

Working Paper

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

by Haroon Chowdry and Barbara Sianesi

Department for Work and Pensions

Working Paper No 96

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

Haroon Chowdry and Barbara Sianesi

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

© Crown copyright 2011.

You may re-use this information (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence.

To view this licence, visit <http://www.nationalarchives.gov.uk/doc/open-government-licence/> or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gsi.gov.uk.

This document/publication is also available on our website at:
<http://research.dwp.gov.uk/asd/asd5/rrs-index.asp>

Any enquiries regarding this document/publication should be sent to us at:
Department for Work and Pensions, Commercial Support and Knowledge Management Team,
Upper Ground Floor, Steel City House, West Street, Sheffield S1 2GQ

First published 2011.

ISBN 978 1 84712 936 9

Views expressed in this report are not necessarily those of the Department for Work and Pensions or any other Government Department.

Contents

Acknowledgements.....	vii
The Authors.....	viii
Abbreviations and acronyms.....	ix
Summary.....	1
1 Background, research questions and overview.....	8
1.1 Background.....	8
1.2 Research questions.....	9
1.3 Overview.....	10
2 Non-participation in the ERA study: The issues.....	12
2.1 How did non-participation come about.....	12
2.2 What is known about non-participation in the ERA study.....	12
2.2.1 <i>Qualitative evidence</i>	12
2.2.2 <i>A quantitative description and assessment of non-participation</i>	15
2.3 A note on the WTC group.....	17
2.4 Experimental analysis and non-participation analysis.....	18
2.5 Summary.....	23
3 Data and sample definition.....	25
3.1 Data.....	25
3.2 Sample.....	25
3.3 Outcomes and observable characteristics.....	28
4 Methodological approaches.....	31
4.1 Analysis framework.....	31
4.2 Bounds without assumptions on the selection process.....	34
4.2.1 <i>Sensitivity analysis</i>	35
4.3 Impact estimates under selection on observables.....	35
4.3.1 <i>Selection on observable characteristics</i>	35
4.3.2 <i>Estimating impacts for the non-participants</i>	36
4.3.3 <i>Analysis of take-up</i>	37
4.4 Impact estimates under selection on unobservables.....	38

5	Implications of non- participation for the four-year experimental impact estimates	40
5.1	Experimental findings.....	40
5.2	Bounds	41
5.3	Selection on observables	43
5.3.1	<i>Sensitivity analysis</i>	45
5.3.2	<i>Analysis of take-up</i>	47
5.4	Selection on unobservables	51
5.4.1	<i>Testing for selection on specific unobservables</i>	51
5.4.2	<i>Control function results as a sensitivity check</i>	53
6	Summary and conclusions.....	56
6.1	Drawing the findings together.....	56
6.1.1	<i>ND25 plus group</i>	56
6.1.2	<i>NDLP group</i>	57
6.2	Conclusions	59
Appendix A	Control function: Methods and results.....	62
Appendix B	District-level results.....	69
Appendix C	Matching diagnostics	80
Appendix D	Power and validity of the instrument	84
References	85

List of tables

Table 1	Summary results of non-participation analysis for main ERA impacts, allowing for selection on observed characteristics	4
Table 2.1	Some causal effects of interest.....	24
Table 3.1	Sample breakdown by intake group	26
Table 3.2	ND25 plus: Sample breakdown by district	27
Table 3.3	NDLP: Sample breakdown by district.....	27
Table 3.4	Non-participation breakdown by district.....	28
Table 3.5	Summary of observed characteristics	29
Table 5.1	Experimental findings over the four-year follow-up period.....	41
Table 5.2	Derivation of bounds.....	42
Table 5.3	Matching four-year impact estimates	44

Table 5.4	Sensitivity analysis: ATE_{θ} , θ from 0.5 to 1.5	47
Table 5.5	Take-up and involvement with Jobcentre Plus in the first year predicted for the non-participants both under ERA and without ERA.....	50
Table 5.6	Differences in outcomes for participants (control group) compared to non-participants with the same observed characteristics during the four-year follow-up	52
Table 6.1	Summary of overall findings by intake group, analysis type and outcome during the four-year follow-up period.....	61
Table A.1	Control function models: ERA impacts on days in employment during the four-year follow-up – ND25 plus.....	66
Table A.2	Control function models: ERA impacts on days in employment during the four-year follow-up – NDLP	67
Table A.3	Control function models: ERA impacts on cumulative earnings	68
Table B.1	Experimental findings for the ND25 plus group during the four-year follow up.....	69
Table B.2	Experimental findings for the NDLP group during the four-year follow up	70
Table B.3	Non-parametric bounds for the ATE – Outcome: ever employed during the four-year follow up.....	71
Table B.4	Matching estimates for the ND25 plus group during the four-year follow-up .	74
Table B.5	Matching estimates for the NDLP group during the four-year follow-up.....	75
Table B.6	Differences in outcomes for participants (control group) compared to non-participants with the same observed characteristics during the four-year follow up: ND25 plus.....	77
Table B.7	Differences in outcomes for participants (control group) compared to non-participants with the same observed characteristics during the four-year follow up: NDLP.....	78
Table C.1	Estimation of the propensity score	80
Table C.2	Covariate balancing indicators before and after matching.....	83
Table D.1	First stage of the instrument.....	84
Table D.2	Share of explained variance accounted for by the instrument in the participation equation (full sample)	84
Table D.3	Testing part of the exclusion restriction	84

List of figures

Figure 2.1	From eligibility to service receipt in the ERA demonstration.....	20
Figure 4.1	Simplified structure of the problem.....	32
Figure 5.1	Sensitivity analysis: ATE_{θ} for ever employed during the four-year follow-up, θ from 0.5 to 1.5.....	43
Figure B.1	ND25 plus: Sensitivity analysis: ATE_{θ} for ever employed, θ from 0.5 to 1.5.....	72
Figure B.2	NDLP: Sensitivity analysis: ATE_{θ} for ever employed, θ from 0.5 to 1.5.....	73
Figure C.1	Common support between non-participants and programme group: Distribution of $P(Q=0 \mid Q=0 \vee R=1, X)$	82

Acknowledgements

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

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

The Authors

Haroon Chowdry is a Senior Research Economist in the Skills sector at the Institute for Fiscal Studies, and has an MSc in Economics from University College London. His recent research has examined the factors affecting educational attainment in secondary school and participation in higher education. He has also worked on a range of evaluations, including welfare-to-work and educational programmes.

Barbara Sianesi is a Senior Research Economist in the Education 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 policies as well as to educational investments.

Abbreviations and acronyms

ASA	Advancement Support Adviser
BIF	Basic Information Form
DWP	Department for Work and Pensions
ERA	Employment Retention and Advancement (study)
HMRC	Her Majesty's Revenue & Customs
IB	Incapacity Benefit
IS	Income Support
JSA	Jobseeker's Allowance
ND25+	New Deal 25 Plus
NDLP	New Deal for Lone Parents
RA	Random assignment
WTC	Working Tax Credit
WPLS	Work and Pensions Longitudinal Study

Summary

Overall summary

The Employment Retention and Advancement (ERA) study is a large-scale randomised evaluation that was run in six UK regions between October 2003 and October 2007 to test the effectiveness of an innovative package of support for individuals starting the New Deal 25 Plus (ND25+) or the New Deal for Lone Parents (NDLP) programmes.

The aim of this report is to explore how the four-year findings from the experimental research relate to the impacts that would have been experienced, on average, by all the people who were eligible for ERA, had they participated in the programme. (Please note that a full ERA evaluation report looking at the five-year impacts, as well as at the costs and benefits associated with the programme will be published by the Department for Work and Pensions (DWP) during spring 2011.)

Overall, the findings either validate the four-year interim experimental results of the main ERA evaluation or indicate that the ERA intervention would have been even more effective for its full eligible population than it was for its actual participants. Specifically, for the ND25+ group the experimental impacts on earnings appear to be representative of the impacts that the full eligible ND25+ population would have experienced under ERA. The experimental findings on employment are, however, found to actually underestimate, by over one-third, the gains that the ND25+ eligible population would have enjoyed under ERA.

A similar story emerged for the NDLP group, but in terms of different outcomes. The NDLP eligible population would have experienced the same (zero) impact as the participants in terms of employment. However, the eligible population would have experienced a significant and positive increase in earnings instead of the absence of any significant impact as experienced by the study participants.

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 ND25+ 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 Great Britain (see e.g. Dorsett *et al.*, 2007, for additional background on ERA).

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 issue

All individuals flowing into ND25+ 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 not all of them 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 eligible individuals 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 ND25+ and 30 per cent of NDLP – were thus not represented in the experiment.

Research objectives

After setting the foundation work for the analysis of non-participation in the ERA study, this report employs an array of non-experimental methods to inform the policymaker on the impact that the full ERA eligible population would have been likely to experience had they been offered the chance to participate in ERA. The report thus aims to:

- explain the subtle issues that non-participation raises for the ERA demonstration;
- introduce the different approaches and methodologies to deal with it;
- present the findings four years after entering the New Deal programme; and
- discuss emerging lessons.

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 the full eligible population compare to the experimental impact estimate obtained for the ERA study participants?
- 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. In other words, 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 eligible individuals 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 the eligible population of making the ERA package available. This **average effect of the ERA offer for all those eligible for ERA 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 eligible individuals, rather than over a potentially adviser-selected and self-selected subgroup of the eligible population.

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 Deal participant would automatically be enrolled upon having found work.² A scenario in which all NDLP and ND25+ entrants are automatically ‘opted in’ for ERA gives direct and high policy relevance to the full NDLP and ND25+ samples, the focus of this report.

The report assesses whether an overall 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

Had the level of non-participation been low, this would be of less interest. It is the fact that non-participation is around a quarter of all those eligible that opens up at least the theoretical possibility that the experimental impacts are biased for the impact that the full eligible population would have experienced.

The ERA study offers the rare chance to look at this issue because: (1) the treatment is the offer of ERA support and incentives; (2) the whole population of ND25+ and NDLP entrants in the six districts was eligible for this offer (and would be eligible under an official policy); and (3) such entrants are identified in the available administrative data.

Types of non-experimental analyses

This study performs different types of non-experimental analyses seeking to recover the potential impact of ERA on the full eligible population (in the six districts) and compares it to the experimental impact estimate for the ERA study participants. In most cases, identifying and estimating the average impact on the eligible population 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.

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 the analyst observes all outcome-relevant characteristics that drive selection into the ERA study (‘matching approach’); and
- impact estimates that allow selection into the ERA study to depend on unobservable factors (‘control function approach’).

The specific nature of the set-up and data – randomisation coupled with administrative outcome data for the non-participants – allows one to perform a number of tests not generally available. Specifically, one can test for the presence of residual selection on some type of unobservable; one can test, to some extent, the validity of the instrument needed for the control function model; and one can test two other features of the performance of the control function model.

² An eligible worker 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 the whole eligible population.

Based on extensive diagnostic and specification tests, as well as on contrasting and cross-checking the findings and evidence from the different methodological approaches, the picture that emerges based on the most robust results is summarised in the following key findings.

Key findings

The analyses in this report have considered the impact that the full population eligible for ERA would have experienced had they been offered the chance to participate in ERA and how this relates to the experimental impact found for the study participants. The story that emerged, summarised in Table 1, differs for the ND25+ and NDLP groups and across outcomes.

Table 1 Summary results of non-participation analysis for main ERA impacts, allowing for selection on observed characteristics

	Based on ERA study participants only	After allowing for effects of non-participation
ND25+ group		
Ever employed **	+2.8** ppts	+3.7*** ppts
Days in employment	+25** days	+34*** days
Earnings 2005–09	+£1,805**	+£1,940***
NDLP group		
Ever employed	-0.8 ppts	0.0 ppts
Days in employment	-4 days	+3 days
Earnings 2005–09 *	+£767	+£1,262**

Notes: Significance * at the 10% level, ** at the 5% level, *** at the 1% level.

Rows marked with asterisks denote a result allowing for non-participation which is significantly different from the experimental one based on participants only.

The second column of Table 1 presents key findings from the experimental research four years after programme start.⁴ No impact could be detected on the employment and earnings outcomes of the NDLP programme group compared to the control group. By contrast, ND25+ study participants were found to enjoy better employment outcomes (a 2.8 percentage point higher probability of being in work and an extra 25 days spent in employment) and higher earnings (an annual increase of £392 and a £1,805 cumulative gain) due to ERA.

As shown in the third column, the findings from analyses in this report either validate the four-year experimental results of the main ERA evaluation or indicate that the ERA intervention would have been even more effective for its full eligible population than it was for its actual participants. Specifically, for the ND25+ group the experimental impacts on earnings appear to be representative of the impacts that the full eligible ND25+ population would have experienced under ERA. The experimental findings on employment are however found to actually underestimate, by over one-third, the gains that the ND25+ eligible population would have enjoyed under ERA.

⁴ A full ERA evaluation report looking at the five-year impacts, as well as at the costs and benefits associated with the programme will be published by DWP during spring 2011.

ND25+ group

What the table clearly shows is that the overall ND25+ pattern of significant increases in earnings and employment is confirmed; both in terms of size and statistical significance, impacts for the full population would have been stronger than for the experimental group of participants. Indeed, for the probability of being ever employed during the four-year follow-up the difference in impacts for the eligible population and for the experimental sample reaches statistical significance.

- If all those eligible to take part in the ERA study had done so, the **employment** effects of ERA in the 48 months since inflow would have been significantly higher. In other words, according to the best estimates derived in this report, the effect on employment outcomes from the experimental analysis is a significant underestimate of the impact which would have been seen if all eligible individuals had participated.

This result is driven by the fact that the employment impact for the non-participants would have been considerably larger than the one for the participants.

- Compared to a 2.8 percentage point increase on the probability of ever being employed over the four-year period found for the participants, it is estimated that the non-participants would have enjoyed a 6.4 percentage point increase, and all those eligible for ERA a 3.7 percentage point increase, with both impacts for the non-participants and for the eligibles being statistically significantly different from the impact enjoyed by the participants.
- Similarly, compared to an increase in employment duration over the four-year period of 25 days found for the experimental sample of participants, the non-participants would have enjoyed a 63 day increase, and all those eligible for ERA a 34-day increase.
- By contrast, the four-year experimental estimates of the impact for study participants are representative of the average impact that the eligible population would have experienced in terms of **earnings**, as measured by the four tax-year earnings post inflow and by cumulative 2005–09 tax earnings. Specifically, the point estimates for the earnings impacts for non-participants as well as for the full eligible population are always larger than the corresponding point estimates for participants, but such differences are far from reaching any statistical significance.

NDLP group

For the NDLP group, the broad pattern of no impact on employment is confirmed, but the previously insignificant positive impact on earnings now becomes a significant one.

- The experimental impact estimates of **employment outcomes** in the four years post-inflow are found to be representative of the average impact that ERA would have had on its eligible population. Specifically, no significant impact was found for the experimental sample on either employment duration or the probability of ever being employed during the four-year follow-up period, and the absence of any significant impact extends to the non-participants, and hence, to the eligible population. Again, it is worth noting, however, that the point estimates for the eligible population and especially for the non-participants are higher and in positive territory compared to the (insignificant but) negative ones for participants.
- By contrast, if all those eligible for ERA had taken part the **earnings** impacts for the NDLP group would have been higher. In other words, excluding the non-participants from the analysis significantly underestimates the impact that ERA would have had on the average earnings of all those eligible for ERA. This is the case for earnings in the 2006/07 and 2008/09 tax years as well as for cumulative earnings (point estimates for the non-participants are strongly significant and larger than the ones for participants in the other two fiscal years as well). Specifically, while ERA has not significantly increased participants' cumulative earnings over the follow-up period (an insignificant £767 point estimate), non-participants would have enjoyed a highly significant £2,393 rise in cumulative earnings and the eligible population a significant £1,262 rise.

In conclusion, for both New Deal groups and in terms of all outcomes considered, non-participants would have enjoyed uniformly higher ERA impacts than do participants. Only for ND25+ employment and NDLP earnings outcomes do such differences reach statistical significance and indeed extend to significant differences between impacts for study participants and impacts for the full eligible ERA population.

Finally, the report has assessed the conjecture that if ERA became an official policy, non-participants would be mostly uninterested in taking up its support and incentives anyway. Exploiting first-year follow-up data on participants' involvement with Jobcentre Plus, the analyses have found no support for this hypothesis for either intake 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 was in the first year only in terms of a couple of measures, and then only marginally so. Indeed, had they become eligible for 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 within their first year, 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 in the first year.

Conclusions and lessons learnt

The picture that emerged from the interim four-year experimental findings is that ERA has not been effective in improving the employment and earnings outcomes of NDLP participants in the four-year follow-up period, but that the intervention has significantly improved both types of outcomes for ND25+ participants.

How has the presence of the non-participants affected the representativeness (or external validity) of such experimental impact estimates?

Overall, the representativeness of the four-year experimental impact estimates was found to depend on the intake group and outcomes considered. As a result, it was found that the four-year experimental findings either do not change or provide an overly conservative picture of the effectiveness that ERA would have had on its full eligible population. Specifically:

- For ND25+ earnings and NDLP employment the four-year experimental results appear to generalise to the full eligible population, leaving the story unchanged. In particular, the four-year experimental findings show that ERA had no effect on NDLP employment, while it has significantly and substantially increased ND25+ earnings. This report shows that what the programme has done for the participants in these dimensions, it would also have done for the non-participants and hence, for the full eligible population.
- By contrast, the representativeness of the ERA study is lower in terms of ND25+ employment and NDLP earnings, with the four-year experimental impact findings underestimating the gains that all eligible individuals would have enjoyed had they been offered ERA services and incentives. These analyses thus do find evidence of non-participation bias (or of some loss in external validity) in the ERA study. For the ND25+ group, the employment impact estimates that rely on experimental data alone (+2.8ppt probability and +25 days) underestimate the likely impact that ERA would have had on all ND25+ entrants (+3.7ppt probability and +34 days). For the NDLP group, it is the earnings experimental impact estimates (no significant increase) that underestimate the likely impact that ERA would have had on all NDLP entrants (+£1,262 cumulatively).

Of course, there is always the issue of how different the estimates for the eligible population and for the experimental sample need to be for one 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 eligible group which is over one-third larger than the experimental estimate or indeed a large, significant impact for the eligible population that surfaces when none was found experimentally can be viewed as findings of substance.

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 Great Britain 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 25 Plus (ND25+) 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 (see, for example, Dorsett *et al.*, 2007, for additional background on ERA).

Since ERA offered a package of support once in work⁶, all individuals flowing into ND25+ 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 not all of them have entered the evaluation sample: some eligible individuals 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 eligible population – 23 per cent of ND25+ and 30 per cent of NDLP⁷ – was thus not represented in the experiment.⁸

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. See Section 2.4 for a more extended discussion.

⁶ Eligible individuals 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 at least 30 hours a week, for a work retention bonus as well.

