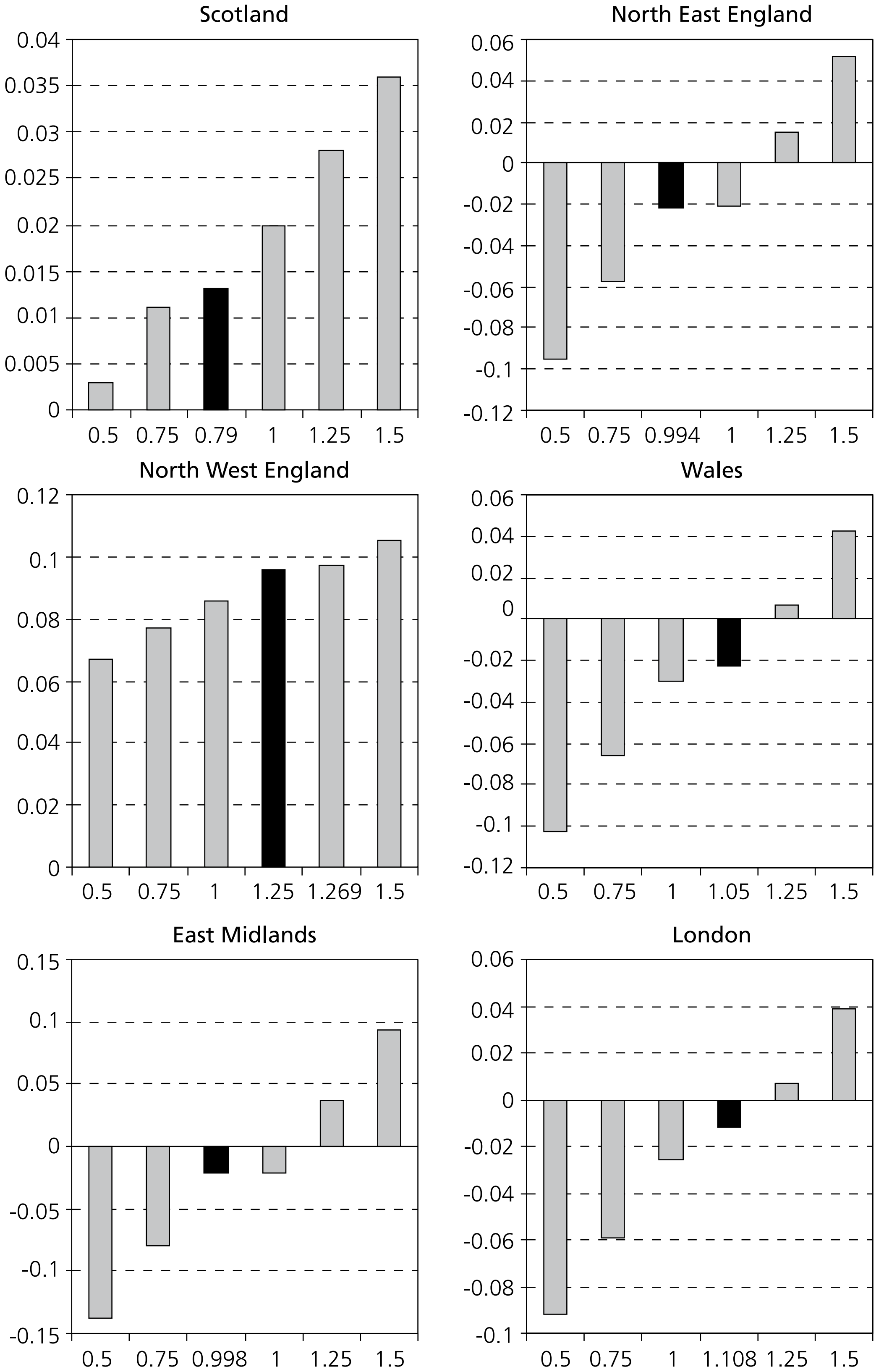
Figure A.1 ND25Plus: Sensitivity analysis: ATE_θ for ever employed, θ from 0.5 to 1.5



In black: experimental impact estimate and corrresponding θ.

Sample sizes: see Table A.10.

Figure A.2 NDLP: Sensitivity analysis: ATE_θ for ever employed, θ from 0.5 to 1.5



In black: experimental impact estimate and corrresponding θ.

Sample sizes: see Table A.10.

Table A.11 Non-parametric bounds for the *ATE* – Outcome: high earnings

	p	ATE_1	$\underline{ATE}$	$\overline{ATE}$	95% CI lower	95% CI upper	N
ND25Plus							
All	0.226	0.029	-0.204	0.249	-0.236	0.280	3,515
Scotland	0.084	0.060	-0.029	0.139	-0.116	0.227	378
North East England	0.341	0.038	-0.316	0.366	-0.387	0.436	675
North West England	0.147	0.082*	-0.077	0.217	-0.156	0.296	487
Wales	0.196	-0.079	-0.259	0.133	-0.346	0.220	389
East Midlands	0.270	0.045	-0.237	0.303	-0.309	0.375	792
London	0.255	0.026	-0.236	0.275	-0.303	0.342	794
NDLP							
All	0.299	0.054**	-0.262	0.337	-0.294	0.369	3,784
Scotland	0.049	0.088	0.034	0.133	-0.087	0.254	273
North East England	0.285	0.056	-0.245	0.325	-0.319	0.399	685
North West England	0.063	0.118**	0.048	0.174	-0.054	0.276	337
Wales	0.227	0.049	-0.189	0.265	-0.279	0.354	403
East Midlands	0.467	0.050	-0.441	0.494	-0.497	0.550	1,248
London	0.305	-0.019	-0.318	0.291	-0.391	0.364	838

Note: Confidence intervals covering the identification region with 95 per cent probability have been derived from 1,000 bootstrap replications following Horowitz and Manski (2000).

