



Designing a Demonstration Project An Employment Retention and Advancement Demonstration for Great Britain

Stephen Morris, David Greenberg, James Riccio, Bikash Mittra,
Hazel Green, Stephen Lissenburgh and Richard Blundell

2nd edition



Designing a Demonstration Project An Employment Retention and Advancement Demonstration for Great Britain

Stephen Morris, David Greenberg, James Riccio, Bikash Mittra,
Hazel Green, Stephen Lissenburgh and Richard Blundell

2nd edition

The report is the 2nd edition of No. 1, GCSRO Occasional Papers Series.

The views in this report are the authors' own and do not necessarily reflect those of the Cabinet Office.

Government Chief Social Researcher's Office

ISBN 0 7115 0445 7

Crown Copyright 2004

CONTENTS

Foreword	2
Executive Summary	3
Chapter 1 - Policy Design	9
Chapter 2 - Site Selection and Target Group Size	18
Chapter 3 - Process Study	25
Chapter 4 - Impact Study	33
Chapter 5 - Cost and Cost-benefit Analysis	69
Chapter 6 - Data Collection	79
Chapter 7 - Programme Cost Estimates	97
Chapter 8 - Bibliography	102
Annex 1 - The Justification and Evidence For Intervention	109
Annex 2 - Key Issues in Assigning Individuals at Random	131
Annex 3 - Measuring Non-experimental Impacts	139
Annex 4 - Description of ERA Demonstration Technical Adviser Role	143

FOREWORD

The Government Chief Social Researcher's Office, located within the Strategy Unit in the Cabinet Office, undertook to design a demonstration project to test new policies to help those on the margins of the labour market retain work and advance in employment. A demonstration project comprises a limited test of a new programme or policy that is subject to a multi-method evaluation, the centrepiece of which is an impact assessment conducted through random assignment (social experimentation).

This report outlines the outcome of the design phase, but acknowledges that the Department for Work and Pensions and Jobcentre Plus, which are responsible for taking this programme forward into its delivery phase, may have good reason to change certain aspects of the design. This project represents a thorough attempt to tackle some of the problems and difficulties associated with evaluating welfare-to-work policies in Great Britain, particularly as they related to the issue of random assignment.

The design process was a collaborate venture. The Design Project Steering Group consisted of representatives from Jobcentre Plus, The Department for Work and Pensions, Her Majesty's Treasury, Inland Revenue, Department for Education and Skills as well as the Cabinet Office. The design phase also enjoyed the support of an Advisory Group consisting of academics, government researchers and policy analysts with an interest in labour markets, employment policy and evaluation design.

The project team would like to thank members of both the Steering Group and Advisory Group for their help and support in facilitating and encouraging the design process. We would particularly like to thank Professor Ron Amann, Sue Duncan and Philip Davies based at the Cabinet Office, Mike Daly, Jane Hall and Geoff Scammell at the Department for Work and Pensions, James Richardson and Margaret McDonald from HM Treasury, Tony Manners, Jan Gregory and Jeremy Barton at the Office for National Statistics, Howard Reed and Mike Brewer at the Institute for Fiscal Studies, Lynda Pickering and Harsha Savani from Indepth Consultants, and finally the project director, Julia Gault, implementation and delivery manager, Nigel Hall and project officer, Paula Lydon, who worked tirelessly on the implementation and policy design, and without whom this project could not have taken place.

EXECUTIVE SUMMARY

This document presents a comprehensive research design for evaluating the *Employment Retention and Advancement (ERA) Demonstration Project*. The demonstration will test a new strategy for improving job retention and advancement for low-wage workers in Britain. It represents a potentially important step in strengthening the country's evolving welfare-to-work and anti-poverty policies. Although a number of current measures help those on the margins of employment retain work and improve their earnings, policy to date has largely focused on reattaching people to the labour market and makes less provision for assisting people once in work.

The ERA Demonstration also promises to make a significant contribution to the process of evidence-based policy making in Britain. It differs from many of the pilot projects seen recently in Great Britain, in that there are no plans at the outset to roll out ERA on a national scale before results of the evaluation are known. Moreover, the effectiveness of the intervention will be tested as a large-scale, multi-site, random assignment social experiment. Testing a cutting-edge - but unproven - policy innovation through a demonstration project utilising random assignment will help determine whether projects of this type can make a useful contribution to devising effective social policy in Great Britain, whatever the substantive policy focus.

As part of its role in encouraging excellence in policy, research, and evaluation design, the Cabinet Office led the effort to design both the intervention policy and the evaluation for the ERA Demonstration. The design team has worked in close collaboration with officials in stakeholder departments- Department for Work and Pensions (DWP), Inland Revenue (IR), Department for Education and Skills (DfES), and HM Treasury (HMT) (who were all represented on the project's Steering and Advisory Groups). DWP will lead the operational phase of the project through Jobcentre Plus and oversee its evaluation.

The ERA target groups and intervention

ERA will be directed at individuals in three different low-income groups known to have difficulty retaining jobs or advancing to better positions:

- those eligible for New Deal 25plus;
- those volunteering for New Deal for Lone Parents; and
- Lone Parents on Working Tax Credit (WTC) working part-time in low-wage jobs.

ERA will offer both pre- and post-employment assistance. For the two New Deal groups, the programme will begin *before* they enter employment; for the WTC groups, it will begin *after* they have started working. Once in ERA, all participants will have access to a combination of *work-related services* and *financial incentives* for a substantial period after employment commences.

Work-related services: One-to-one support for each participant from a dedicated Advancement Support Adviser (ASA) is at the heart of the ERA service. These advisers will attempt to identify and address the problems that are keeping participants from staying employed and advancing to better positions. The ASAs will start to work with the New Deal participants when those individuals are either mandated to join or volunteer for New Deal 25Plus, or volunteer for New Deal for Lone Parents. The ASAs will begin working with the WTC lone parents after they volunteer for ERA.

The ASAs will provide guidance on, or direct assistance with, such issues as: finding a job or a better job; gaining promotion; understanding how much work, or an increase in hours and wages, can 'pay' in terms of net income; dealing with workplace demands and pressures; finding appropriate education and training opportunities that can be combined with part-time or full-time employment; and arranging for support services such as childcare or assistance with personal problems or family circumstances that can impede steady employment. The ASAs will work with participants for a maximum of 33 months. For the New Deal target groups, this allows for up to nine months of pre-employment services and a minimum of 24 months of post-employment assistance.

Financial incentives: The ERA financial incentives are intended to encourage retention in full-time work (which the literature suggests is a better route than part-time work to an eventual escape from poverty – see Annex 1) and the accumulation of skills through advancement-focused training. Thus, the incentives include bonus payments for participants who work full time (i.e., at least 30 hours per week) and for those who combine education or training with work.

Full-time workers will be eligible to receive up to six bonus payments of £400 each (£2,400 maximum) over the course of the demonstration. To be eligible for each payment, a participant must work at least three months out of a set four-month period. Participants can also qualify for up to £1,000 in tuition assistance and a maximum £1,000 training bonus if they take part in approved education or training courses while employed. The training bonus, which will be available only upon successful completion of an approved course, will be paid at the rate of £8 per classroom hour.

Random assignment design

In the ERA Demonstration, random assignment means that individuals eligible for the programme will be given an equal random chance of being assigned to a programme or control group. Those assigned to the control group will *not* have access to ERA services or incentives for the duration of the demonstration. However, they will continue to be eligible for all currently available non-ERA assistance (e.g., New Deal and Jobcentre Plus services and the WTC). As shown in Figure 1 in Chapter 4 of this report, individuals in the New Deal target groups, who will not be employed when they enter the study, will be randomly assigned at the point that they would normally enter the New Deal. In contrast, as depicted in Figure 2, members of the WTC target group, who will already be working, will be randomly assigned after they volunteer for ERA.

Testing ERA through a random assignment research design will provide the most convincing evidence of its quantitative impact on important outcome measures. This is because, when properly executed, random assignment ensures that there are no measurable or unmeasurable pre-programme



differences (e.g., higher motivation, greater skills, better access to opportunities), on average, between the programme and control groups; thus, any subsequent differences in outcomes between them can confidently be attributed to the programme. For example, the ERA evaluation will compare the average number of weeks worked after random assignment among members of the programme group to the average duration of employment among members of the control group. Because the only systematic difference between the two groups will be that the former received ERA services, any difference in the amount of employment between them can be traced directly to ERA.

The ERA Demonstration offers the chance to use a large-scale random assignment social experiment as a tool for policy development. It is, therefore, an objective of this project to test the usefulness of this approach with an eye toward applying it in other carefully constructed pilot evaluations of social policy.

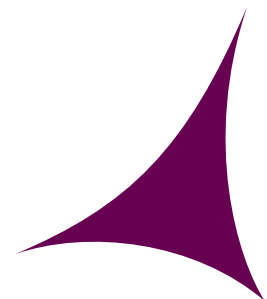
Components of the evaluation

Among the key purposes of the evaluation of ERA are to learn:

- whether the new measures cause improvements in employment stability and advancement above and beyond what would have occurred in their absence;
- what it takes to implement the new measures well; and
- how any economic benefits they generate compare to their costs.

To answer these questions, a comprehensive evaluation design is proposed that includes the following components:

1. An impact study (Chapter 4) will determine the effects of the ERA programme on outcomes related to participants' employment, such as job-entry rates, employment duration, earnings, wage growth, job quality, total income and poverty rates, and on a variety of non-economic outcomes of interest, such as personal and family quality of life, material hardship, and outcomes for children. These outcomes will be measured over a five-year follow-up period for members of the randomly selected programme and control groups (see above). The difference between the groups on each outcome will indicate the programme's effect, or 'impact'. The study will estimate ERA's impacts separately for each target group within each demonstration site where ERA is implemented (see below), and for all of these sites combined. It will also examine impacts for key subgroups within the three target groups (for example, subgroups with different levels of educational attainment before random assignment). The analysis will rely on administrative data on employment, earnings, and WTC receipt maintained by Inland Revenue; DWP administrative records data on transfer benefit receipt; and a multi-wave client survey. The survey will be conducted at 12 and 24 months after entry into the study, and, if response rates are adequate, a five-year follow-up survey will also take place. If response rates were to be inadequate, an evaluation of outcomes after five years could be done using administrative data.
2. A process study (Chapter 3) will determine how the ERA model is adapted to local conditions in each of the selected Jobcentre Plus agencies, and how it differs in practice from what is offered to



New Deal participants and users of Jobcentre Plus. It will also measure how much ERA participants actually take advantage of ERA-provided services and financial incentives as well as any other employment-related services they can access on their own. Importantly, the study will compare the extent of participation in all such activities between the programme and control groups in order to determine whether the ERA causes an increase in skill-building efforts that might contribute to advancement in the labour market. The process study will also examine the kinds of problems that impede participants' steady employment and progression in the labour market, and how staff attempt to respond to these challenges. Data for this study will come primarily from qualitative interviews with ERA and Jobcentre Plus staff, a quantitative staff survey, participation tracking data maintained by the DWP and Jobcentre Plus, a subset of questions in the client survey that focus on service receipt and use of financial incentives, and qualitative interviews with a sample of employers.

3. A cost study (Chapter 5), will determine how much it costs per person to operate ERA and its specific elements (e.g., caseworker services and the financial incentives). The study will rely on DWP expenditure data for Jobcentre Plus and the ERA programme, published and unpublished unit cost information on other education, training, and support services, and programme tracking and client survey data on clients' receipt of ERA, Jobcentre Plus, and other available services.
4. A cost-benefit study (Chapter 5), using data from each of the other studies, will determine the net economic gains and losses generated by ERA over a time horizon of five years or more. The study

will focus particularly on how much participants (who stand to benefit from the programme) and the government budget and taxpayers (who will pay for it) gain or lose financially. Net gains or losses from the perspective of employers and for all these groups combined (i.e., a societal perspective) will also be computed.

Overall, the ERA evaluation will draw on the findings from each of these four studies to provide an integrated body of evidence on the operations, effectiveness, and economic consequences of the ERA programme. For example, the impact study will show whether the programme generates any substantial 'added value' in retention, advancement, and quality-of-life goals compared to what would have been achieved in its absence—that is, by the existing provision of New Deal, Working Tax Credit, and other services and supports for work. The cost-benefit analysis will show the potential economic 'return' on the government's investment in the ERA programme, and the net cost of achieving desirable effects, in addition to retention and advancement (e.g., poverty reductions and improvement in family and child quality-of-life outcomes). If, based on these analyses, Ministers are inclined to replicate ERA nationally, the findings from the process, impact, and cost studies can offer guidance on what programme strategies should be emulated—or avoided—and whether the programme in general ought to be targeted toward certain groups in order to enhance the chances that the replicated version will be well-run, effective, and cost-efficient. Alternatively, if the impact and cost analyses show that the programme does not work as well as expected, or is too costly, the process study may help point to the reasons why and, possibly, how the programme might be improved. In sum, the different components of the evaluation, when



taken together, will offer policymakers and administrators a firm basis of information for deciding whether to institutionalise the programme as a regular feature of Jobcentre Plus nationally, and if so, how best to do this.

Proposed ERA sites

The ERA programme will be set up as a special unit within selected Jobcentre Plus Districts and evaluated as part of a multi-site social experiment. The project design team recommend that six areas (Jobcentre Plus Districts) be included in the demonstration. This number of sites is considered to be a practical number from a cost and management standpoint, and one that would allow the evaluation to determine whether the ERA programme can be operated successfully and achieve positive impacts on participants when operated under a variety of local conditions and by different staff. This test of the programme model's 'robustness' is important for gauging the possible effect of a wider rollout of the new measures, which would be of interest to policymakers if the results of the evaluation were positive. Multiple sites will also provide some opportunities to learn about the 'best practices' for operating an effective ERA programme.

With these considerations in mind, the following criteria were used in identifying specific locations to include in the demonstration:

- Subject to other criteria, the number of potential clients in the three target groups should be sufficiently large to permit reliable site-specific estimates to be made of the programme's impacts on each of those groups.
- The number of potential clients from an ethnic minority background should be sufficiently large, for at least the New Deal 25plus group, across the sites, to permit reliable estimates to be made of the programme's impacts on that key subgroup, although it is very unlikely that the number of ethnic minority customers will be sufficiently large to allow measurement of impacts on individual ethnic groups.
- The socio-demographic profiles, labour market conditions, and other local contextual factors across the six sites should reflect some of the diversity in these conditions across Great Britain.
- The six sites should be spread geographically across Great Britain and include locations in England, Scotland and Wales.

The proposed sites are listed in Table 1.

Table 1: Proposed Demonstration sites

Area	Region ⁽¹⁾
Renfrewshire, Inverclyde, Argyll and Bute	Scotland
South East Wales	Wales
Manchester	North West
East London	LASER ⁽²⁾
Gateshead and Tyneside	North East
Derbyshire	East Midlands

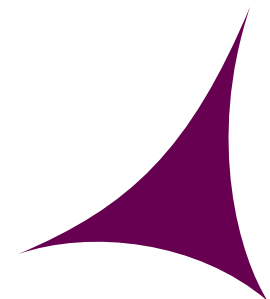
Notes:

(1) Government Office Regions

(2) London and South East region

These districts were selected from a list of 25 candidate areas where fully integrated Jobcentre Plus offices¹ are currently, or will soon be, operating. The project design team sought to avoid areas that are likely to be implementing the fully-integrated Jobcentre

¹ For a description of Jobcentre Plus see <http://www.jobcentreplus.gov.uk/cms.asp?Page=/Home/AboutUs>



Plus model—a substantial administrative reform—at the same time that the ERA programme would need to be implemented. For that reason, consideration was given only to the 25 districts due to have been operating the Jobcentre Plus model for at least six months before the scheduled ERA programme start date of October 2003.

Sample sizes

Recruitment and intake for the ERA programme would last for about one year. It is estimated that during this time, almost 27,000 individuals across these six locations and three target groups could be enrolled in the study sample and randomly assigned to the programme or control group. Half the sample (nearly 13,500) would be assigned to each research group. The total research sample in each site would range from about 2,200 individuals (1,150 in each research group) in South East Wales to about 6,600 individuals (approximately 3,300 in each research group) in Derbyshire.

Administrative records (e.g., earnings and benefit payment data) will be collected on all individuals who are randomly assigned. However, the client surveys will be administered to a representative subsample of these groups, with the total number of respondents projected to be approximately 4,800 across the six demonstration sites for the survey at 24 months.

The final decision on site selection will be made by the DWP and will be one of the crucial early decisions of the implementation phase. In addition to the research criteria specified above, DWP will need to consider each area's management capacity and administrative potential for operating a high quality ERA programme—one that is true to the

programme design, can be implemented on the timetable required by the demonstration, and that merits a rigorous evaluation.

A further consideration is the location of other pilot programmes at the sites.

Timetable

Programme intake is envisaged to begin in October 2003 and continue through to September 2004. All ERA services will cease by the end of June 2007 (33 months after the end of intake in September 2004).

Results from the random assignment impact study could emerge as early as spring 2005. These early findings will be based on administrative data and constitute limited and tentative comparisons of employment rates and levels of benefit claiming. Some partial information may also be available at this time on earnings. Earlier analysis than this of administrative data will be indicative of the general progress of the demonstration, for examples, whether the numbers coming forward and the numbers entering work are within expected ranges, but it will not provide any estimates of net impact.

The first comprehensive survey results should be available by late-summer 2006. In noting these provisional reporting milestones, it is important to bear in mind that retention and advancement are phenomena that can only be measured successfully over the medium to long term. Ideally, outcomes would be measured two to five years after a reasonably large cohort of individuals have entered and been through the programme (i.e., from autumn 2006 for administrative data, and from late 2007 for surveys).

CHAPTER 1 – POLICY DESIGN

The ERA Demonstration aims to test a new policy to help those on the margins of the labour market retain work and advance. The policy combines new and existing services with financial incentives in order to achieve these goals. The underlying rationale involves the extension of services for up to two years to individuals who are either already employed in low-paying jobs, or who have newly entered work from benefits.

This chapter examines the nature of the services to be provided and outlines the groups at which the programme is targeted. In doing so, a clear indication is given of whom the new policy is aimed at and the types of interventions these individuals will receive, thereby describing what it is that the ERA Demonstration is to evaluate.

Target groups

The ERA Demonstration aims to test the effectiveness of services delivered to three groups:

- those eligible for the New Deal for Long-term Unemployed (ND25plus) – this group will include both those who are required to join ND25plus and those who volunteer for the programme;
- those who choose to enter the New Deal for Lone Parents (NDLP); and
- lone parents working part time and claiming Working Tax Credit (WTC).

We look at each of these groups in turn.

New Deal for long-term unemployed

The first group of individuals at which ERA services will be targeted are those who become eligible for ND25plus, both on a mandatory and voluntary basis. As explained below, the Demonstration aims to compare the effectiveness of ERA services with those provided through ND25plus. It does this through diverting, at random, a fraction of the target group into the ERA programme. This allows for a fair comparison of ND25plus with ERA. The random ‘diversion’ or ‘assignment’ takes place when individuals would normally qualify for entry to ND25plus and only in those areas (experimental sites) where ERA services are being tested.

Although there are some individuals who can volunteer for ND25plus, for the majority of eligible individuals’ participation in ND25plus is mandatory. This includes those aged 50 or over, for whom the Gateway period of the ND25plus programme, described below, is mandatory – although the Intensive Activity Period also described below, is not. The intention is that DWP/Jobcentre Plus will put regulations in place to give ERA Advancement Support Advisers comparable powers of compulsion to New Deal advisers. The assumption, therefore, is that participation in the ERA Demonstration is also mandatory for non-volunteers in this group. By implication, individuals required to join the ND25plus cannot therefore object to being randomly assigned. They will, however, along

with the other two target groups, have to give consent to be interviewed and possibly also to allow their administrative records to be analysed as part of the evaluation. If individuals refuse to consent to this, they will not be randomly assigned and therefore will *not* be part of the study, but will enter the New Deal as normal. The hope is that the possibility of receiving financial incentives through the ERA Demonstration, and of receiving payment for participation in surveys (if this measure is adopted), will mean that only a very small number of individuals will refuse to take part in the research (for more detail on payment for participation in surveys see Chapter 6 of this report).

ND25plus provides a range of services for individuals who are registered unemployed and claiming Jobseeker's Allowance, and have been seeking work for 18 months or more over a 21-month period. The programme involves three main programme stages: the 'Gateway', the 'Intensive Activity Period' (IAP) and 'Follow Through'. As the experimental research design involves comparing the effectiveness of ND25plus with ERA services, it is important to outline briefly what the ND25plus programme consists of.

ND25plus commences with a period of treatment referred to as the Gateway. All individuals entering the Gateway are screened for basic skill deficiencies, such as numeracy or literacy problems, as well as for English language needs. In essence, the Gateway comprises a series of 'job-focused' interviews with a personal adviser that can last for up to four months. The objectives of these interviews are to: 1) agree a plan for getting work; 2) provide help for finding work; 3) address barriers to work; and 4) identify any additional help such as specialist support services needed for debt counselling, career advice and so forth.

If an individual who has been through the Gateway remains unemployed, they next enter the 'Intensive Activity Period' (IAP). The IAP provides a mix of help tailored to individual needs and drawn from four main options:

- self-employment support;
- Basic Employability Training (BET);
- Education and Training Opportunities (ETO); and
- flexible packages of support combining work experience, work placements, work-focused training and help with motivational and soft skills.

IAP lasts 13 weeks, although it can be extended for another 13 weeks in some cases and, in rare instances, can last for up to 52 weeks in total. Those in the 25–49 age group can face benefit sanctions if they fail to participate in IAP.

Finally, those who remain without work after IAP enter what is called the 'Follow Through' phase. This involves additional help and support from a New Deal Personal Adviser building on the skills and experience gained during the IAP phase. 'Follow Through' lasts for up to six weeks but can be extended to 13 weeks in total.

From April 2003, both those with and without children in the ND25plus group who enter full-time work (defined as working greater than 30 hours per week) will qualify for Working Tax Credit (HM Treasury 2002). Those receiving ERA services will also qualify for Working Tax Credit upon obtaining full-time work and, as described below, will also receive additional incentives over and above their Tax Credit receipts.



In summary, members of the ND25plus target group will be:

- registered unemployed and claiming Jobseeker's Allowance for at least 18 months over a 21-month period (although there will also be some individuals who volunteer for ND25plus before 18 months);
- live in one of the six ERA experimental sites (see Chapter 2 for further details).

New Deal for Lone Parents

NDLP is a voluntary programme that aims to help lone parents find work. Any lone parent with a dependent child, out of work or working less than 16 hours per week can participate. Half of those who choose to enter NDLP at the six experimental sites where ERA will be operating will be assigned at random to receive ERA services instead. The important point is that only lone parents who decide to join NDLP will be offered the chance to enter ERA.