⁷ The composition of the non-participants varied markedly between the two intake groups, as the bulk of non-participation in the ND25+ group was thus due to formal refusals, while in the NDLP group by diverted customers. Specifically, nine per cent of all ND25+ eligibles appear to have been diverted and 14 per cent formally refused. By contrast, over one quarter (26.4 per cent) of all eligible NDLP entrants in the six districts appear to have been diverted, while only four per cent formally refused.

⁸ Goodman and Sianesi (2007) have explored diversion/refusal in detail; some of their main conclusions and findings are summarised in Chapter 2.

1.2 Research questions

The key objective of the report is to quantify the impact that the full ERA eligible New Deal population would have been likely to experience in the four years since inflow into the New Deal had they been offered the chance to participate in ERA, and to assess how this impact for the full eligible group relates to the experimental impact estimated on the subgroup of study participants.⁹

To achieve this aim, the report first sets out the foundation work for the analysis of non-participation in the ERA study not only by introducing the different approaches and methodologies to deal with non-participation, but also by explaining the various and subtle issues that it raises for the ERA demonstration.

So 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 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 eligible population and on the causal effect for them of making the ERA package available. This **average effect of the offer of ERA for all those eligible for ERA in the six districts** over the study intake 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 full eligible group, rather than over a potentially adviser-selected and self-selected subgroup of the eligible population.

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 Deal participant would automatically be enrolled upon entering work.¹⁰ A scenario in which all NDLP and ND25+ entrants are automatically ‘opted in’ for ERA gives direct and high policy relevance to the full NDLP and ND25+ samples, 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 eligible population. 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?¹¹

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) the whole population of ND25+ and NDLP entrants in the six districts was eligible for this offer (and would be eligible under an official policy); and (3) such entrants are identified in the available administrative data.

⁹ Throughout the report, ‘participants’ refers to everyone who entered the randomisation process, i.e. it includes those receiving ERA services and the control group.

¹⁰ An eligible worker 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.

¹¹ Alternatively, non-participation can be viewed as introducing potential bias in the experimental estimate if interest lies in the impact of ERA on the eligible population (in the six districts).

A previous descriptive report (Goodman and Sianesi, 2007) has explored how representative the group is for whom one 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 general aim of this report is thus to build on these descriptive findings to assess whether an overall non-participation rate of 26.6 per cent is likely to have affected the extent to which the four-year experimental results can be generalised to the full eligible population, and hence, their representativeness and policy relevance. Had the level of non-participation been low, this would be of less interest. It is the fact that non-participation is around a quarter of all those eligible that opens up at least the theoretical possibility that the experimental impacts are biased for the impact that the full eligible population would have experienced.

1.3 Overview

The study performs different types of non-experimental analyses seeking to recover the impact of ERA on the full eligible population (in the six districts) and compares it to the four-year experimental impact estimates for the ERA study participants. In most cases, identifying and estimating the average impact on the eligible population requires first 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.

The report starts by considering analyses that provide bounds for the impact of interest without any assumption on the selection process.

Next, impact estimates are provided under the assumption that one can 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. Furthermore, within this framework one can estimate the type of involvement that the non-participants would have had with ERA and more generally with Jobcentre Plus in the first year had they participated in the evaluation study. This allows one 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.

Finally, approaches are considered 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 available data contains no direct information on educational attainment, which is thus among the 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. The analysis starts with the standard sample selection model, but then extends it in various directions: it relaxes independence between the observed characteristics and the unobservables; it relaxes the normality assumption; and it allows 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 (1979) selection model. In this set-up one is, however, in the rather unique position to observe the (administrative) outcomes of the selected-out sample – the non-participants. Coupled with randomisation, this feature of the data allows one to test for the presence and extent of residual selection on some type of unobservable. The study further exploits it to test part of the assumption needed for the validity of the instrument, as well as two other features of the performance of the model. Specifically, the analysis is 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 that has been 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.

- **Chapter 2** outlines how non-participation in the ERA evaluation has come about and summarises the available qualitative and quantitative evidence. It then briefly discusses the intake group excluded from this analysis (lone parents working part-time and in receipt of WTC), before focusing on placing the experimental and non-participation analyses into proper context.
- **Chapter 3** briefly describes the data and the working definition of ERA eligibility. It also provides sample breakdowns by intake group and district and describes the rich set of variables that have been collated from different sources in order to capture key characteristics relating to the individuals themselves, their office and their local area.
- The methodological approaches and the type of analyses performed in the report are presented in **Chapter 4**. The description is kept as non-technical as its rather technical nature allows.
 - **Section 4.1** starts by formally presenting the analytical framework, together with conditions for the experimental impact estimate to coincide with the average impact for the full eligible population.
 - Bounds which make no assumption on the selection process into the ERA study are discussed in **Section 4.2**. This section also sketches some sensitivity analysis to assess how robust the estimate of the average treatment effect for the full eligible group is to assumptions about the selection process.
 - **Section 4.3** deals with methods relying on the selection-on-observables assumption. It starts by briefly relating the available data to the plausibility of this assumption. It then outlines the approach to estimate the impact on all eligible individuals and suggests simple sensitivity analyses to assess how sensitive the estimates are to straightforward violations of this crucial assumption. 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 and outlines the basic ideas behind the ‘control function method’, as well as tests on whether there are outcome-relevant unobservable differences between ERA study participants and non-participants and for the validity of the instrument needed for this type of model. Appendix A contains a more in-depth description.
- The results of all empirical analyses are presented and discussed in **Chapter 5**.
 - **Section 5.1** starts by presenting the four-year experimental findings concerning the average impact of ERA for the participants.
 - **Section 5.2** reports the findings from the bounds and sensitivity analyses.
 - **Section 5.3** focuses on those arising from the different models based on the selection-on-observables assumption. This section also includes the results of the analysis of take-up of ERA services and involvement with Jobcentre Plus.
 - **Section 5.4** is devoted to presenting and discussing the tests and sensitivity checks relating to the different control function models.
- **Chapter 6** summarises the key results and briefly concludes.
- The appendices provide additional material: **Appendix A** contains a more in-depth description of the control function approach, as well as the full results for these models. **Appendix B** presents the results and summary boxes for the district-level analyses, while **Appendices C and D** contain intermediate diagnostic and estimation results.

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:

- 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
- 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 whilst 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 What is known about non-participation in the ERA study

2.2.1 Qualitative evidence

Qualitative work conducted as part of the ERA evaluation has shed interesting light on the origins and sources of non-participation. In particular, Hall *et al.* (2005) and Walker *et al.* (2006) have looked closely at the assignment and participation process in ERA at selected sites. Based on detailed observations, interviews and discussions with both staff and individuals, the authors have put forward the conjecture that it is quite unlikely for ERA non-participants to be a random subgroup of the two eligible New Deal groups.

Recognising that two parties – the caseworker and the individual – are involved in the decision processes that led to inclusion in the sample of ERA study participants, the discussion of what is known about non-participation from this qualitative work is organised in two parts.

Since the individual can only refuse once having been offered the chance to participate, the individual’s decision has direct bearing on the second choice, i.e. the one between participation and formal refusal. On the other hand, the caseworker can affect both types of outcomes: they basically have sole decision power as to who to offer ERA to, as well as considerable influence in steering the individual’s response to such an offer. In an individual case, it might also be lack of understanding of the process on the part of the adviser, or even the possibility that the New Deal starting dates (which qualify an individual to be offered ERA) as recorded on the system may not be as precisely perceived by staff.

Ensuring that staff randomly assigned all eligible individuals

The six districts could exercise significant discretion in how they organised the ERA recruitment, intake and random assignment processes, so that a number of models ended up being used.¹² Although the expectation in any model was that the intake staff, be it an ERA adviser (ASA) or a New Deal PA, would encourage **all** eligible individuals – and encourage all of them **equally hard** – to consent to be randomly assigned and have a chance to participate in ERA, staff could use discretion on two fronts:

- what individuals to tell about ERA, directly determining the extent of diverted customers; and
- in what terms to present and market ERA to individuals, thus affecting the likelihood that they would become formal refusers.

As to the latter, the abstract notion that staff would use the same level of information and enthusiasm in recruiting all eligible individuals was particularly hard to implement in practice.¹³ Discretion in their choice of marketing strategy could take various forms, e.g.:

- how ‘hard’ to sell ERA;
- what features of the programme to mention – in particular whether and in what terms to mention the retention bonus, or whether to selectively emphasise features (e.g. the training bonus) to make ERA more appealing to the particular situation of a given individual;
- how far to exploit the misunderstanding that participation in ERA be mandatory.

But why and under what circumstances would caseworkers want to apply such discretion?

There could have been situations where the adviser did not deem that the individual would be interested in taking advantage of ERA or would benefit from it.

Furthermore, the Jobcentre Plus target structure gave advisers individual-level targets for how many people they moved into work and accordingly rewarded staff for job entries. This incentive structure seems to have led advisers conducting the intake process to use their own discretion in deciding what individuals to sell random assignment to or how hard to sell it in order to ‘hang onto’ those who they perceived as clearly likely to move into work quickly. The discussion in Walker *et al.* (2006) highlights how job entry targets had an asymmetric influence on incentives of New Deal and of ERA

¹² The model closest to the original plan saw ERA intake and random assignment being undertaken by a specifically allocated intake adviser, who had no vested interest in its outcome. In other districts, it was the New Deal Personal Advisers (PAs) who conducted the intake and randomisation, with the ERA advisers (Advancement Support Advisers (ASAs)), being responsible for working with ERA programme group members only after random assignment had taken place. In yet other districts, the ASAs were also responsible, alongside the New Deal PAs, for conducting intake interviews and randomisation. Typically, ASAs in these districts handed over to the New Deal advisers those individuals allocated to the control group and those who had refused to participate in ERA. These models did not necessarily apply at the district level, since within a particular district, different offices and staff members sometimes used somewhat different procedures. Furthermore, the intake and randomisation procedures varied over time, in the light of experience and depending on the situation and needs of the district or even a single office.

¹³ In addition to discretionary choices about how much information to disclose, it also became apparent that probably owing to their greater knowledge of and enthusiasm for ERA, ASAs tended to give clearer explanations of ERA than PAs (Walker *et al.*, 2006, Appendix F).

advisers: where the intake was conducted by New Deal advisers, job-ready individuals would be more likely to be diverted from ERA; where ERA advisers were doing the intake, they would be less likely to be diverted.¹⁴

It is thus known from this research that ERA non-participants, and especially diverted customers, are not likely to be random subgroups of the eligible population; rather, these are people whom advisers had a vested interest in not subjecting to ERA.

How willing were individuals to be randomly assigned?

Individuals who were given the option to participate in random assignment could **formally refuse** the offer, and thus be excluded from the experimental sample of ERA study participants.

A data-driven qualification to the statement that the individual can only formally refuse relates to the fact that individuals cannot be forced to fill in the Basic Information Form (BIF) in order to continue onto the New Deal programmes. Hence, an adviser might have actively offered ERA to an individual, but if the person was not interested for whatever reason and refused **without** filling the BIF, this person would appear in the data as a 'diverted customer' while in fact being a formal refuser. Although no quantitative or qualitative evidence is available on how frequently such a situation has occurred, it is thus worth keeping in mind that, strictly speaking, the individuals themselves might have partly contributed to determining the extent of 'diversion'.

What is clear from the qualitative work is that recruitment to ERA greatly differed between the two New Deal groups. While lone parents on New Deal for Lone Parents (NDLP) were all volunteers to that programme and thus, mostly responded favourably to ERA too, New Deal 25 Plus (ND25+) participants were more difficult to recruit. The reasons for formal refusal that were identified were:

- being puzzled by how the additional offer of ERA fitted in the mandatory participation in ND25+;
- having been unemployed for long periods of time and thus finding it difficult to envisage what might happen after they obtained a job, an outcome that they and their advisers thought rather unlikely anyway;
- feeling close to getting a job in the near future and not wanting to stay in touch with Jobcentre Plus.

¹⁴ *'Overall, when New Deal Personal Advisers undertook the interviewing, they had reason to encourage people with poor job prospects to join ERA (because in many cases they would move on to ASAs and off their caseloads) and those with good prospects to refuse (because they would keep them on their caseloads and get credit for a placement). When ASAs were involved in conducting intake interviews, they could have benefited from encouraging customers with poor employment prospects to refuse ERA and people with good prospects to join.'* (Walker et al., 2006, p.26). The study concludes on this issue that: *'While [this] incentive structure was real and widely recognised, it is impossible to assess with any degree of precision how strong an effect it had on marketing strategies (and, thus, on the resulting make-up of the groups of customers who ended up being randomly assigned).'* (p.27).

It thus appears that the group of formal refusers, and in particular those amongst the more problematic ND25+ group, might be far from random, and instead selected on (predicted) non-ERA outcomes.¹⁵

Furthermore, some staff identified specific attitudes and traits as good predictors that individuals, particularly among those mandated to start ND25+, would decline participation:

- a strong antipathy to government, feeling alienated from systems of support and governance;
- being resistant to change or taking risks, ‘preferring to stick with what they know’;
- reacting against the labour market, enjoying being able to refuse to do something in the context of a mandatory programme.

A further possible reason for refusal is being engaged in benefit fraud.

Overall, the available qualitative evidence on refusals suggests that those who declined to join may, in fact, differ in important respects from those who agreed to participate. Formal refusers, especially those amongst the more problematic ND25+ group, appeared to have weaker job prospects and poorer attitudes than the average New Deal entrant.

In addition, the refusal rate was observed to fall later on during random assignment, likely due to a combination of enhanced adviser experience at selling ERA and the permission to mention the monetary incentives. The refusal process is, thus, likely to have changed over the intake window, with refusers in later entry cohorts presumably forming quite a selective group.

Finally, as mentioned above, the incentive structure arising from Jobcentre Plus job entry targets had an asymmetric influence on New Deal and on ERA advisers in terms of how hard to sell ERA. Specifically, when New Deal advisers undertook the intake interviews, they could benefit if job-ready individuals refused to participate in ERA and those with bad prospects consented. Conversely, when ERA advisers were leading the intake process, they could benefit if individuals with bad job prospects formally refused to participate, while those with good prospects agreed to participate.

2.2.2 A quantitative description and assessment of non-participation

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 one 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 the 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 ND25+ and NDLP intake 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

¹⁵ Traditionally, the default assumption for a voluntary programme is that those who select into the programme are those most likely to gain from it. So in the ERA case, the exclusion of the formal refusers might be expected to bias the impacts upward. This, however, assumes that individuals can reliably predict not only their non-ERA outcomes but also their outcomes under ERA, which in the case of ERA appears particularly problematic given the limited information potential participants were given about the programme.

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 ND25+ and NDLP intake groups, as well as across districts. As to incidence, non-participation overall was lower amongst the ND25+ group (23 per cent of the eligible group) than amongst NDLP entrants (over 30 per cent). In terms of composition, nine per cent of all ND25+ eligibles appear to have been diverted and 14% formally refused. By contrast, over one quarter (26.4 per cent) of all eligible NDLP entrants in the six districts appear to have been diverted, while only four per cent formally refused. The bulk of non-participation in the ND25+ group was thus due to formal refusals (59 per cent), while in the NDLP group by diverted customers (86 per cent).

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 intake 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 entrants did not take part in ERA, most of them diverted customers. Focusing on the ND25+ 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 eligible ND25+ participants 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 intake 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 an individual's district, office affiliation and inflow month, underscoring the key role played by local practices and constraints. Individual 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 entrants, but to have significantly **worse** employment outcomes among ND25+ entrants.¹⁶

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, intake 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 ND25+ group are significantly easier to employ than the average ND25+ 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.

¹⁶ ND25+ 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/New Deal entrant 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 eligible population. The analyses in the present report aim to formally assess and quantify the loss in external validity, or the amount of non-participation bias.

2.3 A note on the WTC group

This report only considers the issue of non-participation for the two main ERA target groups – the ND25+ and NDLP groups – which represent 83 per cent of all ERA study participants. The third ERA group, lone parents working part-time and in receipt of Working Tax Credit (WTC), is excluded from this analysis for both conceptual and data reasons.

For the two groups considered, prior to being offered the chance to participate in ERA, all individuals were committed to receiving back-to-work support from Jobcentre Plus. Since the ERA offer included all of the same services as New Deal, plus additional help (including the possibility of substantial monetary bonuses) it was expected that the overwhelming majority would agree to participate.¹⁷ The expectation was that, were ERA to be implemented, it would be in place of the existing New Deals, so that all eligible individuals would necessarily participate. Moreover, because those who do not participate are, nevertheless, active benefit recipients and/or programme participants, there is useful data on their activities and outcomes.

By contrast, the lone parents in the WTC group were not currently receiving any employment support from Jobcentre Plus. The intended target group was made up of those lone parents working between 16 and 30 hours who were interested in increasing their hours to above 30. It was always anticipated that at most, 30 per cent of those potentially eligible would take up the offer and indeed the impact of interest has always been the average effect of the ERA offer **for those WTC recipients who have volunteered to receive it**, a potentially selected subgroup of the eligible WTC recipients contacted by letter or phone by Jobcentre Plus. This is thus a different parameter than the average effect of the offer of ERA for all **eligibles**, which was the relevant one for the other two ERA groups. Note, however, that both types of parameter are intention-to-treat parameters, that is, they consider the impact of **offering** ERA unconditional on the decision to take up ERA elements and services.¹⁸

As to the issue of **diversion**, the presumption was that **all** eligible WTC recipients in the six districts were to be contacted by Jobcentre Plus, made aware of ERA and offered the chance to participate in it. It does, however, appear that some eligible names were not called up by offices, although because of lack of data it has not been possible to assess the extent of this. The hope is that if this has indeed happened, it has happened in a random way.

In terms of **formal refusal**, as one might expect given the set-up, only a small minority of those WTC recipients who came to volunteer for ERA at Jobcentre Plus did then actually formally refuse before

¹⁷ In the original ERA design, it was anticipated that 95 per cent of those offered the chance to participate would take it up; the issue of diversion was not anticipated.

¹⁸ No WTC recipient who has come forward for ERA can actually be forced to meet with their adviser, follow their advice, retain work, undertake training, etc. The extent to which a WTC recipient who has volunteered for ERA **actually uses** ERA services is left to the individual's choice, and is not conditioned upon when estimating (experimental) impacts.

being randomly assigned, 0.6% (compared to 4.7% of NDLP and 14.6% of ND25+, all percentages being conditional on ERA offer). It is however noteworthy that a non-negligible share (4.4%) of those who came forward remained undecided, considerably larger than the fraction of undecided among NDLP (1.8 per cent) or ND25+ (1.6 per cent). Overall, what we define as formal refusers represent 5.0 per cent of the WTC recipients who came forward, which is comparable to the corresponding share of NDLP, 6.6 per cent (but much lower than the 16.2 per cent ND25+ share).