Summary Box A.2 Bounds and sensitivity analysis

	ND25Plus	NDLP
Ever employed	positive in Scotland	positive in NW England
• Sensitivity	overall: positive small under most scenarios positive substantial in Scotland positive small in NW England negative in Wales sensitive though positive in E Midlands sensitive in NE England and London	overall: sensitive positive small in Scotland positive substantial in NW England negative in Wales sensitive in E Midlands sensitive in NE England mostly negative in London
High earnings	(positive small in Scotland) all other: non-informative	positive in NW England (positive small in Scotland) all other: non-informative

A.3 Selection on observables

Administrative outcomes

Table A.12 Administrative outcomes: Matching estimates for the ND25Plus group

	p	ATE_1	ATE_0	ATE	$ATE_1 \neq ATE$	N
All						
Ever employed	0.230	0.017	0.056***	0.026**	***	
Days employed	0.230	4.560**	9.984***	5.805***	*	4,831
Days on benefits	0.230	-2.966	8.862**	-0.250	***	
Scotland						
Ever employed	0.087	0.048	0.107*	0.053*		
Days employed	0.087	9.238	12.362	9.509		455
Days on benefits	0.087	-7.156	-4.585	-6.933		
North East England						
Ever employed	0.349	-0.036	0.089**	0.007	***	
Days employed	0.349	-10.245	7.583	-4.022	**	737
Days on benefits	0.349	9.101	-0.008	5.921		
North West England						
Ever employed	0.146	0.034	0.019	0.032		
Days employed	0.146	7.665	6.959	7.562*		932
Days on benefits	0.146	-11.177*	2.249	-9.220		
Wales						
Ever employed	0.207	-0.036	-0.023	-0.033		
Days employed	0.207	-14.038	3.336	-10.442		344
Days on benefits	0.207	29.265***	19.798	27.306**		
East Midlands						
Ever employed	0.275	0.029	0.084***	0.044*	**	
Days employed	0.275	7.587	11.623*	8.696*		1,097
Days on benefits	0.275	-13.594**	10.271	-7.034	***	
London						
Ever employed	0.259	0.023	0.046**	0.029*		
Days employed	0.259	9.057**	15.317***	10.675***		1,266
Days on benefits	0.259	0.875	7.882	2.686		

$ATE_1 \neq ATE$ column: bootstrap-based statistical significance of the difference.
Statistical significance based on bootstrapped bias-corrected confidence intervals (1,000 replications): *** significant at 1%, ** at 5%, * at 10%.

Table A.13 Administrative outcomes: Matching estimates for the NDLP group

	p	ATE_1	ATE_0	ATE	$ATE_1 \neq ATE$	N
All						
Ever employed	0.304	-0.006	0.015	0.000		
Days employed	0.304	-2.208	-1.957	-2.132		4,768
Days on benefits	0.304	-5.078	8.881**	-0.831	***	
Scotland						
Ever employed	0.053	0.037	0.152	0.043		
Days employed	0.053	7.940	54.167	10.379		229
Days on benefits	0.053	2.799	-48.241	0.107		
North East England						
Ever employed	0.292	-0.019	0.031	-0.005		
Days employed	0.292	0.484	7.767	2.613		915
Days on benefits	0.292	-12.118	-5.445	-10.167		
North West England						
Ever employed	0.062	0.067**	-0.067	0.059*		
Days employed	0.062	22.075**	-5.765	20.354**		452
Days on benefits	0.062	-16.027	13.243	-14.218		
Wales						
Ever employed	0.236	-0.044	0.002	-0.033		
Days employed	0.236	-14.381	-13.498	-14.173		419
Days on benefits	0.236	-13.228	10.345	-7.659		
East Midlands						
Ever employed	0.471	-0.033	0.016	-0.010	*	
Days employed	0.471	-14.557**	-4.517	-9.824		1,576
Days on benefits	0.471	0.842	14.500*	7.282		
London						
Ever employed	0.310	0.000	0.003	0.001		
Days employed	0.310	-3.135	-8.047	-4.656		1,177
Days on benefits	0.310	1.614	10.067	4.233		

$ATE_1 \neq ATE$ column: bootstrap-based statistical significance of the difference.
Statistical significance based on bootstrapped bias-corrected confidence intervals (1,000 replications): *** significant at 1%, ** at 5%, * at 10%.

Table A.14 Sensitivity analysis by district: ATE_θ, θ from 0.5 to 1.5: ND25Plus group

Ever employed		Days employed		Days on benefits		Ever employed		Days employed		Days on benefits	
θ	ATE_θ	θ	ATE_θ	θ	ATE_θ	θ	ATE_θ	θ	ATE_θ	θ	ATE_θ
Scotland						**Wales**					
0.50	0.037	0.50	7.239	0.50	-19.931	0.50	-0.064	0.50	-16.087	0.50	0.814
0.75	0.045	0.75	8.374	0.75	-13.432	0.75	-0.049	**0.68**	**-14.038**	0.75	14.060
0.84	**0.048**	**0.94**	**9.238**	**0.99**	**-7.156**	**0.96**	**-0.036**	0.75	-13.264	1.00	27.306
1.00	0.053	1.00	9.509	1.00	-6.933	1.00	-0.033	1.00	-10.442	**1.04**	**29.265**
1.25	0.061	1.25	10.644	1.25	.	1.25	-0.017	1.25	-7.620	1.25	40.552
1.50	0.069	1.50	11.780	1.50	.	1.50	-0.001	1.50	-4.798	1.50	.
North East England						**East Midlands**					
0.50	-0.049	0.50	-13.363	0.50	-39.830	0.50	-0.012	0.50	-1.099	0.50	-42.017
0.62	**-0.036**	**0.67**	**-10.245**	0.75	-16.955	0.75	0.016	0.75	3.799	0.75	-24.525
0.75	-0.021	0.75	-8.693	1.00	5.921	**0.87**	**0.029**	**0.94**	**7.587**	**0.91**	**-13.594**
1.00	0.007	1.00	-4.022	**1.04**	**9.101**	1.00	0.044	1.00	8.696	1.00	-7.034
1.25	0.036	1.25	0.649	1.25	28.797	1.25	0.072	1.25	13.594	1.25	10.458
1.50	0.064	1.50	5.320	1.50	.	1.50	0.100	1.50	18.491	1.50	.
North West England						**London**					
0.50	0.011	0.50	3.762	0.50	-27.771	0.50	-0.005	0.50	3.933	0.50	-31.672
0.75	0.021	0.75	5.662	0.75	-18.495	0.75	0.012	0.75	7.304	0.75	-14.493
1.00	0.032	1.00	7.562	**0.95**	**-11.177**	**0.91**	**0.023**	**0.88**	**9.057**	**0.97**	**0.875**
1.05	**0.034**	**1.01**	**7.665**	1.00	-9.220	1.00	0.029	1.00	10.675	1.00	2.686
1.25	0.042	1.25	9.462	1.25	0.056	1.25	0.045	1.25	14.046	1.25	19.865
1.50	0.052	1.50	11.362	1.50	.	1.50	0.062	1.50	17.417	1.50	.

In bold: experimental impact estimate and corresponding θ. Missing $ATE\theta$ denotes an inadmissible θ value.
Sample sizes: see Table A.12.