The focus of NDLP is getting individuals job-ready and addressing any barriers to work they may have. The programme provides access to a Personal Adviser who helps to identify training and skills needs and provides help with job search and finding appropriate child-care, as well as assistance with claiming in-work benefits such as the Working Tax Credit (from April 2003). One of the key features of the programme is the provision of a better-off calculation, which shows clients whether they are likely to be better off in a specific job compared to remaining on benefit.

Currently, lone parents who enter work after participating in NDLP can claim Working Families' Tax Credit, subject to certain conditions. From April 2003, this will be replaced by two credits –

the Working Tax Credit (WTC) and the Child Tax Credit (CTC) (HM Treasury 2002). The CTC does not require lone parents to be in work and is means tested – the value of the credit falls as income rises. Lone parents assigned to receive ERA services will be eligible to receive WTC and CTC as well as additional financial incentives through the ERA Demonstration.

In a summary, members of the NDLP target group will be:

- a lone parent with at least one dependent child;
- out of work or working less than 16 hours per week;
- have chosen to participate in NDLP; and
- live in one of the six ERA experimental sites.

Lone Parents working part time and claiming WTC (WTC LPs)

The final group, for which ERA services will be tested, are lone parents currently working part time, that is, for more than 15 but less than 30 hours a week, and claiming Working Tax Credit. Unlike the two New Deal eligible target groups discussed above, ERA services will be offered to the stock of claimants that meet the inclusion criteria, as well as to the flow of new claims during the ERA intake period.

To summarise, an individual is eligible in this case if they are:

- living in an area designated an experimental site;
- have a live Working Tax Credit claim;
- are a lone parent with a least one dependent child; and are
- working at least 16 hours but less than 30 hours per week.

Eligible individuals must have a live WTC claim at a specified date prior to the beginning of the ERA Demonstration intake period, which is scheduled to run for 12 months from October 2003. In addition, any new claim for WTC lodged during the intake period that meets these conditions will also be eligible to enter the ERA Demonstration.

WTC LPs who meet these criteria will be approached during the intake period by ERA Demonstration WTC Recruitment Officers and encouraged to take part. Taking part will be optional. Of those who agree to enter the Demonstration, half will be assigned to a control group and continue to receive WTC. The other half will receive ERA services and the ERA financial incentives in addition to WTC.

Employment retention and advancement services

This section outlines the services to be tested through the ERA Demonstration. Services can be thought of as comprising two broad components – caseworker services provided by an Advancement Support Adviser and financial incentive payments.

Advancement Support Adviser

The ERA Demonstration Project caseworker – Advancement Support Adviser (ASA) – will help the customer devise an Advancement Action Plan (AAP) and take the steps to help the customer achieve the goals outlined within the AAP.

Those who are jobless when they are assigned to the ASA will be given job-search help using a ‘step down’ approach – looking for the best job match to start with but quickly moving down to less optimal options if that is unsuccessful. The initial job will be the first stepping stone in their advancement plan.

The ASA will continue to support the individual (for up to two years) while they are in work to help them make further progress. This extended support is a unique feature of ERA and will be important in helping the ASA maintain an effective relationship with the customer. The support will include help in finding new work if the customer loses their job, or is working part time but would like full-time work, or is ready to advance.

Advancement Action Plan

The AAP will focus on the steps needed to help the customer work steadily and then progress in employment. It will be developed in consultation with the customer. The AAP will be a ‘living document’ containing short, medium- and long-term goals and will be revised subject to the customer’s progress and other changes in circumstance. The AAP may include agreed actions for the ASA, as well as actions for the customer. Depending on the customer’s needs, the plan can include:

- action related to job search (researching local vacancies, CV writing, making applications, etc);
- action related to overcoming practical barriers (researching local childcare provision, researching transport arrangements, etc);
- action related to human capital building (training, exploring training opportunities with current employer, researching for non-work opportunities to develop and practice new, job-relevant skills, etc).

To ensure that the customer’s AAP is grounded in the reality of the local job market, the ASA will work with Jobcentre Plus Local Account Managers (who are responsible for identifying local job opportunities and liaising with local employers) to provide information on employment opportunities for ERA customers.



This information will need to include details covering local career ladders and the skill credentials valued by local employers for posts above entry level. For the ND25plus group, non-compliance with agreed AAP actions would attract sanctions for those who, in the normal course of events, would be mandated to join the ND25plus programme.

Accessing other support

The ASA will also help the customer access local support to address practical barriers that stand in the way of the customer's employment retention and advancement. For example, the ASA may bring in the Jobcentre Plus Childcare Partnership Managers to help participants develop robust childcare arrangements, or may facilitate contact with the local Citizens Advice Bureau to help with financial/debt advice.

Intensive pre-employment support

The ASA will provide intensive pre-employment support for up to nine months; after that time, if the customer has not obtained work, the AAP will be revised to place more responsibility on the individual to find suitable employment quickly. In effect, the period of intensive pre-employment support will be equivalent to ND25plus' Gateway and IAP combined, with the added flexibility of no particular time limit. If, for example, the ASA thought that it would benefit a customer to be on the equivalent to Gateway for nine months, this would be allowed. Conversely, if they thought it appropriate to send a customer on IAP type provision from day one of the customer entering ERA, this would also be permitted. This pre-employment support could also include training courses lasting up to 12 months.

The incentive to enter full-time work after nine months in the ERA programme will be reinforced by the availability of the retention and advancement bonus (see 'Financial Incentives' section below). To qualify for the maximum six payments of this bonus, the customer will need to enter full-time work within nine months of entering the ERA programme. ND25plus customers who fail to obtain work after 27 months will remain in the ERA programme for a further six months so that they receive the maximum 33 months of ERA service possible but will revert back to intensive pre-employment support for this six month period.

Learning to continue advancement

During the in-work period, participants will be encouraged to develop the skills to continue working on their own advancement after the ASA's support has stopped. This may include coaching from the ASA and workshops with other ERA customers.

Providing services to those in work

As ASAs will offer support to customers who are in work for up to two years, ASAs will need to be available to meet with customers outside standard working hours. This has implications for the organisation of the work. Services available from 7am to 9pm should provide sufficient opportunity for most customers to make contact at a time that fits in with their work and other responsibilities. Consideration should also be given to offering services at the weekend – perhaps on Saturday mornings.

To provide sufficient cover for these extended hours of operation, ASAs will work in teams. Customers should have one ASA as their main point of contact, who will arrange meetings with the customer at a mutually convenient time. The need for ASA support may, however, arise outside those planned meetings. If a customer calls the ERA office

and their ASA is unavailable, another ASA in the team will be able to help with their case. ASAs will need to have time to discuss individual cases regularly to ensure the effectiveness of this 'team-working' approach.

Financial incentives

The ASA will also be responsible for administering the two new financial incentives available under ERA: the retention and advancement bonus and the training bonus. They will be responsible for explaining the eligibility rules for these incentives, checking entitlement, and authorising payment.

As previous experience has shown that caseworkers can have problems engaging with participants who are in work, the two new financial incentives available to ERA customers are structured to encourage both:

- customer behaviour most likely to result in retention and advancement; and
- continued customer contact with the ASA.

Retention and Advancement Bonus

As employment retention is a pre-requisite for advancement, there should be a financial incentive to encourage retention. In addition, there is evidence that part-time workers are likely to experience more difficulty achieving advancement (see Annex 1), partly through the more limited availability of opportunities, but also because employers are less likely to offer them training opportunities. Thus, this bonus should be limited to the retention of full-time work. In this way, the bonus partially offsets incentives under the WTC that tends to encourage part-time work.

The detailed considerations that underpin this bonus design are set out in this section. In summary, to encourage retention and advancement:

- a retention and advancement (R&A) bonus should be payable only for full-time work (of at least 30 hours per week);
- ERA customers would qualify for the R&A bonus by working for a minimum of any 13 weeks in a set 17-week period, which starts from the day they first enter work;
- the R&A bonus is paid at a rate of £400 per 17-week period;
- entitlement is to be assessed and paid every 17 weeks by the ASA, based on evidence provided by the customer;
- if the customer loses their job and becomes ineligible for this bonus within that 17-week period, a new 17-week period will start when the customer begins a new job. This provision maximises the incentive to re-enter and sustain work.

A key aim of the R&A bonus is to encourage ERA customers to develop the habit of steadier work patterns. The plan is designed to obtain the optimum balance between a structure that is easy to understand and administer, and one that maximises the reward for behaviour that is beneficial to the customer.

The R&A bonus is intended to reinforce the other elements of the policy package offered to customers under the ERA Demonstration Project. To encourage customers to attend face-to-face contact with their Adviser (ASA), the ASA will be responsible for authorising payment. To encourage customers to meet with their ASA a minimum of once every four months, the R&A bonus payments



will be assessed on a pattern of rolling 17-week periods and customers cannot receive payment, even if they are eligible in all other respects, unless they attend scheduled meetings with their ASA.

In order to encourage steadier work patterns, customers can qualify for a bonus payment if they work any 13 weeks during the 17-week period. The simpler option of making the payment dependent on working 13 consecutive weeks would have the serious disadvantage of not rewarding reasonably steady work patterns. For example, an individual laid off after 12 weeks, who found work within two weeks, would not qualify for a payment, even though their work pattern was reasonably steady. In all probability, the individual would seek to hide the short period of unemployment from the ASA, which would work against developing an appropriate relationship of trust between them.

To address this problem, payment can be received if any 13 weeks over the 17-week period is spent in full-time work. This would not just encourage customers to retain their first job, but also to seek quick re-employment (within four weeks) if this first job ended.

To ensure that a customer who is unemployed for more than four weeks in any 17-week period retains the incentive to seek quick re-employment, a new 17-week assessment period will start on the first day of their new job. The four-monthly face-to-face meetings will not be the only contact between ASA and customer as the ASA will place special emphasis on contacting the customer while they are unemployed. Thus, the re-setting of the 17-week period should not interfere unduly with the cycle of meetings between them.

Although paying the bonus at a weekly rate, dependent on the number of weeks worked in the 17-week period was considered, it appeared to be simpler and more cost-effective to pay the bonus at a flat rate for the successful completion of a total of 13 weeks of full-time work in the 17-week period. In determining the level of bonus that might produce the most cost-effective result, account was taken of micro-simulation findings produced by the Institute for Fiscal Studies as well as what might be financially sustainable for the taxpayer should the policy be considered for national roll out. The conclusion was that the R & A bonus be paid at a flat rate of £400 per 17-week period. Thus, no extra bonus would be earned for working more than 13 weeks any 17-week period because it is anticipated that most customers who have worked the first 13 weeks of a 17-week period will continue to work the remaining four weeks.

It is not the intention that the R&A bonus should be a permanent addition to the customer's income. The incentive will be provided for a limited period to support the development of new working habits. During the ERA Demonstration Project, customers will be offered a maximum of 24 months of in-work support, during which time the R&A bonus will be available to them. By working steadily (13 weeks in 17), for an average of 30 or more hours a week, for the full 24-month period, ERA customers can qualify for the maximum of six R&A bonus payments (i.e. a total of £2,400).

In order to authorise payment, the ASA will need evidence that the customer has worked full time (an average of at least 30 hours a week) for at least 13 weeks in a 17-week period. However, existing evidence such as payslips, may not contain all the necessary

information, especially on hours worked. To avoid placing an additional burden on employers to certify hours worked, the customer will be required to make a signed statement of the number of weeks of full-time work completed in the 17-week period and provide payslips to support this. The ASA will then make simple checks to verify this statement – e.g. checking that the implied average hourly wage rate is at least as high as the minimum wage. A small percentage of client statements will be checked with employers. This is similar to the approach that worked effectively in administering a welfare-to-work financial incentive programme in Canada (the Self Sufficiency Project).

Training bonus

As lack of human capital can be an important barrier to achieving employment advancement, ERA includes a training bonus to encourage ERA customers who are working to improve their human capital by participating in training linked to their AAP. Like the R & A bonus, the training bonus is intended to reinforce ASA support, giving the ASA a tangible way of encouraging customers to commit to training that is not provided by the customer's employer, where such training is needed to help the customer meet their longer-term aspirations. To qualify for payment, the training must be work-related and successfully completed. Moreover, the customer must be working 16 hours or more a week to qualify for the bonus.

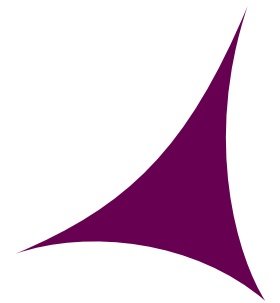
The ASA will, in addition, be able to pay tuition fees of up to £1,000 per customer, where funds for this are not available from another source. This £1,000 is a maximum amount and, if the customer takes a number of courses, the combined tuition fees cannot exceed this limit. In addition, the ASA will

be able to use their discretionary fund to contribute towards other incidental expenses such as travel costs or course materials relating to the agreed training.

To be effective, the training bonus should aim to reward proportionately the effort and the loss of time for other activities that training entails. It is difficult to envisage a simple mechanism that could directly reward effort, as it will vary between individuals undertaking the same training. It has been suggested that a two-tier system (say £500 and £1,000) might give the ASA some flexibility to do this. The ASA would need detailed guidance, however, about how to decide which level of bonus to apply and would need to take into account such factors as the length of the course. It appears easier to base the payment on a simple proxy for effort – such as the course length – to avoid developing complex guidance and to ensure that discretion is being exercised reasonably, and broadly similar decisions are being made in similar circumstances.

Thus, the ERA training bonus will simply be paid at a rate based on the hours of classroom time involved. This information is readily available. The ASA and customer will discuss the time commitment entailed before the training becomes an agreed part of the AAP. The customer will be expected to research the training opportunities and determine (e.g. through prospectuses) the course length. The customer will demonstrate successful completion through showing the ASA a qualification certificate or certificate of course completion.

On the assumption that for every hour of classroom time, the customer will need to put in an hour of home study, the training bonus will be paid at the rate of £8 per classroom hour. The £8 an hour is broadly



twice the minimum wage rate – rounded down for ease of administration. If the individual does not use their allotted £1,000 training bonus entitlement on one course, the remainder will remain available for payment for any additional training that is taken up during the 33 months the customer can participate in ERA. As in the case of tuition fees, however, the training bonus will be capped at £1,000.

Participants, who retain full-time work and take up training as well, would qualify for both bonuses. Participants who work part-time can qualify for the training bonus, but not the R&A bonus.

Both bonuses will be paid via the ASA. Participants should have regular meetings (at least every four months) with their ASA to review progress against their AAP. Successful completion of a training course is also a natural point at which next steps should be considered. Making payment of both bonus payments dependent on customers' meetings with their ASA should give participants an additional incentive to attend.

CHAPTER 2 – SITE SELECTION AND TARGET GROUP SIZE

In this chapter, six experimental sites where the ERA Demonstration might be tested are discussed. The criteria used to select the proposed experimental sites, their locations, and estimates of the size of the three target groups at each site are presented. Selecting a limited number of specified sites allows evaluation of new ERA services and financial incentives without incurring the cost of introducing them across the entire country. The ideal situation would be to select a large number of experimental sites at random throughout Great Britain, as this would provide the best basis on which to generalise findings from the evaluation nationally. However, due to budgetary and practical constraints, this option is not feasible.

Each of the chosen six experimental sites is equivalent to a Jobcentre Plus District. The aim of the demonstration is to compare ERA services with those provided through non-ERA Jobcentre Plus. For these reasons, the choice was restricted to districts that are due to have been operating the new Jobcentre Plus service delivery model for at least six months before the scheduled ERA start date of October 2003, because they should be relatively stable in administrative terms by the time ERA commences. According to the Department for Work and Pensions' (DWP) current rollout schedule, there are 25 Districts that will have introduced Jobcentre Plus by April 2003, and thus satisfy this criterion. The six Districts in Table 2 were chosen from among these 25.

Table 2: Experimental sites for the ERA Demonstration

Area	Region ⁽¹⁾
Renfrewshire, Inverclyde, Argyll and Bute	Scotland
South East Wales	Wales
Manchester	North West
East London	LASER ⁽²⁾
Gateshead and Tyneside	North East
Derbyshire	East Midlands

Notes:

(1) Government Office Regions

(2) London and South East region

Selection criteria

In addition to the need to avoid Districts undergoing substantial administrative change, there were a number of other criteria applied to the selection of experimental sites. These are set out below.

Criterion: The number of customers that are expected to enter ND25plus and NDLP

In choosing these six areas, an overriding consideration was the number of customers, according to DWP projections, that are expected to enter the ND25plus and the NDLP in the areas in 2002-2003. Particular importance was attached to this factor because the feasibility of detecting ERA programme impacts on a site-specific basis will depend partly on there being a relatively large number of customers available for random assignment. The numbers entering



ND25plus and NDLP took precedence over the eligible numbers likely to be found in the WTC target group, because the latter comprises a far larger potential sample than the former two groups. Although the number of potential customers in an area is of considerable importance, the need to obtain relatively large sample sizes during the 12-month intake period per area was moderated by the requirement to select areas in Scotland and Wales, where available areas tend to be smaller than ideal.

Criterion: The proportion of ND25plus and NDLP entrants from an ethnic minority background²

One factor to which considerable weight was attached was the proportion of ND25plus and NDLP entrants who were from an ethnic minority background. This is obviously an important factor in view of policy considerations, but another reason why it was seen as particularly important for site selection is that the number of ethnic minority entrants cannot be altered through data manipulation. With other subgroups of interest, such as level of education, work history, benefit record, age of youngest child, age and partnership status, there is either some scope to vary the cut-off point so that there are large enough numbers in the desired categories, or the distribution is sufficiently balanced that sample numbers are unlikely to be a problem. Manchester and East London were chosen as sites, in part, because they are projected to have large numbers of ND25plus and NDLP entrants from ethnic minority backgrounds.

Criterion: Region

To attempt to be representative of Great Britain geographically and politically, it was seen as necessary to have one area from Scotland, one from Wales, and no more than one from any English region. Among the 25 Jobcentre Plus Districts under consideration, only two were from Scotland and only two from Wales. This obviously meant that the choice of experimental sites from these regions was severely restricted. Under the circumstances, it seemed most sensible to select the available Scottish and Welsh districts with the largest numbers of ND25plus and NDLP entrants. These were Renfrewshire, Inverclyde, Argyll and Bute in Scotland and South East Wales. However, South East Wales and Renfrewshire, Inverclyde, Argyll and Bute are likely to have too few individuals in the NDLP target group (see Table 3 below). Careful monitoring of the build-up of sample sizes in all areas will be required during the programme intake period, but particularly in those areas where samples are expected to be relatively small. If the sample proves to be insufficient for the purposes of detecting programme impacts in a particular area, the programme intake period may have to be extended in these areas. Monitoring the build-up of sample will be the Technical Assistance function's responsibility (for more details see Annex 4).

Criterion: Balance between urban, semi-urban and rural areas

While it is clearly not possible with only six experimental sites to achieve a nationally representative selection of areas, it is nevertheless important to avoid choosing a group of sites that have a socio-demographic profile at odds with the broad national picture. The emphasis on securing a high number of ND25plus and NDLP entrants,

² Information on the number of ethnic minority of customers by Jobcentre Plus district was obtained from the New Deal Evaluation Database.

especially from ethnic minorities, led to the selection of two sites from conurbations (Manchester and East London). Selecting one district from both Scotland and Wales, where the available districts were predominantly rural, implied a further restriction on the final two selections – the need to avoid selecting further rural areas.

The final two areas selected, Derbyshire (East Midlands) and Gateshead and South Tyneside (North East), are the largest remaining areas (in terms of ND25plus and NDLP intake) that are not predominately rural (Derbyshire is partly rural but has some large towns within it), are not parts of conurbations, and are not in any of the regions covered by the previous four selections.

Target Group Size

Table 3, presents estimates of the size of the three ERA Demonstration target groups at each experimental site proposed, over a 12-month period. The table shows that, in total, some 52,000 individuals might be eligible for ERA services across the six sites. The bulk of this group, some 34,000 individuals, consists of lone parents working part time and claiming WFTC (WTC from April 2003). Of the remainder, approximately 5,600 lone parents are projected to enter

NDLP and around 12,000 individuals are projected to enter ND25plus.

The estimates for the NDLP and ND25plus groups are based on projections for the year 2002/03 and were obtained from the Department for Work and Pensions (DWP). For the WTC LP group, estimates come from a scan of all WFTC records undertaken in November 2001 by the Inland Revenue. For the purpose of estimating the annual flow of WTC LP cases, the scan of WFTC records also recorded the numbers for whom a WFTC claim had been in place for a year or longer.

These estimates are subject to appreciable levels of uncertainty. For instance, by the time ERA services actually commence, the new Working Tax Credit will be in place, with obvious implications for these estimates if the take-up of the Tax Credit proves to be substantially different than that for WFTC. In the case of New Deal eligible target groups, the estimates cover the financial year 2002/03, while the ERA Demonstration intake period is due to run from October 2003 to September 2004. Finally, all the projected target group sizes are highly dependent on general economic circumstances, and specifically on local labour market conditions at the ERA experimental sites.



Table 3: Target Group Size Estimates for the Selected ERA Experimental Sites

	South East Wales	Derbyshire	East London	Gateshead and South Tyneside	Manchester	Renfrewshire, Inverclyde, Argyll and Bute	Total
<i>Number of eligible individuals</i>							
ND25plus	729	2,736	3,065	1,808	2,307	1,325	11,970
NDLP	396	1,395	1,088	757	1,415	600	5,651
WTC LPs	3,793	8,961	4,701	4,807	5,846	5,758	33,866
Stock	2,703	6,328	3,197	3,447	4,255	4,188	24,118
Flow	1,090	2,633	1,504	1,360	1,591	1,570	9,748
Total	4,918	13,092	8,854	7,372	9,568	7,683	51,487
<i>Number of individuals randomly assigned ⁽¹⁾</i>							
ND25plus	693	2,599	2,912	1,718	2,192	1,259	11,373
NDLP	376	1,325	1,034	719	1,344	570	5,368
WTC LPs	1,138	2,688	1,410	1,442	1,754	1,727	10,159
Total	2,207	6,612	5,356	3,879	5,290	3,556	26,900
<i>Number of individuals assigned to the programme group ⁽²⁾</i>							
ND25plus	347	1,300	1,456	859	1,096	629	5,687
NDLP	188	663	517	360	672	285	2,685
WTC LPs	569	1,344	705	721	877	864	5,080
Total	1,104	3,307	2,678	1,940	2,645	1,778	13,452

Notes: Estimates of ND25plus and NDLP target group sizes come from Department for Work and Pensions' projections for the year 2002/03 and represent estimated caseload sizes for the New Deal. DWP inform us that the NDLP estimates may be revised downwards. Estimates for the WTC target group are based on the total number of lone parents working part time and claiming WFTC by Local Authority District at November 2001. The estimated flow of WTC claims is based on the number of new WFTC part-time claims made by lone parents in the year to November 2001. These estimates were then mapped approximately on to Jobcentre Plus Districts.