It is important to note that for this third customer group, there is an **additional** potential source of bias in the composition of the experimental sample compared to the scenario in which ERA is run as an official policy. The limited information about ERA conveyed to potential participants is most likely to have affected the composition of WTC recipients **deciding to volunteer** for ERA under these circumstances. In other words, the WTC group who made a special journey to Jobcentre Plus offices (an organisation they may not have had any previous dealings with) to volunteer for a 50/50 chance to join a vaguely described intervention is most likely to differ from the WTC group who would come forward in a full-information situation. Again, it was the requirement not to disappoint the randomised-out controls that led the offices to explain very little (at least initially) about the actual generosity of ERA support and incentives. In a situation of normal operation of ERA, many more – and most likely quite different – WTC recipients would have come forward to participate in the ERA study.¹⁹ The experimental impact on the smaller group who did come forward may, thus, not reflect what the impact would have been on the wider group who would have come forward in a normal setting.

As discussed, the issue of non-participation conceptually pertains to the WTC group too, but it has more layers to it and is even more complex given the element of individual choice that defines the population of interest (i.e. those eligible WTC recipients who have come forward to volunteer for ERA). Any analysis of non-participation for this group would need to speculate about what the WTC group's participation in ERA in the six evaluation districts would have been if ERA had been an official policy requiring eligible WTC recipients to actively apply for it.

Moreover, many WTC recipients eligible for ERA have neither been on benefits nor on programmes; for them there is no Department for Work and Pensions (DWP)/ Work and Pensions Longitudinal Study (WPLS) data at all.

For both conceptual and data reasons the issue of non-participation among the WTC group is, thus, not considered in this analysis.

2.4 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:

¹⁹ A WTC recipient who decided to come forward would immediately qualify for free training and the training bonus, plus have immediate access to the In Work Emergency Fund. Coming forward would thus entail an immediate and certain benefit (as opposed to the New Deal entrants who, being still in unemployment, might view such incentives as more remote).

1 Satisfy the criteria for ERA eligibility

Starting the ND25+ or 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 eligible population (9.4 per cent of eligible ND25+ and 26.4 per cent of eligible NDLP).

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 Deal participant would automatically be enrolled; in this case, no formal application process would be necessary. Alternatively,
- b New Deal participants would need to make their eligibility operational by registering with the ERA programme to be allocated to an adviser.

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.²⁰

4 Decide on the take-up of services

ERA was a voluntary programme in the sense that it was up to the participating individuals to decide whether, and to what extent, to avail themselves of the ERA elements. Specifically, eligible individuals 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 intake 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 ND25+) 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 ERA services and incentives.

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

Furthermore, the ERA intervention itself represents an **offer** of support and incentives. The experimental estimator is thus perfectly suited to recover the impact of **offering** ERA services and

²⁰ Those consenting to take part 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.'* Those formally refusing would by contrast sign that *'I do not consent to taking part in this research scheme or to being randomly assigned.'*

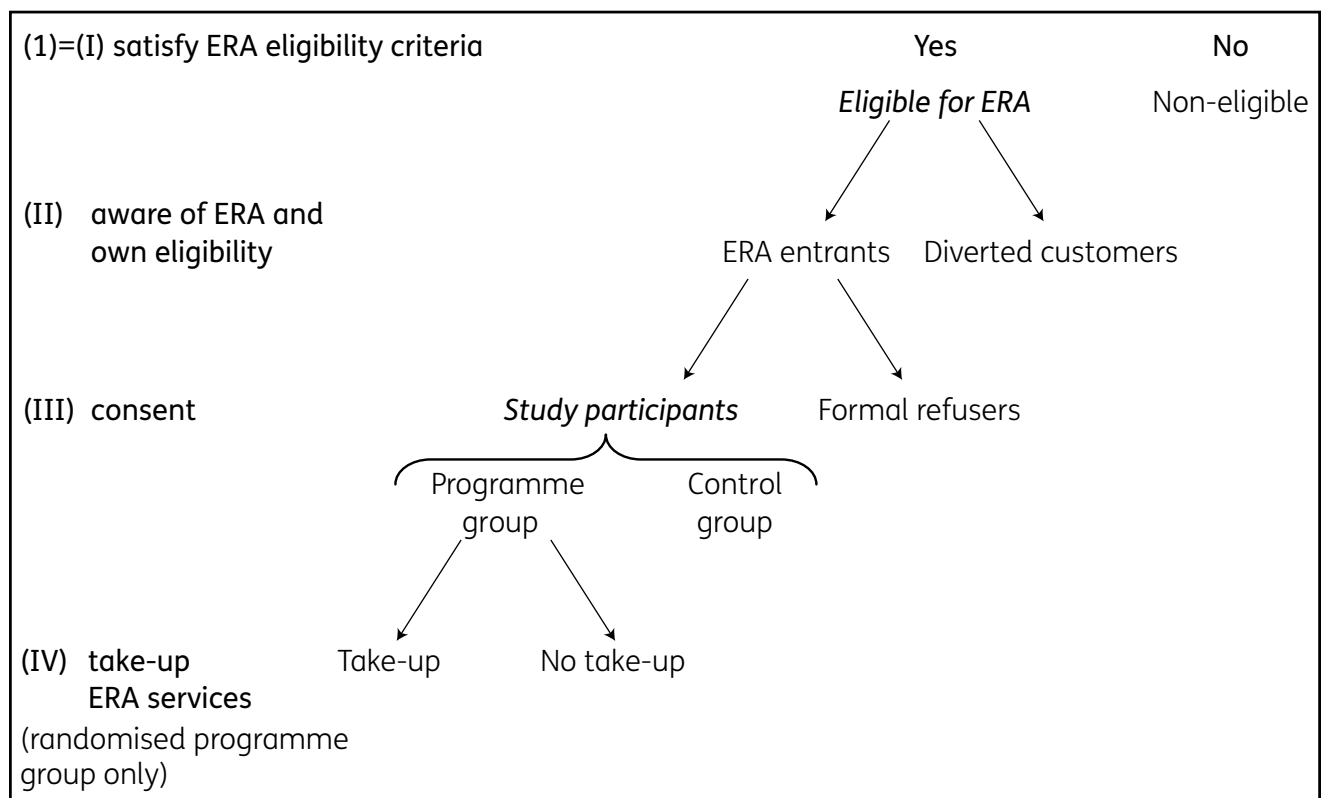
incentives. As mentioned, it is unconditional on the actual take-up of the services or even actual knowledge of the services and incentive structure.

Thus, coming back to the first issue the section 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** for 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).²¹

To better understand where the experimental parameter fits into this discussion, Figure 2.1 shows the structure of selection into the ERA study group, where the Roman numerals 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 ‘official-policy running’ means a situation in which ERA would be available to all ND25+ 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 versus comparison area-based evaluation scheme). In the demonstration, only parts of the target population entered the evaluation sample: some eligible individuals actively refused to be randomly assigned and 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



²¹ 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).

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 it was an official policy which required individuals to actively apply for it (Stage 3b).

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. On 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 3b.

Consider first the diverted customers, eligible New Deal entrants 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 the unemployed themselves after those sessions, not much.²² If ERA were an official policy, there would be no need to severely restrict information on the actual extent of ERA support in order to prevent disappointment among the control group²³ (nor in fact would there be a need to perform randomisation²⁴). It is highly plausible that even under full information some refusers, especially among the ND25+ group, would still have been reluctant to prolong contact with Jobcentre Plus, all the more likely if they did not intend to be especially proactive in looking for work²⁵. 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 (Stage 3b), there still would have been some eligible individuals 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.

²² 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).

²³ 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'.

²⁴ Formally, formal refusers were signing that they did 'not consent in taking part in this research scheme or to being randomly assigned'.

²⁵ Although there would be nothing to lose to become formally eligible by registering with the programme (as one can then always decide to refuse to take up its services), the qualitative analysis has highlighted that especially among ND25+ entrants there is often a tendency to resist any involvement with Jobcentre Plus beyond what is minimally necessary.

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 the type at Stage 3b, 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 eligible individuals to apply for it. Furthermore, if ERA had been an official policy superimposed by default on the New Deals (i.e. of the type at Stage 3a), 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 one can assess whether the non-participants are individuals who even if offered ERA services would not take them up. Estimates of 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 are presented in Section 5.3.2.

The interest of the current report in the full eligible population 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 (the type at Stage 3b). A policymaker can only make the ERA support package available, but cannot force eligible individuals to apply for it or to take up its services. Hence, the causal effect for the eligible population 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 those eligible for ERA in the six districts, irrespective of: how well informed they are about ERA; whether they realise their eligibility or not; whether they apply or not; and 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 the full eligible group, rather than over an adviser-selected and self-selected subgroup of the eligible population.

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 of Stage 3a policy, that is, as a seamless next stage of the New Deal programmes in which **any** New Deal participant would automatically be enrolled. In other words, the unemployed 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.²⁶ 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.²⁷

The ERA experiment was carefully planned and designed: assignment to the control or to the programme group has taken place after the individual 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

²⁶ An eligible worker 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.

²⁷ It was a conscious design decision to make the ERA offer at the point of joining New Deal, rather than making the offer at the point at which individuals moved into work. Partly this was so that the adviser could offer advice about retention and advancement at the earliest possible stage, but mostly it was to offer a seamless service that would contain drop-out rates.

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. The analysis moves 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 for ERA. The problems investigated in this report are, thus, circumscribed to an issue of **external validity**, or the inference that can be validly drawn from the experimental set-up for the eligible population (in the six evaluation districts).

An alternative view to consider this issue defines the parameter of interest as the effect for those eligible for ERA (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 eligible group 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 necessarily somewhat 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 local labour market conditions. Ideally, it would also try to take into account entry effects (e.g. more lone parents volunteering for the 'NDLP with ERA' package, or some long-term unemployed delaying their job entry to become eligible for the 'ND25+ with ERA' package), as well as general equilibrium effects.

2.5 Summary

To summarise the discussion in this chapter (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:
 - the eligible population (the focus of the present report);
 - the ERA study participants (the focus of the experimental evaluation);
 - 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);
 - the eligible population 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 the eligible population.

Table 2.1 Some causal effects of interest

Mean impact of offering ERA for:		
• those eligible for ERA	Stage 1=I	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 eligible group than to ERA study participants
• those eligible for ERA if as an official policy it were an integral component of the New Deal programmes	Stage 1=I	Current report

3 Data and sample definition

3.1 Data

A number of data files have been put together for this analysis. The administrative data held by the Department for Work and Pensions (DWP) on New Deal 25 Plus (ND25+) and New Deal for Lone Parents (NDLP) entrants provide the sampling frame. Files were extracted 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. The New Deal extract files have further been exploited for information about past programme participation as well as a number of other relevant individual characteristics.

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

The administrative data has further been combined 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 entrants who were approached for recruitment into ERA, including the identifier of those who formally refused to participate. This data is used mainly to draw information on individuals’ decisions about participation in ERA, as well as the outcome of random assignment (control/programme group) for those who agreed to participate in the study.

Finally, local-area level data (travel-to-work and super-output area data) has been merged in.

Section 3.3 summarises the extensive variables that have been selected and derived from all of these sources.

3.2 Sample

To perform the analyses aiming at estimating the impact of ERA for the full ERA eligible population, one obviously needs to start by clarifying exactly what is meant by ‘ERA eligibility’. This is a conceptual issue which requires the analyst to decide on who should count as eligible. For such a definition to be operational, the criteria that determine ERA eligibility must also enable identification of the relevant individuals in the data.²⁸

For the purposes of this analysis, which relates to the current experimental evaluation, the following groups are, thus, considered as **eligible** for ERA²⁹:

²⁸ See Goodman and Sianesi (2007) for a description of how problem cases were handled and what adjustments were performed on the ERA experimental sample.

²⁹ The rationale underpinning the following definition of ERA eligibility is that Jobcentre Plus staff were instructed that those who became mandatory for ND25+ or expressed an interest in NDLP during the random assignment window would be eligible for ERA. Those among these groups 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.

- those who have become mandated for ND25+ 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
- those lone parents who were told about NDLP (had a WFI and/or expressed an interest in NDLP) during the period when the respective district was conducting random assignment **and** who subsequently also volunteered for NDLP still within the relevant random assignment intake window.

The random assignment window (or sample intake window) is actually district- and intake group-specific, since one district started conducting random assignment later than the others and some districts stopped conducting random assignment for some groups earlier. To identify the eligible population, the period when each district was conducting random assignment 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 following tables provide various sample breakdowns by participation status, separately for the two intake groups and by district. Section 2.3 has already provided a more detailed discussion of non-participation patterns; the following just summarises the main points.

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 amongst the ND25+ group (23 per cent of all eligible entrants) than amongst NDLP entrants (over 30 per cent).

For both intake groups, Scotland saw the lowest proportion of non-participation, with around nine per cent of ND25+ and five per cent of NDLP eligible entrants not participating in ERA; North West England is not far behind at around 15 per cent of all ND25+ eligible individuals and six per cent of all NDLP eligible individuals not participating. South East Wales and North East London saw closer to average levels of non-participation, whilst East Midlands and North East England saw the highest non-participation levels amongst both intake groups. In particular, in East Midlands almost half (47 per cent) of all eligible NDLP entrants 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).

Table 3.1 Sample breakdown by intake group

	ND25			NDLP		
	N	%	%	N	%	%
Eligible population	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 ND25+: Sample breakdown by district

	Scotland		North East England		North West England		Wales		East Midlands		London	
	N	%	N	%	N	%	N	%	N	%	N	%
Eligible population	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	703	65.1	1377	85.4	456	79.3	1245	72.5	1480	74.1

Table 3.3 NDLP: Sample breakdown by district

	Scotland		North East England		North West England		Wales		East Midlands		London	
	N	%	N	%	N	%	N	%	N	%	N	%
Eligible population	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	983	70.8	759	93.8	514	76.4	1131	52.9	1252	69.0

Table 3.4 Non-participation breakdown by district

	Overall		ND25+		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

This report assesses the impacts of ERA on employment and earnings outcomes during a 48-month follow-up period. The analyst can crucially exploit administrative records for the full sample of those eligible for ERA in the six evaluation districts, i.e. for the participants and, most importantly for this analysis' purposes, for the non-participants too.

For the employment measures the 48-month follow-up period is counted from the moment individuals flowed in (i.e. from the moment ND25+ entrants started the Gateway, or lone parents volunteered for NDLP), and consider both the probability of having ever been in employment and the total number of days in employment during that period.

The earnings outcomes considered are taken from P14 tax records. Focus is on cumulative earnings between the 2005/06 and 2008/09 tax years, as well as on the most recent earnings, those for the 2008/09 tax year. The main analysis also looks at tax year earnings for previous post-random assignment years. Depending on when during the tax year a person flowed into the sample, these earnings measures will pertain to X to $(X+1)$ years post inflow as follows:

P14 earnings for the:	Correspond to:
2005/06 tax year	1-2 years post-random assignment
2006/07 tax year	2-3 years post-random assignment
2007/08 tax year	3-4 years post-random assignment
2008/09 tax year	4-5 years post-random assignment
2005/06 to 2008/09 tax years	1-4 or 2-5 years post-random assignment

An extensive collection of variables has been put together 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 the non-participants, which entails that such information has to be derived from administrative data sources alone.

Table 3.5 groups and summarises the various observable factors used in the analysis; the table also briefly comments on the variables and lists the omitted category for discrete or categorical variables. Section 4.3.1 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 ND25+ Gateway or volunteered for NDLP
Demographics	
Female (compared to male)	
Age at inflow and age squared and missing age	
Ethnic Minority (compared to white)	
Disability indicator Missing disability status (compared to non-disabled)	Disability indicator: if has a disability at inflow and/or if claiming IB at inflow
Has partner Missing partner information (compared to having no partner)	For ND25+
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 ND25+ and starting the Gateway (for ND25+ group), or between being told about NDLP and volunteering for it (for NDLP group)
Early entrant into ND25+ programme	For ND25+ Spent <540 days on JSA before entering ND25+

Continued

Table 3.5 Continued

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, Work Based Learning for Adults (WBLA) or Outreach
Past participation in ND25+ programme	For ND25+
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, ND25+, 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 IS and IB
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 Super Output Area (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. It then briefly outlines, 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, the discussion is kept as informal as its rather technical nature allows. For those who would still rather skip this chapter, the following is a very straightforward summary of the three approaches that have been adopted, which should be sufficient to allow one to follow the results in Chapter 5:

a A simple bounds analysis

Given that the outcomes the non-participants would have experienced had they participated in Employment Retention and Advancement (ERA) and been assigned to the treatment group are unobserved, one can look at what would have happened had they participated and all been in employment (the upper bound of the overall impact) and if they had participated and none of them got a job (the lower bound).

b Looking at observable differences between participants and non-participants

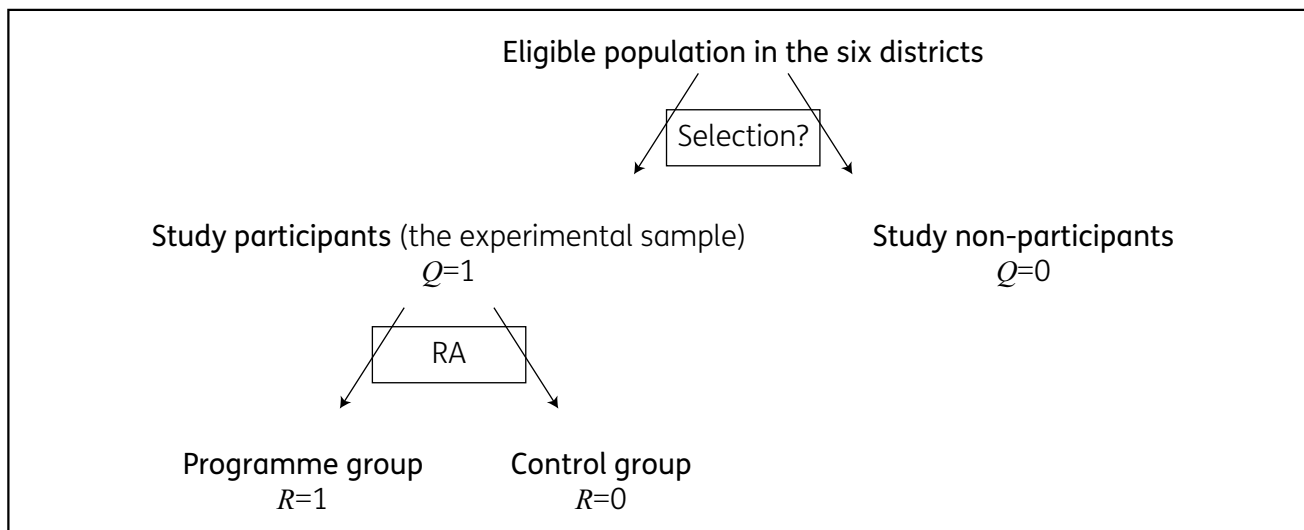
If one assumes that the only outcome-relevant differences between participants and non-participants are those which are captured in the many characteristics observed in the data, such as age, gender, duration on benefit, labour market history, etc., then by looking at the ERA outcomes for programme group members observationally similar to the non-participants, one can estimate what the outcomes of the latter would have been had they participated. (One can similarly estimate which aspects of the ERA service the non-participants would have been expected to take up.)

c Recognising that non-participants may differ from participants in important ways that the analyst cannot directly observe, one can test whether this is in fact the case, and experiment with a control function approach to allow for this residual bias.