Table A.15 Sensitivity analysis by district: ATE_θ, θ from 0.5 to 1.5: NDLP group

Ever employed		Days employed		Days on benefits		Ever employed		Days employed		Days on benefits	
θ	ATE_θ	θ	ATE_θ	θ	ATE_θ	θ	ATE_θ	θ	ATE_θ	θ	ATE_θ
Scotland						**Wales**					
0.50	0.026	0.50	6.927	0.50	-5.654	0.50	-0.111	0.50	-31.415	0.50	-26.474
0.75	0.034	**0.65**	**7.940**	0.75	-2.774	0.75	-0.072	0.75	-22.794	0.75	-17.066
0.82	**0.037**	0.75	8.653	1.00	0.107	**0.93**	**-0.044**	**0.99**	**-14.381**	**0.85**	**-13.228**
1.00	0.043	1.00	10.379	**1.23**	**2.799**	1.00	-0.033	1.00	-14.173	1.00	-7.659
1.25	0.051	1.25	12.105	1.25	2.987	1.25	0.006	1.25	-5.551	1.25	1.748
1.50	0.060	1.50	13.831	1.50	5.868	1.50	0.046	1.50	3.070	1.50	11.156
North East England						**East Midlands**					
0.50	-0.086	0.50	-15.837	0.50	-38.536	0.50	-0.134	0.50	-35.698	0.50	-43.460
0.75	-0.045	0.75	-6.612	0.75	-24.352	0.75	-0.072	0.75	-22.761	0.75	-18.089
0.91	**-0.019**	**0.94**	**0.484**	**0.97**	**-12.118**	**0.91**	**-0.033**	**0.91**	**-14.557**	**0.94**	**0.842**
1.00	-0.005	1.00	2.613	1.00	-10.167	1.00	-0.010	1.00	-9.824	1.00	7.282
1.25	0.036	1.25	11.838	1.25	4.017	1.25	0.052	1.25	3.114	1.25	32.652
1.50	0.076	1.50	21.063	1.50	18.201	1.50	0.114	1.50	16.051	1.50	58.023
North West England						**London**					
0.50	0.040	0.50	15.814	0.50	-20.709	0.50	-0.073	0.50	-21.049	0.50	-32.214
0.75	0.049	0.75	18.084	0.75	-17.464	0.75	-0.036	0.75	-12.853	0.75	-13.991
1.00	0.059	1.00	20.354	**0.86**	**-16.027**	**0.99**	**0.000**	1.00	-4.656	**0.96**	**1.614**
1.22	**0.067**	**1.19**	**22.075**	1.00	-14.218	1.00	0.001	**1.05**	**-3.135**	1.00	4.233
1.25	0.069	1.25	22.624	1.25	-10.973	1.25	0.038	1.25	3.540	1.25	22.457
1.50	0.078	1.50	24.894	1.50	-7.727	1.50	0.075	1.50	11.736	1.50	40.680

In bold: experimental impact estimate and corresponding θ. Missing $ATE\theta$ denotes an inadmissible θ value.
Sample sizes: see Table A.13.

Summary Box A.3 Selection on observables – administrative outcomes

	ND25Plus	NDLP
Employment	better effect for Q=0 driven by Scotland NE England East Midlands London	same – zero – impact except in NW Eng (worse for Q=0) E Midl (better for Q=0)
	ATE_1 *underestimates* ATE for employment outcomes	**ATE_1 *representative* of ATE for employment outcomes**
• Sensitivity	ATE positive but small overall driven by Scotland NW England East Midlands London ATE negative in Wales not informative for NE England	not informative overall except in Scotland (positive) NW England (positive) Wales (mostly negative) E Midl (mostly negative)
Benefits	worse effect for Q=0 driven by NW England East Midlands except Wales: better effect for Q=0	worse effect for Q=0 driven by East Midlands
	ATE_1 *overestimates* ATE for benefit outcomes	**ATE_1 *overestimates* ATE for benefit outcomes**
• Sensitivity	ATE mostly negative driven by Scotland NW England East Midlands ATE positive in Wales not informative for NE England and London	not informative overall except in NW England (negative)

Survey outcomes

Table A.16 Weighting and matching estimates of the average ERA impact on earnings for all eligibles accounting for non-response

	Weighting						Matching				
	$\Delta_{S=1}$	ATE	$E(Y_1)$	$E(Y_0)$	$P_{R=1,S=1\|Q=1}$	$P_{R=0,S=1\|Q=1}$	ATE	$E(Y_1)$	$E(Y_0)$	$P_{R=1,S=1}$	$P_{R=0,S=1}$
ND25											
All	393.2*	559.9**	2,772.3	2,212.3	0.162	0.159	580.2***	2,779.6	2,199.4	0.126	0.123
Scotland	497.9	704.7	2,429.7	1,725.0	0.228	0.206	620.8*	2,587.7	1,966.9	0.209	0.189
North East England	429.7	428.6	2,926.0	2,497.4	0.233	0.233	176.3	2,864.7	2,688.4	0.154	0.154
North West England	467.0	20.9	2,227.0	2,206.1	0.106	0.106	446.3	2,476.7	2,030.4	0.090	0.090
Wales	-1,146.5**	-988.2	2,191.4	3,179.6	0.313	0.313	-966.9*	2,275.0	3,241.9	0.252	0.252
East Midlands	869.4*	896.2	3,451.8	2,555.6	0.143	0.149	907.7***	3,411.0	2,503.3	0.104	0.109
London	1044.2	1,137.1	2,852.5	1,715.4	0.114	0.106	742.6	2,591.2	1,848.6	0.085	0.079
NDLP											
All	730.2***	644.7**	3,557.9	2,913.2	0.189	0.177	718.2***	3,509.2	2,791.1	0.133	0.124
Scotland	1,443.3**	1,402.4	4,105.6	2,703.2	0.324	0.332	1,343.0***	4,320.4	2,977.4	0.308	0.315
North East England	561.7	809.4	3,524.8	2,715.3	0.172	0.153	766.3*	3,422.6	2,656.4	0.123	0.110
North West England	680.0	653.4	3,518.3	2,864.9	0.212	0.184	738.5	3,699.3	2,960.8	0.198	0.172
Wales	1,080.0**	1,221.4*	3,789.6	2,568.1	0.304	0.278	935.5**	3,757.7	2,822.2	0.235	0.215
East Midlands	491.4	641.5	3,126.3	2,484.9	0.147	0.138	720.5	3,300.1	2,579.6	0.078	0.073
London	310.0	606.0	3,618.0	3,012.0	0.138	0.135	596.6	3,673.5	3,076.9	0.096	0.094

Notes: $\Delta_{S=1}$ is the experimental estimate ignoring potential non-response bias;
ATE is the average ERA impact for all eligibles;
$E(Y_1)$ are average earnings of all eligibles under ERA treatment; $E(Y_0)$ are average earnings for all eligibles without ERA treatment;
$P_{R=1,S=1|Q=1}$ ($P_{R=0,S=1|Q=1}$) are the responding programme (control) group members as a fraction of all ERA study participants.
$P_{R=1,S=1}$ ($P_{R=0,S=1}$) are the responding programme (control) group members as a fraction of all eligibles.
Matching estimator: kernel matching with epanechnikov kernel (bandwidth of 0.06), common support imposed separately for each term.
Statistical significance based on bootstrapped bias-corrected confidence intervals (1,000 replications for the weighting estimator, 500 for the matching estimator): *** significant at 1%, ** at 5%, * at 10%.
In this table, $\Delta_{S=1}$ is never statistically significantly different from the ATE according to bootstrap-based statistical significance of the difference.
Sample sizes: see Table A.17.