1. This assumes that for the ND25plus group, participation in ERA depends on securing an individual's agreement to take part in the research. This means agreeing to have their contact details passed to a survey research organisation and possibly also, have their administrative records made available for analysis. Those who refuse to take part in the research will not be randomly assigned and will therefore not be able to receive ERA services, but will enter the New Deal as normal. The expectation is that around five per cent of ND25plus group members will refuse to take part in ERA for this reason. For the NDLP and WTC LP groups, participation (meaning being randomly assigned) is entirely voluntary. The assumption is that five per cent of the NDLP target group and 70 per cent of the WTC target group will refuse to participate. This could be because they refuse to take part in the research, or because they refuse to give consent to random assignment.
2. 50 per cent of those randomly assigned are assigned to the programme group.

In addition to presenting estimates of the number of individuals eligible for ERA services, Table 3 also shows the number to be randomly assigned and the number assigned to the programme group. It is assumed that around five per cent of those eligible for ND25plus will refuse to give their consent to take part in the research. Chapter 6, later in this report, explains that eligible individuals from all target groups, who refuse to participate in the research, will not be randomly assigned. Therefore, their probability of entering ERA will be zero.

For the NDLP target group, whose participation is entirely voluntary, the assumption is that five per cent of those who put themselves forward for the New Deal *will not* be randomly assigned. They might refuse either because they do not wish to be randomly assigned, or because they object to taking part in the research.

For the WTC group, who are in work, it is assumed that 70 per cent of the identified caseload will either refuse to participate, be ineligible for the programme because they are no longer working part time, refuse to give consent to either random assignment or participation in the research, or have already been randomly assigned through prior membership of the NDLP or ND25plus samples. It is generally thought that it will be more difficult to convince individuals already in work to take part in the programme than those out of work and to receive services from Jobcentre Plus.

Broadly, it is expected that the offer of ERA services, particularly financial incentives, and that individuals will be compensated for taking part in survey interviews, if payments for survey participation are made, will ensure that refusals to participate and to be randomly assigned will be relatively low

among the New Deal eligible target groups. It should be noted however, that the level of refusals will also depend on how well the Intake Clerks and other programme staff 'sell' the programme to those eligible to participate.

Ideally, the minimum sample sizes in each of the six sites suggested should be around 1,000 customers per target group, per site, in order to ensure that the impacts of ERA in that particular site can be detected using administrative data. Because of the need to select Scottish and Welsh sites, referred to above, two sites have smaller numbers than ideal. Moreover, it is fairly difficult to select reasonably varied sites in terms of location, that are also able to yield samples of those eligible for NDLP of a sufficient size. The size of the ERA target groups can also be affected by other concurrent pilot projects at the selected sites that target the same groups.

If the target group size estimates presented in Table 3 prove to be under-estimates, or the flow of individuals being randomly assigned is lower than expected, remedial action will be required. In order to ensure that enough individuals are randomly assigned in such circumstances, it may be necessary to extend the intake period to allow more individuals to flow into the programme. Such remedial action will, however, delay the availability of findings. Action of a similar nature might also be required to ensure that subgroups of a sufficient size are available (see following paragraph). It will be a function of the Technical Assistance role to monitoring the build-up of cases and recommend remedial action when it is appropriate. The Technical Assistance function is outlined in detail in Annex 4 of this report.



Table 4: Estimated size of subgroups – experimental sites combined

Total subgroup size (Individuals randomly assigned)				
Subgroup	ND25plus	NDLP	WTC LP	
<i>Education⁽¹⁾ (includes vocational quals.)</i>				
Qualifications	7,165	3,221	6,701	
No Qualifications	4,208	2,147	3,458	
<i>Benefit claim history</i>				
Claim >3 years	3,525	–	–	
Claim <3 years	7,848	–	–	
<i>Ethnic group⁽²⁾</i>				
White	9,780	4,885	9,550	
Non-white	1,593	483	609	
<i>Partnership status⁽³⁾</i>				
Partner	3,525	–	–	
No Partner	7,848	–	–	
<i>Age</i>				
< 50 years	8,643	–	–	
> 50 years	2,730	–	–	
<i>Work history previous 3 years</i>				
Some work	–	3,382	–	
No Work	–	1,986	–	
<i>Age of youngest child⁽⁴⁾</i>				
> 5 years	–	2,738	2,947	
< 5 years	–	2,630	7,212	
<i>Length of current claim⁽⁴⁾</i>				
> 24 months	–	2,362	–	
< 24 months	–	3,006	–	
<i>Length of current claim⁽⁵⁾</i>				
> 12 months	–	–	4,063	
< 12 months	–	–	6,096	

Notes

1. Estimates for WFTC groups come from an analysis of all lone parents claiming FC in 1999, Marsh, McKay, Smith and Stephenson (2001) Table 2.23 p61
2. Estimates of ethnicity come from the New Deal Evaluation Database for NDLP and ND25plus groups. For the WTC group, estimates come from an analysis of all LP FC claims recorded on SOLIF, see Marsh, McKay, Smith and Stephenson (2001) p42
3. Estimates from the evaluation of ND25plus, see Lissenburgh (2001)
4. From the evaluation of the national NDLP, see Lessof C *et al*, 2000, Tables 8.2.4 p 76 & Table 8.3.1 page 77, for the WTC group, estimates come from analysis of all LP FC claims in SOLIF, see Marsh, McKay, Smith & Stephenson (2001), p41
5. Estimated from WFTC 100% scan at November 2001 – part-time lone parents claiming WFTC only

The Impact Study design presented later in this report involves the estimation of programme impacts for key subgroups. Subgroup impacts are to be obtained by merging samples of those randomly assigned to either the programme or control group across the six sites, in order to form a 'pooled' sample large enough to detect programme impacts of a reasonable magnitude. Table 4 presents estimates of the numbers likely to be randomly assigned in each target group, within key subgroups, across all experimental sites and thus indicates the likely size of pooled administrative data samples for these subgroups. This also gives an indication of the total numbers of individuals from which subgroup booster survey samples can be drawn.

CHAPTER 3 – PROCESS STUDY

ERA is an ambitious policy intervention, offering a variety of services and financial incentives for work and training over a relatively long period of time. Implementing it will be challenging, and successful implementation cannot be taken for granted. It is, therefore, important to ask – what happens when this policy is tried in the ‘real world’? Do programme staff and managers succeed in constructing an intervention that is true to the original vision of ERA? What operational challenges do they encounter, and how do they address them? Does the targeted population actually take advantage of what the programme has to offer? And how do such factors as local labour market conditions, the local service environment, and customers’ background characteristics and circumstances affect how staff operate ERA and the extent to which targeted customers participate in it?

These are the concerns of the evaluation’s implementation and process study (referred to from hereon as the ‘process study’). The findings from this study will be critical for assessing the overall feasibility of ERA, for understanding the nature of the intervention being tested by the impact and cost–benefit analyses, and for explaining why the programme did or did not achieve its hoped-for effects, and how its operation might be improved. The findings will also help in explaining any substantial variation in programme costs and impacts observed across sites and across different customer subgroups. Finally, the process study will inform lessons on ‘best practices’ for

operating an effective ERA programme, should the impact and cost-benefit analyses find it successful and the decision is taken to roll-out the programme nationally. These lessons will offer guidance to Ministers, policymakers and programme administrators in any effort to replicate ERA across the country. The next few sections discuss the specific objectives, topics, analytical strategies, and data for this part of the evaluation.

Objectives

The primary goals of the process study will be to:

- describe what the ERA policy design ‘looks like’ when put into practice within Jobcentre Plus;
- learn about customers’ barriers to employment retention and advancement, and how well ERA responds to these;
- learn about customers’ participation in programme activities, take-up of the financial incentives, and interactions with their ASAs;
- show how the programme group’s receipt of services and financial incentives differ from the control group’s use of any comparable services available to them outside of the ERA programme (e.g., through the New Deal or other existing programmes);

- explain how and why staff practices in operating ERA vary across locations;
- learn how those alternative practices affect participants' responses to the programme; and
- identify what programme practices ought to be emulated and which ones ought to be avoided in any effort to replicate the programme in other locations across the country.

To achieve these objectives, the process analysis will examine seven broad topic areas: the local context, the organisational structure of the ERA programme, the operating strategies of the frontline staff, programme strategies, local office management, customers' receipt of services and financial incentives, the factors affecting customers' receipt of those services and incentives, and common impediments that customers' experience that act as a barrier to their retention and advancement. These areas of inquiry are described next.

Areas of inquiry

Local context

The local context at each site will profoundly shape the service and employment opportunities available to ERA participants and to members of the control group. These conditions can thus affect greatly the implementation and impacts of the ERA intervention at the site. Most important will be the labour market context (e.g., job availability, characteristics of available jobs), which will influence how easy or hard it is for members of the programme and control groups to find and keep jobs, especially good jobs with opportunities for advancement. Access to transport, both public and private, as well as provision of education, training,

and child care services, will also have a bearing on how ERA operates and on how much and in what ways the control group receive services outside of the ERA programme.

Organisational structure

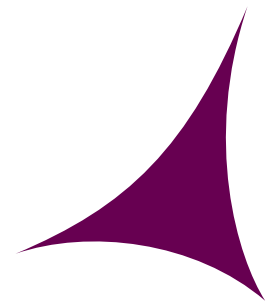
The ERA programme will not operate as an entirely free-standing programme. Instead, it will be housed as a unit within a Jobcentre Plus office(s) at each demonstration site. In addition, it will rely on various outside organisations to provide education, training, and support services that Advancement Support Advisers (ASAs) and participants agree to as part of participants' Advancement Action Plans (AAP). These features make it important to understand the organisational structure of the programme in each location. As indicated by the following list of questions, this includes learning about ERA's relationship to other parts of Jobcentre Plus, and the programme's staffing patterns, caseload sizes, and linkages to external service providers and employers.

- ***ERA integration into the Jobcentre Plus agencies***

In what ways does ERA function as an independent unit within the Jobcentre Plus? What resources, management structures, and organisational procedures does it share in common with other parts of the agency?

- **Staffing**

How are ERA roles and functions allocated across different staff positions? What types of people (with what kinds of skills or experience) are recruited to fill these positions, and how do they compare to the roles and functions of Jobcentre Plus staff?



- **Caseload size**
How many ERA customers, on average, are assigned to each ASA? How does this compare with the customer-to-staff ratios for New Deal Personal Advisers?
- **Changes over time**
How does the organisational structure vary over time? How do ASAs cope with the transition from dealing with mainly pre-employment clients in the first few months, more of a mix of pre- and post-employment clients for a while, and then with a larger fraction of post-employment clients for the final two years?
- **Provider networks**
What kinds of service providers are included in the networks of agencies to which ERA refers participants for employment-related training and services? How are these providers selected?
- **Services**
What kinds of education, training, and support services (e.g., child care, counselling) are made available to ERA participants through these networks?
- **Employer networks**
What special linkages, if any, are developed with employers or employer associations?
- **Random assignment**
The process study may also address the issue of random assignment, how it was implemented and the reaction of customers to being randomly assigned. Whilst the technical functioning of the random assignment process will be monitored by Technical Advisers (see Annex 4), the process study might explore the response of customers and administrators to its operation and, through this, it may be possible to examine whether some of the assumptions underpinning random assignment hold.

Programme strategies

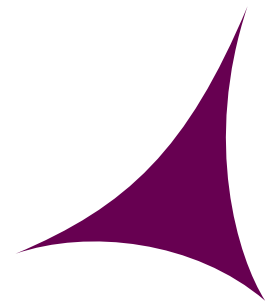
Much of what a participant experiences in ERA will depend on how frontline workers – especially the ASAs – execute their roles and responsibilities. What the frontline staff tell customers when constructing AAPs, when reviewing and updating these plans at face-to-face meetings, and during informal contacts, will be an important part of the ERA programme. For example, the advice the ASAs give on the kinds of jobs to pursue, their guidance on whether to combine work and training, and their explanations of how much ‘work pays’ when the Working Tax Credit and the ERA Retention and Advancement (R&A) bonus are considered, may influence the choices that participants make concerning work and training. The intensity and persistence with which staff monitor participants’ progress, deliver the programme’s messages about work and training, help participants deal with problems that arise at their jobs or in their personal lives, and help them look for and find opportunities for advancement are also important aspects of the ERA programme that may influence participants’ decisions.

Understanding these and other behaviours on the part of ERA frontline staff, starting at the time that customers are recruited for the programme, will thus form a critical part of the process study’s efforts to describe and analyse the ‘programme theory’. The study will be guided in this area by the following specific questions:

- **Outreach and recruitment strategies**
Through what procedures are eligible individuals in each target group identified, and what are they told about the programme (and random assignment) to encourage them to apply?



- Messages and guidance on initial job choices
How much do ASAs encourage non-working participants to be selective about the kinds of jobs they take against taking 'any job'? How do they use the 'step-down' concept to adjust their advice the longer a participant remains out of work? How strongly do they encourage full-time over part-time work? In general, what considerations influence the advice they offer to different types of participants?
- Identifying job opportunities for placement and advancement
How do ASAs work with Jobcentre Plus local employer account managers (job developers) to help non-working or already-working participants identify job openings at specific employers? What strategies do the local account managers use to find jobs that offer better opportunities for retention and advancement? How do ASAs and other staff help participants improve their own job search skills for finding better jobs?
- Messages and guidance on mixing work and training
How strongly do ASAs encourage participants to enter education or training programmes while they are working? Do they offer different kinds of advice to customers from different target groups? How do they try to ensure that customers make the most appropriate choices about the specific education and training opportunities they want to pursue (e.g., how do staff assess participants' potential to benefit from certain kinds of training, and how do they stay informed about the quality of the available training providers)?
- Explaining how much 'work pays'
How do ASAs help participants understand how much better off financially they can become by working full time rather than part time, or by advancing to better-paying jobs? What efforts do they make to use 'better-off' calculations to demonstrate the payoff of different employment choices and to highlight the important contributions of the Working Tax Credit?
- Explaining the added value of the ERA financial incentives
How do staff present information about the ERA bonuses to participants? How clearly and thoroughly do they describe the conditions attached to these incentives, and how often over the course of the follow-through period are participants reminded about the potential value of these incentives? How clearly do the ASAs understand the incentives?
- Ongoing personalised guidance on employment retention
In what ways, and with how much effort, do staff attempt to learn in advance about emerging problems that working participants are facing in their personal lives or at their jobs, that could result in job loss or failure to advance? How intensively do staff attempt to help individuals avoid job loss?
- Ongoing personalised guidance on employment advancement
How pro-active are staff in helping working participants identify and pursue better job opportunities and career ladders with their current employer or with a new employer? What efforts do they make to learn about the particular circumstances of a participant's current job and the potential for advancement with the current employer or elsewhere?

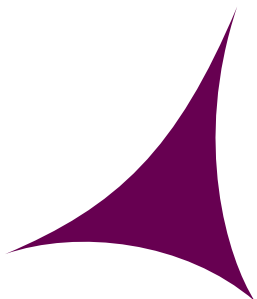


- **Nature and intensity of follow-up contacts**
How often do ASAs contact participants in between their regularly scheduled meetings? To what extent do they rely on phone calls, home visits, worksite visits, or meetings in other locations in the community? How often do they have contact with working participants outside of normal business hours (e.g., in the evenings and on weekends)? Do the nature and intensity of the contacts vary for different types of participants?
- **Enforcing participation requirements for the New Deal 25plus target group**
How do the ASAs use the added leverage derived from these requirements to encourage members of this target group to look for work? How strongly do they enforce the participation requirements, and how, if at all, do their efforts differ from the enforcement efforts of New Deal Personal Advisers?
- **Intervening with employers**
Under what circumstances and how often do staff contact participants' employers to address issues concerning job retention and advancement?

Office management to promote retention and advancement

How ERA frontline staff perform their critical functions will depend, in part, on the kind of the training they are given, the procedures they are asked to follow, the standards of performance to which they are held, and the management guidance and support provided to them. The process study will investigate several important aspects of local office management:

- **Staff training**
How are staff trained for their ERA roles, and how does this training differ from what is normally provided to New Deal personal advisers?
- **Staff performance assessment**
How is the performance of ASAs and other ERA staff assessed? What do staff need to do or accomplish in order to earn a high performance rating?
- **Office performance targets**
What outcome targets (e.g., job placement rates, advancement rates), if any, are set for local ERA offices to try to achieve? How is 'success' against these goals measured and reported to staff?
- **Participant tracking systems**
What adaptations in the Jobcentre and New Deal data systems are made to help ERA staff track participants' employment progress, the frequency of their contacts with their ASAs and the programme office, their service assignments and participation in the activities to which they are referred, and their receipt of the ERA bonuses? How well do these data systems function in helping the ASAs monitor the overall progress of their customers?
- **Team work**
To what extent do ERA staff within an office review each other's cases, advise each other on how best to respond to particularly challenging cases (e.g., through regular case conferencing), and, in general, share ideas on how to improve the overall performance of the programme? What management strategies encourage (or discourage) this kind of team work?



Participants' use of services and incentives

In addition to describing and analysing the organisational structure of the ERA programme and the practices of managers and frontline staff, the process study will measure participants' understanding and responses to those efforts, and their use of the programme's services and supports. This area of inquiry will include four dimensions.

First, the study will focus on what members of each target group hear and know about the programme and what it offers and encourages. As previously noted, programme messages can be an important part of the ERA programme and can shape customers' behaviour. But this will depend, in part, on whether those messages actually get through to participants, and how participants interpret them.

Second, this part of the process study will measure what participants do and what 'dose' of service they receive – that is, what proportion make use of various services and supports and for how long. For example, the study will focus on participation in education and training activities, receipt of counselling, financial incentive take-up rates, use of childcare, etc. It will also determine how this participation varies among key subgroups of participants, which may have systematically different needs, interests, and capacities. In addition, the study will measure how long participants remain in contact with their ERA personal advisers (a significant number may choose not to remain in touch for the entire two-year in-work follow-up period).

Third, the study will investigate how participants view the services, incentives, and supports that ERA offers them. What value, if any, do they see in these? What kinds of assistance do they find most or least helpful?

And why do some individuals choose not to take advantage of the assistance the programme offers?

Finally, the study will measure the difference in the extent to which the programme and control groups use various kinds of services and financial incentives. It will be recalled that, although members of the control group will not have access to ERA, those in the New Deal target groups will be eligible for regular New Deal services. Furthermore, all three target groups will be permitted to seek education, training, and job search assistance on their own from other agencies in the community. Controls, like members of the programme group, will also be entitled to receive the Working Tax Credit. This comparison between the programme and control groups is important because it pertains to a central hypothesis behind the programme design: that ERA will lead to a substantial increase in the receipt of employment-related services, incentives, and supports, even after participants are working, and that this increase, in turn, will affect the net cost of ERA and, possibly, the magnitude of the programme's impacts on retention and advancement.

Explaining office variation in participant responses to ERA

How local programme staff translate the ERA programme model, as it is described 'on paper,' into an actual functioning programme 'on the ground' will no doubt vary across local offices. This variation may lead to different customer responses, including variation in participants' rates of service receipt and use of incentives across offices and districts. The process study will explore whether different organisational structures and staff practices encourage or discourage participants' use of services,



incentives, and supports, and help explain any variation in programme-control group differences on these measures by site.

As part of this effort, the process study will try to quantify various dimensions of programme implementation so that all offices and districts in the study can be ranked relative to the others in terms of such criteria as how much certain key practices are emphasised (for example., the intensity of staff interactions with participants, and how strongly staff encourage mixing education and training), and the degree to which key organisational conditions (e.g., low caseload sizes) are present. The focus will be on programme practices and conditions that are hypothesised to be important factors in affecting programme performance. The analysis will then measure the correlation of these implementation factors with the programme group's participation outcomes, with the programme-control group differences in participation outcomes, and, in turn, with office-level and district-level impacts. In this way, the overall ERA evaluation will attempt to link implementation factors with impacts across locations, and to draw cross-cutting lessons about the relative effectiveness of particular strategies. Although it will not be possible to disentangle completely the factors that are driving differences in impacts across locations, the analysis will identify factors that at least appear to be linked to success, as well as those that are not.

Impediments to retention and advancement

In addition to studying the structure and operation of ERA and customers' responses, the process study will try to develop a deeper understanding of the kinds of problems that impeded retention and advancement among members of the three target groups. It will

do so by learning from programme staff, outside service agencies, and employers about the kinds of problems they observe or that customers tell them about, and through direct interviews with customers themselves. The study will also assess the programme's handling of these problems.

Key questions for this part of the study will include: What are the most common challenges to retention and advancement that ASAs are called upon to address (e.g., family problems, difficulty complying with workplace norms, transportation or child care problems, employer discrimination, lack of particular kinds of skills, mental health problems, substance abuse, low expectations, or limited job opportunities)? How well suited are the ASAs to handling these problems, and how adequate for doing so are the tools and resources that the ERA programme makes available to them? Are there any types of difficulties experienced by customers that the programme does not address?

Data sources and data collection

The process study will utilise a mix of qualitative and quantitative data (See Table 5). In-depth field research will provide rich, descriptive information about how the ERA programme is structured and implemented at the local level, how frontline staff perform their roles, how customers view and experience the programme, and what kind of impediments to employment retention and advancement customers face.

This qualitative data collection will be done through on-site observation, in-depth structured interviews, and focus groups with ERA programme managers and frontline staff and with Jobcentre Plus staff. Data for constructing quantitative measures of

programme practices, which will be used in comparing and ranking offices and districts, will come from a survey of ERA staff using closed questions. Information from programme manuals and documents will supplement these data. To describe the local context, the process study will use published documents that provide information on local labour market conditions and the provision of local social services, education and training opportunities.

The same customer surveys (see Chapter 6) that will be used for the impact study will also be used to collect information on programme-control group differences in service receipt and incentive take-up rates, as well as quantifiable evidence on the programme group's experiences in, and

responses to, ERA. These surveys will also be used to gather quantitative data on customers' retention and advancement barriers. Programme tracking data will provide additional evidence on the use of ERA services and on the control group's use of regular New Deal services.