4.1 Analysis framework

The section starts by setting up the framework and introducing some basic notation. Figure 4.1 highlights the structure of the problem that needs to be address, while Box 4.1 summarises the notation.

The population of interest are those **eligible** to be offered 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 which was not ($R=0$).

Figure 4.1 Simplified structure of the problem**Box 4.1 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 the eligible population
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 those eligible for ERA (parameter of interest)
ATE_1	average ERA effect on ERA study participants (experimental estimate)
ATE_0	average ERA effect on non-participants

The problem here arises due to changes in the 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 eligible population would have experienced.

Let p be the probability of non-participation among the eligible group. This is directly identified in the data by the proportion of non-participants among the eligible population (see Tables 3.1-3.3).

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).

The parameter of interest 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 all those eligible for ERA if they were offered ERA services compared to the average outcome for all those eligible for ERA if they were not offered ERA services:

$$ATE_1 \equiv E(Y_1 - Y_0)$$

What one can however directly identify from the available experimental data is the average effect of ERA for participants in the experiment, $ATE_1 \equiv E(Y_1 - Y_0 | Q=1)$. 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 eligible individuals (i.e. on the non-participants) by

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

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

$$ATE = (1-p) \cdot ATE_1 + p \cdot ATE_0 \tag{1}$$

Equation (1) simply states that the parameter of interest, i.e. the average impact of ERA on all eligible individuals in the six districts, is given by a weighted average of the parameter one 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)).

Previous work (Goodman and Sianesi, 2007) has 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 the eligible population cannot thus be directly identified, unless additional assumptions are made.

Note that in this set-up one can exploit administrative employment and earnings data on all those eligible for ERA. This means that one does observe an outcome for the non-participants: the no-treatment outcome. The average outcome that the non-participants would have experienced had they not been offered ERA thus coincides with their average observed outcome:

$E(Y_0 | Q=0) = E(Y | Q=0)$. Hence to answer the question of how different the average impact for the eligible population would have been compared to the experimental impact for the participants, the only unobserved term is the outcome that the non-participants would have experienced, on average, had they been offered ERA services, $E(Y_1 | Q=0)$. Equation (1) thus becomes:

$$ATE = (1-p) \cdot ATE_1 + p \cdot (E(Y_1 | Q=0) - E(Y | Q=0)) \quad (1a)$$

This report assesses how different the average impact on participants is from the average impact on the eligible population based on the following three identification strategies for $E(Y_1 | Q=0)$ and hence for ATE :

- bounds for the ATE that can be obtained without having to make any assumption on the selection process;
- identification of the ATE under the assumption of selection on observable characteristics;
- identification of the ATE allowing for selection on unobserved characteristics and impacts.

Before turning to a description of these strategies, let us consider under what conditions the average impact for those taken through random assignment would be the same as the average impact for the full eligible population even in the presence of a non-negligible share of non-participants. For this to be the case, it is necessary that the decision (either by the adviser or by the individual) to participate in ERA is not correlated with the individual gain from receiving ERA services. (Note that a trivial example of this is where the programme effect is exactly the same for everybody in the population, regardless of their characteristics, which would be an exceedingly implausible assumption.)

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 they live 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 eligible population.

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).

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 ERA is being evaluated in terms of employment probability, so that the outcome Y is bounded between 0 and 1.

From equation (1a), the lower bound for the effect on the eligible population is obtained by assuming that none of the non-participants would have gone into employment from the programme (even if some of them are observed to have entered work without going through ERA), the upper one by assuming that all non-participants would have been in work had they received ERA:

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

The width of the bound is given by p , the proportion of non-participants in the eligible group.

4.2.1 Sensitivity analysis

One can further explore how sensitive the estimate of the effect on the eligible population 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), the effect for the eligible group can then be calculated as a function of θ :

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

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

To clarify with a numerical example, suppose that 30 per cent of the non-participants and 30 per cent of the ERA control group were employed, and that 40 per cent of all those assigned to the ERA programme group were employed. The impact on participants' employment rate is therefore 10 percentage points. Setting $\theta=0.5$ would amount to assuming that the non-participants would have had an employment rate of $0.5 \cdot 40$ per cent, i.e. 20 per cent, had they participated. With a 25 per cent non-participation rate, the average treatment effect for all eligibles would thus be $0.75 \cdot 10 + 0.25 \cdot (20 - 30) = 5$. If on the other hand $\theta=1.5$, i.e. it is assumed that non-participants would have had an employment rate of 60 per cent had they participated, then the overall treatment effect on the eligibles would be $0.75 \cdot 10 + 0.25 \cdot (60 - 30) = 15$.

4.3 Impact estimates under selection on observables

4.3.1 Selection on observable characteristics

This and the next section describe two sets of methods aimed at arriving at a point estimate of the effect for the eligible population. 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 eligible group 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 the eligible population) 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 available data; in the following, its content in relation to the estimation problem to be addressed is briefly addressed.

All the outcomes of interest – employment probabilities and durations as well as earnings – are related to labour market performance. As listed in Table 3.5, an extensive collection of individual, office and local area characteristics has been assembled that are most likely to affect individuals' labour market performance, and that might potentially have affected participation in 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), the data include summary information on an individual's 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 of whether New Deal 25 Plus (ND25+) entrants volunteered for the Gateway ahead of time.

Three years' worth of labour market history has also been constructed, with variables summarising the proportion of time employed and the proportion spent on benefits, separately on active benefits (JSA and compensation whilst on a labour market programme) and inactive benefits (Income Support (IS) and Incapacity Benefit (IB)). The data also include variables capturing the extent of past participation in voluntary employment programmes (as a crude indicator of willingness to improve one's circumstances), in ND25+ (a mandatory programme) and in Basic Skills (a programme designed to address basic literacy, numeracy and IT skills).

Furthermore, information has been collected on local labour market conditions (specifically, travel-to-work area unemployment rates), as well as on the deprivation of the area the individual lives in (index of local deprivation). Additionally, information at the office level (total New Deal caseload and share of lone parents in such a caseload) has been constructed to capture 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 the available data, together with 'innate ability', discipline or work commitment. The previous literature has, however, indicated the potential for detailed labour market histories (like those in the data exploited in this report) to help proxy 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).³⁰

4.3.2 Estimating impacts for the non-participants

Previous work (Goodman and Sianesi, 2007) reviewed in Section 2.2.2 has shown the extent to which outcome-relevant observed characteristics X of the participants and non-participants differ.³¹ One 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 the eligible population.

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

³⁰ For their main analysis of the NDLP programme, Dolton *et al.* (2008) rely on the same administrative data used in this report. 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.

³¹ Note that one can test whether the two groups significantly differ in terms of observables; one can only speculate about whether such observables are likely to affect impacts.

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:

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

Assumption (A1) thus requires that for the eligible population, 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), one also needs 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.

As for implementation, each non-participant is matched to one or more similar programme group member(s) based on the propensity score (the probability that an eligible 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, one can explore how sensitive the estimate of the impact for the eligible population is to straightforward violations of assumption (A1) by relaxing it to:

$$(A1') \quad E(Y_1 | Q=0, X) = \theta E(Y_1 | Q=1, X)$$

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 . As the outcomes of interest (employment probability, days employed and earnings) are favourable, $\theta > 1$ implies negative selection into the sample participating in ERA (i.e. the non-participants would have enjoyed higher employment/earnings outcomes under ERA than actual ERA participants with the same observed characteristics, so that those who select or are selected into ERA have below average outcome expectation), while $\theta < 1$ positive selection.

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, this type of analysis aims 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?

One 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 individual 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, one needs 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, the selection-on-observables assumption (A1) is again invoked, which requires that, once conditioning on the 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, this assumption rules 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, one 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 the 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

The 'control function models' are a class of models which allow selection into the group of ERA study participants to depend on outcome-relevant **unobservables**.

These models, which build upon the classical sample selection model introduced by Heckman (1979), are framed within a formal set-up and require technical conditions for identification and at times quite complex estimation methods. In the following only a few features of this approach are thus highlighted, leaving it to Appendix A to provide a more in-depth description.

As always, the key issue is that treatment outcomes are only observed for the participants (as represented by the programme group), but are unobserved for the non-participants.

The control function model tries to tackle this issue whilst allowing the unobserved determinant of participation to be potentially correlated with unobserved individual characteristics as well as with idiosyncratic ERA impacts. The model thus allows for selection into the ERA study based on both unobserved characteristics and unobserved individual-specific ERA impacts.

The model crucially relies on an 'exclusion restriction' – an observable variable which affects the decision to participate in the study, but does not affect potential ERA outcomes directly.

The basic model also imposes that the unobservables are jointly normal and homoskedastic (where the latter basically amounts to imposing that the selection process into the ERA study is the same for eligible individuals with different observed characteristics). Such parametric assumptions can, however, be relaxed and thus tested. For this analysis, both assumptions have in fact been relaxed; in addition, the outcome has been allowed to be censored (at zero in the case of employment duration or earnings) in both the treatment and no-treatment state.

A convenient feature of these types of models is that they allow one to separately test for selection into the ERA study based on unobserved characteristics and based on unobserved individual-specific ERA impacts, evidence which can be of interest in its own right.

Due to the unique set-up of randomisation coupled with administrative outcome measures that cover the selected-out (i.e. non-participant) sample as well, one is, however, in the rather exceptional position of being able to perform four additional types of tests.

The first 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, which thanks to administrative data is observed for both groups.

1 Testing part of the exclusion restriction of the instrument

The strength (“power”) of the instrument in affecting the choice among eligible individuals of participating in the ERA study is, as usual, a testable condition. In the case of administrative outcomes, however, one can assess whether the exclusion restriction holds in terms of non-ERA outcomes. 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. If this shows a significant association between the instrument and the outcomes, then it is not a valid instrument – it does not meet the exclusion restriction.

2 Testing for the presence of residual selection on unobservables related to non-ERA employment or earnings outcomes

One 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.

The test can be performed by running a regression, on the pooled sample of controls and non-participants, of observed outcomes on the observables X , plus a dummy variable for whether or not the individual participated. If the coefficient on the dummy variable is significant, this indicates that there is selection on unobservables.

The results of this test are not just informative in themselves, but they lend themselves to construct an important specification check for any of the control function models.

Exploiting administrative outcomes, one can construct two specification tests to assess – and order – the performance of the different control function models:

3 Testing how well the various control function models capture the presence and direction of the selection on unobservables uncovered by test (2)

The analyst is in the unusual position of being able to choose between different specifications of the control function based on how closely a given model matches the difference in adjusted observed outcomes between the control group and the non-participants (reflecting the results from the test of selection on unobservables).

4 Testing how well the various control function models predict the (observed) non-ERA outcome for the non-participants

One can further choose between different specifications of the control function based on how closely a given model matches the average predicted and observed (no-treatment) outcomes of the non-participants.

5 Implications of non-participation for the four-year experimental impact estimates

This chapter presents the empirical results. The analyses have always been performed separately for the two intake groups, New Deal for Lone Parents (NDLP) and New Deal 25 Plus (ND25+). For all estimation methods except the control function models, they have been performed both overall and by district. In the following, focus is on the overall findings, with district-level ones being mentioned only if worthy of special note. The reader is referred to Appendix B for the tables with all the district-level results and corresponding summary boxes.

The chapter starts with the benchmark experimental findings that omit the non-participants.

5.1 Experimental findings

This section presents the experimental findings concerning the average impact of Employment Retention and Advancement (ERA) for the participants on a series of outcomes measured over a four-year follow-up period. 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').³² Controlling for such characteristics 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, as the 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. Specifically, although the programme and control groups can be expected to be more or less identical on average, in practice they are very unlikely to be precisely identical – in the same way as with 100 tosses of a coin, while one might expect to get around 50 heads, one is in fact very unlikely to get precisely 50 (the probability of this event being only about 0.08). Regression adjustment allows one to control for these chance differences in observables.

A positive effect of ERA on employment outcomes has been uncovered for the ND25+ group, both in terms of a 2.8 percentage point increase in the probability of being ever employed during the four years after inflow and in terms of a 25-day increase in the time spent in employment over that period. These employment effects appear to be driven by impacts in London (+4.1 percentage points) and Scotland (+74 days).

No employment impact could, by contrast, be detected for NDLP participants. Although the probability of being employed in the 4 follow-up years and days in employment remained unaffected overall, interesting impacts have been uncovered at the district level. Specifically, participants in North West England enjoyed substantial increases in employment chances (+6.8 percentage points) and duration (+70 days), while participants in Wales were severely negatively affected by a reduction in employment probability of 5.6 percentage points and in employment durations of 97 days.

³² Background characteristics have all been derived from the administrative data (see Section 3.3).

The same overall story was found to apply when assessing the impact of offering ERA on participants’ earnings: a significantly positive impact for ND25+ participants and no impact for NDLP participants.

In particular, the employment boost received from ERA translated in a substantial rise in earnings for the ND25+ programme group, who enjoyed a £1,805 increase cumulatively and an increase of between £400 and £560 in each tax year. These earnings impacts were found to be driven first and foremost by the district which also enjoyed the most substantial employment boost: Scotland, where participants enjoyed a £1,508 increase in earnings in the latest tax year and a remarkable £3,817 increase cumulatively. The East Midlands contributed to the overall effect on cumulative earnings with a £2,662 rise, although in this district the earnings impact in the latest year was insignificant.

The substantial and highly statistically significant ERA impact on earnings for the ND25+ group contrasts sharply with the absence of any significant impact for the NDLP group, other than the earnings impact in the very early year (2005/06). Indeed, the ERA intervention has significantly reduced, by almost £1,600, the earnings in the latest tax year (2008/09) for participants in Scotland.

Table 5.1 Experimental findings over the four-year follow-up period

	Raw		Adjusted	
	Effect	Std.Err.	Effect	Std.Err.
ND25+				
Ever employed	0.027**	0.013	0.028**	0.012
Days employed	24.6**	11.2	25.0**	10.6
Earnings 2005/06 (£)	4,450**	190	438**	188
Earnings 2006/07 (£)	456**	194	415**	191
Earnings 2007/08 (£)	578***	213	560***	210
Earnings 2008/09 (£)	410**	199	392**	195
Earnings 2005–09 (£)	1,894***	702	1,805***	687
NDLP				
Ever employed	-0.002	0.012	-0.008	0.011
Days employed	2.0	14.7	-4.0	13.9
Earnings 2005/06 (£)	420**	176	376**	173
Earnings 2006/07 (£)	260	195	218	192
Earnings 2007/08 (£)	226	214	185	211
Earnings 2008/09 (£)	21	230	-13	227
Earnings 2005–09 (£)	927	737	767	722

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

Sample sizes: 6,006 for ND25+ and 5,052 for NDLP.

5.2 Bounds

The bounds analysis aims at bounding the average ERA impact on the eligible population without making any assumption on the participation process in the ERA study. This analysis confirms some of the four-year experimental 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. Specifically, the impact of ERA on the

probability of employment of the eligible population is positive in Scotland for ND25+ and in North West England for NDLP, while most likely negative in Wales 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 intake groups overall, for whom the impact on the employment probability of those eligible for ERA is bounded between -8 and +15 percentage points (ND25+) and -23 and +8 percentage points (NDLP). Table 5.2 shows how these bounds are derived.

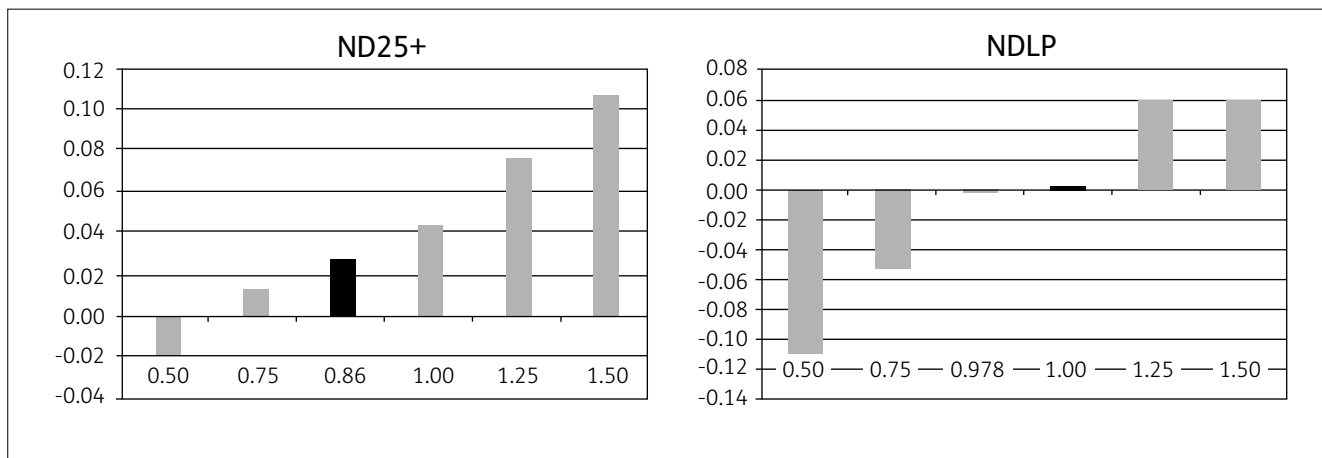
Table 5.2 Derivation of bounds

		ND25+	NDLP
(a)	Participation rate	0.77	0.70
(b)	Non-participants' employment rate	45	74
(c)	Control group's employment rate	52	76
(d)	Programme group's employment rate	55	76
(e)	Experimental impact ATE_1	(d-c)	3
			0
Bounds for impact on eligibles ATE			
	Lower bound	$(a \cdot e) + ((1-a) \cdot (0-b))$	-8
	Upper bound	$(a \cdot e) + ((1-a) \cdot (100-b))$	15
			8

The sensitivity analysis proved to be often 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 ND25+ group overall, the average effect remains positive and small under the most scenarios, 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).

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 intake 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 ND25+. Specifically, for the NDLP group, the experimental estimate would recover the average effect under the assumption that the non-participants did not select into the ERA study based on ERA outcomes. For the ND25+ group, by contrast, to take the experimental impact as representative of the impact on the eligible population, one would need to assume that non-participants among the group of eligible ND25+ would have experienced much **lower** employment probabilities had they been offered ERA services than what actual participants receiving ERA are observed to experience. Overall, for the experimental estimate to be an unbiased estimate for the ATE , the non-participants should have experienced a 14 per cent lower employment probability under ERA than does the actual ERA programme group.