Table A.17 Matching estimates of the average ERA impact on earnings for all eligibles

	$\Delta_{S=1}$	*ATE* allowing for non-response, separate CS	*ATE* ignoring non-response, separate CS	*ATE* ignoring non-response, joint CS	% lost to CS	N
ND25						
All	393.2*	580.2***	442.8*	443.5*	0.8	7,399
Scotland	497.9	620.8*	474.1	467.5	16.7	784
North East England	429.7	176.3	448.2	438.4	12.7	1,041
North West England	467.0	446.3	467.6	467.4	3.7	1,487
Wales	-1,146.5**	-966.9*	-1,142.3*	-1,191.2**	15.6	556
East Midlands	869.4*	907.7***	722.5 (*)	728.4 (*)	2.7	1,639
London	1,044.2	742.6	974.4	972.3	5.2	1,892
NDLP						
All	730.2***	718.2***	662.8***	660.4**	1.0	6,809
Scotland	1,443.3**	1,343.0***	1361.0**	1,345.0**	40.0	406
North East England	561.7	766.3*	740.1	728.5	12.7	1,323
North West England	680.0	738.5	614.3	642.5	26.5	777
Wales	1,080.0**	935.5**	1,230.0***	1,171.2**	24.4	595
East Midlands	491.4	720.5	540.0	548.4	8.7	2,015
London	310.0	596.6	644.1 (**)	655.7 (**)	13.8	1,693

Notes: Statistical significance based on bootstrapped bias-corrected confidence intervals (500 repetitions): *** significant at 1%, ** at 5%, * at 10%.
(*): statistically different from experimental estimate ignoring potential non-response bias ($\Delta_{S=1}$) at the *% level.
Kernel matching with epanechnikov kernel (bandwidth of 0.06).
Separate CS: common support imposed on the non-participants separately for each term; Joint CS: estimates pertain to those non-participants satisfying both support conditions.
When ignoring non-response, $\Delta_{S=1}$ is assumed to be equal to ATE_1.

Summary Box A.4 Selection on observables – earnings

	ND25Plus	**NDLP**
Allowing for non-response	Impact on earnings for the responding experimental group **underestimates** the average impact for the full eligible population overall Scotland (East Midlands) Wales (reduction in negative impact)	Impact on earnings for the responding experimental group is **representative** of the average impact for the full eligible population (though point estimates often go down)
Ignoring non-response	Results for the *ATE* ignoring non-response are much closer to the experimental estimates than those allowing for it	
Ignoring non-response: separate vs joint support	Same evidence as above	

A.4 Testing for selection on specific unobservables

Table A.18 Differences in outcomes for non-participants compared to participants (control group) with the same observed characteristics

		OLS	FILM	Matching	N
ND25Plus					
All	Ever employed	**-0.044*****	**-0.057*****	**-0.056*****	
	Days employed	**-7.9*****	**-9.8*****	**-9.5*****	4,755
	Days on benefits	**-10.1*****	**-9.7****	**-9.2****	
Scotland	Ever employed	-0.042	-0.050	-0.032	
	Days employed	-8.2	-9.9	-6.8	432
	Days on benefits	1.1	-0.9	-0.5	
North East England	Ever employed	**-0.148*****	**-0.155*****	**-0.171*****	
	Days employed	**-23.3*****	**-26.0*****	**-26.6*****	720
	Days on benefits	12.1	14.1	15.1	
North West England	Ever employed	0.011	0.011	0.010	
	Days employed	-1.4	-3.4	-2.7	915
	Days on benefits	**-17.2***	**-17.4****	**-17.4***	
Wales	Ever employed	-0.017	-0.075	0.005	
	Days employed	-16.6	-25.6**	-7.2	350
	Days on benefits	-1.5	7.8	-13.8	
East Midlands	Ever employed	**-0.060****	**-0.070*****	**-0.075****	
	Days employed	-5.2	-6.6	-8.8	1,092
	Days on benefits	**-24.0*****	**-25.9*****	**-22.4*****	
London	Ever employed	-0.010	-0.016	-0.013	
	Days employed	-3.9	-4.2	-3.8	1,246
	Days on benefits	-11.0	**-12.1***	-12.3	
NDLP					
All	Ever employed	**-0.041*****	**-0.045*****	**-0.041****	
	Days employed	**-10.3*****	**-11.5*****	**-11.4****	4,702
	Days on benefits	**-8.2****	**-9.3****	**-9.6***	
Scotland	Ever employed	-0.068	-0.069	-0.066	
	Days employed	**-70.5*****	**-74.4*****	**-76.8****	230
	Days on benefits	**60.8****	**60.1*****	**61.2***	
North East England	Ever employed	**-0.062****	**-0.064****	**-0.079****	
	Days employed	-13.9	**-17.5***	**-20.2***	880
	Days on benefits	-5.6	3.2	3.8	
North West England	Ever employed	**0.127***	**0.094****	0.123	
	Days employed	**33.4***	**24.3****	31.8	407
	Days on benefits	-17.6	-10.6	-17.9	

Continued

Table A.18 (Continued)

		OLS	FILM	Matching	N
Wales	Ever employed	-0.054	-0.047	-0.025	
	Days employed	-5.4	-9.4	0.6	413
	Days on benefits	**-32.0****	**-27.8***	-27.9	
East Midlands	Ever employed	**-0.049****	**-0.041***	-0.042	
	Days employed	**-11.1***	-10.5	-10.5	1,573
	Days on benefits	**-12.5***	**-16.9****	**-16.8****	
London	Ever employed	-0.025	**-0.047***	-0.031	
	Days employed	-3.2	-7.9	-6.0	1,199
	Days on benefits	-12.1	-13.4	-10.9	

Significance based on robust standard errors for OLS and FILM, and on approximate standard errors for kernel matching. ***: significant at 1%,**: at 5%, *: at 10%.