Finally, structured qualitative interviews with employers will be used to collect information on the perspectives, experiences, and roles of employers – for example, their experiences with caseworkers and participants, their perceptions about the reasons for instability or lack of upward mobility, and their perceptions about how the intervention affects their 'bottom line' (e.g. turnover, recruitment and training costs).

Table 5: Data source for each process study topic

Topic	Data sources
Local context	Qualitative interviews with local managers; programme documents; published labour market and service delivery information
Organisational structure	Qualitative interviews with local managers and staff; programme documents
Programme strategies and management	Qualitative interviews with local managers and staff; survey of frontline staff; written programme procedures
Participants' use of services and incentives and views of ERA	Programme tracking records; customer survey; customer focus groups
Participants' barriers to retention and advancement	Qualitative interviews with local managers and employers; customer survey; customer focus groups

CHAPTER 4 – IMPACT STUDY

The impact study considers whether the ERA programme achieves important objectives. It attempts to provide evidence of whether new services and financial incentives delivered through the ERA Demonstration programme have led to improvements in retention and advancement. It asks whether these improvements have occurred across the demonstration as a whole and at each experimental site, for each of the three target groups. In addition, the impact study considers whether differences in impacts can be found across subgroups within each target group, such as those with no qualifications, or those from non-white ethnic groups³.

Estimates of programme effects are known as ‘impacts’ and, in order to measure the impacts of ERA services⁴, a random assignment research design is recommended. Random assignment involves assigning eligible individuals at random to either a programme group or a control group. Those in the programme group receive ERA services, while those in the control group receive existing services only.

Random assignment is seen as the most effective way of obtaining unbiased estimates of programme impacts. The impact of ERA services is estimated as the difference between the average values for outcomes of interest measured across the programme group, and those observed in the control group. Impacts measured in this way are ‘*internally*’ valid or unbiased, because the two research groups are statistically

equivalent in terms of both the observed and unobserved characteristics of the individuals in them. The only difference at the point of assignment is that members of the programme group go on to receive ERA services, while those in the control group do not.

This chapter commences with a discussion of the main research questions and detailed hypotheses the impact study will seek to address. Following this, the random assignment design is set out, with specific attention paid to how individuals are recruited into the ERA programme, and when in this process they are randomly assigned. Then the estimated sizes of the target groups and subgroups presented in Tables 3 and 4 above, along with other information, are used to give a rough indication of the likely size of programme impacts the impact study might be able to detect. The chapter ends with a discussion of the types of analyses that might be carried out and the issues associated with making generalisations from the results of the impact study.

Research questions, experimental hypotheses and outcomes

The random assignment design, set out in the following section, enables some important questions about the effectiveness of ERA services to be addressed. Here, higher-level research questions are considered and some more tentative hypotheses are examined.

³ Although, as mentioned in Chapter 2 above, it is extremely unlikely that sample sizes will be large enough to allow measurement of the impact of ERA among non-white lone parents.

⁴ Hereafter the term ‘ERA services’ is taken to include financial incentives as well as caseworker services.

The hypotheses and questions set out here are based on experimental comparisons. In some cases further exploration might be required to consider the impact of the programme on members of the programme and control groups who get jobs or who are in work at a specific point in time. Because of the process of selection into jobs, impacts estimated on the basis of these comparisons must be estimated non-experimentally, taking account of selection bias. It is particularly important to take account of this selection bias when considering the programme's impact on advancement, although some aspects of the programme's impact on advancement can be estimated experimentally. A discussion of non-experimental impact estimation is provided later in this chapter.

In essence, the impact study will reveal whether ERA services result in greater levels of retention and advancement than would otherwise be the case. It aims to answer this question on a number of levels. First, it seeks to uncover whether ERA services have improved levels of retention and advancement for each target group separately, across the Demonstration as a whole. Programme impacts estimated in this way are referred to as *pooled impacts*, as they are estimated on combined data from the six experimental sites⁵. For each outcome of interest, the aim is to estimate a pooled impact for each target group.

Second, the impact study aspires to measure impacts for each target group at each site separately – these impacts are known as *site-specific impacts*. The extent to which these impacts can be measured depends crucially on the number of individuals who are randomly assigned at each site and

whether there are suitable administrative data with which to measure outcomes. Surveys can also be used to measure *site-specific impacts* across a wider range of outcomes, though sample size constraints might frustrate the ability to measure site-specific impacts with any acceptable level of statistical precision. Surveys might only be able to detect impacts for individual target groups at those individual sites where impacts are large.

The impact study aims to be able to estimate pooled impacts for various subgroups – for example, to answer the question: did the impact of the programme differ for those with qualifications compared to those without qualifications? These impacts are referred to as *pooled subgroup impacts*. An important function of the Process Study (see Chapter 3) is to explore reasons for variations in *site-specific* and *pooled subgroup impacts*, should they occur. This chapter also presents a brief discussion on the extent to which it will be possible to compare programme impact estimates between sites.

The main research questions to be addressed, for each target group, through the impact study are:

To what extent do services and financial incentives delivered through the ERA Demonstration improve the work retention and employment advancement, as well as other outcomes, of those assigned to receive them?

To what extent do services and financial incentives delivered through the ERA Demonstration improve work retention and employment advancement among different subgroups of those assigned to receive them?

⁵ It is worth pointing out that because the experimental sites were selected purposively and *not* at random, the pooled-impact estimates represent the true impact of the programme at the six sites only. In other words, if a random sample of programme participants of the same size from across the six experimental sites were to be drawn 100 times, and the programme impacts estimated each time, 95 of the 100 resultant confidence intervals would contain the true programme impact for the six sites. These results could not be used to infer whether, and with what frequency, the true programme impact for the whole eligible population across the country appears in these confidence intervals. This means that the impact estimates are internally valid; whether they are externally valid, or generalisable is a separate question. A discussion of the extent to which results from the impact study can be generalised is set out later in this chapter.



To what extent does the impact of services and financial incentives for those assigned to receive them differ from site to site?

These broad research questions lead to the formulation of specific hypotheses that can be tested. These specific hypotheses reflect different outcomes that are indicative of improvements in retention and advancement.

One key outcome is earnings. It might be expected that individuals who retain work and advance might earn more. Earnings, however, are a composite outcome because increases in earnings might reflect working longer hours or earning a higher hourly wage, or some combination of the two. It is debatable as to whether an individual who is earning more in total but receiving the same hourly wage can be said to have advanced. In that sense, the measurement of earnings has limitations when it comes to determining the extent of advancement. Earnings also have limits in terms of measuring retention. For example, gains in earnings could occur either through increases in hourly wages or total weekly hours, or some combination of the two, without any improvement in job retention. With this in mind, random assignment will allow exploration of the following hypothesis for each of the three target groups:

Hypothesis: *Those assigned to receive ERA services and incentives have higher annual net earnings, on average, than those assigned to the control group.*

The extent of advancement is better determined through comparing wage rates between programme and control groups. Many in each group, however, will record zero hourly wages, as they do not enter employment during the study period. As a result, studying certain hypotheses about

the impact of the ERA Demonstration on hourly wage rates can only be achieved non-experimentally (see below). This is because a comparison of the hourly wage rate of members of the programme group in work, with the hourly wage rate of members of the control group in work, is required. As already mentioned, because these groups are 'selected' into jobs, they are not statistically equivalent. There are, however, some hypotheses about wage rates that can be tested experimentally to provide evidence of whether advancement has occurred, for example:

Hypothesis: *The proportion of those assigned to receive ERA services and incentives earning hourly wages above a specified threshold (for example, £7.00 per hour) is higher than for those assigned to the control group.*

Such hypotheses could be explored through estimating pooled and site-specific impacts. The Process Study would explore the reasons for any statistically significant differences in these impacts across sites. Furthermore, these hypotheses can be examined for key subgroups (see Table 4 and Table 13 for suggested subgroups), for example:

Hypothesis: *Those without qualifications assigned to receive ERA services and incentives have higher annual net earnings, on average, than those without qualifications assigned to the control group.*

Another set of outcomes of importance is employment and work retention. If ERA services were successful, the expectation is that a higher proportion of the programme group will be in work and work, on average, for longer periods of time. This implies the following hypothesis might be tested for each target group:

Hypothesis: *Those assigned to receive ERA services and incentives are more likely to be employed, at a given point in time, than those assigned to the control group.*

A more specific measure of job retention might be obtained through considering the following hypothesis:

Hypothesis: *Those assigned to receive ERA services and incentives remain employed for longer periods of time, on average, than those assigned to the control group.*

Because the programme aims to test the importance of working full time through the provision of a retention and advancement bonus for retaining full-time work, the following hypothesis might be tested:

Hypothesis: *Those assigned to receive ERA services and incentives are more likely to work full time than those assigned to the control group.⁶*

This hypothesis might be usefully tested alongside a similar hypothesis related to the proportions in programme and control groups working part time. Moreover, one might wish to explore whether those in full-time work retain employment and advance in work more than those in part-time work. Such a hypothesis would, in effect, test the rationale for the retention and advancement bonus discussed earlier, but would require the application of non-experimental methods, similar to those outlined later in this chapter.

If individuals in the programme group are more likely to be in work and to remain in work longer than their counterparts in the control group, it follows that they are less likely to be claiming out-of-work benefits. In considering benefit dependency, a testable hypothesis might be formulated as follows:

Hypothesis: *Those assigned to receive ERA services and incentives are more likely to be off benefits at a given point in time, than those assigned to the control group.*

In the case of pooled-subgroup impacts, for the ND25plus target group, the impact of the programme on older workers might be examined through testing the hypothesis that:

Hypothesis: *Those aged 50 or over assigned to receive ERA services remain in work for longer periods of time, on average, than those aged 50 or over assigned to the control group*

Table 13 in Chapter 6, indicates that the ability to test such a hypothesis is likely to be limited by the expected sample size of the ND25plus target group aged 50 and over. Later in this chapter, adding a booster sample of older workers to the ND25plus core sample at the 24-month follow-up survey is addressed as a way of countering this problem.

It is possible to test additional hypotheses through the random assignment design. For example, advancement might comprise an improvement in terms and conditions independent of increases in earnings or wage rates. As a result, a hypothesis to be tested might be:

Hypothesis: *Those assigned to ERA services enjoy higher levels of non-pecuniary work-related benefits than those assigned to the control group.*

These non-pecuniary benefits might include, for example, paid holiday and employer childcare facilities. Differential receipt of other pecuniary benefits between programme and control groups might also be tested, for example, the receipt of pension contributions. These hypotheses, however, would need to be tested non-experimentally.

⁶ It is possible that providing an additional incentive to work at least 30 hours could lead to those in receipt of the incentive working fewer hours above the 30-hour threshold, on average, than they would have in the absence of the programme. This is the case where those in receipt of the retention and advancement bonus substitute work hours above 30 hours for leisure in response to bonus, known as an income effect. This income effect would not affect the full-time/part-time impact being tested.



Other issues that might be explored through random assignment are family income. A hypothesis to address this question might be formulated as follows:

Hypothesis: *Those assigned to ERA services and incentives record higher levels of family income than those assigned to the control group.*

In other words increases in earnings are not offset by reduced working among other family members and a decline in the receipt of benefits.

Measuring outcomes

In order to test the above hypotheses as well as others, outcomes need to be defined and measured for all members of the study sample. Some examples of a range of outcomes that might be measured are presented here. Outcomes identified as a priority will inform the design of survey instruments, though the length of survey interviews and the availability of administrative data will limit the number of outcomes that can be measured. Chapter 6 of this report outlines the design of sample surveys and the structure and content of the administrative data sources available for the impact study.

Table 6: Examples of important outcome measures and data sources

Outcome	Source	
	Administrative data	Survey data
Proportion not claiming out-of-work benefits	✓	✓
Proportion in employment of 16 hours a week or more	✗	✓
Mean or median net earnings over 12 months	✓	✓
Mean number of weeks in employment of 16 hours a week or more over 12 months	✗	✓
Mean net hourly wage	✗	✓

Notes:

✓ indicates that the relevant outcome is available from the indicated data source, ✗ indicates that the outcome is not available from the indicated data source.

Table 6 sets out a selection of some of the main outcomes that the evaluation will seek to measure for each target group. Administrative data sources can be used to determine the proportion of individuals in the programme and control groups claiming out-of-work benefits at a given point. Collecting the relevant administrative data over a sustained period of time will enable comparisons of levels of benefit dependency. Similarly, using both administrative and survey data, comparisons can be made at given intervals between the proportion of programme and control group members in

employment. Having access to both survey and administrative data will allow cross-checking of each data source and also allow detailed non-response analysis.

Looking at advancement, the outcome measures to be collected reflect the notion that advancement primarily manifests itself for individuals in the form of higher-paying jobs. The two key outcomes are average hourly net wages reported by programme and control groups, and mean net annual earnings.

Job retention will be measured by recording the proportion of weeks over a 12-month period spent in employment of 16 hours a week or more.

There are other outcomes that measure 'job quality' or advancement that can be collected through surveys. For example, these might include:

- proportion of all jobs classified as part-time;
- proportion of all jobs classified as temporary;
- proportion of all jobs providing occupational pension coverage;
- proportion of all jobs providing paid holiday and sick leave; and
- proportion of individuals engaged in work-related training over a specified time period.

In terms of family and child welfare, the following outcomes might be measured through surveys:

- total annual family income;
- proportion of individuals residing in owner-occupied housing; and
- proportion of individuals in families or households experiencing hardship.

It may also be possible to look at outcomes for children, either through collecting data from school and health records, or eliciting this information from parents. The types of outcomes that could be measured include:

- performance in school tests;
- levels of truancy; and
- levels of mental wellbeing.

The ability to detect differences across the outcomes mentioned above, at standard levels of statistical precision, will depend on the size of impacts, the expected variance of each outcome and the size of the research sample. The minimum detectable effects that an evaluator can reasonably expect to be able to measure for a selection of major outcomes are set out below.

Random assignment design

This section considers the random assignment research design and particularly focuses on the relationship between intake into the programme and the process of assigning individuals to programme and control groups.

The random assignment design for the New Deal eligible target groups is illustrated by Figure 1 below. The New Deal eligibles consist of those who opt-into the New Deal for Lone Parents (NDLP) and those who are mandated to join, or volunteer for, the New Deal for the Long-term Unemployed (ND25plus). Figure 2 illustrates the random assignment design for the third target group – lone parents claiming the Working Tax Credit (WTC) and working part-time. Both designs involve the random assignment of eligible individuals into a control group and a single programme group⁷.

The randomised design for each target group is considered next.

⁷ A design comprising the creation of two-programme groups plus a control, for both New Deal Eligible and WTC target groups, was also considered. The purpose of this design was to measure the impact of different combinations of programme elements separately. The two-programme group design involved the random assignment of individuals to receive either: (1) a financial incentive combined with a caseworker service; (2) a financial incentive alone; or (3) existing services (the control group). Two main issues made such a design unattractive. First, it was felt that the modest financial incentives proposed for the ERA Demonstration would not generate impacts of a sufficient magnitude without the input of a caseworker. Second, the increased complexity associated with having two programme groups, as opposed to one, would make such a design more difficult to administer and implement.

Figure 1: Random assignment design for New Deal eligibles



New Deal eligibles – New Deal for Long-term Unemployed (ND25plus) target group

The intake process for the ND25plus eligible target group usually commences when individuals who have been claiming JSA for 18 months or more, within a 21-month period, receive a letter from Jobcentre Plus requiring them to join the New Deal.⁸

In the sites where the ERA Demonstration is operating, all eligible individuals coming forward to enter ND25plus will be referred to an ERA Intake Clerk who will check the individual's eligibility for the programme. In addition, the Intake Clerk will help the customer complete a Basic Information Form (or BIF – see Chapter 6 for further details) before taking the customer through the rest of the intake process. The BIF will record the basic characteristics of the individual. This feature of the intake process, in combination with the use of administrative records, will enable comparisons to be made between those who enter the Demonstration and those who refuse to participate in the research and, thereby as a result, do not enter the Demonstration.

The Intake Clerk will then inform the individual that ERA services, including financial incentives, are available in addition to ND25plus services, explaining both programmes. Individuals will be informed that whichever of the two services they receive will be determined through random assignment and that they have an equal probability of receiving either service.⁹ Following this explanation, individuals will be asked to participate in the Demonstration

and give their consent to take part in the research. This consent will be recorded on the BIF and signed off by the individual. Refusal to consent, together with the reason(s) for refusal, will also be recorded on the BIF. It is extremely important that participation in the Demonstration is presented in a very positive manner by the Intake Clerk and that numbers refusing to take part in the research are kept very low. Intake staff will be provided with scripts to help them sell participation in the programme and the research to customers as well as help allay any fears customers may have.

Because participation in ND25plus is mandatory for those who have been unemployed for 18 months or more, it is assumed that participation in the ERA Demonstration will also be mandatory.¹⁰ However, individuals who refuse to consent to participation in the research (that is, those who refuse to participate in survey interviews, or to make available their administrative records for analytical purposes if this is required) will not be randomly assigned.¹¹ They will not join the Demonstration, but instead enter the New Deal as normal.

Once the BIF is complete, the Intake Clerk will telephone the ERA Database Controller who will be located in a central office and be part of the research/evaluation function. The ERA Database Controller will be responsible for randomly assigning individuals from all target groups and maintaining the ERA Evaluation Database.

⁸ However, there can be early entrants into ND25plus on a voluntary basis – they too will be eligible to join ERA. These volunteers include customers aged 50 or over and other types of individuals such as the homeless or ex-offenders. Such volunteers constitute only a small minority of total ND25plus numbers.

⁹ In the normal course of events random assignment proceeds on the basis of an equal probability of assignment. However, if some customers are assigned to a non-research group in order to control the flow of numbers into the Demonstration, this equal probability will be less than 50% – see Annex 2 on the design of the random assignment algorithm.

¹⁰ But not for volunteers for ND25plus.

¹¹ Consent to participate in research includes consent to random assignment when pertaining to volunteers for ND25plus. If consent to random assignment is refused, these individuals will also not be part of the ERA Demonstration but will receive ND25plus services as normal.



For the random assignment to proceed, the Intake Clerk will need to provide the ERA Database Controller with certain information. First, the Intake Clerk will provide the individual's postcode to the ERA Database Controller, who will enter the postcode into a database that will display the corresponding range of correct postal addresses. The ERA Database Controller will then ask the Intake Clerk to read out the individual's address and check to see if it is compatible with the range of addresses displayed. If the address and postcode provided are incompatible, the Intake Clerk will need to check the address¹² with the individual.

In addition to postcode and address information, the ERA Database Controller will ask the Intake Clerk for the serial number pre-printed on the individual's BIF (assuming the BIF is a paper form), the individual's National Insurance Number (NINO), and their surname. These three pieces of information are required to ensure that every randomly assigned individual can be properly identified.¹³ In addition to this, the ERA Database Controller will ask several questions that confirm the individual's eligibility for the programme, as well as ask the Intake Clerk for the answer to one or two questions on the individual's BIF at random. This is to ensure that the Intake Clerk has an incentive to complete the entire BIF for each individual prior to random assignment.

The ERA Database Controller will enter the address information (including postcode and telephone number), NINO, serial number, surname and other information on to the ERA Evaluation Database. An automatic check of existing records will immediately take place in order to determine whether an individual is already part of the Demonstration, either as a member of the programme or control group. If they are not, a random assignment algorithm contained on the database will be triggered and the individual will be assigned to either the programme or control group.

Once the algorithm has randomly assigned the individual, the ERA Database Controller will place a marker on the database recording the individual's assigned status (whether they are in the programme or control group) and will inform the Intake Clerk accordingly. The Intake Clerk will, in turn, inform the individual, as well as mark the individual's BIF with their status. It is extremely important that all staff at participating Jobcentre Plus experimental sites are fully aware of how the Demonstration operates and have access to records that will enable them to determine quickly whether a particular customer is participating in the Demonstration, and if so, to which group they are assigned.

¹² Obtaining a correct address for each person randomised is a high priority in order to provide accurate contact information for the purpose of sampling individuals for survey interviews. The objective is to minimise survey non-response resulting from poor quality contact details.

¹³ Achieving a positive identification for each randomised individual can be compromised by errors in recording and relaying information, particularly NINOs, and complicated by the existence of shared or duplicate NINOs. Chapter 6 of this report discusses this issue in greater detail.

Individuals assigned to the programme group will then be given an appointment to see their Advancement Support Adviser and start to receive services. As indicated in Figure 1, services will remain available to individuals after they have entered work. Individuals assigned to the control group will receive the same services they would have in the absence of the ERA Demonstration. The random assignment design allows a comparison of retention and advancement for those receiving services through the ERA Demonstration (the programme group) with those participating in the New Deal and assigned to the control group.

The research design seeks to ensure that random assignment takes place prior, yet as close in time as possible, to the receipt of services. If random assignment were to take place some time before individuals were exposed to ERA services, a substantial fraction might never actually enter the programme. This means that the programme's impact would be diluted and a somewhat larger sample would be required, than that outlined here, to detect programme effects.

New Deal eligibles – New Deal for Lone Parents target group

The second target group comprises those who choose to participate in the New Deal for Lone Parents. The intake process is similar to that outlined for the ND25plus group; hence, Figure 1 again illustrates the key features of the research design as it relates to this group.

Lone parents with dependent children can volunteer to participate in NDLP at anytime. Most lone parents learn of the opportunity to participate in NDLP at a mandatory work-

focused interview when starting a new claim for Income Support. At the sites where ERA services are being tested, everyone volunteering to participate in NDLP will be referred to the ERA Intake Clerk.¹⁴

The process will then be the same as for the ND25plus group. The Intake Clerk will check the customer's eligibility and help each customer complete a BIF.¹⁵ The Intake Clerk will inform the lone parent that new ERA services are available to them and that they have an equal chance of being assigned to receive ERA services, including extra financial incentives, in place of NDLP. The Intake Clerk will explain the ERA programme to the lone parent and how random assignment works and invite the lone parent to take part in the Demonstration. If the lone parent agrees to participate, his/her written consent to participating in the Demonstration, which involves agreement to being randomly assigned and taking part in the research, will be recorded on the BIF. For those declining to give their consent, the reason(s) for refusal will be noted on the BIF. Individuals who refuse to give consent can enter NDLP as normal, but they will not be part of the Demonstration and will not receive ERA services. Again, it is important that participation in the Demonstration is presented to customers in a highly positive manner and that very few customers refuse consent. Scripts will be available to programme staff to help them to do this. After consent is obtained, individuals will then be randomly assigned in the same way as the ND25plus group. After random assignment, the Intake Clerk will inform the individual as to whether they are in the programme or control group, mark the BIF accordingly, and set up the appropriate appointment.