Figure 5.1 Sensitivity analysis: ATE_{θ} for ever employed during the four-year follow-up, θ from 0.5 to 1.5



In black: experimental impact estimate and corresponding θ .

As a reminder, a value of 0.5 for θ means that non-participants would be half as likely to achieve an outcome if they participated and were in the programme group, as those who actually were in the programme group

5.3 Selection on observables

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

Table 5.3 presents the matching results for ND25+ and NDLP overall, while Appendix B 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 C).

Starting with the results for the ND25+ group, once differences in observed characteristics between participants and non-participants are corrected, the analysis finds that the non-participants would have experienced a considerably larger ERA impact on employment outcomes than participants, the difference between the two impacts being statistically significant at all levels. In particular, had they been offered ERA services, the group of non-participants would have spent 63 days longer in employment (significant at the 1% level) during the four follow-up years than if they had not been offered ERA. By contrast, participants are found to spend 25 days more in employment (significant at the five per cent level) thanks to ERA. The ERA impact on the eligible population is estimated as a highly statistically significant 34 days. The same story applies when considering the probability of being ever employed during the four years post-inflow: compared to an ERA-led increase for participants of 2.8 percentage points (significant at the five per cent level), non-participants would have enjoyed a 6.4 and the full eligible population a 3.7 percentage points increase (both impacts significant at the one per cent level and statistically different from the impact for participants).

Table 5.3 Matching four-year impact estimates

	ATE_1	ATE_0	ATE	$ATE_1 \neq ATE$	$ATE_1 \neq ATE_0$
ND25+					
Days employed	25.0**	62.5***	33.6***		***
Ever employed	0.028**	0.064***	0.037***	**	**
Earnings 2005/06 (£)	438**	463*	444**		
Earnings 2006/07 (£)	415**	819***	508***		
Earnings 2007/08 (£)	560***	627*	575***		
Earnings 2008/09 (£)	392**	482*	413**		
Earnings 2005–09 (£)	1,805**	2,392**	1,940***		
NDLP					
Days employed	-4.0	17.9	2.7		
Ever employed	-0.008	0.020	0.000	*	*
Earnings 2005/06 (£)	376**	550***	429***		
Earnings 2006/07 (£)	218	627***	342**	*	*
Earnings 2007/08 (£)	185	525**	289*		
Earnings 2008/09 (£)	-13	691***	202	***	***
Earnings 2005–09 (£)	767	2,393***	1,262**	*	*

Notes:

ATE_1 is the (experimental) impact for the participants; ATE_0 is the impact for the non-participants; ATE is the impact for all eligibles.

Statistical significance based on bootstrapped bias-corrected confidence intervals (1,000 replications); $ATE_1 \neq ATE$ and $ATE_1 \neq ATE_0$: bootstrap-based statistical significance of the difference.

*** significant at 1%, ** at 5%, * at 10%.

Sample sizes: 7,796 for ND25+ and 7,261 for NDLP.

The finding that ERA employment impacts for the ND25+ non-participants and for the ND25+ eligible population would have been **consistently better** than those experimentally estimated for the subgroup of participants point to the possibility that the ND25+ non-participants might in fact be easier to help back into the labour market than the average ND25+ entrant.

As to ERA's impacts on earnings, the experimental estimate of the impact for ND25+ study participants was found to be representative of the average impact for all those eligible for ERA, both in terms of the four tax-year earnings post-inflow and in terms of cumulative tax earnings. Specifically, the point estimates for the earnings impacts for non-participants as well as for the eligible population are always somewhat larger than the corresponding point estimates for participants, but such differences never reach any statistical significance.

For the NDLP group, 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 and for the full eligible group as it was for the experimental group.

Two districts stand out from this general pattern in terms of employment outcomes: North West England and Wales. In North West England ERA was found to enhance participants' employment (by a weakly significant 70 days increase); non-participants, however, are estimated to be unaffected by ERA (with a negative point estimate), resulting in an ATE for the eligible population slightly smaller than the experimental estimate (+58 days) but no longer significant (non-participation in this district is only 6%). In Wales by contrast, the situation is reversed: in this district, participants experience

significantly fewer days (-97) in employment due to ERA, while non-participants (who represent almost half of the eligible population) would not have been significantly affected by the programme (indeed the point estimate is positive) and the eligible population would thus have experienced no significant adverse ERA impact.

As to ERA’s impacts on earnings for the NDLP group, the evidence in terms of point estimates, their statistical significance as well as formal tests of the difference, tells a consistent story: the earnings impacts estimated on the experimental group systematically **underestimate** the average impacts that the programme would have had on the non-participants and on the full eligible population. Specifically, while ERA failed to significantly increase participants’ cumulative earnings over the four post-inflow years (an insignificant £767 point estimate), non-participants would have enjoyed a highly significant £2,393 rise in cumulative earnings and the eligible population a significant £1,262 rise. For the first tax year (2005/06), the point estimates of the impacts for the non-participants and for the eligible population are strongly significant and larger than the £376 increase (significant only at the five per cent level) estimated for the participants. For the following three tax years, while the experimental impact for the participants has dropped in size and become indistinguishable from zero, non-participants would have enjoyed a highly significant £500-600 increase in tax-year earnings and all those eligible for ERA a mostly significant £200-340 rise.

In conclusion:

- For the ND25+ intake group, the experimental impact estimate of ERA underestimates the contribution that the programme would have given to all those eligible for ERA in terms of improving their employment chances, while it is representative of the expected impact that ERA would have had on their earnings. Overall though, the broad story remains unchanged: significant impacts on both employment and earnings for the experimental sample as well as for the full eligible population.
- By contrast, for the NDLP group, the broad story does change. In the experimental analysis, there is no impact on employment, and only a modest and temporary impact on earnings. After allowing for selection on observables, there is still no employment impact for all eligibles, but the earnings impact is strengthened, and there is now a significant impact on overall earnings for the full eligible population.

5.3.1 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:

$$(A1') \quad E(Y_1 | Q=0, X) = \theta E(Y_1 | Q=1, X)$$

Since outcomes of interest are favourable outcomes such as employment probability, days employed and earnings, $\theta > 1$ implies negative selection into the sample participating in ERA³³, $\theta < 1$ positive selection; while for $\theta = 1$, one obviously obtains the matching estimates discussed above.

In line with the bounds analysis in Section 5.2, the sensitivity analysis in Table 5.4 is quite informative for the ND25+ group and clearly paints a rather favourable picture for the impact that ERA would have had on the eligible population – both in terms of employment and earnings outcomes.

³³ If $\theta > 1$, this means that the non-participants would have enjoyed higher employment/earnings outcomes under ERA than actual ERA participants with the same observed characteristics, so that those who select or are selected into ERA have below average outcome expectation.

In particular, the employment effect of ERA for the eligible group in terms of both the chance of ever being employed and the number of days employed would have been positive under all scenarios considered bar the most extreme one of $\theta=0.5$ (corresponding to a situation in which, had they received ERA, the non-participants would have experienced half of the employment probability and duration than those programme group members with the same observed characteristics). Indeed, the impact on earnings outcomes for the eligible population would have been positive under all selection scenarios considered.

In contrast to what has been found for the ND25+ group, 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 the eligible population would range from substantial and negative to substantial and positive in terms of both employment and earnings outcomes.

Table 5.4 also displays the value of θ for which the experimental estimate coincides with the average impact for the whole eligible population.

An interesting finding is that for both groups and all outcome measures considered there seems to be consistency in the value of θ required for the experimental impact to be an unbiased estimate of the average effect for the full eligible population. In particular, such a value is always smaller than 1. This means that in order to take the experimental impact as representative of the impact on the eligible population, one would need to assume that had they been offered ERA services and incentives, the non-participants would have experienced lower employment and earnings outcomes than actual programme group members with the same observed characteristics.

More specifically, in order to ignore non-participation one needs to assume only a slightly favourable selection into the ERA study in terms of the two earnings measures for ND25+ and the two employment measures for NDLP (the required scenario being that had they received ERA, non-participants would have experienced 96-97 per cent of the earnings or employment probability and duration of observationally equivalent participants). By contrast, one would need to assume a stronger positive selection into the ERA study in order to ignore non-participation in terms of employment duration for ND25+ and earnings in the latest tax year for NDLP (non-participants being required to experience under ERA only 86-87 per cent of the employment duration and earnings enjoyed by those programme group members with their same observed characteristics).

Table 5.4 Sensitivity analysis: ATE_{θ} , θ from 0.5 to 1.5

Ever employed		Days employed		Earnings 2008/09		Earnings 2005/06–2008/09	
θ	ATE_{θ}	θ	ATE_{θ}	θ	ATE_{θ}	θ	ATE_{θ}
ND25+							
0.50	-0.022	0.50	-0.5	0.50	21	0.50	411
0.75	0.007	0.75	16.6	0.75	217	0.75	1,176
0.93	0.028	0.87	25.0	0.97	392	0.96	1,805
1.00	0.037	1.00	33.6	1.00	413	1.00	1,940
1.25	0.066	1.25	50.6	1.25	609	1.25	2,704
1.50	0.095	1.50	67.7	1.50	805	1.50	3,468
NDLP							
0.50	-0.116	0.50	-88.2	0.50	-564	0.50	-1,500
0.75	-0.058	0.75	-42.8	0.75	-181	0.75	-113
0.96	-0.008	0.96	-4.0	0.86	-13	0.91	767
1.00	0.000	1.00	2.7	1.00	202	1.00	1,262
1.25	0.058	1.25	48.1	1.25	584	1.25	2,636
1.50	(1)	1.50	94	1.50	967	1.50	4,011

In bold: experimental impact estimate and corresponding θ .

(1) With 76% of the NDLP programme group having been employed at some time during the subsequent 4 years, a value of θ of 1.5 would imply an employment rate for non-participants above 100%, so this entry is left blank.

Sample sizes: 7,796 for ND25+ and 7,261 for NDLP.

5.3.2 Analysis of take-up

Although as argued in Section 2.4, 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 5.5 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 within the first year since inflow:

- measures of presence, type and intensity of contact with Jobcentre Plus staff (any contact, the individual 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 individual was not working (staff offered any help/advice, performed a Better Off Calculation, suggested individual attends a Jobclub/Programme Centre, arranged an education or training course, offered advice without being requested);
- measures of the individual’s assessment of the advice received; and
- for the programme group analysis only, measures directly linked to knowledge of ERA features (whether the individual has heard of the employment and 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 considered.

First, the analysis estimates the take-up that the non-participants would have exhibited in their first year 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.³⁴

For the ND25+ 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 in their first year on ERA 85 per cent of the programme group has received help or advice from Jobcentre Plus staff while not working, the 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 ND25+ non-participants would have been quite heavily involved with ERA and Jobcentre Plus in their first year. And although 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.

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, in their first year the non-participants would have been over 3 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 of receiving help or advice from Jobcentre Plus when not working. As was the case for ND25+, it is again the programme group which 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 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 ND25+ entrants 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 ND25+ non-participant

³⁴ Again, note that if some eligible individuals 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 eligible population.

group is made up of individuals who would shun involvement with Jobcentre Plus at all costs. This supposition is not borne out in the data: had they been assigned to the control group, the involvement that the ND25+ non-participants would have had with Jobcentre Plus in their first year would not have been statistically different from the one displayed by the actual control group in any of the dimensions considered.

As opposed to ND25+, NDLP entrants 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 amongst 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 5.5, 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 5.5 Take-up and involvement with Jobcentre Plus in the first year predicted for the non-participants both under ERA and without ERA

	ND25+				NDLP			
	ERA outcome Programme group	Non-participants	Non-ERA outcome Control group	Non-participants	ERA outcome Programme group	Non-participants	Non-ERA outcome 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 5.4.2, the findings from the different control function models that have been estimated, Section 5.4.1 reports the results from testing for selection in terms of no-treatment unobservables and presenting the chosen instrument together with evidence on its power and validity.

5.4.1 Testing for selection on specific unobservables

Section 4.4 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 earnings outcomes. Specifically, this involves assessing whether, once controlling for observable characteristics, the non-ERA outcomes of the participants (as represented by the control group) differ, on average, from those of the non-participants.

Table 5.6 reports the results of this test, which for robustness was carried out via simple Ordinary Least Squares (OLS), fully interacted regression (FILM), matching and Tobit or Probit. Simple OLS regression can suffer from misspecification bias if observed characteristics affect the outcome in a non-linear fashion and/or affect the way in which the control group differs from the non-participants in terms of outcomes. A fully interacted linear model (FILM) relaxes the latter restriction, while non-parametric matching relaxes both. Furthermore, OLS regression (simple or interacted) does not take into account the potentially censored or binary nature of the outcome variable. Tobit and Probit models take account of, respectively, censoring and a binary outcome variable, but do so based on a distributional assumption.

Interestingly, non-participants are found to be subject to the same type of selection on unobservables in both intake groups.

In terms of four-year employment outcomes and with weak evidence in terms of earnings, non-participants have unobservables leading them to experience worse (non-ERA) outcomes than observationally similar participants. Specifically, compared to participants with the same observed characteristics, ND25+ and NDLP non-participants:

- are three to four percentage points less likely to have been ever employed and have spent around 40 fewer days in employment in the four years post-inflow; and
- earned roughly £400 less in 2008/09 and around £1,000 less between 2005 and 2009 (according to Tobit only and then significant at the ten per cent level).

While the selection story that emerged is the same for both intake groups, this overall result was found to be driven by different districts within the two groups.

For the ND25+ intake group, adverse selection in terms of employment was found in East Midlands, North East England, Wales and London; adverse selection in terms of earnings was found in North East England only; and no selection on unobservables was found to have taken place for ND25+ non-participants in North West England and Scotland.

For the NDLP group, non-participants in Wales, Scotland and possibly East Midlands have unobservables that caused them to experience worse employment outcomes than the participants; non-participants in Scotland and Wales experienced lower earnings; non-participants in London and North East England have not been subject to any residual selection on unobservables; and non-participants in North West England have indeed been subject to favourable selection, experiencing better employment and possibly better earnings outcomes than observationally equivalent participants.

Table 5.6 Differences in outcomes for participants (control group) compared to non-participants with the same observed characteristics during the four-year follow-up

Outcome	Method	α	Given observables, participants:
ND25			
Days employed	OLS	50***	spend more days employed
	FILM	34**	
	Matching	37***	
	Tobit	45***	
Ever employed	OLS	0.042***	are more likely to be employed
	FILM	0.026	
	Matching	0.033*	
	Probit	0.050***	
Earnings 2008/09	OLS	330	do not earn more in 2008/09 (except according to Tobit)
	FILM	72	
	Matching	103	
	Tobit	385**	
Earnings 2005–09	OLS	972	do not earn more between 2005/06 and 2008/09 (except weak Tobit evidence)
	FILM	138	
	Matching	163	
	Tobit	1,210*	
NDLP			
Days employed	OLS	35**	spend more days employed
	FILM	26	
	Matching	24	
	Tobit	40**	
Ever employed	OLS	0.035***	are more likely to be employed
	FILM	0.027*	
	Matching	0.031*	
	Probit	0.035***	
Earnings 2008/09	OLS	379	do not earn more in 2008/09 (except weak Tobit evidence)
	FILM	416	
	Matching	326	
	Tobit	428*	
Earnings 2005–09	OLS	1,196	do not earn more between 2005/06 and 2008/09 (except weak Tobit evidence)
	FILM	1,201	
	Matching	741	
	Tobit	982*	

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+ and 4,702 for NDLP.

In conclusion, though, when selection on unobservables has been uncovered, the picture that emerges during the four-year follow-up period is one of worse employment outcomes and possibly lower earnings for the non-participants. Non-participants, thus, seem to be less attached to the labour market – experiencing shorter employment durations, smaller incidence of employment and (hence?) somewhat lower earnings. The unobservables characterising the non-participants might thus relate to those more on the fringe of the labour market.

Before concluding this section it is important to note that even if the data has failed to pass this test in terms especially of employment outcomes, this would not *per se* invalidate the matching estimates of the impacts for the non-participants and for the eligible population which rely on the selection-on-observables assumption (Section 5.3). Even if after controlling for observable differences, control group participants have significantly different outcomes from the non-participants, this only relates to the no-treatment outcome. With administrative outcomes being available for all those eligible for ERA, the matching methods do not need to predict the no-treatment outcome, as these are observed. Matching methods by contrast need to predict the average treatment outcome for the non-participants, and do so using observably similar programme group members.

5.4.2 Control function results as a sensitivity check

Before discussing the results from the extensive search for an appropriate control function model to take account of the residual selection uncovered in Section 5.4.1, it is helpful to remind readers of the key terminology:

an **instrument** is an observable factor which is found to affect the probability of participating in ERA, but does not directly affect ERA outcomes;

the **exclusion restriction** is the second part of the above definition – that there is no effect on the outcomes in question;

the **power** of the instrument relates to the first part of the definition – the extent to which the instrument affects ERA participation; the greater this is, the more precise the results.

Defending the choice of instrument and presenting evidence on its power and validity is crucial, as without a sensible and strong instrument, no control function model could be robustly identified.

Motivated by the idea that for both intake groups the observed fall over time in 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), a promising instrument would be the elapsed number of days since random assignment started in an individual's district and for that individual's intake group. This measure is thus relative to random assignment start in each district and for each New Deal intake group, and is conditional on controlling for calendar time using three dummies for five-month periods.³⁵

This instrument-based on the increased persuasiveness of the advisers and the greater promotion of the ERA bonuses does indeed look like a very promising one, both in terms of its relevance and validity (see Appendix C). Specifically, for both intake groups it displays a very powerful first stage

³⁵ Originally the authors 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, this instrument does not pass the exclusion restriction test.

(that is, it greatly contributes in explaining whether an eligible individual participates in the ERA study or not) and it passes the exclusion restriction test in terms of non-ERA outcomes at any significance level.³⁶

Tables A.1, A.2 and A.3 present the findings from the four types of control function models that have been 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 different estimations have been performed, including non-linear and interaction terms in the first-stage Probit ('interactions') or not ('no interactions').

To preview the conclusions on this part of the analysis, the strong demands placed on the data to identify and estimate the various models in a robust way have proved difficult to meet. Several of the models failed to pass the strict specification tests, and often provided very noisy and unstable estimates. Nonetheless, the results that have emerged can be viewed as useful sensitivity checks that have indeed mostly corroborated the evidence emerging from the analysis based on controlling only for observables (Section 5.3).

Specifically, for the ND25+ intake group, six out of seven of the control function models for employment (Table A.1) do not uncover any significant selection on unobservables, and the one which does confirms the previous finding that for employment outcomes, the experimental impact significantly underestimates the impact for the non-participants and indeed for the eligible population.