Summary Box A.5 Selection on unobservables

	ND25Plus		NDLP
All		worse employment outcomes fewer days on benefits	
North East England		worse employment outcomes	
East Midlands		worse employment outcomes fewer days on benefits	
North West England	fewer days on benefits		better employment outcomes
Wales	*no selection*		fewer days on benefits
London		*no selection*	
Scotland	*no selection*		worse employment outcomes more days on benefits

Appendix B
Matching diagnostics

Table B.1 Estimation of the propensity score

	ND25Plus	NDLP
Scotland	-0.256***	-0.383***
North East England	0.109***	-0.019
North West England	-0.133***	-0.393***
Wales	-0.081***	-0.128***
East Midlands	0.025	0.175***
2nd month of RA	-0.080**	-0.066
3rd month of RA	-0.057	-0.045
4th month of RA	-0.084**	-0.075**
5th month of RA	-0.084**	-0.087**
6th month of RA	-0.109***	-0.081**
7th month of RA	-0.118***	-0.045
8th month of RA	-0.129***	-0.062
9th month of RA	-0.112***	-0.108***
10th month of RA	-0.159***	-0.150***
11th month of RA	-0.109***	-0.099***
12th month of RA	-0.157***	-0.139***
13th month of RA	-0.217***	
Female	-0.014	-0.002
Missing gender	-0.064	-0.081
Age at inflow	-0.027***	0.005
Age squared	0.000***	-0.000
Missing age	-0.361***	0.068
Ethnic Minority	0.043**	-0.016
Missing ethnicity	0.024	0.038

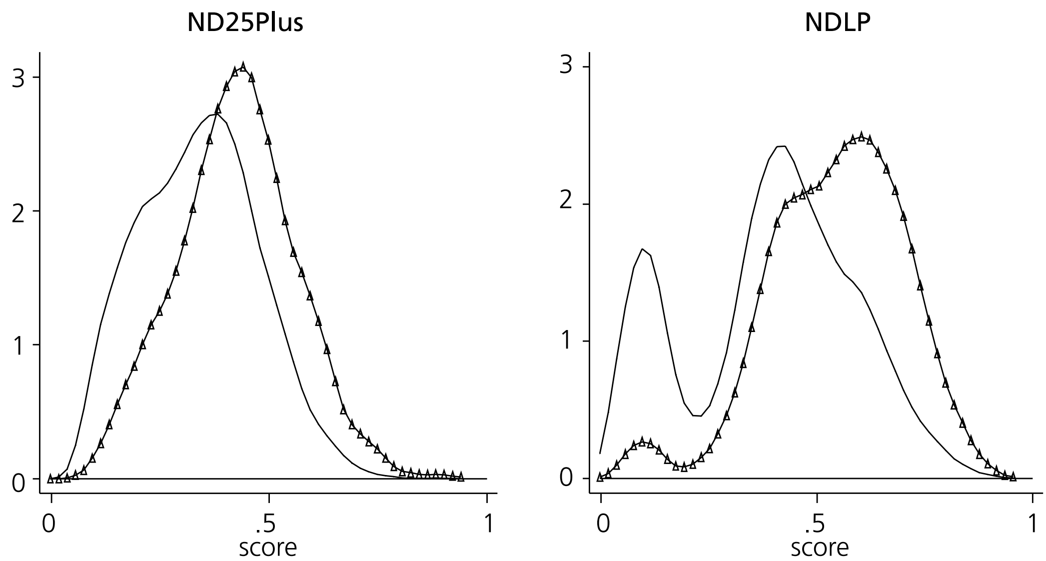
Continued

Table B.1 (Continued)

	ND25Plus	NDLP
Has disability/claims IB at inflow	0.023	-0.008
Missing disability status		0.014
2 children, NDLP		-0.006
≥3 children, NDLP		-0.043*
Missing child info, NDLP		0.018
Youngest child <1 at inflow, NDLP		-0.039
Youngest child 1-5 at inflow, NDLP		0.012
Age youngest child missing, NDLP		-0.017
Has partner, ND25+	-0.025	
Missing marital status, ND25+	-0.063*	
Early entrant - ND25+	-0.036	
Not on benefits at inflow		0.102***
Employed at inflow	0.055*	0.150***
Show up same day	0.060*	0.061
Show up w/in 30 days	-0.022	-0.083***
Past participation in basic skills	0.016	-0.025
Past participation in ND25+ program	0.027***	
Past participation in voluntary programs	-0.061***	0.081***
Spent <50% of past 3 yrs on active benefits	0.003	
Spent >50 & <100% of past 3 yrs on active benefits	-0.005	
Spent 0% of past 3 yrs on active benefits, NDLP		-0.091
Spent >0 & <50% of past 3 yrs on active benefits		-0.084
Spent 0% of past 3 yrs on inactive benefits	-0.024	-0.047
Spent >0 & <50% of past 3 yrs on inactive benefits	-0.001	0.003
Spent >50 & <100% of past 3 yrs on inactive benefits	-0.069	-0.032
Spent >0 & <25% of past 3 yrs in employment	-0.025	-0.003
Spent ≥25% and <50% of past 3 yrs in employment	-0.031	-0.020
Spent ≥50% of past 3 yrs in employment	-0.093**	-0.053**
Total ND caseload at office (100)	-0.003	-0.006***
Share of LP in ND caseload at office	0.048	-0.065
Bottom quintile of local deprivation	0.048	-0.018
2nd quintile of local deprivation	0.034	0.062
3rd quintile of local deprivation	0.028	0.037
4th quintile of local deprivation	0.018	-0.025
TTWA-level unemployment rate	0.963	-1.472
Postcode missing or incorrect	0.493***	0.001
Observations	4,829	4,766

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%.

Figure B.1 Common support between non-participants and programme group: Distribution of $P(Q=0 \mid Q=0 \vee R=1, X)$