¹⁴ It is possible, however, that the adviser who interviews the lone parent during their work-focused interview may undertake the ERA induction process if this proves more convenient for the customer.

¹⁵ The BIF used for lone parent target groups will need to be slightly different in layout and content to that used for the ND25plus target group. This is because it is assumed that lone parents can refuse to participate in the Demonstration for a wider range of reasons than the ND25plus target group: for example, because they object to random assignment.



The random assignment design for the NDLP target group seeks to compare the extent of work retention and employment advancement for the ERA programme group with a control group that receives standard services through NDLP. As with the ND25plus target group, random assignment will take place as close as possible to the receipt of services after lone parents *have* agreed to join the programme and be randomly assigned. If random assignment were to occur before an individual agreed to participate in any programme, large numbers of randomly assigned lone parents might opt-out of services. The random assignment design aims to minimise this and maximise the numbers of randomly assigned individuals actually exposed to services.

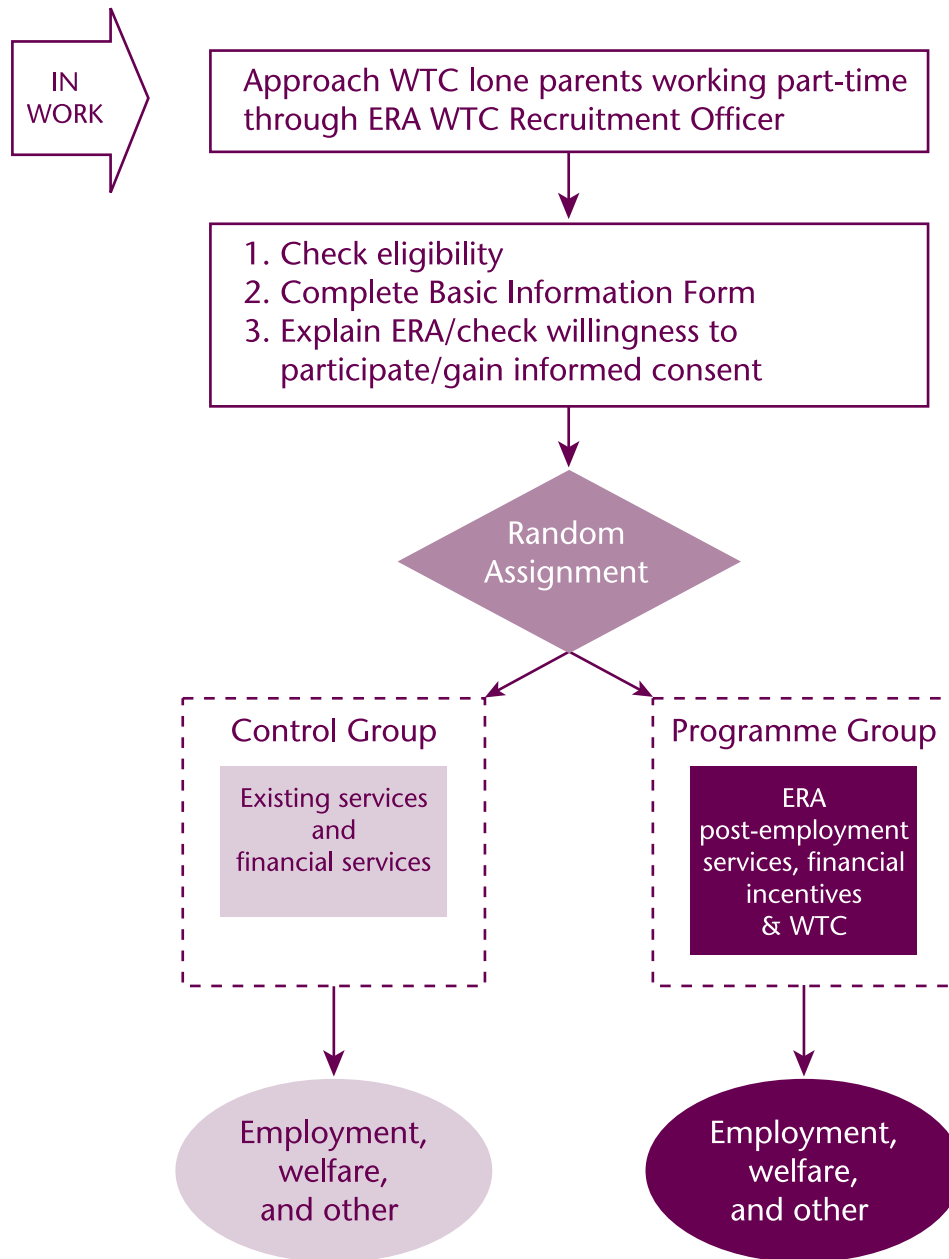
Under an alternative approach, one that is not recommended, random assignment would take place prior to lone parents indicating their willingness to join *any* type of programme. For example, individuals might be randomly assigned prior to

attending a Work Focused Interview.

Those assigned to the programme group would then be offered ERA services, while those assigned to the control group would be offered NDLP services. Such a design would involve randomly assigning the entire eligible caseload, rather than just those who had expressed an interest in participating. From an impact study perspective, such a design would allow the evaluator to compare the rate at which individuals offered the two sets of services chose to participate in them, as well as enable the estimation of programme impacts. Random assignment would ensure that all these estimates were unbiased, as long as impacts were estimated by comparing outcomes between the entire programme group and the entire control group, including those who have opted-out. The major drawback of this design, however, is that it would require randomly assigning a very large number of people and involve selecting many more sites than envisaged here, at considerable extra cost.¹⁶

¹⁶ A similar design could be used for the WFTC LP target group.

Figure 2: Random assignment design for WTC LPs





WTC **Part-time** Lone Parents target group

Figure 2 on the previous page, illustrates the random assignment design for the WTC lone parent target group. The intake process and research design for this target group are notably different from that outlined for the two New Deal eligible target groups. The most important points of difference are:

- this group will be in part-time work (16-29 hours per week) at the time of random assignment;
- many WTC LPs will have no existing point of contact with Jobcentre Plus; and
- WTC LPs will be recruited into the programme by peripatetic ERA Demonstration WTC Recruitment Officers.

The intake of WTC LPs into the ERA programme commences with a sample of 'stock' cases drawn from WTC records. In addition, in each quarter during the intake period, lone parents working part time and making new claims for WTC will become eligible for ERA services.

Each individual sampled from WTC records will be selected so that they meet all the following criteria:

- a *new* WTC claim during the ERA Demonstration intake process, *or* an *existing* WTC claim at the start of the programme;
- a lone parent with at least one dependent child;
- live in an area where the ERA Demonstration is running; and
- working more than 16 hours a week, but less than 30 hours a week (that is part time).

Once individuals meeting these criteria are identified, they will be written to and told about the ERA Demonstration and informed that a WTC Recruitment Officer will make contact with them to invite them to take part. This approach letter is important and should present the Demonstration in a highly positive light, emphasising the benefits of taking part. ERA WTC Recruitment Officers will, after the despatch of the letter, attempt to visit each eligible individual in their own home. During the home visit, the WTC Recruitment Officer will:

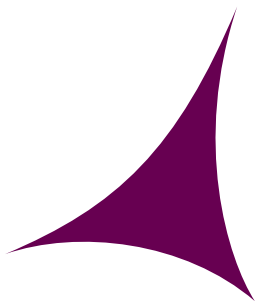
- check the individual's eligibility for ERA services – only lone parents working part time and meeting the criteria above can enter the programme¹⁷;
- help the individual complete a BIF – the BIF will record the individual's consent to enter the programme; and
- explain the programme and random assignment to eligible individuals, and try to encourage and/or convince them to take part.¹⁸

For those individuals who give their consent, the WTC Recruitment Officer will then use a mobile telephone to call the ERA Database Controller. In a similar manner to the New Deal eligible target groups, individuals will be randomly assigned and their details added to the ERA Evaluation Database. The WTC Recruitment Officer will inform the individual of their random assignment status and note this on the BIF. If the individual is assigned to the programme group, the WTC Recruitment Officer will set-up an appointment for the individual to meet their ASA.

¹⁷ Those whom the WTC Recruitment Officer finds to be working full time will not be eligible for ERA services.

Those who are found to be out of work will be encouraged to enter the New Deal where appropriate.

¹⁸ Those who refuse to take part in the research will not be allowed to enter the ERA programme and be randomly assigned.



ERA services for the WTC LP target group are aimed at those already in work and place a strong emphasis on job advancement. Those assigned to the control group will be able to access existing services, *but will not* receive services through the ERA programme. The research is designed to compare retention and advancement for those in the ERA programme group with that observed among the control group, many of whom will receive no in-work support other than Tax Credits. Those in the ERA programme group will receive Tax Credits, as well as ERA services and incentives.

In a manner similar to the New Deal eligible groups, the random assignment design for WTC LPs seeks to assign individuals after they have chosen to participate in the programme and agreed to be randomly assigned as well as take part in the research.

Assigning individuals randomly – the assignment algorithm

Several different approaches can be used to construct a random assignment algorithm to assign individuals to programme or control status. These alternative approaches are discussed in some detail in Annex 2. For reasons outlined in Annex 2, the alternative that seems to make the most sense for the ERA Demonstration is to establish a single sequence of blocks of a random length, which is used to assign individuals from each site and target group.

To illustrate, the sequence of blocks under this approach might look something like this:

PPCC, CP, PCPCPC, CPCP, CCpp, CPPCPC, CP,...

where 'P' represents an assignment to the programme group and 'C' represents assignment to the control group. Both the

ordering of the Ps and the Cs within each block and the length of each block would be randomly determined, but the number of Ps and Cs within each block would always be equal.

As pointed out in Annex 2, because neither Jobcentre Plus staff nor members of the ERA target population would have knowledge of the sequence of block lengths or the order of Ps and Cs within each block, it would be virtually impossible to manipulate the assignment process. The same sequence could be used for all six sites, but each site would begin at a different block.

The same sequence would be used to assign individuals randomly from all three target groups. To illustrate, imagine that the first three individuals who are randomly assigned at a particular site are from the New Deal 25plus target group, the next two are from the NDLP target group, and the next two are from the New Deal 25plus target group. Using the illustrative sequence of blocks appearing above, four of the five New Deal 25plus individuals would be assigned to the programme group, while both of the NDLP individuals would be assigned the control group. Given the 'Law of Large Numbers', however, it is likely that by the end of the intake process, after hundreds of individuals from each group have been randomly assigned, the numbers assigned to the programme and control groups, within each target group, at each site, will be approximately in balance. It is unlikely, however, that an exact 50:50 ratio will be obtained.

Annex 2 also discusses possible adjustments to the random assignment process if the flow of clients into the programme group at a particular site is greater than anticipated and, as a consequence, the site's ERA programme



is at risk of becoming overwhelmed by the number of entrants. The suggested method for handling this problem, if it actually occurs, is to randomly exclude some specified proportion of the intake from the study sample. This so-called ‘non-research group’ would receive the same services as the control group, but outcome data would not be collected on those in the group. It is recommended that the potential for establishing a non-research group be built into the random assignment process.

Minimum Detectable Effects

This section explores the size of impacts that will be detectable given the research design and the size of survey samples set out in Chapter 6 of this report. The concept of a Minimum Detectable Effect (MDE) is used to explore the size of impacts that are likely to be observed or detected for a selection of outcomes. Orr (1999) defines an MDE as:

“the smallest true impact that would be found to be statistically significantly different from zero at a specified level of significance with specified power” (Orr 1999: 112).

The focus here is on the MDEs that can be obtained from survey data. As a result, the MDEs reported here are larger than those that would be obtained from similar analyses using administrative data (because administrative samples are larger), and thus present a conservative estimate of the size of impacts the research design is capable of detecting. Of the three survey waves specified in Chapter 6 of this report (12, 24 and 60 months after random assignment follow-up surveys), MDEs for the 24-month follow-up survey are considered here. As a result of sample attrition, the achieved sample at 24 months will be smaller than that achieved at 12 months. The sample achieved at the 60-month follow-up will be

even smaller. Unlike the 24-month survey, however, there is little information available that can help in estimating the likely response rates at 60 months. Taking account of these issues, it appears that MDEs estimated on the expected size of the survey sample at 24 months will be most informative in helping judge the overall sensitivity of the research design. All the estimates presented here are subject to an appreciable level of uncertainty, however, and should be considered indicative rather than exact.

The following information is required to estimate MDEs for each outcome:

- an estimate of the outcome’s variance;
- type of statistical tests that will be applied to the data (for example, will impacts be estimated through a simple comparison of mean differences, or estimated using a linear regression model);
- proportion of the sample assigned to the programme group;
- total sample size (programme and control groups combined); and
- levels of statistical significance and statistical power required, as well as whether a one- or two-tailed test will be applied to the data.

The random assignment design, discussed above, requires that half the individuals entering the demonstration are assigned to the programme group. The sizes of the survey samples are set out in Chapter 6 of this report. This section discusses the required levels of statistical significance and power, the nature of the statistical tests used to estimate impacts, and the estimation of outcome variances.

An equation for the calculation of an MDE for a given impact can be written as:

$$\text{MDE} = z \sqrt{\frac{\sigma^2(1-R^2)}{p(1-p)n}}$$

Where:

'z' represents the sum of the Z values, drawn from a cumulative normal distribution of mean zero, for one minus the assumed level of statistical significance of the test and one minus statistical power;

' σ^2 ' the population variance of the outcome;

' R^2 ' the estimated explanatory power of the linear regression model used to estimate programme impacts (it is assumed that all estimated impacts will be regression adjusted)¹⁹;

'p' the proportion of the sample assigned to the programme group; and

'n' the total sample size – programme and control groups combined.

The ideal is to get an MDE with a value as small as possible or at least a value that corresponds to an impact of a reasonable magnitude, given what is known about the programme and target groups. The formula shows that the MDE varies directly with ' σ^2 ' and inversely with 'n'. In other words, if the variance is small, the minimum detectable effect will be smaller than if the variance had a larger value. Conversely, a larger sample size will result in a smaller MDE than a smaller sample.

Given the random assignment research design, consisting of 50:50 assignment to a programme and control group, 'p' will be equal to '0.5'. As discussed below, for MDEs calculated assuming pooled impact estimates

on data collected at the 24-month survey, 'n' will equal 1,600 (see Chapter 6).

In what follows, ' R^2 ' is set at 0.2 for all experimental impacts which are assumed to be regression adjusted.

The next element of the calculation involves determining the value of 'z'. To do this, the following decisions have to be made:

- What is an acceptable rate for a Type 1 statistical error? In other words, the highest acceptable probability of wrongly inferring that the programme has an impact when, in fact, it has none.
- What is an acceptable rate for Type 2 statistical error? That is, the highest acceptable probability of wrongly inferring that the programme has not had an impact, when in fact it has.
- Whether the statistical impact estimates are based on a one- or two-tailed statistical test.

The choice of statistical error rates implies a trade off between making Type 1 and Type 2 errors. Accepting an increased risk of making either type of statistical error reduces, *ceteris paribus*, the estimated MDE. It is common practice among researchers to accept a five per cent risk of making a Type 1 error. Assuming the costs associated with making either type of error are equal, this implies setting both error rates to five per cent.²⁰

It is unlikely, however, that policy-makers will assign equal risk in cost terms to making Type 1 and 2 errors. This is particularly the case in the context of a demonstration project or policy pilot, where the objective is to test services that can then be introduced more widely if found to be cost effective. Making a Type 1 error involves incurring the cost of introducing a programme that does not work and therefore wasting resources.

¹⁹ For the purpose of estimating MDEs for outcomes measured as proportions, a linear probability model is assumed.

²⁰ Note that an error rate of five per cent for a Type 1 error implies statistical significance of 95 per cent, while an error rate of five per cent for a Type 2 error implies statistical power of 95 per cent.



If one were to make a Type 2 error and not introduce a new policy that does work, the benefits of the programme will not be enjoyed, but at the same time programme costs will not be incurred. This implies acceptance of a higher risk of Type 2 error, because making such an error is less costly (Orr 1999). The MDE calculations below assume a five per cent level of statistical significance and an 80 per cent level of statistical power (20 per cent probability of Type 2 error). In other words, the accepted risk of making a Type 2 error is four times that of making a Type 1 error. Error rates of this magnitude are typical of those assumed in many North American welfare-to-work experiments.

Statistical tests can be conducted assuming either a one- or two-tailed test. As Bloom states, however:

"the main goal of a program impact study should be to determine whether or not a program produced the results it was intended to" (Bloom 1995: 554)

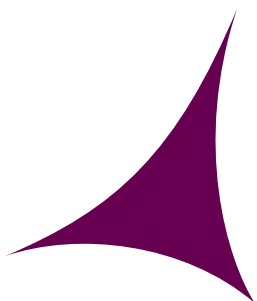
ERA services aim to improve retention and advancement. On this basis, the alternative hypothesis to the null or 'zero finding' hypothesis, is that ERA services have led to changes in outcomes consistent with improvements in retention and advancement. This implies the use of a one-tailed statistical test, as interest is in determining whether ERA services have met their objective. One-tailed tests have the advantage of being able to detect smaller MDEs and thus have greater statistical power for a given sample size.

In the following section, two broad types of estimates are discussed: site-specific MDEs, where a site is equivalent to a Jobcentre Plus District and pooled MDEs. Site-specific MDEs are calculated using estimates of the sample

sizes expected at each site and assuming that variances are the same from site to site. Pooled MDEs are calculated on the basis of the expected size of the total sample across all sites. In order to estimate MDEs using the equation above, an estimate of ' σ^2 ' the outcome's population variance, is required for each impact that is being measured. Ideally, each estimate of ' σ^2 ' (that is, the estimate for a particular outcome, for a target group, at either a single site or where data are pooled across sites) should be as close as possible to the estimate that will be obtained from the actual 24-month follow-up survey data. Obviously, the variances that will be obtained from the survey data cannot be known in advance. Consequently, existing data sources are used to estimate the size of outcome variances.

Estimated variances for each main outcome of interest for each target group have been obtained from a number of sources. For the ND25plus target group, these estimates come from surveys carried out as part of the evaluation of the ND25plus pilots. For the WTC LP group, variance estimates have been obtained from an analysis of the first two waves of the Survey of Low Income Families (SOLIF). For the NDLP target group, estimates were derived from an analysis of data collected through the Lone Parent Personal Adviser Meetings evaluation.

In considering how reliable these estimated variances are, it is useful to speculate on the sources of variance that will affect site-specific and pooled impacts at the 24-month follow-up survey. Before doing so, however, it is worth pointing out that because the variance estimates used here are estimated on existing survey data assuming a simple random sample design, when in fact the design of these surveys is more complex, it is likely that, from the outset, the variance estimates used here are under-estimates.



Thinking about site-specific impacts, it is likely that individuals within each target group, at each site, will be relatively homogenous group and that samples will be drawn so that each randomised individual in a site has a known and equal chance of selection for the survey. These features would suggest that the variance estimates used here for site-specific outcomes are likely, in most cases, to be over-estimates. The variance estimates used are based on survey data drawn from a variety of geographical locations, that is, from a more heterogeneous population than one would expect to encounter at a typical ERA Demonstration site. Moreover, the samples upon which variance estimates used here have been calculated are much smaller for WTC LPs than the expected achieved site sample for the actual impact study at 24 months, and slightly smaller for NDLP eligibles. The reverse is the case for ND25plus eligibles. Taken together, these factors suggest that the site-specific MDEs presented below are conservative in the case of WFTC LPs and NDLP target groups and potentially a little optimistic for the ND25plus eligibles.

For pooled impacts, the situation is more complex. The within-site variance at each site and the variance across sites will together determine the variance of outcomes estimated on pooled data. As already noted, the variance estimates obtained here are based on samples drawn from a variety of geographical locations and, for each target group, based on smaller samples than the expected size of the pooled sample. These factors suggest that the variance estimates are probably larger than those likely to be encountered in the actual impact study at 24-month follow-up survey. This is true for all three target groups. On the other hand, the estimated variances used here have not been adjusted to account for the fact that

individual cases in the pooled 24-month follow-up survey data will not be sampled with equal probabilities. This feature works in the opposite direction. Taking into account these sources of variance, it is likely that the MDEs for pooled impacts presented below are over-estimates and thus reasonably conservative.

Minimum Detectable Effects for experimental impacts

This section presents MDE estimates for impacts estimated on survey data collected 24 months after random assignment. The MDEs presented are those for impacts estimated experimentally; that is, impacts estimated on the whole experimental sample, comparing average outcomes in the programme group with those in the control group. As previously discussed, some key impacts are best measured non-experimentally, for example, differences in hourly wage rates. A separate section below considers the best approach to measuring outcomes in this way. The focus here is on outcomes such as employment rates, annual earnings and weeks employed; in other words, outcomes that lend themselves to being estimated experimentally.

The survey design set out in Chapter 6 of this report suggests an expected pooled sample size of some 4,800 individuals at 24 months after randomisation, based on an assumed minimum acceptable response rate of 65 per cent. It is also assumed that equal numbers are sampled from each target group. On this basis, it is expected that the final 24-month sample will contain approximately 1,600 individuals from each target group. As Chapter 6 explains, however, the objective is to achieve a response rate somewhat in excess of 65 per cent.



Table 7 sets out the MDEs for each target group, for three outcomes, measured at the 24-month survey, for pooled impacts. The impact of ERA services can be measured by comparing the proportion of programme members who are in work with the proportion in the control group. For all three target groups, it is predicted that the sample at 24 months will be sufficient in size to detect programme impacts on being in work of five percentage points. In other words, an impact where the proportion in work in the programme group is five percentage points higher than the proportion in the control group can be detected. An impact smaller than five percentage points will not be detectable at 95 per cent level of statistical significance. Likewise, the sample is sufficient to detect an increase in earnings of £211 per annum in the programme group, above earnings in the control group, for the WTC LP target group. Similarly, for the ND25plus and NDLP target groups, programme impacts of at least £346 and £248 are detectable (however, see note 2 in Table 7).

In a similar manner to Table 7, Table 8 sets out MDEs for site-specific impacts. The same variance estimates are used here as for the pooled impacts. The only difference between the two sets of calculations is that for the site-specific MDEs, a sample size at the 24-month follow-up of 270 achieved interviews in each target group is assumed. The sample size estimates presented in Chapter 6 of this report, show that approximately 270 individuals at each experimental site, in each target group, are expected to respond to the 24-month survey (assuming equal numbers are sampled at each site).