As for cumulative earnings (Table A.3), the control function model fails to detect any significant selection on unobservables, as well as any significant difference between the impact for participants on one hand, and the impacts for the eligible population and for non-participants on the other. As was the case for the selection-on-observables findings, the point estimates (and their significance) indicate a possibly larger impact for the latter two subgroups, though, as mentioned, such differences never reach statistical significance.

³⁶ One caveat concerns the – untestable – part of the exclusion restriction which relates to ERA outcomes. If there were a relationship between becoming more persuasive in convincing people to participate in the ERA study and helping them in the labour market, the instrument would not be a valid one. It has to be kept in mind though that no such relationship was detected in terms of non-ERA outcomes; in this case, (New Deal) advisers becoming more persuasive in getting people into the study did not become more helpful for the controls. Finally, a strong factor driving the instrument is the permission for advisers to mention ERA's bonuses, which is arguably random.

For the NDLP group, all employment models (Table A.2) fail to detect any significant residual selection on unobservables and all models but two confirm the findings in Section 5.3 that the four-year 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.³⁷ Since the models detect no significant selection, the results of the models suggesting increased impacts should not be given too much weight – but it is nevertheless reassuring that they do show the same direction of difference. For the earnings outcome (Table A.3), the control function model does uncover weak evidence of selection on unobservables and points to estimates for the impact on the non-participants (and hence, on the eligible population) many times larger than the experimental estimate. However, none of these three estimates is statistically significant, and neither are the differences in impacts between participants and the other two subgroups. This model also fails to pass all of the five specification tests, which means that its findings have to be viewed as indicative at best.

³⁷ Two models for employment do find significant differences between the experimental impact and the impacts for participants and for the eligible population despite the absence of evidence of selection on unobservables. The estimated impacts for the non-participants appear, however, implausibly large. The average effect for the eligible group is correspondingly large and significant. These large effects easily become statistically different from the zero effect for participants. These two models, thus, seem to imply that the experimental effect on employment underestimates the impact for the non-participants and for the eligible population. Such findings should, however, be interpreted with extreme care, especially given how inadequately one of the models (the censored model) performs in terms of the two criteria.

6 Summary and conclusions

6.1 Drawing the findings together

This concluding section starts by drawing together the findings from the different types of analyses that have been performed, the results of which are summarised in Table 6.1 for the two intake groups.

- The issue of non-participation is potentially a serious one, both given the extent of non-participation (with over one-quarter of the eligible population not participating in the Employment Retention and Advancement ERA study) and given important selective differences between study participants and non-participants.
- In the absence of ERA, the study participants of both the New Deal for Lone Parents (NDLP) and New Deal 25 Plus (ND25+) groups experience better employment outcomes – and possibly somewhat higher earnings – over the four-year follow-up period than non-participants with the same observable characteristics, including extensive labour market histories. The adviser- and self-selected individuals making up the group of the non-participants are, thus, characterised by unobservable factors that make them, on average, more detached from the labour market.
- This study has extensively explored whether, and how much, the experimental impacts in the four years since inflow into the New Deal are representative of the potential impact of offering ERA services and incentives to the entire population of New Deal entrants, that is, to the full group of those eligible for ERA, 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, the story that emerges appears to be quite different for the ND25+ and NDLP groups and across outcomes.

6.1.1 ND25+ group

- If all those eligible to take part in the ERA study had done so, the **employment** effects of ERA in the four years since programme start would have been significantly higher. In particular, the effect on employment probability from the experimental analysis which ignores the non-participants is a significant **underestimate** of the impact that the full eligible population would have experienced.
- This result is driven by the fact that the employment impact for the non-participants would have been considerably larger than the one for the participants. Compared to a 2.8 percentage point increase on the probability of being ever employed over the four-year period found for the participants, the non-participants would have enjoyed a 6.4 percentage point increase, and all those eligible for ERA a 3.7 percentage point increase, with both impacts for the non-participants and for the eligibles being statistically significantly different from the impact enjoyed by the participants.

Similarly, compared to an increase in employment duration over the four-year period of 25 days found for the experimental sample of participants, the non-participants would have enjoyed a 63-day increase, and all those eligible for ERA a 34-day increase.

- As the non-participants were found to be more detached from the labour market – experiencing worse employment outcomes and at times lower earnings – these findings might indicate that for these more labour-market detached ND25+ entrants some extra help in the form of advice and financial incentives might be particularly helpful in improving their labour market situation.
- By contrast, the four-year experimental estimates of the impact for study participants are **representative** of the average impact that the eligible population would have experienced in terms of **earnings**, as measured by the four tax-year earnings post New Deal entry and by cumulative 2005–09 tax earnings. Specifically, the point estimates for the earnings impacts for non-participants as well as for the full eligible population are always larger than the corresponding point estimates for participants, but such differences are far from reaching any statistical significance.

6.1.2 NDLP group

- The experimental impact estimates of **employment** outcomes in the four years post-inflow are found to be **representative** of the average impact that ERA would have had on its eligible population. Specifically, no significant impact was found for the experimental sample on either employment duration or the probability of being employed during the four-year follow-up period, and the absence of any significant impact extends to the non-participants, and hence, to the eligible population. Again, it is worth noting, however, that the point estimates for the eligible population and especially for the non-participants are higher and in positive territory compared to the (insignificant but) negative ones for participants.
- By contrast, if all those eligible for ERA had taken part, the **earnings** impacts for the NDLP group would have been higher. In other words, excluding the non-participants from the analysis significantly **underestimates** the impact that ERA would have had on the average earnings of all those eligible for ERA. This is the case for earnings in the 2006/07 and 2008/09 tax years as well as for cumulative earnings (the point estimates for the non-participants are strongly significant and larger than the ones for participants in the other two fiscal years as well). Specifically, while ERA has not significantly increased participants' cumulative earnings over the follow-up period (an insignificant £767 point estimate), non-participants would have enjoyed a highly significant £2,393 rise in cumulative earnings and the eligible population a significant £1,262 rise, where the impacts for the non-participants and for the eligibles are statistically significantly different from the impact for the participants.

In conclusion, for both New Deal groups and in terms of all outcomes considered, non-participants would have enjoyed uniformly higher ERA impacts than do participants. Only for ND25+ employment and NDLP earnings outcomes do, however, such differences reach statistical significance and indeed, extend to significant differences between impacts for study participants and impacts for the full eligible population.

- How do these results four years after inflow compare with the previous findings by Sianesi (2010) relating to the first year?
 - As can be seen from Box 6.1, the story has remained the same at one and four years after inflow for employment outcomes: the experimental impact underestimates the impact for the non-participants and for the eligible population for the ND25+ group, while the absence of any experimental impact for NDLP participants would extend to the non-participants and to the eligible population.

- By contrast, the story has changed in terms of earnings, notably with the finding that four years post-New Deal start, the experimental impact on NDLP participants is no longer representative of (and would indeed underestimate) the impact on non-participants and on the eligible population.

It has to be noted that the previous report could only use earnings outcomes collected from the first-wave survey of a sample of ERA participants, while the current report was able to exploit administrative earnings data on all those eligible for ERA. Despite the fact that these data were measured differently and for different samples (the latter requiring the use of a different methodology to identify and estimate impacts), the findings in the first year since inflow reassuringly coincide. Specifically, both in terms of survey earnings measures from the previous report and in terms of administrative earnings measures from the current one, the positive impact on earnings estimated for the NDLP participants was found to be representative of the impact that the non-participants and the full eligible population would have experienced under ERA in their first year since inflow. Similarly for the ND25+ group, first-year earnings impacts from both survey measures, ignoring non-response and from administrative data, point to the experimental estimates being representative of the impact on the full eligible population.

Box 6.1 Summary of year-one results

ND25+		NDLP
Employment		
Participants	none; +4.6 days	No impact for participants, non-participants and eligible population
Non-participants	+5.6pp; +10 days	
Eligible group	+2.6pp; +5.8 days	
Experimental impact	underestimates the effect for the eligible population	Experimental impact is representative of the effect for the eligible population (i.e. none)
Earnings		
Impact for the responding participants	is representative of (possibly underestimates when allowing for non-response) the effect for the full eligible population	Impact for the responding participants is representative of the impact for the full eligible population

Notes:

Employment outcomes are the probability of having ever been employed during the 12 months since inflow and the number of days employed during the 12 months since inflow.

Earnings are total yearly earnings in the 12 months since random assignment and are taken from the first-wave survey of a sample of ERA participants.

For the full set of results and their in-depth discussion, see Sianesi (2010).

- Finally, no support has been found for the conjecture that had the non-participants been offered ERA, they would have been mostly uninterested in effectively taking up its support and incentives. In fact, the results show that for both intake groups overall, those eligible for ERA 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 in the first year since inflow. 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 for 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 in

their first year. 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 Conclusions

The picture that emerged from the interim four-year experimental findings is that ERA has not been effective in improving the employment and earnings outcomes of NDLP participants in the four-year follow-up period, but that the intervention has significantly improved both types of outcomes for ND25+ participants.³⁸ How has the presence of the non-participants affected the representativeness (or external validity) of the experimental impact estimates?

In the descriptive examination of the non-participation problem by Goodman and Sianesi (2007), it had been speculated that it would be hard for the non-participants to give rise to an estimate for the eligible population 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 by ‘relatively large’ or ‘relatively small’).

The findings in this report have, however, proved this previous supposition partly wrong.

Indeed, the report has found that the picture emerging from the four-year findings on ERA at times **does** change, painting a more favourable picture of the effectiveness that the ERA intervention would have had on its full eligible population.

This is the case in terms of employment chances for the ND25+ group and especially in terms of earnings for the NDLP group. In both of these instances, the experimental impact estimates for the participants underestimate the gains that the eligible population would have enjoyed had it been offered ERA services and incentives.

For the ND25+ group, the experimental 2.8 percentage point increase in employment probability and 25 day increase in employment duration for the participants underestimate the likely impact that ERA would have had on all ND25+ entrants by almost one percentage point and ten days. Of even more substance is the fact that for the NDLP experimental group no ERA impact was found on earnings beyond the first year, whereas the earnings impact that ERA would have been likely to have had on all NDLP entrants remains significant in most subsequent years and totals a statistically significant £1,262 gain cumulatively.

For other combinations of outcomes and intake groups, the picture emerging from the four-year findings on ERA remains completely unchanged. In particular, the four-year experimental results of a substantial increase in ND25+ earnings and of no change in NDLP employment appear to generalise to the full eligible population.

Overall, the representativeness/external validity of the experimental impact estimates was, thus, found to depend on the intake group and outcome considered. As outlined, for ND25+ employment and NDLP earnings, evidence of non-participation bias (or of some loss in external validity) has been uncovered.

Of course, there is always the issue of how different the estimates for the eligible population and for the experimental sample need to be for one to view the issue as a particularly important one.

³⁸ A full ERA evaluation report looking at the five-year impacts, as well as at the costs and benefits associated with the programme will be published by the Department for Work and Pensions (DWP) during spring 2011.

Randomised experiments are, however, conceptually designed to provide, with accuracy, the ‘true’ answer to the evaluation question. Hence, an effect for the eligible population which is over one-third larger than the experimental estimate or indeed a large, significant impact for the eligible population that surfaces when none was found experimentally, can be viewed as findings of substance.

This report has not only extensively assessed the external validity of the four-year ERA findings, but it has set the foundation work and developed a sound and thorough methodological framework for the analysis of non-participation in experimental studies. Given that in many evaluation settings the problem of non-participation is an empirically relevant one (see, for example, Kamionka and Lacroix, 2005), the framework developed in this report 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 intake group, analysis type and outcome during the four-year follow-up period

	ND25+	NDLP
Employment		
Experimental impact for participants	Positive impact (2.8 pts, 25 days)	No impact on probability nor duration
Bounds and sensitivity analysis	Impact on the eligible population is positive and small under most scenarios	Impact is sensitive to the scenario assumed but more likely to be negative rather than positive
Selection on observables analysis	Stronger effect for non-participants (6.4 versus 2.8pts, 63 versus 25 days): significantly higher than experimental impact, which is an underestimate of the impact on all those eligible for ERA (3.7pts and 34 days)	Weak evidence that effect on non-participants' probability of employment is higher than experimental impact estimate; impact on participants still representative of the impact on all those eligible for ERA
Selection on observables: sensitivity	Impact on the eligible population is positive (both for days and probability)	Non-informative
Selection on unobservables analysis	Supports evidence from selection on observables analysis	
Earnings		
Experimental impact for participants	Positive effect (£392 annually, £1,805 cumulatively)	No significant impact (annually or cumulatively)
Selection on observables analysis	Impact for participants is representative of the impact for the eligible population	Large impact for non-participants: significantly higher than the experimental impact, which underestimates the impact for the eligible population
Selection on observables: sensitivity	Impact on the eligible population is positive overall (both annually and cumulatively)	Non-informative
Selection on unobservables analysis	Supports evidence from selection on observables analysis	

Appendix A

Control function: Methods and results

A.1 Impact estimates under selection on unobservables

This appendix sketches a class of models which allow selection into the group of Employment Retention and Advancement (ERA) study participants to depend on outcome-relevant unobservables. All of these models fall within the family of ‘control function models’ and build upon the classical sample selection model introduced by Heckman (1979). The current set-up with administrative outcomes, however, places the analyst in the rather unique position of being able to observe the outcomes of the selected-out sample. Together with randomisation, this feature of the data is exploited to test:

- (a) part of the exclusion restriction of the instrument;
- (b) for the presence of residual selection on unobservables related to no-treatment employment or earnings outcomes;
- (c) how well the various control function models capture the presence and direction of the selection on unobservables that has thus been uncovered; and
- (d) 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. The section, thus, starts by presenting tests (a) and (b), then moves on to sketch the various models, outlining the idea behind tests (c) and (d).

A.1.1 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, one needs 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:

- (CF) (a) $P(Q=1 | X, Z)$ is a non-trivial function of Z
- (b) $E(Y_1 | X, Z) = E(Y_1 | X)$

The strength (‘power’) of the instrument in affecting the choice among eligible individuals to participate in the ERA study, i.e. condition (CF.a), is, as usual, a testable condition. In this case, however, of modelling administrative outcomes one can test part of the exclusion restriction as

well (condition CF.b), specifically one can test whether this condition holds in terms of non-ERA outcomes, i.e. whether $E(Y_0 | X, Z) = E(Y_0 | X)$. 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 this unique set-up, one is 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 earnings outcomes. This can be accomplished 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.³⁹ 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 shown below, they lend themselves to construction of an important specification check for any of the control function models.

A.1.2 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.

This is a rather formal set-up, requiring technical conditions for identification and at times quite complex estimation methods. The description provides the least detail which is necessary to appreciate the assumptions underlying the estimates and to interpret the output presented below.

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:

$$\begin{aligned} Y_0 &= \beta_0 X + u & u &\sim N(0, \sigma_u^2) \\ Y_1 &= \beta_1 X + u + b & b &\sim N(0, \sigma_b^2) \end{aligned}$$

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 ($Q=0$). Let the observability rule for Y_1 be:

³⁹ 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. Furthermore, the qualitative evaluation found no evidence to support the possibility that the process of holding out the chance to become eligible to substantial financial bonuses and then informing the controls that they would not in fact be eligible may have had some impact on their motivation and subsequent actions.

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

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

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):

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

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

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:

- (CF) (a) $P(Q=1 | X, Z)$ is a non-trivial function of Z
 (b) $E(Y_1 | X, Z) = E(Y_1 | X)$

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

Apart from the parametric choice of the distribution of the unobservables implied by condition (CF.c) (in particular, joint normality and homoskedasticity), the control function model crucially relies on an exclusion restriction. Specifically, one needs 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 CF.a), but is not contained in X , i.e. does not affect potential ERA outcomes directly (condition CF.b).

As discussed in Section A.1.1, condition (CF.a) and condition (CF.b) in terms of non-ERA outcomes Y_0 can be tested. Also, the parametric assumptions in (CF.c) can be relaxed (and thus tested), as shown in the next subsection.

Under the assumptions of the model, one can derive the exact form of the expected unobserved treatment outcome for each individual non-participant with a given set of characteristics W .

A convenient 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.

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

In particular, it is possible 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 A.1.1. 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, one can use the model to recover the predicted no-treatment outcomes for the non-participants, which can then be compared to the average **observed** no-treatment outcome for the non-participants.

The analyst is, 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 the test of selection on unobservables), as well as the average predicted and observed (no-treatment) outcomes of the non-participants.

A.1.3 Extensions to the standard control function model

The standard model has been extended in two broad directions.

First, the parametric assumptions on the unobservables have been relaxed in terms of both the **independence** and **normality** implied by condition (CF.c). 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 entrants 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.

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, one can perform a number of ‘tests’ on the performance of the model. In particular, one 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, the model is used 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. The average effect for the participants using the extended model is estimated, and this estimate is then compared to the experimental one.

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

	How well the model...	Should be...
$(\alpha - \text{selection terms})$	captures the actual extent of selection	zero
$Q=0$: observed–predicted Y	predicts (no-treatment) outcomes for non-participants	zero
$R=0$: observed–predicted Y	predicts (no-treatment) outcomes for the control group	zero
$R=1$: observed–predicted Y	predicts (treatment) outcomes for the programme group	zero
$E(Y R=0)-E(Y_0 R=1)$	predicts no-treatment outcomes for the programme group	zero
$ATE_{1,\dots}$	predicts the average impact for participants	ATE_1 (experimental)

Finally note that the estimate of the average ERA impact for the eligible population 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.