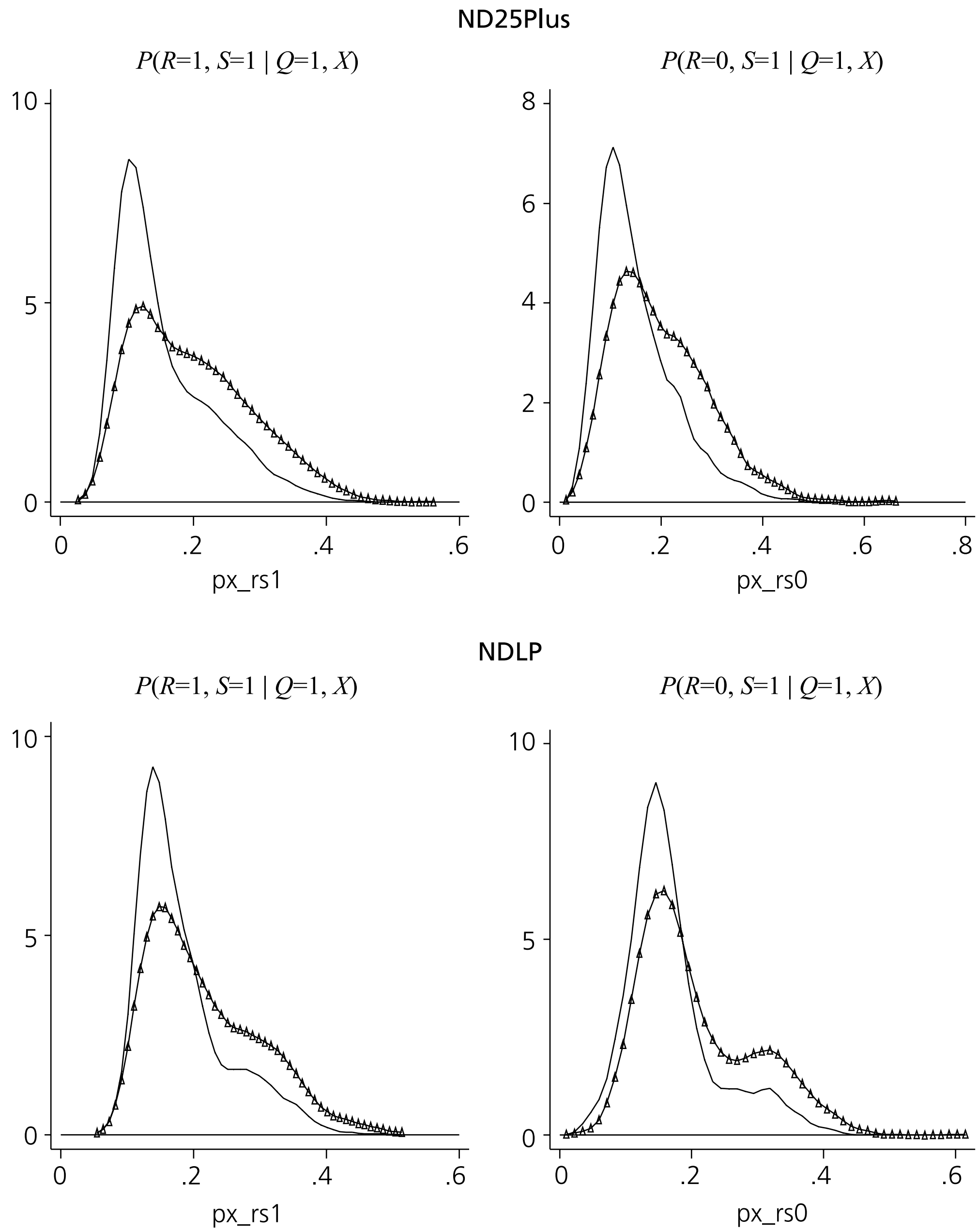
Note: $Q=0$ denotes participants, $R=1$ the programme group.

Table B.2 Administrative outcomes: Covariate balancing indicators before and after matching

	Prob>chi		Pseudo R^2		Median bias		% lost CS
	Before	After	Before	After	Before	After	
ND25							
All	0.000	1.000	0.069	0.001	4.2	0.6	0.2
Scotland	0.005	1.000	0.170	0.011	13.8	2.7	4.3
North East England	0.000	1.000	0.102	0.006	7.8	1.3	4.2
North West England	0.013	1.000	0.064	0.004	5.6	1.0	1.3
Wales	0.000	1.000	0.189	0.030	10.8	3.4	5.0
East Midlands	0.004	1.000	0.048	0.004	4.2	1.5	0.4
London	0.000	1.000	0.061	0.002	4.5	1.3	1.0
NDLP							
All	0.000	1.000	0.121	0.001	3.8	0.8	0.2
Scotland	0.798	1.000	0.240	0.140	10.1	7.2	13.0
North East England	0.002	1.000	0.065	0.003	5.0	1.2	1.2
North West England	0.542	1.000	0.135	0.015	6.4	4.0	2.0
Wales	0.001	1.000	0.149	0.015	8.3	3.2	3.1
East Midlands	0.000	1.000	0.046	0.002	5.6	1.2	1.2
London	0.000	1.000	0.123	0.006	7.7	2.0	3.2

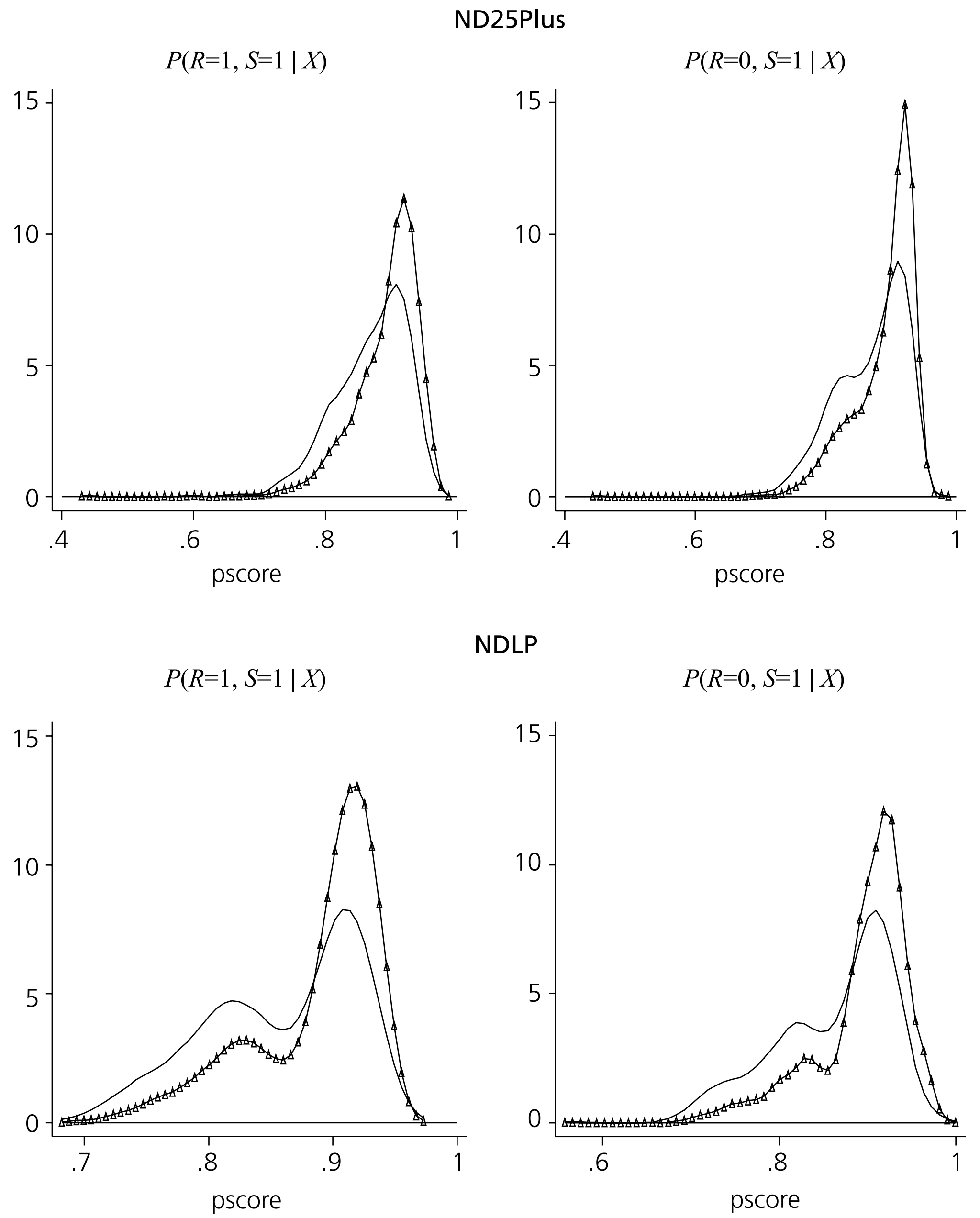
Notes:
Prob>chi: p-value of the likelihood-ratio test before (after) matching, testing the hypothesis that the regressors are jointly insignificant, i.e. well balanced in the two (matched) groups.
Pseudo R^2: from probit estimation of the conditional probability of being a non-participant (before and after matching), giving an indication of how well the observables explain non-participation.
Median bias: median absolute standardised bias before and after matching, median taken over all the regressors. Following Rosenbaum and Rubin (1985), for a given covariate, the standardised difference **before** matching is the difference of the sample means in the non-participant and participant subsamples as a percentage of the square root of the average of the sample variances in the two groups. The standardised difference **after** matching is the difference of the sample means in the matched non-participants (i.e. falling within the common support) and matched participant subsamples as a percentage of the square root of the average of the sample variances in the two original groups.
% lost to CS: Share of the group of non-participants falling outside of the common support.