Table 7: Minimum Detectable Effects for selected impacts estimated on pooled data (sample size = 1,600 for each target group)

	Proportion in work	Mean annual net earnings (£s)	Mean number of weeks employed (Weeks)
New Deal 25plus target group ⁽¹⁾			
Standard deviation	0.42	£3,111.77	17.5 weeks
Mean	0.23	£1,073.59	7.7 weeks
Minimum Detectable Effect	0.05	£346	2.0 weeks
New Deal for Lone Parents target group ⁽²⁾			
Standard deviation	0.47	£1,117.76	9.5 weeks
Mean	0.32	£587.05	6.3 weeks
Minimum Detectable Effect	0.05	£248	2.0 weeks
Lone Parents working part-time on WTC ⁽³⁾			
Standard deviation	0.45	£1,899.54	14.8 weeks
Mean	0.72	£3,999.57	45.5 weeks
Minimum Detectable Effect	0.05	£211	1.6 weeks

Notes: The estimates presented for earnings and weeks worked include those who did not work during the period of observation in the sample

1. Data come from the evaluation of the New Deal for Long-term Unemployed prototype evaluation. The proportion of individuals in work is measured approximately 12 months after the sample became eligible for the New Deal. (Lissenburgh, 2000: 35, Table 3.1) Variances for earnings and weeks employed are estimated over a 12 month period.
2. Data come from the evaluation of lone parent Personal Adviser (PA) meetings. The proportion in work is measured six months after individuals attended a PA meeting. Variances for earnings and weeks employed are estimated over a six month period. As a rough approximation, the MDEs presented are doubled (not the standard deviations) to reflect the fact that the variances were measured over six months rather than 12.
3. The proportion in work at 12 months is an approximation, estimated on benefits data provided by DWP. Estimates for earnings and weeks employed come from an analysis of the first two waves (12 months apart) of the Survey of Low Income Families.



Table 8: Minimum Detectable Effects for main impacts estimated on site-specific data (sample size = 270 for each target group per site)

	Proportion in work	Mean annual net earnings (£s)	Mean number of weeks employed (Weeks)
New Deal 25plus target group	0.11	£842	4.7 weeks
New Deal for Lone Parents target group ⁽¹⁾	0.13	£606	5.1 weeks
Lone Parents working part time on WTC	0.12	£514	4.0 weeks

Note: Variances used here are the same as those used for the pooled MDE estimates at Table 7.

1. As with Table 7, variance estimates are for a six-month period. In order to obtain roughly comparable estimates with those for the WTC and ND25plus group, the estimated MDEs for earnings and weeks employed have been doubled.

The site-specific MDEs are much larger than those calculated for the pooled impacts, because survey sample sizes are much smaller at the site level. For example, the ND25plus target group, site-specific MDEs for earnings are over twice those calculated for the pooled sample. A similar difference in the magnitude of MDEs can be seen for each target group across each outcome. The important question is whether ERA services can be expected to generate impacts of this size. It may be that one or two sites might produce impacts that are substantially larger than the average across all sites. This means that estimating site-specific impacts on survey data is worth doing in order to ascertain whether certain sites are producing unusually large impacts, even though the MDEs in Table 8 are large compared to those for the pooled sample. It should also be born in mind that administrative data samples will be much larger at the site level than the site-specific survey sample sizes reported here. This means that impacts estimated on the basis of administrative samples will have lower MDEs, albeit for a narrower range of outcomes.

Subgroups

Chapter 6 in this report provides estimates of the size of key subgroups expected at the 24-month follow-up survey (see Table 13). These estimated sample sizes are used to calculate MDEs for two key pairs of subgroup comparisons: those with and without qualifications, and white and non-white groups. Administrative data will be required to measure subgroup impacts in most cases, though where a wider range of outcome measures are required, subgroup impacts will need to be estimated using survey data.

Table 9 reports the MDEs for earnings and weeks worked. There are a number of issues that need to be considered. First, the variance estimates come from the same sources as those used in Table 7. Second, as previously mentioned (note 2, Table 7), those presented for the NDLP target group are estimated over a six-month rather than 12-month period. As a result, the MDEs



reported for the NDLP target group have been doubled. Lastly, the variance estimates presented in this table for the non-white subgroup are estimates on a very small sample and are therefore unlikely to be reliable.

Policy interest is likely to focus on the impacts of the programme on those without qualifications and the impact on those from non-white ethnic groups. Table 9 shows that, because of low sample numbers, difficulties will be encountered detecting programme impacts for those with no qualifications, particularly for earnings in the ND25plus target group. As a result, impact estimates for those with no qualifications are likely to be best estimated using administrative data for a limited range of outcomes. Alternatively, a booster sample for all three target groups could be selected. Setting aside the fact that the variance estimates for the non-white ethnic groups are probably unreliable, the MDEs reported for this group are consistently larger than ideal and points to the need for an ethnic minority booster sample.

To detect a programme impact of £300 in net annual earnings for the non-white subgroup in the ND25plus target group, a sample of approximately 900 ethnic minority respondents would be required at the 24-month follow-up survey. Assuming that the estimated variances for non-white ethnic groups are valid (and there is doubt about this), it should be possible to detect an impact of 2.2 additional weeks worked for the non-white ND25plus group with a sample of 900 individuals. For a sample of the same size for the NDLP target group, it should be possible to detect annual earnings impacts of £337 and three additional weeks worked. Achieving 900 interviews for non-

white members of the ND25plus and NDLP target groups will require booster samples of 675 and 756 individuals respectively.

Table 3, which is found in Chapter 2, indicates that approximately 11,373 individuals in the ND25plus target group will be randomly assigned. This means that ethnic minorities in the ND25plus group will be sampled from a total population of 1,593 (see Table 4). Thus it will be necessary to select approximately 1,385 of these individuals to achieve 900 interviews, given assumed survey response rates.

Following a similar strategy for the NDLP target group presents a problem. It is estimated that the NDLP target group will consist of a total of 483 randomised individuals from non-white groups. Sampling all these individuals at 24 months would achieve some 314 interviews with non-white lone parents, short of the target of 900. The low numbers of non-white lone parents in the experiment suggests that it is only realistic to draw a booster sample of ethnic minorities from the ND25plus target group and highlights the difficulties likely to be experienced in measuring impacts for non-white lone parent subgroups even using administrative data.

Finally, the issue of older workers needs to be considered. There is particular policy interest in encouraging older workers, who are long-term unemployed, to re-engage with the labour market. For this reason, Table 4 reports the number of individuals in the ND25plus target group over 50 years old. At the 24-month follow-up, it is anticipated that survey interviews will be achieved with 385 individuals in this group (see Table 13).



Table 9: Minimum Detectable Effects for selected subgroups and outcomes,
New Deal eligible target groups

	ND25plus target group		NDLP target group ⁽¹⁾	
	Mean annual net earnings (£s)	Mean number of weeks employed (Weeks)	Mean annual net earnings (£s)	Mean number of weeks employed (Weeks)
Educational qualifications				
Qualifications				
Sample size N=	1011	1011	963	963
Mean	£1,223.58	8.4 weeks	£786.27	7.0 weeks
Standard deviation	£3,487.28	18.3 weeks	£1,329.84	9.6 weeks
MDE	£488	2.6 weeks	£381	2.8 weeks
No qualifications				
Sample size N=	594	594	642	642
Mean	£795.79	6.4 weeks	£285.75	5.2 weeks
Standard deviation	£7,011.88	15.8 weeks	£569.24	9.3 weeks
MDE	£1,280	2.9 weeks	£200	3.3 weeks
Ethnic group				
Non-white				
Sample size N=	225	225	144	144
Mean	£610.81	5.1 weeks	£685.50	7.0 weeks
Standard deviation	£2,001.72 ⁽²⁾	14.8 weeks ⁽²⁾	£1,136.25 ⁽²⁾	10.0 weeks ⁽²⁾
MDE	£594	4.4 weeks	£842	7.4 weeks
White				
Sample size N=	1,380	1,380	1,461	1461
Mean	£1,103.37	7.9 weeks	£578.63	6.2 weeks
Standard deviation	£3,159.70	17.6 weeks	£1,118.88	9.5 weeks
MDE	£378	2.1 weeks	£260	2.2 weeks

Notes: See Table 13 for estimated subgroup sizes and sources. Variance estimates come from the same data sources as those used for estimated variances and MDEs presented in Table 7.

1. Because variance estimates for the NDLP target group were estimated over a six-month period, the MDEs for the groups presented in this table have been doubled (standard deviations remain those estimated over a six month period).

2. Variance estimates are probably unreliable as they are based on a sample of less than 50 cases.

Table 10: Minimum Detectable Effects for under and over 50 subgroups – ND25plus target group

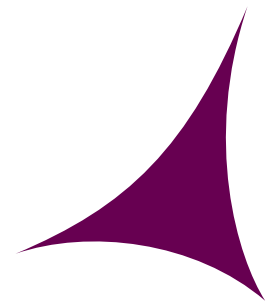
	ND25plus target group		
	Proportion in work ⁽¹⁾	Mean annual net earnings (£s)	Mean number of weeks employed (Weeks)
Age			
50 and over			
Sample size N=	385	385	385
Mean	0.21	£930.56	8.8 weeks
Standard deviation	0.41	£2,495.64	18.6 weeks
MDE	0.09	£566	4.2 weeks
Under 50 years			
Sample size N=	1,220	1,220	1,220
Mean	0.19	£1,119.02	7.3 weeks
Standard deviation	0.39	£3,284.29	17 weeks
MDE	0.05	£418	2.2 weeks

Notes: Variance estimates obtained from survey data collected as part of the evaluation of the New Deal for Long-term Unemployed Pilots.

1. Variance for the proportion in work measured at 12 months

Table 10, sets out the MDEs estimated for older workers and shows that they are quite large, again implying the need for a booster sample. Detecting a difference in rates of employment of around five to six percentage points between those aged 50 and over in the programme group and their counterparts in the control group, would require a total sample of some 1,100 individuals. This would imply achieving an additional 715 interviews with those aged over 50 at the 24 month follow-up and sampling around 1,700 in total from this group. Such a booster sample, as in the case of the ethnic minorities' booster sample, has cost implications.

One final point on the MDEs is worth making. As mentioned in Chapter 2, the minimum sample sizes to detect site-specific MDEs should ideally be around 1,000 customers per target group per site, in order to ensure that impacts of a reasonable size at that particular site can be detected. The constraints referred to in Chapter 2, particularly the need to select a Scottish and Welsh site, means that administrative sample sizes will be below this ideal minimum in some sites. Additionally, other pilot projects within the selected sites that target the same groups as ERA could also affect sample numbers. If sample sizes prove to be substantially less than anticipated, then it may be necessary to run the project intake process for a longer period of time in order to increase the inflow of participants into the Demonstration.



Comparing impacts across sites

As previously discussed, impacts will be computed separately for each of the six ERA programme sites, as well as for samples that are pooled over the six sites, for each target group and for subgroups. It is likely that the impact estimates for the individual sites will vary from one another. Thus, it will be of interest to ask whether the differences in impacts among the sites are statistically significant or due only to sampling error. If the differences in impacts between sites are statistically significant, then it will be of interest to explore the sources of such difference, for example, variability in the state of local labour markets, in the characteristics of participants and different approaches to programme implementation and operation. The process study described earlier will focus on explaining such differences.

The meta-analysis literature provides a very simple method for examining whether observed differences in impacts among sites are statistically significant, which is known as the 'homogeneity test' (for example, see Hedges, 1984 or Lipsey and Wilson, 2001). A homogeneity test relies on the Q statistic, where Q is the weighted sum of squares of the estimated programme impacts for the individual sites, ' E_i ', about the weighted mean effect, ' \bar{E} ', and where the weights are the inverse of the square of the standard errors (or sample variances) of the estimated programme effects, or ' $\frac{1}{w_i}$ '. Thus the formula for Q is:

$$Q = \sum \frac{1}{w_i} (E_i - \bar{E})^2$$

where:

$$\bar{E} = \sum \frac{1}{w_i} E_i / \sum \frac{1}{w_i}$$

and:

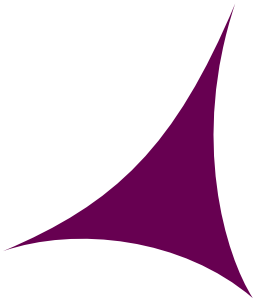
$i=1,2,3,4,5,6$ where there are six sites.

'Q' is distributed as a chi-square statistic with degrees of freedom equal to one less than the number of programme effect estimates—thus five in the case of the six ERA sites.

If 'Q' is below the critical chi-square value, then the distribution in the programme impact estimates around their mean is no greater than that expected from sampling error alone. If the null test of homogeneity is rejected (i.e. 'Q' exceeds the critical value), this implies that there are differences among the programme impact estimates that are due to systematic differences among the sites (for example, differences in programme or participant characteristics or site environmental characteristics), and is not due just to sampling error.

If the null test of homogeneity were rejected with all six ERA sites then the test should be repeated with the site for which $(E_i - \bar{E})$ is largest (i.e. the most heterogeneous site) excluded. If the null test of homogeneity were again rejected, then the test should again be repeated but with only the four most homogeneous sites included, and so forth. Ultimately, these homogeneity tests would indicate which sites, if any, have produced an impact estimate that differs statistically from the mean effect.

The formula for the Q test can be used to address a question that is analogous to the minimum detectable effect (MDE) discussed above: what is the minimum detectable difference in effects (MDDE) among the sites?



To make this issue tractable, it is helpful to make two simplifying assumptions: (1) the estimated programme effects for five of the six sites are equal to one another, but the estimated effect for the sixth site is larger than the others; (2) both the sample sizes and the standard errors (SE) of the impact estimates for all six sites are identical.

Given these two assumptions, and the fact that the ' $w_i = SE^2$ ':

$$Q = \left(\frac{1}{SE^2} \right) (E_i - \bar{E})^2$$

$$Q = \left(\frac{1}{SE^2} \right) (D^2 + 5(D/5)^2)$$

$$Q = \left(\frac{1}{SE^2} \right) (D^2 + 0.2D^2)$$

$$Q = \left(\frac{1}{SE^2} \right) (1.2D^2)$$

where ' D ' is the difference between the effect estimate for the outlier site, ' E_0 ', and the weighted mean for all six sites (i.e. $D = E_0 - \bar{E}$).

Setting ' Q ' equal to 11.07, which at the five per cent level of statistical significance is the critical value of the chi-square statistic with five degrees of freedom, and solving for ' D^2 ':

$$D^2 = (11.07/1.2)SE^2$$

$$D = 3.04(SE)$$

' D ' is the difference between the impact estimate for the outlier site and that for the weighted mean for all six sites, which is a bit smaller than the difference between the outlier site and the other five sites—that is the value of the MDDE. To compute the value of the MDDE, it is necessary take account of the fact that because the effect estimate for outlier site is above the mean for the six sites by ' D ', the effect estimates

for the other five sites (which by assumption are equal) must each be below the mean by one-fifth of this amount. Thus:

$$MDDE = D + D/5$$

$$MDDE = 3.04SE + (3.04SE)/5$$

$$MDDE = 3.61SE$$

To see what this means, imagine that there is interest in determining whether programme impacts, in terms of the proportion in work vary among ERA sites by an amount greater than sampling error. It is assumed that in five of the sites the programme produces a two percentage point impact in the percentage in work in the programme group compared to that in the control group, and at each of these five sites, 23 per cent of the control group are in work. Taking the case of survey data collected 24 months post random assignment, we can compute the MDDE between these five sites and a sixth outlier site.

The equation below allows the standard error of the impact estimate at one of the five sites to be computed, where σ^2 represents the variance of outcome, p the probability of being assigned to the programme group and n the total sample size at the 24 month survey, programme and control groups combined. It is assumed that the standard errors of the impacts are equal across the six sites.

$$SE_i = \sqrt{\frac{\sigma^2}{p(1-p)n}}$$

$$SE_i = \sqrt{\frac{0.1771}{0.5(1-0.5)266}}$$

$$= 0.052$$



Multiplying this value by 3.61, as set out above, yields an MDDE of 0.19. This means that unless the outlier site records an impact of 21 percentage points, or greater, in the percentage in work, comparisons between the outlier site and the other five sites will not yield a difference that is statistically significant at the 95 per cent level.

This is clearly not a very sensitive test. The reasons are twofold. First, the 24-month survey sample is expected to achieve interviews with around 260 individuals in each target group at each site – a relatively small sample. Second, the ERA Demonstration comprises only six sites and the power of this statistical test could be improved by adding further sites. Unfortunately, given budgetary constraints there is no scope for adding further sites. However, a comparison of impacts across sites could be made using larger the administrative samples that will be available to the evaluation (see Chapter 6).

Table 3, in the Site Selection chapter above, shows that a total administrative sample of some 27,000 individuals across the three target groups and six sites is expected. The numbers in this sample by target group and site vary considerably. On average, however, we can expect approximately 750 administrative records for the programme group and another 750 for the control group at each site for each target group. In these circumstances, the standard error computed above as 0.052, would fall to 0.022, and the MDDE to eight percentage points, or an impact at the outlier site of 10 percentage points. This is still quite a large impact and therefore a less sensitive test than ideal.

Estimating experimental impacts

Impacts on standard outcomes

As previously discussed, the analysis will require estimation of impacts for a number of standard labour market outcomes. These outcomes include: proportions in work, total number of weeks employed during the follow-up period, hours employed per week or month, average earnings, hourly wage rates while employed, and the receipt of fringe benefits while employed. In addition, it will be important to estimate programme impacts on the receipt of government transfer payments.

Using the baseline data (from BIFs and administrative sources), these impact estimates will be regression-adjusted to reduce random error due to differences among programme and control groups that occur by chance, thereby increasing the statistical precision of the impact estimates.²¹ For impacts on continuous outcomes, such as mean earnings and transfer payments, regression models of the following form would be estimated, using ordinary least squares.

$$Y_i = \alpha + \beta_o P_i + \sum_j \beta_j X_{ji} + \varepsilon_i$$

where:

Y_i = the outcome measure for sample member i ,

P_i = one for programme group members and zero for control group members,

X_{ji} = background characteristic j for sample member i ,

ε_i = a random error term for sample member i ,

β_o = the impact of ERA on the average value of the outcome,

α = the intercept of the regression, and

β_j = the regression coefficient for background characteristic j .

²¹ Given random assignment, regression analysis increases the statistical power of impact estimates beyond that for a simple programme/control group difference of means or proportions; but it does not change the expected values of the impact estimates.



Ordinary Least Squares will be estimated for impacts on continuous outcomes such as earnings; and logistic regression models based on maximum likelihood methods for impacts on binary outcomes such as employed or not.

Impacts on retention and advancement

There are two major challenges involved in measuring these critical outcomes:

- defining measures of retention and advancement, which have not been commonly used in prior random assignment experiments; and
- developing appropriate analytic strategies to estimate impacts.

Also, as has been discussed, a key distinction must be made between target groups that are working at the time of random assignment and target groups that are not working at that time.

Various measures of retention and advancement will be considered during the ERA evaluation. Two examples of possible measures of post-employment experience are: (1) the extent to which sample members remain employed, either with one employer (*job retention*) or with one or more employers over time (*employment retention*); and (2) the extent to which improvements occur in the quality of their jobs, either with one employer (*job advancement*) or with one or more employers (*employment advancement*).

It is useful to think of two different ways to measure these outcomes:

- ‘fixed-period’ measures; and
- ‘spell-based’ measures.

Fixed-period measures of retention or stability indicate the amount of time (months or

weeks) that sample members are employed – and the pattern of their employment (and non-employment) – *during a specified follow-up period* (for example, 12 months). Fixed-period measures of advancement indicate the change in the quality of jobs (e.g., the hourly wage rate) that occurs during a specified follow-up period. Spell-based retention or stability measures indicate the *duration* of specific jobs or employment spells. Spell-based measures of advancement indicate the corresponding change in the quality of jobs.

With experimental data, where programme and control group members do not differ systematically at baseline, *fixed period* measures of job or employment retention or advancement are relatively straightforward. For example, employment retention can be measured in an unbiased manner as the regression-adjusted difference between programme and control group members in the number of weeks or months employed during the post-random assignment follow-up period.

With experimental data and when a programme enrolls only working people, such as the WTC LP group, it is also relatively straightforward to estimate *spell-based* impacts for the duration of the first employment spell at time of random assignment. An example is the regression-adjusted difference in the mean duration of the first job held by programme and control group members employed at baseline. For second and subsequent employment spells, however, impacts are not so easily derived because the programme intervention may affect the time at which second and subsequent spells begin. However, as discussed below, alternative impact measures can be used to capture behaviour during these subsequent spells.



Even the first job spell could extend beyond the follow-up period. As a consequence, its full duration will be unknown for some sample members, and, therefore, hazard rate models (Allison, 1995) are needed to account for these incomplete or censored spells. Discrete-time and continuous-time hazard rate models are used to study the timing of spells when data for the end of these spells are missing for some but not all cases. This is a common situation for studies that measure programme impacts on the duration of activities (e.g., employment or benefit receipt), the duration of a condition (e.g., living in poverty), or the elapsed time before an event occurs (such as beginning a job). For some cases, the event that completes the spell occurs after the follow-up period and its timing must be predicted from the time pattern of events observed for other cases.

For individuals *not employed at baseline*, such as New Deal eligibles, more complex non-experimental methods must be used to estimate statistically valid impacts on employment retention. However, simple methods, though biased, may nevertheless be informative. For example, suppose the programme succeeds in speeding up the entry into employment of more disadvantaged members of the caseload. In this case, on average, employed programme group members will be more disadvantaged than employed control group members. Because more disadvantaged persons might be expected to have shorter spells of employment than less disadvantaged persons, the programme/control group differences in length of the first employment spell after baseline would tend to be negatively biased. If the programme/control group difference in retention were positive,

that would be evidence that the true impact is even more positive.

When the programme enrolls people who are not employed at baseline, measures of retention will reflect two different effects: (1) the effect on the rate and speed of job-taking and (2) the effect on job (or employment) stability *once employed*. To help pull apart these effects, while retaining the strengths of an experimental design, measures that have been developed in previous evaluations can be used. These measures first assess the employment effect, and then estimate how much of the effect is due to stable (or unstable) employment.

For example, consider the findings in Table 11 from recent evaluations of the NEWWS programmes in Portland and Riverside. To study employment retention, the evaluators first examined the effect of the programmes on the percentage of sample members that were employed *at any time* during the first two years after random assignment. They then separately looked at the effect of the programme on the percentage of *all sample members* that had stable employment and the effect of the programme on the percentage that did not have stable employment.²² As can be seen, both programmes increased employment substantially. However, their impacts on stable versus unstable employment were quite different. Portland had a large impact on stable employment (10.3 percentage points) and almost no effect on unstable employment. This is consistent with the programme's emphasis on 'quality jobs.' In contrast, Riverside increased unstable employment by more than it increased stable employment. This is consistent with its emphasis on taking any job and to enter employment as quickly as possible.