A.2 Control function results

Table A.1 Control function models: ERA impacts on days in employment during the four-year follow-up – ND25+

	Standard		No independence		No normality		Censoring	
	No interactions	Interactions	No interactions	Interactions	No interactions	Interactions	No interactions	No interactions
Selection on unobserved characteristics	No	No	No	No	Yes**	No	No	No
Selection on unobserved ERA gains	No	No	No	No	No	No	No	No
(α - selection terms)	312	197	-673	-210	-3,192***	-1,332***	54	54
$Q=0$: observed-predicted Y	411	39	696	207	3,248***	1,333***	-80	-80
independence: Y_0			No**	No***				
independence: Y_1			No**	No*				
normality: Y_0					No**	Yes		
normality: Y_1					No*	Yes		
$R=0$: observed-predicted Y								-7
$R=1$: observed-predicted Y								-9
$E(Y R=0)-E(Y_0 R=1)$								-8
ATE_1	25**	25**	24**	25**	25**	25**	26	26
ATE_0	343	220	57	195	1,930**	791***	133	133
ATE	98	70*	32	64	462**	201***	51*	51*
$ATE_1 - ATE$	-73	-45	-8	-39	-437**	-176***	-25	-25
$ATE_1 - ATE_0$	-318	-196	-33	-171	-1,905***	-766***	-107	-107

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

Selection: α	(OLS)	50***	(Tobit)	45***
ATE_1 (experimental)	(OLS)	25**	(Tobit)	25**
Non-participants' observed Y		234		
N				7,796

Table A.2 Control function models: ERA impacts on days in employment during the four-year follow-up – NDLP

	Standard		No independence		No normality		Censoring	
	No interactions	Interactions	No interactions	Interactions	No interactions	Interactions	No interactions	No interactions
Selection on unobserved characteristics	No	No	No	No	No	No	No	No
Selection on unobserved ERA gains	No	No	No	No	No	No	No	No
(α - selection terms)	-58	81	185	166	375	188	204***	
$Q=0$: observed-predicted Y	122	-29	-232	-180	-390	-194	-158***	
independence: Y_0			Yes	Yes				
independence: Y_1			Yes	Yes				
normality: Y_0					Yes	Yes		
normality: Y_1					Yes	No*		
$R=0$: observed-predicted Y								5
$R=1$: observed-predicted Y								6
$E(Y R=0) - E(Y_0 R=1)$								0
ATE_1	-4	-4	-2	-2	-4	-3	-5	
ATE_0	-87	68	120	72	604	580***	130***	
ATE	-29	18	35	21	181	174***	36***	
$ATE_1 - ATE$	25	-22	-37	-23	-185	-177***	-41***	
$ATE_1 - ATE_0$	83	-72	-122	-74	-607	-583***	-135***	

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

Selection: α	(OLS)	35**	(Tobit)	40**
ATE_1 (experimental)	(OLS)	-4	(Tobit)	-6
Non-participants' observed Y		580		
N		7,261		

Table A.3 Control function models: ERA impacts on cumulative earnings

	Censoring	
	ND25+	NDLP
Selection on unobserved characteristics	No	Weak: $\rho_{uv} = 0.975^*$
Selection on unobserved ERA gains	No	No
(α - selection terms)	49,660**	33,449***
$Q=0$: observed-predicted Y	-51,102**	36,172***
$R=0$: observed-predicted Y	-1,343	-3,139*
$R=1$: observed-predicted Y	-2,592	-1,278***
$E(Y R=0)-E(Y_0 R=1)$	-1,339	-3,229*
ATE_1	2,512	-155
ATE_0	20,916	10,313
ATE	6,738*	3,029
$ATE_1 - ATE$	-4,226	-3,185
$ATE_1 - ATE_0$	-18,404	-10,469
Note: Statistical significance based on bootstrapped bias-corrected confidence intervals (500 replications): *** significant at 1%, ** at 5%, * at 10%.		
	ND25+	NDLP
Selection: α	(OLS)	(OLS)
ATE_1 (experimental)	972 (Tobit)	1210* (Tobit)
Non-participants' observed Y	1803*** (Tobit)	1666*** (Tobit)
N	10,898	15,699
	7,796	7,261
		1196 (Tobit)
		762 (Tobit)
		982*
		846

Appendix B

District-level results

B.1 Experimental findings

Table B.1 Experimental findings for the ND25+ group during the four-year follow up

	Raw		Adjusted		N
	Effect	Std.Err.	Effect	Std.Err.	
Ever employed					
All	0.027**	0.013	0.028**	0.012	6,006
Scotland	0.054	0.036	0.053	0.035	745
North East England	0.011	0.037	-0.005	0.035	703
North West England	0.021	0.027	0.028	0.025	1,377
Wales	-0.034	0.047	-0.038	0.046	456
East Midlands	0.042	0.028	0.040	0.026	1,245
London	0.034	0.026	0.041*	0.024	1,480
Days employed					
All	24.6**	11.2	25.0**	10.6	6,006
Scotland	80.1**	31.6	73.9**	30.7	745
North East England	7.1	34.6	3.4	34.4	703
North West England	29.2	22.5	28.0	21.4	1,377
Wales	-48.7	39.8	-45.3	39.6	456
East Midlands	38.0	25.9	32.5	24.9	1,245
London	12.0	21.6	17.0	20.4	1,480
Earnings (2008/09)					
All	410**	199	392**	195	6,006
Scotland	1,662***	537	1,508***	540	745
North East England	-85	721	11	720	703
North West England	258	372	182	369	1,377
Wales	-285	527	-332	534	456
East Midlands	452	388	378	385	1,245
London	294	449	217	444	1,480
Earnings (2005–09)					
All	1,894***	702	1805***	687	6,006
Scotland	4,472***	1,639	3,817**	1,627	745
North East England	550	2,305	580	2,302	703
North West England	1,582	1,756	1,368	1,753	1,377
Wales	-1,573	1,891	-2,255	1,949	456
East Midlands	2,970**	1,422	2,662*	1,406	1,245
London	1,585	1,324	1,372	1,291	1,480

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

Robust standard errors for ever employed.

Table B.2 Experimental findings for the NDLP group during the four-year follow up

	Raw		Adjusted		N
	Effect	Std.Err.	Effect	Std.Err.	
Ever employed					
All	-0.002	0.012	-0.008	0.011	5,052
Scotland	-0.020	0.037	-0.014	0.038	413
North East England	-0.021	0.027	-0.018	0.027	983
North West England	0.078***	0.029	0.068**	0.028	759
Wales	-0.051	0.033	-0.056*	0.032	514
East Midlands	-0.013	0.026	-0.019	0.025	1,131
London	-0.007	0.026	0.000	0.025	1,252
Days employed					
All	2.0	14.7	-4.0	13.9	5,052
Scotland	16.7	51.9	8.4	53.7	413
North East England	-23.1	33.5	-15.7	32.3	983
North West England	101***	36.6	70*	35.7	759
Wales	-68.4	45.3	-97**	45.2	514
East Midlands	-7.0	31.1	-11.3	30.0	1,131
London	-8.8	30.0	-0.2	28.3	1,252
Earnings (2008/09)					
All	21	230	-13	227	5,052
Scotland	-1,308*	694	-1,583**	710	413
North East England	132	430	333	432	983
North West England	217	501	-115	503	759
Wales	-243	535	-432	538	514
East Midlands	533	439	523	441	1,131
London	-43	628	-189	631	1,252
Earnings (2005–09)					
All	927	737	767	722	5,052
Scotland	-12	2,026	-356	2,113	413
North East England	189	1,324	627	1,305	983
North West England	1,927	1,641	536	1,626	759
Wales	316	1,685	-362	1,648	514
East Midlands	1,612	1,434	1,630	1,429	1,131
London	1,000	2,041	588	2,042	1,252

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

Robust standard errors for ever employed.

Summary Box B.1 Four-year experimental findings for the ATE_1

	ND25+	NDLP
Employment	Positive effect on probability (2.8 ppts) <ul style="list-style-type: none"> • Driven by London Positive effect on days (25 days) <ul style="list-style-type: none"> • Driven mainly by Scotland (74 days) • Negative but insignificant impact in Wales 	No overall significant effect on probability <ul style="list-style-type: none"> • Positive impact in North West England (6.8 ppts) • Negative impact in Wales (-5.6 ppts) No overall significant effect on days <ul style="list-style-type: none"> • Positive impact in North West England (70 days) • Negative impact in Wales (-97 days)
Earnings	Positive effect on earnings (£392 annually, £1,805 cumulatively) <ul style="list-style-type: none"> • Driven by Scotland (£1,508 annually, £3,817 cumulatively) and East Midlands (£2,662 cumulatively) 	No significant effect on annual nor cumulative earnings <ul style="list-style-type: none"> • Significant negative effect in Scotland (-£1,583)

B.2 Bounds

Table B.3 Non-parametric bounds for the ATE – Outcome: ever employed during the four-year follow up

	p	ATE_1	ATE	\overline{ATE}	95% CI lower	95% CI upper	N
ND25+							
All	0.230	0.028**	-0.081	0.148	-0.102	0.169	7,796
Scotland	0.087	0.053	0.011	0.098	-0.056	0.166	816
North East England	0.349	-0.005	-0.149	0.200	-0.202	0.252	1,080
North West England	0.146	0.028	-0.050	0.095	-0.096	0.141	1,612
Wales	0.207	-0.038	-0.126	0.081	-0.202	0.156	575
East Midlands	0.275	0.042	-0.105	0.170	-0.149	0.213	1,717
London	0.259	0.041*	-0.074	0.185	-0.113	0.224	1,996
NDLP							
All	0.304	-0.008	-0.227	0.077	-0.247	0.097	7,261
Scotland	0.053	-0.014	-0.065	-0.012	-0.134	0.057	436
North East England	0.292	-0.018	-0.240	0.052	-0.283	0.094	1,389
North West England	0.062	0.068***	0.023	0.085	-0.030	0.138	809
Wales	0.236	-0.056*	-0.226	0.010	-0.278	0.062	673
East Midlands	0.471	-0.019	-0.357	0.115	-0.390	0.148	2,140
London	0.310	-0.000	-0.220	0.090	-0.259	0.129	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 B.1 ND25+: Sensitivity analysis: ATE_0 for ever employed, θ from 0.5 to 1.5

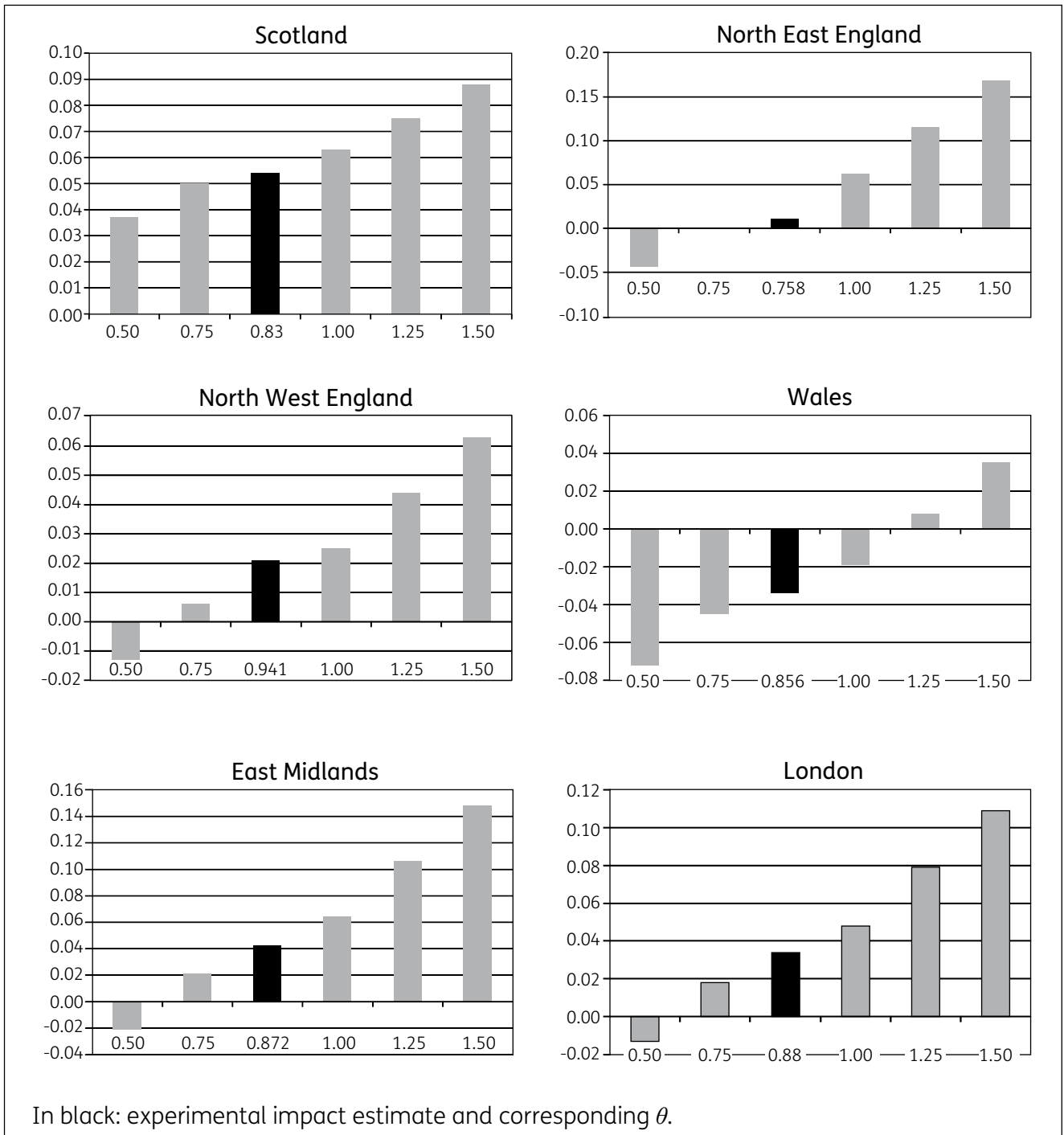
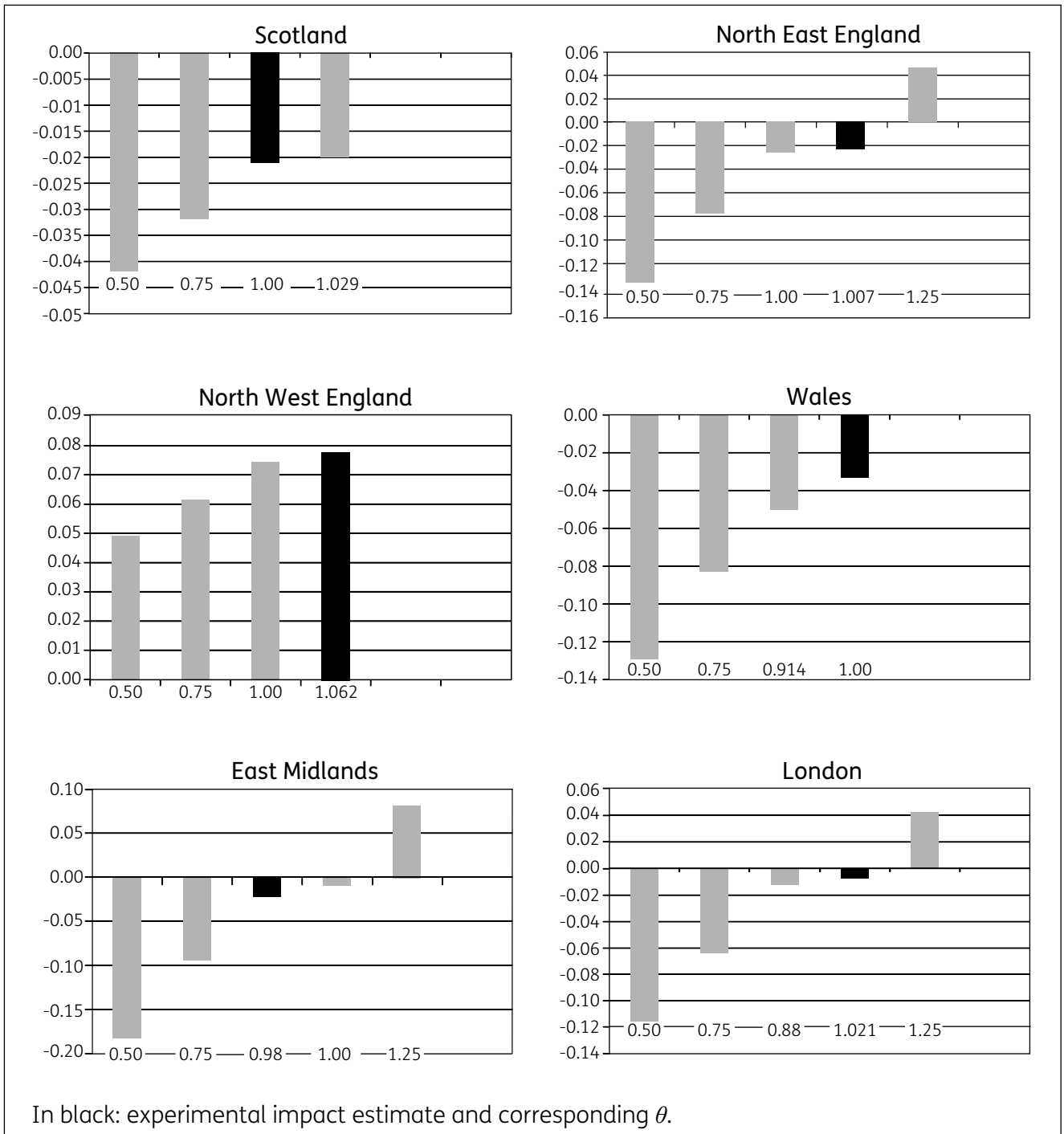


Figure B.2 NDLP: Sensitivity analysis: ATE_0 for ever employed, θ from 0.5 to 1.5



Summary Box B.2 Bounds and sensitivity analysis for ever employed during the four-year follow up

	ND25+	NDLP
Bounds	Impact is positive in Scotland	Impact is positive in NW England and small and negative in Wales
Sensitivity	Overall: positive small under most scenarios Positive substantial in Scotland Positive and small under most scenarios in NE England, NW England, E Midlands and London Negative and small in Wales	Overall: sensitive Positive small in NW England Negative in Scotland and Wales Mostly negative in London, E Midlands and NE England

B.3 Selection on observables

Table B.4 Matching estimates for the ND25+ group during the four-year follow-up

	<i>p</i>	<i>ATE₁</i>	<i>ATE₀</i>	<i>ATE</i>	$\frac{ATE_1 \neq ATE_0}{ATE_0}$	$\frac{ATE_1 \neq ATE_0}{ATE_0}$	<i>N</i>
All							
Days employed	0.230	25.0**	62.5***	33.6***		***	
Ever employed	0.230	0.028**	0.064***	0.037***	**	**	4,831
Earnings 2008/09	0.230	392**	482*	413**			
Earnings 2005–09	0.230	1,805**	2,392**	1,940***			
Scotland							
Days employed	0.087	73.9**	28.2	70.0			
Ever employed	0.087	0.053	0.068	0.054			
Earnings 2008/09	0.087	1,508***	-14	1,375			455
Earnings 2005–09	0.087	3,817**	1,634	3,627			
North East England							
Days employed	0.349	3.4	74.9	28.3	**	***	
Ever employed	0.349	-0.005	0.056	0.017			737
Earnings 2008/09	0.349	11	1,225	434	*		
Earnings 2005–09	0.349	580	5,374	2,254	**	**	
North West England							
Days employed	0.146	28.0	2.5	24.3			
Ever employed	0.146	0.028	-0.009	0.023			932
Earnings 2008/09	0.146	182	280	196			
Earnings 2005–09	0.146	1,368	-534	1,090			

Continued

Table B.4 Continued

	<i>p</i>	<i>ATE₁</i>	<i>ATE₀</i>	<i>ATE</i>	<i>ATE₁ ≠ ATE₀</i>	<i>ATE₁ ≠ ATE₀</i>	<i>N</i>
Wales							
Days employed	0.207	-45.3	27.5	-30.3			344
Ever employed	0.207	-0.038	0.034	-0.023			
Earnings 2008/09	0.207	-331	297	-201			
Earnings 2005–09	0.207	-2,255	-618	-1,900			
East Midlands							
Days employed	0.275	32.5	91.7	48.7	**	**	1,097
Ever employed	0.275	0.042	0.08	0.051			
Earnings 2008/09	0.275	378	238	340			
Earnings 2005–09	0.275	2,663*	2,217	2,540			
London							
Days employed	0.259	17.0	69.3	30.5	**	**	1,266
Ever employed	0.259	0.041*	0.092	0.054	*	*	
Earnings 2008/09	0.259	217	804	368			
Earnings 2005–09	0.259	1,372	3,412	1,899			

ATE₁ ≠ ATE and *ATE₁ ≠ ATE₀* columns: bootstrap-based statistical significance of the difference.