Figure B.2 Survey outcomes and weighting: Common support



Note: Q=1 denotes participants, R=1 the programme group, R=0 the control group and S=1 the responding participants.

Figure B.3 Survey outcomes and matching: Common support



Note: R=1 denotes the programme group, R=0 the control group and S=1 the responding participants.

Table B.3 Survey outcomes: Matching quality allowing for non-response, separate common support

	Eligibles vs responding programme group							Eligibles vs responding control group						
	Prob>chi		Pseudo R2		Median bias		% lost to CS	Prob>chi		Pseudo R2		Median bias		% lost to CS
	Before	After	Before	After	Before	After		Before	After	Before	After	Before	After	
ND25														
All	0.000	0.000	0.030	0.005	4.2	1.3	0.3	0.000	0.000	0.033	0.006	3.9	1.4	1.2
Scotland	0.976	1.000	0.029	0.007	4.4	1.3	1.8	0.896	1.000	0.037	0.005	5.3	1.9	4.0
North East England	0.387	1.000	0.045	0.003	5.3	1.0	11.5	0.628	0.969	0.041	0.009	5.0	1.8	3.5
North West England	0.915	0.610	0.031	0.009	7.0	2.7	3.2	0.675	0.022	0.038	0.015	6.1	2.3	1.7
Wales	0.930	1.000	0.043	0.004	6.1	1.5	9.9	0.994	1.000	0.032	0.002	5.2	0.8	6.3
East Midlands	0.888	0.352	0.028	0.010	4.4	2.0	1.4	0.775	0.077	0.030	0.012	3.1	1.4	1.3
London	0.461	0.000	0.037	0.017	5.0	2.1	1.4	0.269	0.090	0.044	0.011	7.2	2.1	7.4
NDLP														
All	0.000	0.000	0.036	0.006	2.9	1.1	0.1	0.000	0.000	0.042	0.008	3.4	1.1	0.6
Scotland	1.000	1.000	0.034	0.004	6.3	1.3	2.7	0.985	1.000	0.046	0.003	6.8	1.5	14.2
North East England	0.988	0.987	0.025	0.007	4.0	1.4	6.9	0.990	1.000	0.028	0.005	5.9	1.1	7.6
North West England	0.975	1.000	0.033	0.004	6.0	1.6	5.2	0.804	1.000	0.047	0.009	7.8	1.4	5.4
Wales	0.950	1.000	0.043	0.004	5.1	1.5	3.9	0.752	1.000	0.054	0.005	5.6	1.7	9.1
East Midlands	0.974	0.447	0.027	0.009	5.4	1.6	1.7	0.980	0.258	0.027	0.010	3.5	1.6	9.7
London	0.953	0.002	0.029	0.017	4.7	2.4	3.0	0.849	0.978	0.034	0.007	7.4	2.0	10.6

Notes:
Prob>chi: p-value of the likelihood-ratio test before (after) matching, testing the hypothesis that the regressors are jointly insignificant, i.e. well balanced in the two (matched) groups.
Pseudo R^2: from probit estimation of the conditional probability of being a non-participant (before and after matching), giving an indication of how well the observables explain non-participation.
Median bias: median absolute standardised bias before and after matching, median taken over all the regressors. Following Rosenbaum and Rubin (1985), for a given covariate, the standardised difference **before** matching is the difference of the sample means in the non-participant and participant subsamples as a percentage of the square root of the average of the sample variances in the two groups. The standardised difference **after** matching is the difference of the sample means in the matched non-participants (i.e. falling within the common support) and matched participant subsamples as a percentage of the square root of the average of the sample variances in the two original groups.
% lost to CS: Share of the group of non-participants falling outside of the common support.
Sample sizes: see Tables A.12 and A.13.