²² The table note describes the definition of stable and unstable employment that was used. In Table 11, the percentage in unstable employment, plus the percentage in stable employment sum to the percentage of all sample members who were employed, and when added to the percentage not employed account for 100 per cent of all sample members.

By comparing the relative impacts of programmes on stable versus unstable employment, one can obtain strong evidence about their impact on employment retention.

The approach described above for estimating impacts on retention can also be used to estimate impacts on job advancement for a programme that enrolls people who are not employed at the time of random assignment. Table 11 could be changed to reflect persons

Table 11: Estimated impacts of two Welfare-to-Work programmes on employment retention

Employment outcome (%)	Programme group	Control group	Difference (impact)
<i>Riverside LFA</i>			
Percent not employed	40.6	54.9	-14.3***
Percent employed	59.4	45.1	14.3***
Unstable employment	36.7	26.8	9.9***
Stable employment	22.7	18.3	4.4***
<i>Portland</i>			
Percent not employed	29.7	39.8	-10.1***
Percent employed	70.3	60.2	10.1***
Unstable employment	33.6	33.8	-0.2
Stable employment	36.7	26.4	10.3***

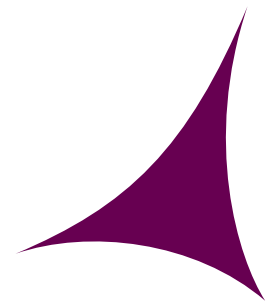
Notes: For this table, stable employment is defined as working in the first or second year after random assignment and working in at least six quarters during years three and four. Unstable employment is defined as working in the first two years after random assignment but not having stable employment.

*** indicates that the impact estimate is statistically significant at the 1 percent level.

Job advancement reflects changes over time in job quality, where quality might be defined in terms of wage rates or fringe benefits. Spell-based measures of advancement refer to changes in job characteristics that occur during a specific job spell or continuous employment spell. If the target population is working at random assignment then changes in job characteristics can be measured from the beginning to the end of the job or the follow-up period, whichever occurs first.²³

whose employment improved versus those whose employment did not improve. There are also experimental methods for estimating programme impacts on wages and earnings that can serve as a proxy for assessing employment quality (see, for example, Card and Robins, 1996 and Lin *et al.*, 1998). These measures construct distributions of earnings, including those with no earnings, to determine whether more of the programme group is working in higher earnings brackets.

²³ This analysis would be less meaningful for fringe benefits, which are unlikely to change in a given job, than for wage rates, which are more likely to change.



Measuring impacts non-experimentally

Although experimentally based estimates of impacts – that is, direct comparisons of outcomes for randomly assigned programme groups with outcomes for the randomly assigned control groups – provide by far the most reliable measures of programme impacts, there will be occasions in the ERA evaluation when obtaining such estimates will not be feasible. For example, in evaluating ERA, it is important to determine whether the programme has had a positive impact on wage rates and wage progression. Because wage rates are only available for individuals who work, the programme group-control group comparisons of wage rates must be limited to such persons. Indeed, examinations of wage progression must rely on individuals in the sample who worked at two separate points in time. Because the ERA programme may influence who it is that works, the characteristics of those in the programme group with jobs might systematically differ from the characteristics of those in the control group with jobs. If so, the comparison between the two groups will not be a randomised comparison; it will instead be made non-experimentally.

Another important non-experimental comparison will probably be required to examine whether full-time work results in greater opportunities for advancement than part-time employment. The provision of R&A bonus payments to those who work full time but not part time, was motivated by the hypothesis that advancement for disadvantaged individuals would be better enhanced if, instead of working part time, they worked full time. However, this hypothesis can only be tested by comparing full-time workers in the programme group with part-time workers in the programme

group in terms of various outcomes that should be associated with full-time work – for example higher wage rates, better fringe benefits, and greater job satisfaction. These comparisons will again be non-experimental, as full-time workers seem likely to differ systematically from part-time workers.

The key problem with non-experimental comparisons is *selection bias* – that is, the possibility that outcomes differ between the groups being compared because their characteristics systematically differ, rather than because of differences resulting from programme effects. For example, if the ERA programme helps those least employable find and maintain employment, this will reduce the average wage rate of the programme group relative to the control group, because the programme group will, on average, have characteristics that are less attractive to employers. This could be due to differences in either ‘observables’ (i.e. characteristics such as age, race, or education) or ‘unobservables’ (characteristics such as motivation and self-esteem).

Approaches that might be used to correct for selection bias in making non-experimental comparisons are described in some detail in Annex 3, with particular emphasis given to how they might be used examine ERA impacts on wage rates and wage progression. Three alternative approaches are suggested:

Assume balancing biases

Here, it is assumed that biases result from restricting the analysis to only members of the sample who work, because such individuals differ from those who do not work. However, it is further assumed that the biases are similar for the programme and control groups. Thus, in comparing the working members of the two groups, the

biases are offsetting and cancel out.

As pointed out in Annex 3, the balancing biases assumption is probably untenable because the ERA programme means that individuals in the programme group face a different set of circumstances than persons in the control group.

Assume that there is selection on the observables, but not the unobservables or, alternatively, that biases resulting from unobservables balance once the observables are taken into account

If this rather strong assumption holds, it is possible to correct for any differences between working members of the programme and control group statistically in a regression framework because the sources of the differences (i.e. the observables) can be measured.

Assume that there is selection on both the observables and the unobservables

In this case, it is necessary to correct for both types of bias. As mentioned under (2), differences between working members of the programme and control group that result from observables can be corrected statistically in a regression framework. As discussed in Annex 3, it might also be possible to correct for differences between the two groups that result from unobservables, by adding a selection term of the sort described by Heckman (1979) to the regression. The selection term itself would be derived from separate probit regression equations, in which employment status is regressed against a set of explanatory variables, which differ from the set of explanatory variables included in the wage rate and the wage progression regressions. The success of this approach depends on how well a set of fairly strong assumptions is satisfied.

Generalising from impact study findings

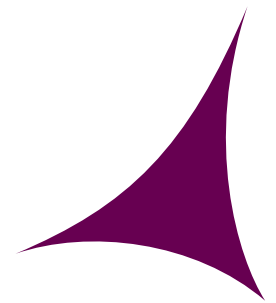
A critical issue in the evaluation of government programmes is 'external validity' – the extent to which estimated programme impacts can be generalised to different locations and populations, to different time periods, and to different variants of the programme being studied. To some extent, the degree to which impact estimates from the ERA Demonstration will be generalisable has been partially addressed in Chapter 2 of this report, which examines the criteria for selecting experimental sites. A number of other important issues related to generalisability are considered here.

Extrapolation to different times and places

This is a serious, if obvious, problem. Social attitudes, government institutions, the business cycle, the relative demand for unskilled and skilled labour, and other relevant factors may change in the years following an evaluation. Likewise, different locations may have dissimilar social attitudes, local government institutions, labour market conditions, and so forth. Moreover, the characteristics of programme participants could differ as well.

Scale bias

The external validity of pilot tests of policy innovations may be compromised by 'scale bias'. Manski and Garfinkel (1992) and Garfinkel, Manski, and Michalopoulos (1992) suggest that when pilot tests are scaled up to universal participation, this could change community norms or combine with patterns of social interaction or information diffusion in ways that will feed back and influence the success of the programme. These community or 'macro' effects, they



argue, will be absent in small-scale pilot programmes or partially scaled programmes. In addition, testing a programme on a small scale may cause the composition of the programme participants to differ from what it would be in a programme after it was rolled out nationally by inhibiting diffusion of information about the programme to potential applicants. Alternatively, the composition of programme participants could be affected in a demonstration such as ERA by discouraging risk-averse individuals from applying to a programme when they could be randomly assigned to the control group (see Heckman 1992, Heckman and Smith 1995, and Manski 1993, 1995).

At present, little is known about the practical importance of these effects. Although the possibility of bias caused by distortion of the participant sample in small-scale pilot tests has strong theoretical appeal, its empirical importance is yet to be demonstrated. This issue is further discussed below in considering 'entry effects'.

One non-experimental approach for avoiding biases caused by testing policy innovations on a small scale is to implement them on a site-wide, fully scaled basis in some locations and, for comparison, use other sites (perhaps statistically matched) that have not adopted the innovation. Although this 'saturation' evaluation design does, in principle, allow feedback effects to be captured, the programme may have to be kept in place for many years, with firm guarantees of permanency, before these effects reach full potency. Moreover, cross-site comparison designs will produce unreliable estimates of programme effects if the programme and comparison sites differ in ways that are inadequately controlled for in the estimation of impacts.

Services received by control group members

It is often the case that some members of control or comparison groups receive services similar to those received by programme group members. For example, in the case of the ERA evaluation, members of New Deal target groups will receive help in securing employment regardless of whether they are assigned to the programme or control group, although this help will differ in some respects. Under these circumstances, estimates of programme impacts do not measure the pure effect of participating in the evaluated programme versus the absence of receiving any similar services at all. Rather, they measure the incremental effect of whatever additional services the programme provides. For example, the ERA programme group will receive two years of post-employment casework services, as well as financial incentives that encourage full-time stable employment and participation in training while working; but the control group will not.

The fact that the services received by the programme and control groups overlap to some degree does not distort programme evaluation findings. As long as the services received by the latter are representative of the 'true counterfactual' 'internal validity' is maintained. If they are, the resulting impact estimates will be clearly policy-relevant. However, the overlap is a source of at least two potential threats to 'external validity'. First, not only will the evaluated programme differ over time or from one place to another, but the array of activities available to comparison group members will also differ, complicating the problem of generalising from the evaluation results. Second, the very existence of the programme being evaluated might change the scale of



services available to the control group.

This second threat to external validity, which Heckman and Smith (1995) call 'substitution bias', could occur if, for example, ERA absorbs resources that would otherwise be available to members of the control group. Alternatively, if, as a result of serving some customers who would otherwise enter the New Deal, ERA frees up resources that can then be used to serve those who enter the New Deal and are assigned to the control group, external validity is also challenged.

Hawthorne effects

The behaviour of participants in a demonstration test of a programme or policy could be influenced by knowledge that they are part of a policy demonstration and not only by the receipt of the tested services – a so-called 'Hawthorne effect'. For example, if ERA participants know that their labour market performance will be measured in terms of certain outcomes, such as stable work patterns, some of them might attempt to succeed in terms of these outcomes.

There is virtually no information about whether Hawthorne effects bias findings from social experiments. It seems possible that members of both the programme and treatment groups could respond similarly to being part of a social experiment. If so, such effects will cancel out in measuring impacts, and there would be no bias. However, if members of the control group respond to the fact that they were assigned to the control, rather than the programme group (for example, by becoming discouraged and putting less effort into finding a job), there would be a bias.

Entry effects

If the services provided by a programme are perceived as beneficial, then some individuals, who are initially ineligible to participate, may adopt behaviours needed to qualify (an 'entry' effect). On the other hand, in the case of mandatory work or training requirements for benefit recipients, individuals might leave benefits when they are informed that they will be subject to the newly-established requirements (an 'exit' effect). Similarly, some individuals who might otherwise have joined the benefit rolls may decide not to do so if they will be required to meet work or training requirements (a 'deterrent' effect).

Manski and Garfinkel (1992) and Moffitt (1992, 1996), among others, have argued that programme entry, exit and deterrent effects could be substantial. However, findings from non-experimental attempts to measure these effects, which have generally relied on aggregate-level time series studies of programme applications, are mixed and inconclusive (for examples, see Johnson, Klepinger, and Dong 1990, Wissoker and Watts 1994, Chang 1996, Phillips 1993, and Schiller and Brasher 1993). There has been only one attempt to use experimental methods to measure entry effects – an evaluation of a Canadian programme that provided very generous earnings supplements to lone parents on welfare who worked full time (Berlin *et al.* 1998). Newly enrolled benefit recipients, who were allocated at random to a programme group, were told that if they remained on welfare for the next 12 months, they would subsequently qualify for earnings supplements provided they then worked full time. The control group was not given this information, as they were not eligible



for the earnings supplement. After a year, about 2.5 per cent more of the programme group were still on welfare compared to the control group.

If rolled out nationally, the ERA programme could potentially cause important entry effects among members of each of the three target groups. First, while individuals must participate in New Deal 25plus after they have been on Jobseeker's Allowance for 18 months, they can volunteer before then. Although not many individuals in receipt of Jobseeker's Allowance currently volunteer, this may change if the opportunity exists to qualify for the financial incentive payments provided by ERA. Second, the New Deal for Lone Parents is a voluntary programme for lone parents who are either not working or working fewer than 16 hours a week. The financial incentives offered by ERA could induce more such individuals to volunteer. Third, lone parents who work part time (between 16 and 30 hours a week) will be able to qualify for ERA incentive payments, but those working full time (over 30 hours) will not. Thus, there will be incentives for lone parents who are currently working full time to temporarily reduce their hours to part time in order to qualify because even if they later increase their hours to above 30, they will continue to qualify for ERA.

None of these entry effects are likely to be important in the pilot test of ERA. Because it will be run in only six sites and enrolment into the pilot test will be limited to a year in most cases, relatively few individuals who qualify for the ERA Demonstration will be sufficiently knowledgeable about it to change their behaviour accordingly. However, this would no longer be the case if the programme were rolled out nationally on a permanent basis. Thus, if entry effects are

important, findings from the pilot test may not generalise to a permanent programme. However, a national roll out of ERA may be accompanied by rules that are specifically designed to minimise entry effects. For example, ERA entry could be restricted to unemployed persons who have been receiving Jobseeker's Allowance for at least 18 months. A similar restriction could be imposed on WTC lone parents who are working part time. Of course, some unemployed persons and part-time workers who desire full-time work might wait for 18 months before taking such jobs. However, the evidence mentioned above for the Canadian programme suggests that this effect is likely to be small. If rules that succeed in limiting entry effects were made part of a national ERA programme, findings from the ERA Demonstration are likely to be better generalise to the permanent programme.

General equilibrium effects

The ERA Demonstration may have important effects on the wellbeing of individuals who are not enrolled in the programme, or at least this would be the case were the programme rolled out nationally. Two such effects are equilibrium wage effects and substitution effects. Empirical evidence about the magnitude of both of these effects is quite limited.

If participants in a programme search harder for jobs or work more weeks or hours than they would otherwise, the resulting increase in labour supply will tend to lower the equilibrium wage within the labour markets in which they work. Thus, workers who are employed in the same labour markets as programme participants could receive lower wages. For this effect to be very large, however, three conditions must hold: (1) the minimum wage must not constrain



downward movements in wage rates; (2) programme participants must account for a fairly large share of the workers in the relevant labour markets; and (3) programme effects on job search and weeks and hours worked must be fairly large.

Even if rolled out nationally, ERA seems unlikely to have substantial equilibrium wage effects. It is anticipated that, at least initially, most participants would be employed in low-wage labour markets. Thus, at least to some degree, the minimum wage would probably constrain reductions in equilibrium wages. Moreover, the ERA target groups are limited to the long-term unemployed and lone parents. The long-term unemployed participate in labour markets with other unemployed persons and with those who are currently employed, and lone parents participate in labour markets with married and childless persons. Therefore, the ERA target groups account for only a fairly small fraction of the total supply population in a given labour market. Finally, ERA's impacts are expected to be moderate at best.

Substitution effects occur if participants in a programme hold jobs that individuals who do not participate would have otherwise held (Johnson 1979). If these non-participants become unemployed or accept lower-wage jobs as a result, then their earnings fall. Despite these potential adverse effects, there is very little research quantifying the magnitude of substitution effects. However, a recent evaluation of the New Deal for Young People provides a preliminary analysis of substitutions effects that suggests they could be modest (Blundell *et al.* 2002).

In the case of ERA, substitution effects would occur if the intervention has a positive impact on the job retention or job advancement of those in the target groups, and, as a result, fewer job vacancies or opportunities for advancement are available to those who are not in the target groups. The magnitude of this potential substitution effect is likely to depend on the state of the local labour markets in the programme pilot sites. If a local labour market is tight, then alternative job opportunities are likely to be available to those outside the target group; but if it is loose, then the cost of substitution to those affected could be substantial.

It is also possible that, as a result of its emphasis on advancement, ERA will help some participants to leave slack occupational labour markets for tight ones – for example, through encouraging training. If this occurs, ERA would decrease the competition for job vacancies in the slack markets, making it easier for those who remain in these markets to find jobs. In theory at least, this could produce a result that is the exact opposite of a substitution effect: total employment among those not participating in ERA could actually increase.

An important focus of the ERA evaluation will be to establish which of the likely factors affecting the capacity of evaluators to make generalisations about the effectiveness of the ERA Demonstration predominate.

CHAPTER 5 – COST AND COST-BENEFIT ANALYSIS

Cost analysis

A cost analysis will be a key component of the ERA evaluation. For example, if the impact analysis implies that the ERA programme has positive benefits, the cost analysis will indicate whether the values of these benefits exceed the costs of producing them. If they do, the cost analysis will also provide information on how expensive it would be to adopt the successful intervention more widely. In addition, a cost analysis can be used to make adjustment, even while the evaluation is being conducted. For example, if an intervention is much more expensive than anticipated, it might be necessary to modify the programme design or scale back the services provided, or reduce the number of participants.

The cost estimates need to be inclusive of both running and fixed costs – for example, salary and fringe benefit costs and overhead and administration costs. (Set-up costs should also be estimated because it may be important to determine the cost of implementing the evaluated intervention elsewhere, but they should be recorded separately from other costs.) Costs should be included regardless of the sources of funding (the central government, local governments, charities, and so forth) used to cover them.

As much of the cost information will be obtained directly from the administrative records of Jobcentre Plus, it will be important to work closely with Jobcentre Plus at the beginning of the evaluation to ensure that its

cost-accounting system provides the required information. It is acknowledged, however, that the administrative records of Jobcentre Plus will be imperfect, and that evaluators will have to work with the information that is available.

For purposes of the cost analysis, it will be important to distinguish between gross cost and net cost. Gross cost is the cost of the services received by the programme group; net cost is gross cost less the costs incurred by the control group for the receipt of services similar to those received by the programme group. Net cost provides the best measure of the increase in costs resulting from ERA interventions, and, as such, is the most relevant measure of cost.

As discussed below, it will be important for certain aspects of the cost analysis to obtain information on the number of contacts each member of the programme group has with their ASA. Similar information will also be required for each member of the control group. This later information is currently collected by Jobcentre Plus advisers and should be available from the Labour Market System (LMS), the computer system used for case management. However, to activate certain components of the LMS software, it is sometimes necessary that advisers record interviews with clients that do not actually occur. It is important that this software be corrected prior to the initiation of ERA so that the number of adviser contacts can be accurately measured for purposes of the cost analysis.

In conducting the cost analysis, the cost of each programme components will need to be estimated separately, as the methods used to estimate costs will vary for different components of the programme being tested.

Programme impacts are naturally measured in terms of differences between their *average* value for the programme group and their average value for the control group. It is important to obtain similar estimates of net costs so that they will be comparable to impact estimates. Thus, costs will also be estimated in terms of their average values for each group. However, cost information on some programme components will be extracted from administrative records and, as a consequence, will be aggregated across individuals. Thus, to obtain measures of total gross cost *per* programme group member and *per* control group member, it will be necessary to divide by the number of persons in each of these groups.

The cost of providing a particular programme component for an average member of the programme group will be estimated. If control group members receive similar services outside the programme, the cost of providing these services will be estimated for an average member of the control group. The estimates for the different components can then be summed to determine total gross cost per programme group member and *total gross cost* per control group member. *Total net cost* per programme group member (i.e., the additional cost resulting from the ERA programme) can then be determined by subtracting total gross cost per control from total gross cost per programme group member.

Each of the programme components are listed below, along with a discussion of how the cost incurred in providing that component can be estimated for an average member of the programme group. When appropriate, estimating the cost incurred by providing similar services to an *average* member of the control group is also discussed. However, it should be borne in mind that some of the cost components listed below apply only to the programme group; controls will not incur some costs.

Use of Advancement Support Adviser's time

The cost of a Advancement Support Adviser's (ASA) time includes salary, fringe benefits, and overheads. These costs can be allocated among the three target groups on the basis of the adviser contact data described above. Note, however, that in using these data, it will be necessary to assume that the average length of the contacts is similar for the three target groups.

Separate information must be collected for the programme and the control groups. It is anticipated that the ASAs responsible for the programme group will devote their time at work exclusively to the ERA project. Thus, the direct cost of employing these persons (i.e., salary and fringe benefits) can probably be readily extracted from administrative records.

Unlike the programme group ASA, the advisers assigned to New Deal eligible controls may spend part of their time on tasks that are not related to serving these individuals. (WTC lone parents who are assigned to the control group will not typically be served by Jobcentre Plus advisers.) Thus, it will be necessary to determine the proportion of their time that is spent on serving controls, so that the cost



of employing them can be apportioned appropriately. One approach that might be used to do this would rely on the adviser contact data described above. For example, if, on average, members of the control group contact their advisers one-third as often as members of the programme group, it might be reasonable to assume that the cost of providing adviser services to controls is also one-third of that for the programme group. Additional adjustments should be made if advisers responsible for controls receive lower salaries than do those responsible for the programme group.

It will be important to set a mechanism in place prior to the initiation of the ERA Demonstration to track the number of ASAs who are dedicated to programme group clients at various points in time. This information can be used to determine the aggregate gross cost of employing ASAs. As previously indicated, to determine the aggregate gross cost for each target group, this figure would then be allocated according to the relative frequency with which the clients in each target group contact their ASAs, assuming that, on average, the duration of these contacts are similar across the target group. Gross cost per client would be determined for a particular target group by dividing the aggregate gross cost figure for the target group by the number of ERA clients in the target group.

An overhead rate will be needed to measure the full cost of employing ASAs. This rate should measure the cost of space, furniture, telephone, information technology, and support staff and supervision as a fraction of the total cost of operating the Jobcentre Plus demonstration sites. Use of an overhead rate will allow the cost analysis to take account of the additional need for physical capacity and supervision resulting from ERA.

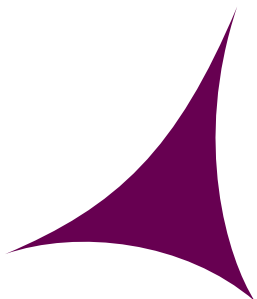
These additional resources will be required because caseloads for the programme group will be smaller than for the control group, because New Deal eligibles in the programme group will be served by advisers over a longer period of time than New Deal eligibles in the control group, and because WTC lone parents in the programme group will be served by advisers but WTC lone parents in the control group will not.