Statistical significance based on bootstrapped bias-corrected confidence intervals (1000 replications): *** significant at 1%, ** at 5%, * at 10%.

Table B.5 Matching estimates for the NDLP group during the four-year follow-up

	<i>p</i>	<i>ATE₁</i>	<i>ATE₀</i>	<i>ATE</i>	<i>ATE₁ ≠ ATE₀</i>	<i>ATE₁ ≠ ATE₀</i>	<i>N</i>
All							
Days employed	0.304	-4.0	17.9	2.7			4,768
Ever employed	0.304	-0.008	0.020	0.000	*	*	
Earnings 2008/09	0.304	-13	691***	202	***	***	
Earnings 2005–09	0.304	767	2,393***	1,262**	*	*	
Scotland							
Days employed	0.053	8.4	184.1	17.6			229
Ever employed	0.053	-0.014	0.003	-0.013			
Earnings 2008/09	0.053	-1,583**	2,214	-1,400	*	**	
Earnings 2005–09	0.053	-356	8261	98		**	
North East England							
Days employed	0.292	-15.7	67.8	8.7			915
Ever employed	0.292	-0.018	0.007	-0.011			
Earnings 2008/09	0.292	333	910	502			
Earnings 2005–09	0.292	627	3,449	1,452			

Continued

Table B.5 Continued

	<i>p</i>	<i>ATE</i> ₁	<i>ATE</i> ₀	<i>ATE</i>	<i>ATE</i> ₁ ≠ <i>ATE</i> ₀	<i>ATE</i> ₁ ≠ <i>ATE</i> ₀	<i>N</i>
North West England							
Days employed	0.062	70.0*	-111.0	58.3	**	**	452
Ever employed	0.062	0.068**	0.026	0.066			
Earnings 2008/09	0.062	-115	-1,300	-186			
Earnings 2005–09	0.062	536	-2900	325	*	*	
Wales							
Days employed	0.236	-97**	45.8	-63.3	*	*	419
Ever employed	0.236	-0.056*	0.039	-0.033	*	*	
Earnings 2008/09	0.236	-432	1,150	-58	**	**	
Earnings 2005–09	0.236	-362	5,038	914	**	**	
East Midlands							
Days employed	0.471	-11.3	10.4	-1.1			1,576
Ever employed	0.471	-0.019	0.021	0.000			
Earnings 2008/09	0.471	523	812	659			
Earnings 2005–09	0.471	1,630	2,418	2,001			
London							
Days employed	0.310	-0.2	-0.9	-0.4			1,177
Ever employed	0.310	0.000	0.009	0.003			
Earnings 2008/09	0.310	-189	818	123			
Earnings 2005–09	0.310	588	2,992	1,332			

*ATE*₁ ≠ *ATE* and *ATE*₁ ≠ *ATE*₀ columns: 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%.

Summary Box B.3 Selection on observables

	ND25+	NDLP
Employment	better effect for $Q=0$ <ul style="list-style-type: none"> driven by NE England, East Midlands and London <i>ATE</i>₁ underestimates <i>ATE</i> for employment outcomes	better effect for $Q=0$ (but weak significance) <ul style="list-style-type: none"> driven by Wales except in NW England (better effect for $Q=1$) <i>ATE</i>₁ representative of <i>ATE</i> for employment outcomes
Earnings	Not significantly different effect for $Q=0$ <ul style="list-style-type: none"> except in NE England (better for $Q=0$) <i>ATE</i>₁ representative of <i>ATE</i> for earnings outcomes	better effect for $Q=0$ <ul style="list-style-type: none"> driven by Wales and Scotland except in NW England (better effect for $Q=1$) <i>ATE</i>₁ underestimates <i>ATE</i> for earnings outcomes

B.4 Testing for selection on specific unobservables

Table B.6 Differences in outcomes for participants (control group) compared to non-participants with the same observed characteristics during the four-year follow up: ND25+

	OLS	FILM	Matching	N
All				
Days employed	50***	34**	37***	
Ever employed	0.042***	0.026	0.033*	4,755
Earnings 2008/09	330	72	103	
Earnings 2005–09	972	138	162	
Scotland				
Days employed	0.9	41.1	33.9	
Ever employed	0.030	-0.022	0.034	432
Earnings 2008/09	-904	459	-984	
Earnings 2005–09	-951	4,445	979	
North East England				
Days employed	95***	89**	97***	
Ever employed	0.079**	0.037	0.051	720
Earnings 2008/09	1,370**	1,559*	1,507**	
Earnings 2005–09	5,060**	6,162***	5,628**	
North West England				
Days employed	-12.9	-21.0	-27.9	
Ever employed	-0.02	-0.007	-0.009	915
Earnings 2008/09	136	-137	37	
Earnings 2005–09	-2,200	-3,800	-2,300	
Wales				
Days employed	121**	131**	136**	
Ever employed	0.038	0.035	0.068	350
Earnings 2008/09	123	-798	-76	
Earnings 2005–09	1,510	-1,900	1,115	
East Midlands				
Days employed	62**	56**	54*	
Ever employed	0.057**	0.046	0.041	1,092
Earnings 2008/09	275	90	184	
Earnings 2005–09	804	-122	335	
London				
Days employed	61***	76***	68***	
Ever employed	0.050*	0.060**	0.054*	1,246
Earnings 2008/09	553	796	640	
Earnings 2005–09	2,028	2,838*	2,629	

Table B.7 Differences in outcomes for participants (control group) compared to non-participants with the same observed characteristics during the four-year follow up: NDLP

	OLS	FILM	Matching	N
All				
Days employed	35**	25.5	23.8	
Ever employed	0.035***	0.027*	0.031*	4,702
Earnings 2008/09	379	416	326	
Earnings 2005–09	1,196	1,201	741	
Scotland				
Days employed	205*	261***	289**	
Ever employed	-0.037	-0.007	0.145	230
Earnings 2008/09	1,952	6,658***	3,977***	
Earnings 2005–09	6,471*	22,000***	11,000***	
North East England				
Days employed	59.812*	54.2	54.5	
Ever employed	0.019	0.01	0.014	880
Earnings 2008/09	110	188	287	
Earnings 2005–09	1,568	1,472	1,700	
North West England				
Days employed	-188**	-164**	-300**	
Ever employed	-0.007	0.008	-0.07	407
Earnings 2008/09	-670	-768	-2,900*	
Earnings 2005–09	-2,000	-2,600	-8,900*	
Wales				
Days employed	116**	150**	130*	
Ever employed	0.117***	0.141***	0.142***	413
Earnings 2008/09	666	982	679	
Earnings 2005–09	2,830	4,468**	3,722	
East Midlands				
Days employed	24.4	29.3	29.3	
Ever employed	0.042*	0.044*	0.046*	1,573
Earnings 2008/09	184	275	250	
Earnings 2005–09	634	758	696	
London				
Days employed	-3.1	-8.3	-7.2	
Ever employed	0.009	0.008	0.008	1,199
Earnings 2008/09	474	597	285	
Earnings 2005–09	-67	-19	-728	

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 B.4 Selection on unobservables

	ND25+	NDLP
All	Worse employment outcomes (<i>Lower earnings</i>)	
	Worse employment outcomes	
East Midlands	Worse employment	(<i>Worse employment</i>)
North East England	Worse employment Lower earnings	No selection
Wales	Worse employment	Worse employment Lower earnings
London	Worse employment	No selection
North West England	No selection	Better employment (<i>Higher earnings</i>)
Scotland	No selection	Worse employment Lower earnings

Note: in italics and brackets: weak evidence.

Appendix C

Matching diagnostics

Table C.1 Estimation of the propensity score

	ND25+	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
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	

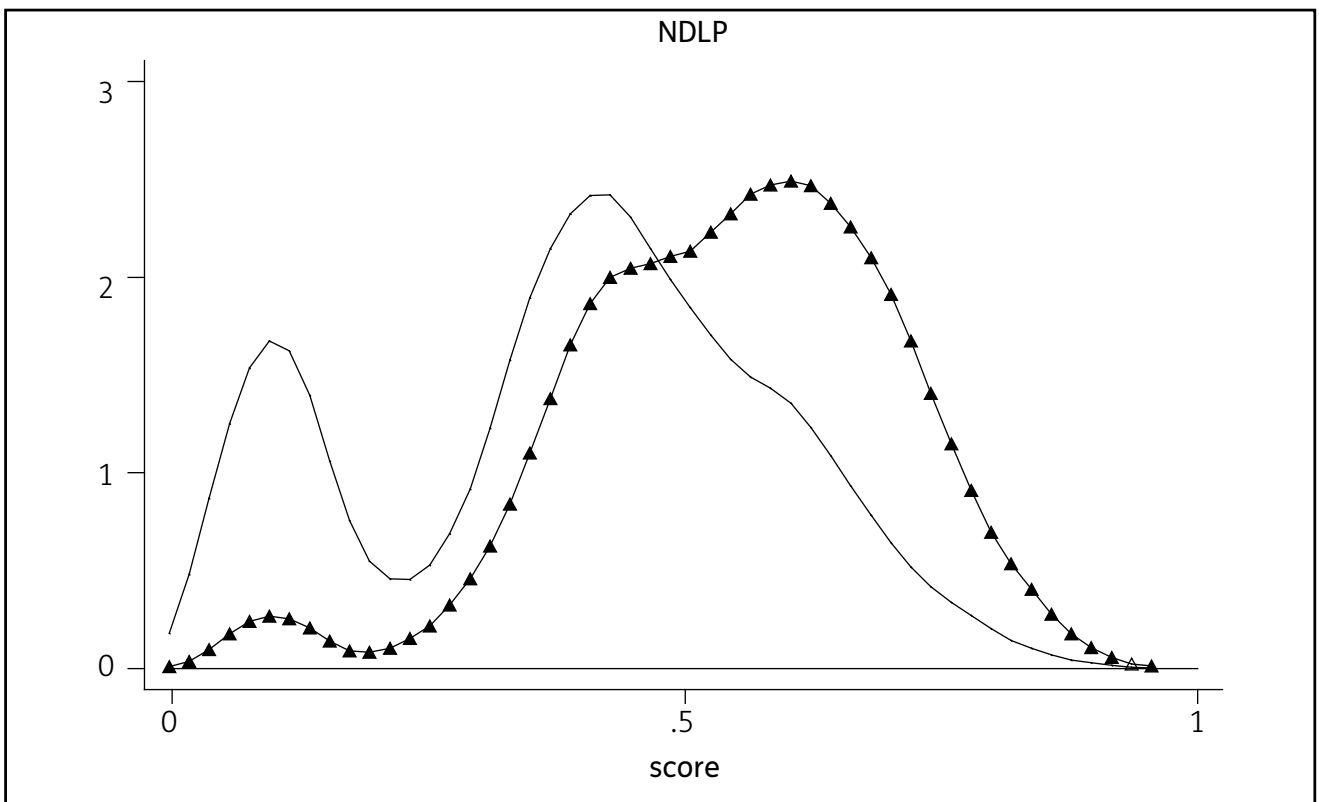
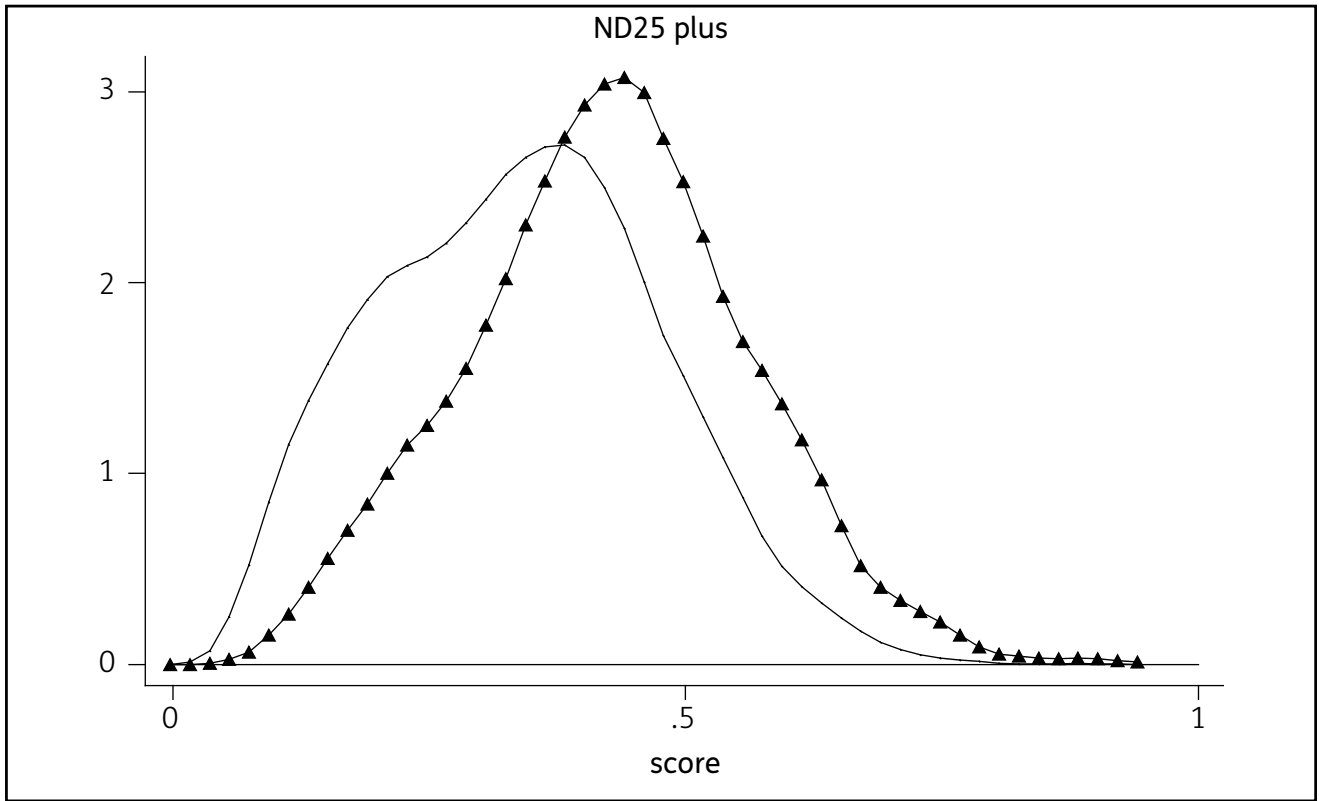
Continued

Table C.1 Continued

	ND25+	NDLP
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 C.1 Common support between non-participants and programme group:
Distribution of $P(Q=0 \mid Q=0 \vee R=1, X)$**



$Q=0$ denotes non-participants, $R=1$ the programme group.

Table C.2 Covariate balancing indicators before and after matching

	Prob>chi		Pseudo R2		Median bias		Percentage lost to 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 R2: 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 3 of the square root of the average of the sample variances in the two original groups.

Percentage lost to CS: Share of the group of non-participants falling outside of the common support.

Appendix D

Power and validity of the instrument

Table D.1 First stage of the instrument

	ND25+		NDLP	
	F- statistic	p-value	F- statistic	p-value
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
Sample size	7,796		7,261	

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

	share	p-value	(Pseudo)-R2
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 D.1.

Table D.3 Testing part of the exclusion restriction

	ND25+		NDLP	
	F- test	p-value	F- test	p-value
Earnings 2005/06	0.0	0.829	0.1	0.795
Earnings 2006/07	1.0	0.318	0.2	0.683
Earnings 2007/08	0.0	0.826	0.0	0.881
Earnings 2008/09	0.5	0.496	0.0	0.903
Cumulative earnings	0.1	0.802	0.1	0.807
Days employed (48m)	0.6	0.431	0.8	0.372
Ever employed (48m)	3.1	0.079	0.5	0.485

Note: Sample sizes: see Table D.1.

References

- Dolton, P., Smith, J. and Azevedo, J.P. (2008). *The Impact of the UK New Deal for Lone Parents on Benefit Receipt*. Mimeo.
- Dorsett, R., Campbell-Barr, V., Hamilton, G., Hoggart, L., Marsh, A., Miller, C., Phillips, J., Ray, K., Riccio, J., Rich, S. and Vegeris, S. (2007). *Implementation and first-year impacts of the UK Employment Retention and Advancement (ERA) demonstration*. Department for Work and Pensions Research Report No. 412.
- Heckman, J.J. (1979). 'Sample Selection Bias as a Specification Error', *Econometrica*, 47, 153-161.
- Heckman, J., Ichimura, H., Smith, J. and Todd, P. (1998). 'Characterising Selection Bias Using Experimental Data.' *Econometrica*, 66, 1017-1098.
- Heckman, J., LaLonde, R. and Smith, J. (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.
- 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., Hoggart, L., Marsh, A., Phillips, J., Ray, K. and Vegeris, S. (2005). *The Employment Retention and Advancement scheme – the early months of implementation: Summary and conclusions*. Department for Work and Pensions Research Report No. 265.
- Kamionka, T. and Lacroix, G. (2005). *Assessing the External Validity of an Experimental Wage Subsidy*. IZA Discussion Paper No. 1508.
- Sianesi, B. (2010). *Non-participation in the Employment Retention and Advancement Study: Implications on the experimental first year impact estimates*. Department for Work and Pensions Working Paper No. 77.
- Walker, R., Hoggart, L. and Hamilton, G., with Blank, S. (2006). *Making random assignment happen: Evidence from the UK Employment Retention and Advancement (ERA) demonstration*. Department for Work and Pensions Research Report No. 330.

The Employment Retention and Advancement (ERA) programme has been evaluated by random assignment. Over one-quarter of the eligible New Deal population was, however, not represented in the experiment.

The aim of this report is to explore how the four-year findings from the experimental research relate to the impacts that would have been experienced, on average, by all the people who were eligible for the programme, had they participated in the study.

If you would like to know more about DWP research, please contact:
Kate Callow, Commercial Support and Knowledge Management Team,
Upper Ground Floor, Steel City House, West Street, Sheffield, S1 2GQ.
<http://research.dwp.gov.uk/asd/asd5/rrs-index.asp>

DWP Department for
Work and Pensions

Published by the
Department for Work and Pensions
March 2011
www.dwp.gov.uk
Working Paper no. 96
ISBN 978-1-84712-936-9