Table B.4 Survey outcomes: Matching quality not allowing for non-response, joint common support

	Non-participants vs responding programme group						Non-participants vs responding control group						
	Prob>chi		Pseudo R^2		Median bias		Prob>chi		Pseudo R^2		Median bias		% lost to CS
	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After	
ND25													
All	0.000	1.000	0.094	0.003	4.5	1.2	0.000	1.000	0.098	0.004	5.3	1.4	0.8
Scotland	0.573	1.000	0.148	0.043	13.3	4.1	0.667	1.000	0.142	0.040	10.4	3.9	16.7
North East England	0.006	1.000	0.108	0.018	9.5	2.3	0.004	0.997	0.112	0.022	11.1	3.6	12.7
North West England	0.264	1.000	0.093	0.014	7.2	3.6	0.273	1.000	0.095	0.023	7.7	3.5	3.7
Wales	0.659	1.000	0.116	0.043	9.8	5.9	0.456	1.000	0.128	0.035	9.4	4.6	15.6
East Midlands	0.411	0.999	0.059	0.015	5.5	2.8	0.048	0.996	0.078	0.017	6.9	3.8	2.7
London	0.176	0.999	0.067	0.013	5.6	2.6	0.016	0.999	0.090	0.013	10.2	2.2	5.2
NDLP													
All	0.000	0.997	0.182	0.005	3.5	1.2	0.000	0.740	0.193	0.008	4.7	2.3	1.0
Scotland	0.999	1.000	0.216	0.255	10.7	6.5	0.779	0.458	0.358	1.000	20.5	13.5	40.0
North East England	0.576	1.000	0.063	0.018	6.4	3.0	0.685	0.999	0.066	0.022	5.8	4.0	12.7
North West England	0.997	1.000	0.113	0.029	8.8	5.1	0.641	1.000	0.209	0.092	14.0	7.8	26.5
Wales	0.937	1.000	0.089	0.026	5.1	4.8	0.171	1.000	0.164	0.066	10.0	6.1	24.4
East Midlands	0.785	1.000	0.044	0.008	6.7	1.9	0.828	0.953	0.043	0.013	3.8	1.8	8.7
London	0.094	1.000	0.084	0.013	5.7	2.5	0.052	0.973	0.090	0.024	9.2	4.3	13.8

Notes:
Prob>chi: p-value of the likelihood-ratio test before (after) matching, testing the hypothesis that the regressors are jointly insignificant, i.e. well balanced in the two (matched) groups.
Pseudo R^2: from probit estimation of the conditional probability of being a non-participant (before and after matching), giving an indication of how well the observables explain non-participation.
Median bias: median absolute standardised bias before and after matching, median taken over all the regressors. Following Rosenbaum and Rubin (1985), for a given covariate, the standardised difference **before** matching is the difference of the sample means in the non-participant and participant subsamples as a percentage of the square root of the average of the sample variances in the two groups. The standardised difference **after** matching is the difference of the sample means in the matched non-participants (i.e. falling within the common support) and matched participant subsamples as a percentage of the square root of the average of the sample variances in the two original groups.
% lost to CS: Share of the group of non-participants falling outside of the common support.
Sample sizes: see Table A.17.

Appendix C
Power and validity of the instrument

Table C.1 First stage of the instrument

	ND25Plus		NDLP	
	F- statistic	*p*-value	F- statistic	*p*-value
Full sample (administrative outcome)				
Non-interacted linear model	11.8	0.001	19.0	0.000
Interacted non-linear model				
All Z terms	2.0	0.000	2.3	0.000
Z, Z^2, Z^3	1.4	0.231	3.8	0.009
XZ interactions	1.7	0.007	2.1	0.000
Survey-eligible sample (survey outcome)				
Non-interacted linear model	6.1	0.013	10.8	0.001
Interacted non-linear model				
All Z terms	2.0	0.000	2.2	0.000
Z, Z^2, Z^3	1.7	0.157	4.0	0.007
XZ interactions	1.8	0.002	2.1	0.000

Note: Sample sizes for full sample: 7,796 for ND25 Plus and 7,261 for NDLP.

Sample sizes for survey-eligible sample (i.e. sample of both participants and non-participants eligible for survey): 7,399 for ND25 Plus and 6,809 for NDLP.

Table C.2 Share of explained variance accounted for by the instrument in the participation equation (full sample)

	Share	*p*-value	(Pseudo)-R^2
ND25			
Logit	12.4	0.001	0.062
Regression	14.0	0.001	0.065
NDLP			
Logit	5.8	0.000	0.111
Regression	7.2	0.000	0.122

Note: Sample sizes: see Table C.1.

Table C.3 Testing part of the exclusion restriction

	ND25		NDLP	
	F-test	*p*-value	F-test	*p*-value
Full sample				
Days employed	0.8	0.364	0.5	0.502
Ever employed	0.1	0.731	0.0	0.998
Days on benefits	0.0	0.899	1.4	0.238
Survey-eligible sample				
Days employed	1.6	0.201	0.8	0.377
Ever employed	0.0	0.955	0.1	0.796
Days on benefits	0.0	0.915	2.3	0.129

Note: Sample sizes: see Table C.1.

References

Blundell, R. *et al.* (2005). Evaluating the effect of education on earnings: models, methods and results from the National Child Development Survey. *Journal of the Royal Statistical Society: Series A*, 168, 473-512.

Dolton, P. *et al.* (2008). *The Impact of the UK New Deal for Lone Parents on Benefit Receipt*, mimeo, March.

Dorsett, R. *et al.* (2007). *Implementation and first-year impacts of the UK Employment Retention and Advancement (ERA) demonstration*. Department for Work and Pensions Research Report No. 412, February.

Heckman, J.J. (1979). Sample Selection Bias as a Specification Error, *Econometrica*, 47, 153-161.

Heckman, J. *et al.* (1998). Characterising Selection Bias Using Experimental Data. *Econometrica* 66, 1017-1098.

Heckman, J. *et al.* (1999). The Economics and Econometrics of Active Labor Market Programs, in Orley Ashenfelter and David Card (eds.), *Handbook of Labor Economics, Volume 3A*. 1865-2097.

Heckman, J. and Smith, J. (1999). The Pre-Programme Dip and the Determinants of Participation in a Social Programme: Implications for Simple Programme Evaluation Strategies. *Economic Journal,* 109, 313-348.

Horowitz, J.L. and Manski, C.F. (2000). Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data, *Journal of the American Statistical Association*, 95, 77-84.

Goodman, A. and Sianesi, B. (2007). *Non-participation in the Employment Retention and Advancement Study: A Quantitative descriptive analysis*. Department for Work and Pensions Working Paper No. 39.

Hall, N. *et al.* (2005). *The Employment Retention and Advancement scheme – the early months of implementation. Summary and conclusions*. Department for Work and Pensions Research Report No. 265, August.

Kamionka, T. and Lacroix, G. (2005). *Assessing the External Validity of an Experimental Wage Subsidy*. IZA Discussion Paper No. 1508.

Imbens, G.W. and Manski, C.F. (2004). Confidence Intervals for Partially Identified Parameters, *Econometrica*, 72, 1845-1857.

Jawitz, J. (2004). Moments of truncated continuous univariate distributions. *Advances in Water Resources*, 27, 269-281.

Rosenbaum, P.R. and Rubin, D.B. (1985). Constructing a comparison group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39, 33–8.

Walker, R. *et al.* (2006) *Making random assignment happen: Evidence from the UK Employment Retention and Advancement (ERA) demonstration*. Department for Work and Pensions Research Report No. 330, March.