Use of Intake Clerk's time

Like ASAs, it is anticipated that Intake Clerks who are responsible for the programme group will devote their time at work exclusively to the ERA project, but that their counterparts, who are responsible for the control group, will also deal with individuals who are not in the research sample. Determining the gross and net cost of the time of Intake Clerks can be done in exactly the same way as outlined above for determining the cost of the ASAs' time.

Special training for ASAs

This cost includes the value of the time the ASAs spend in training as well as payments to trainers. Most of the information needed to measure ASA training costs should be available from administrative records. ASA training cost can be allocated among the three target groups on basis of the ASA contact data described above. It might be necessary to amortise ASA training costs because training would only need to be given once during each ASA's career in an on-going ERA programme. However, if turnover is sufficiently frequent among ASAs, as it may well be, amortisation may not be necessary.



Use of Job Developer's time

The cost of employing a typical job developer (i.e., salary, fringe benefits, and overheads) can be obtained from administrative records. Jobcentre Plus already employs job developers, known as Local Account Managers.

Presumably, however, ERA will require hiring additional Local Account Managers because job developers will serve New Deal eligibles in the ERA programme group for a longer period of time as a result of ERA, and WTC lone parents in the programme group will only receive such a service as a result of ERA. Moreover, in order to promote job advancement, it may be necessary to develop somewhat different types of jobs than those that are currently being developed.

At the same, because Local Account Managers will develop jobs taken by members of both the programme group and the control group, it is not obvious how best to allocate job developer costs between those incurred by the programme group and those incurred by controls. If advisers make more or less formal requests of Local Account Managers for information about jobs, one possible approach is for job developers to maintain records on how frequently ASAs and control group advisers make use of job developer services. This information could then be used to allocate job developer costs between the programme and control groups. Costs can be allocated among the three target groups on the basis of the proportion of programme and control group members drawn from each of the target groups.

If this approach is not feasible, an alternative method might be to determine how many additional slots for Local Account Managers are created during the demonstration period at the ERA Demonstration sites. A measure of net cost could then be obtained by assuming that all these additional slots were required to

serve the programme being tested.

Gross cost could then be estimated by adding the cost of job developers during the pre-demonstration period to the estimate of net cost.

Use of Outreach Worker's time

A mechanism for determining how many of these individuals are employed should be established prior to the beginning of the demonstration. Once this is done, the cost of employing each outreach worker can be obtained from administrative records. Because the work of the individuals filling these positions will be devoted exclusively to the ERA project and exclusively to those in the WTC target group, their entire cost can be assigned to the WTC programme group. Controls and members of the New Deal programme groups will not incur any costs.

ASA special fund

As this fund will be dedicated to the ERA project, its cost can be directly obtained from administrative records. Moreover, the cost can be allocated among the three target groups on the basis of administrative records. Because the ASA special fund will be new, the appropriate record keeping system will have to be established before the demonstration begins. This system would presumably be similar to the one that already is being used in Jobcentre Plus to manage and monitor the Advisers' Discretionary Fund. As discussed further below, care must be taken not to double count ASA special funds that are used to pay for outside services such as training and transportation.



ERA financial incentives

This expenditure will also be solely on members of the programme groups and, hence, can be readily determined from administrative records. However, as in the case of the ASA special fund, the appropriate record keeping system will need to be established before the demonstration project begins. This should be readily accomplished, as Liberator, a separate facility management company, would probably be contracted to issue the financial incentive cheques. The distinction between gross and net cost can presumably be ignored in the case of financial incentives because controls will be ineligible for them. The cost of paying for financial incentives can be allocated among target groups on basis of administrative record information.

Client participation in training

One objective of the ERA programme is to increase substantially client participation in training of various sorts. If the programme accomplishes this, it will be important to measure the resulting net cost. To prevent double counting, any training costs paid directly out of the ASA special fund or through an ERA training incentive will have to be subtracted from estimates of the cost of training the programme group. To estimate training costs, it will be necessary to assume that any expenditures on training correspond to the value of the resources that are used in the training. Because training institutions do not operate in perfectly competitive markets, this assumption is unlikely to be entirely valid.

If the training is directly provided by Jobcentre Plus or provided by outside contractors who are directly paid by Jobcentre Plus, it will be possible to use the Jobcentre Plus Contracting & Funding System to monitor the receipt of such

training by individuals in the programme and control groups, and to determine the amount expended on their training. It will be important to track these costs separately for each target group, and within each target group separately for the programme and control groups, so that the net cost of training for each of the target groups can be estimated. Preparation to obtain the necessary data should commence prior to the initiation of the Demonstration.

Training that takes place outside Jobcentre Plus will be more difficult to track and to cost out. For example, local Further Education colleges provide some training under contract to Jobcentre Plus, but other courses are funded through Learning & Skills Councils in England, Wales, and Scotland. Determining the net cost of training that takes place outside Jobcentre Plus will probably have to rely, in part, on survey information on participation in such training. However, in many instances, survey respondents are unlikely to know whether Jobcentre Plus paid for the training that they received or it was funded by outside sources. Thus, the survey should obtain information on all training received and, whenever possible, also obtain information on how the training was funded. In instances when the funding source cannot be determined, it will be necessary to compare the Jobcentre Plus Contracting & Funding System records with the responses of survey respondents. If the respondent indicates that he or she has received training that does not appear in the Jobcentre Plus Contracting & Funding System records, then it can be assumed that Jobcentre Plus did not fund the training.

If there is little difference between the receipt of outside funded training by the programme and control groups, then additional work on this cost component would be unnecessary,

because the net cost of training that takes place outside Jobcentre Plus would be close to zero. However, if (as anticipated) a substantial difference exists, then cost calculations would need to be done separately for each target group and, within each target group, separately for the programme group and the control group. To make these cost calculations, the fraction of each group that participates in training would be multiplied by the number of weeks of training participants actually receive. This figure would then be multiplied by the average cost of a week of training. In the case of training provided by the Further Education colleges, it might be possible to obtain estimates of the cost of an average week of training from the Learning & Skills Councils.

Client receipt of miscellaneous services

‘Miscellaneous services’ include the receipt of all services not specifically discussed above – for example, help in job search, drug and alcohol counselling, transport, day care, and perhaps others. It is only necessary to measure the net cost of specific miscellaneous services if the ERA programme substantially increases client receipt of the services. As indicated below, however, measuring the net cost of certain miscellaneous services is likely to be the most challenging part of the cost analysis. To prevent double counting, any payments made out of the ASA’s special fund to help provide miscellaneous services should be subtracted from the estimates for the programme groups.

As in the case of training cost, it will be possible to use administrative records to determine some of the cost of certain miscellaneous services—for example, payments made under the WTC for day care.

Administrative records cannot be used in many instances, however. For example, Jobcentre Plus will not fund most drug and alcohol rehabilitation that members of the programme and control groups will receive. Perhaps, more importantly, job search services are often provided by outside vendors. Because the job search vendor is likely to serve both members of the programme group and members of the control group, as well as other Jobcentre Plus clients, it is difficult to determine the proportion of payments to the vendor that should be allocated to the programme and control groups.

The first step in determining the net cost of miscellaneous services that are provided outside of Jobcentre Plus is to use survey information to determine whether the programme group differs importantly from the control group in their receipt of each of these services. When they do not differ, the net cost of the service is approximately zero, and further cost estimation will be unnecessary. For each service for which they do differ, the fraction of each group that received the service would need to be multiplied by the ‘units’ of the service that they received (e.g., hours of drug counselling, days spent with job search vendors). This figure would then be multiplied by the average cost of a unit of the service.

Information on the number of units of each service received would have to be obtained from survey respondents. The extent to which this survey information would be accurate is highly problematic, although it should be possible to determine whether individual respondents received each service. It is not currently known how estimates of the cost of a unit of various services would be obtained, but developing these estimates can



wait until it is determined that they are actually needed. As a practical matter, it probably will not be possible to obtain the unit costs of services that are provided by volunteer organisations, as some may well be.

Cost-benefit analysis

The cost-benefit analysis will integrate the estimates of the costs of the interventions with results from the impact and process analyses to draw conclusions about the effectiveness of ERA at meeting its goals. The objective of the cost-benefit analysis will be to determine if the value of the benefits of ERA service exceeds its costs. In addition, the analysis will provide information about how the benefits and costs of ERA services are distributed among participants, the government, employers, and society as a whole.

Design

As already suggested, the cost-benefit analysis will depend heavily on estimates of net costs obtained from the cost analysis and estimated benefits obtained from the impact analysis. Estimates of particular importance from the latter include impacts on earnings, whether transfer benefits were received, and the amounts of receipts of such benefits. Because the cost and impact values that will be used in the cost-benefit analysis will all be estimated as differences between programme and control groups, they should be net of deadweight, although it will be important to ensure that the control groups are uncontaminated. Net benefit estimates should be obtained for each target group and subgroups of interest. The analyses should rely on costs and benefit estimates that pertain to the period after the tested interventions have reached a steady state, so as to minimise the influence of various start-up adjustments on the findings.

A very simplified and preliminary accounting framework for use in conducting the cost-benefit analysis appears in Figure 3. The plus and minus signs indicate whether each item is expected to be a benefit (+) or cost (-) from the perspective of four groups: programme clients, employers, the exchequer or government (which is defined to include the transfer benefits system), and the whole of society. In the final step of the cost-benefit analysis, benefits and costs will be summed to determine the net present value of the programme from each perspective. It is obviously only possible to do this if all the benefits and costs are valued in pounds. As discussed below, there are some benefits from, and costs of, ERA for which this will not be possible. These benefits and costs are not listed in Figure 3.

As indicated, benefits and costs to society are simply the algebraic sum of benefits and costs to the first three groups. Thus, the framework implies that if an intervention causes a decline in benefit payments received or taxes paid by clients, this should be regarded as a cost to clients, a saving or benefit to the government, and as neither a benefit nor a cost to society, but simply a transfer of income from one segment to another. Similarly, the payment of financial incentive bonuses would be a transfer from the government to clients and, hence, would again be neither a benefit nor cost to society as a whole. Government expenditures on training (including tuition payments for training) and miscellaneous services for ERA clients are not transfers between segments of society, as they involve the use of 'real' resources. If effective, these expenditures should result in an increase in client earnings.

All the benefits and costs listed in Figure 3 should be adjusted for inflation so that they reflect the value of pounds at the same point in time.

Figure 3: Stylised cost-benefit accounting framework for the ERA project

	COST ANALYSIS	IMPACT ANALYSIS	PROCESS ANALYSIS	
Cost-benefit analysis: Financial gains and losses				
	ERA clients	Employers	Gov't	Society ^(a)
<i>Earnings:</i>				
Earnings	+	0	0	+
Fringe benefits	+	0	0	+
Tax payments	–	0	+	0
Work-related expend. ^(b)	–	0	–	–
<i>Transfer benefits:</i>				
Out-of-work benefits	–	0	+	0
In-work benefits	+	0	–	0
Financial incentives	+	0	–	0
<i>Employers' costs:</i>				
Employers' costs ^(c)	0	+	0	+
ERA costs				
Net training costs	0	0	–	–
Net costs of misc. services	0	0	–	–
Net costs of special funds	0	0	–	–
Net ERA operating costs	0	0	–	–
Net effects				
Net ERA benefits ^(d)	?	+	?	?

Notes

(a) Column sum (b) For example, childcare and transportation (c) For example, hiring and training costs (d) Row sum

Figure 3 does not treat the earnings and fringe benefits received by clients as a cost to employers because it is assumed that these payments are offset by the value to the employer of the output produced by its employees. Including the perspective of employers in Figure 3 is an innovative approach, as most cost-benefit analyses of social programmes simply consider

‘employers’ to be a subset of ‘taxpayers’. In the case of ERA, it is especially important to make the employer perspective explicit because if ERA services lead to lower employee turnover, they may in turn reduce recruitment and training costs for employers and produce a net gain for society. However, it will probably not be possible to place a precise value on this benefit in practice.



Nevertheless, through interviews with employers, it should be possible to determine the extent to which employers consider ERA valuable from their perspective.

Estimation of observed benefits

Although many programme impacts may be viewed as either positive or negative, depending upon the perspective being considered, in general, impacts that result from positive changes in behaviour by ERA participants are best regarded as programme 'benefits.' Benefits can be divided into those that are financial and those that are non-financial, or not easily measurable in pounds.

With the possible exception of employer recruitment and training cost, the analysis of each of the *financial* benefits listed in Figure 3 should estimate the net gain or loss in pounds for the programme group relative to the control group. Many impact measures used in computing financial benefits – e.g., impacts on earnings or welfare payments – are naturally expressed in pounds. In other cases, additional computations will be necessary. For example, fringe benefit impacts can be estimated by multiplying earnings by a fringe benefit rate derived from published data. (Long and Knox, 1985, provides more detail about this method.)

A potential *fiscal* cost of the ERA programme that is not listed in Figure 3, because it is rarely estimated in cost-benefit studies of welfare-to-work programmes, is the value of non-work time that will be relinquished by ERA clients if the Programme is successful in increasing work retention. If the time that people spend away from work has value to them, then this loss in non-work time will offset part of the benefits resulting from any ERA-induced increases in earnings. A procedure that can be used to obtain an

estimate of the value of lost non-work time in pounds is described in detail in Greenberg, 1997.

A second potential *fiscal* cost of the ERA intervention that also does not appear in Figure 3 would result if there were programme substitution effects. Substitution effects would occur if an ERA intervention has a positive impact on the job retention or job advancement of those who are included in the target groups, and, as a result, fewer job vacancies or opportunities for advancement are available to those who are *not* included in the ERA target group. The magnitude of this potential substitution effect is likely to depend on the state of the local labour markets in the demonstration sites. If a local labour market is tight, then alternative job opportunities are likely to be available to those outside the target group; but if it is loose, then the cost of substitution could be substantial. Evaluators of employment and training programmes have been aware of the possibility of substitution for many years, but, as indicated by Friedlander, Greenberg, and Robins (1997), there have been virtually no successful attempts to estimate the likely scale of substitution effects. Nonetheless, if unemployment is high in several of the demonstration sites, methods for taking costs incurred through substitution effects into account should be explored in conducting the cost-benefit analysis.

Costs and benefits can only be directly compared and a bottom line net benefit estimate can only be obtained if all values are expressed in pounds. Thus, potential non-financial benefits from ERA are not listed in Figure 3. However, such effects might well change the net programme benefit estimate if they could be measured in pounds. Thus, appropriate ways of incorporating them into the cost-benefit analysis should

be considered. Because some *non-financial* benefits are potentially quantifiable (e.g. reduced parental stress or improved child well-being), they will be easier to consider in the analysis, than others that are difficult to measure (e.g. the psychological benefit to programme participants of more stable employment).

Calculating future benefits and costs

Given the costs of prolonging an evaluation for a lengthy period, ERA programme impacts are likely to last beyond the period for which benefits and costs can be directly observed. To account for such long-term effects, programme impacts will need to be extrapolated into the future using an algorithm that includes a *time horizon*, a *projection period*, a *base period* impact estimate (representing the final 6- or 12-months observed), a *decay rate*, and a *discount rate*. The time horizon for the cost-benefit analysis is simply the observation period (i.e. the period over which data are collected on the individuals in the ERA evaluation sample) plus the projection period.

To extrapolate the base period impacts, an assumption must be made about whether these impacts decay or increase over the projection period. This assumption can be based on how the observed ERA impacts change over time and on information from other studies (see, for example, Greenberg, Ashworth, Cebulla, and Walker, 2003 and Greenberg, Michalopoulos, and Robins 2003).

Discounting is necessary because benefits and costs that occur in the future are not valued as highly as benefits and costs that occur in the present. Discount rates of around five per cent have typically been used in cost-benefit analyses of welfare-to-work

programmes. However, there is considerable controversy about the appropriate value of the discount rate to use in cost-benefit analysis. (See Boardman, Greenberg, Vining and Weimer 2001, Chapter 10 for a discussion of this controversy and a description of discounting procedures.) Guidance on this issue is also provided by the HM Treasury in the form of its Green Book.

Extrapolation is particularly important in the case of the ERA interventions because prior evidence suggests that benefits from services that lead to career advancement may not decay over time; instead, they may remain steady or even grow (Friedlander, Greenberg, and Robins 1997). For this reason, even though previous evaluations of employment and training programmes have typically used a short time horizon such as five years (even for programmes that increase human capital or provide work-related benefits indefinitely), a longer time horizon may be appropriate in evaluating the ERA interventions. It will also be important in conducting the ERA cost-benefit analysis to use appropriate methods for testing the sensitivity of the results to the assumptions about the decay rate and the discount rate.

CHAPTER 6 – DATA COLLECTION

This chapter describes the information that will be required from administrative databases and surveys at different stages of the project from sample selection through to the final measurement of outcomes. The focus of this chapter is on data collection as it relates primarily to the estimation of programme impacts. Other elements of data collection, such as qualitative interviews and surveys of employers are not dealt with here (see Chapters 3 and 5).

The two sources of data discussed, administrative records and survey data are complementary. The administrative records have the advantage of being available for all cases and less costly to collect, but the content is limited and, as already noted the case of Jobcentre Plus records, the data will inevitably be imperfect to some degree. Survey data, on the other hand, provide more detail about customers' characteristics and experiences but are subject to non-response.

The impact study design describes how, in practice, the target group samples will be identified, how the intake procedure works, and how entrants will be randomly assigned. This process is described again here but with emphasis on the information being collected at each stage.

Generating the samples for the evaluation

New Deal 25plus

Customers aged 25 or over who have been unemployed and claiming JSA for 18 months within a 21-month period have to join the ND25plus programme (although certain categories of JSA recipients may volunteer to join the programme before the 18 months stipulation – they too will be eligible to join ERA). Jobcentre Plus staff are alerted to the fact that a customer is due to start the New Deal through a marker on their record on the LMS database, and a letter is sent requesting them to attend a meeting at a Jobcentre.

At the meeting, the customer will be referred to the ERA Intake Clerk who will check the eligibility of the customer and enter details about the customer on the Basic Information Form (described below). The Intake Clerk will then explain the ND25plus programme, and also explain the ERA programme as an alternative to ND25plus. If the customer agrees to take part in the ERA research²⁴, the Intake Clerk will initiate the random assignment (RA) process. If the customer refuses to take part in the research, the Intake Clerk will note the reason(s) for refusal on the BIF as far as the customer is willing to divulge the information.

In addition, extract files from the Department for Work and Pensions' (DWP) database will be produced for all cases identified as having entered the ND25plus and ERA programmes

²⁴ For volunteers for ND25plus, agreement to participate in ERA also includes agreement to be randomly assigned, as well as to take part in research and surveys

over a reference period at the six sites. These files will contain information about benefits received and other relevant characteristics. They will be matched with the BIF records on the ERA database and will provide information that supplements BIF data. ND25plus entrants with no BIF will be identified and followed up, as will those with an a BIF but no administrative record, providing a check on how well the administrative systems are performing and allowing missing cases to be picked up. Sampling will operate, in most cases, for one year and will cover those joining ND25plus between October 1st 2003 and September 30th 2004.

New Deal for Lone Parents

Entry to the NDLP programme is entirely voluntary. Most lone parents volunteer for the programme after being informed of it during their (mandatory) work-focused interview as part of their Income Support (IS) claim²⁵, but they can opt not to participate. As with the ND25plus group, lone parents volunteering for NDLP will be referred to the Intake Clerk who will explain the ERA programme as an alternative to NDLP and ask them to participate. The clerk will also explain about the survey and, possibly, financial inducements (respondents to ERA Demonstration surveys could be eligible for a payment or inducement – more details are given below). A BIF will be completed for all, including refusals, and the RA process will then be initiated for those choosing to take part in the study. Because of their voluntary status, lone parents agreeing to participate in ERA must consent to being randomly assigned as well as to taking part in the research.

²⁵ The overwhelming majority of lone parents will be on IS. However, as NDLP is open to all lone parents who are not working or working less than 16 hours per week, there can be lone parents who are not claiming IS. There will obviously not be an administrative record for these non-IS lone parents.

²⁶ WTC will be introduced in April 2003 and the size of the credit awarded will be based on family income during the previous tax year. This award will continue for one year. Recipients of WTC will need to notify the IR only of changes in circumstances relating to household composition and childcare costs. A change of employment status will be picked up automatically. Claimants will not need to report changes in income during the year but any increases in income amounting to £2,500 or less during the year are disregarded when the award is reassessed at the end of the tax year. Any other changes affecting eligibility, which are not recorded during the year, will also be taken into account at the reassessment stage.

Again, as with the ND25plus group, extract files from DWP's database will be produced to supplement BIF data and to serve as a check on missing BIFs and the administrative system's performance. Likewise, sampling in most cases will operate for one year and will cover those joining NDLP between October 1st 2003 and September 30th 2004.

Working Tax Credit Lone Parents

Information about all eligible WTC cases will be produced by the Inland Revenue from their database. This file will provide information about the full eligible population that can be approached. It will be used to provide the sample for the WTC Recruitment Officers who will recruit customers to the study. A letter will be sent to eligible lone parents explaining the ERA programme and informing them that a WTC Recruitment Officer will be in contact with them to give further details. For those who go on to join the study, additional information about their characteristics contained on the individual's tax credit record²⁶ will be added to their BIF record on the ERA database.

The initial WTC LP sample will consist of lone parents working part time (16-29 hours per week) and receiving WTC at a fixed point in time (the stock sample) plus new claimants who make a claim during the ERA intake period (the flow sample). It is suggested that four equal-sized samples of the stock are taken during the intake period of one year, in order to even out the caseload of WTC Recruitment Officers. For the second and subsequent samples, cases already selected would need to be removed prior to selection.



The WTC Recruitment Officers will contact the customers at home and check that they are still lone parents working part time and receiving WTC. Those lone parents who have stopped working would be directed to NDLP (and so could enter ERA via the NDLP group). If a lone parent meets the eligibility criteria, the WTC Recruitment Officers will explain about the ERA programme and attempt to recruit them into the Demonstration. For lone parents who put themselves forward for participation in ERA, the WTC Recruitment Officer will initiate the random assignment process.

WTC Recruitment Officers will keep a record of their contacts and the final outcome. They will complete a BIF for all those who agree to participate in the study and a short non-response form giving the outcome code for all other cases (refusals, non-contacts, and ineligible). The non-response forms could be electronic or paper-based. In the latter case, responses would need to be keyed. The data will then be sent to the ERA database controller and matched onto the original sample database for non-response analysis. Because of their voluntary status, WTC lone parents agreeing to participate in ERA must consent to being randomly assigned as well as to taking part in the research.

The Basic Information Form (BIF)

The BIF will be completed, as far as possible, for all persons in all target groups who are approached for recruitment into ERA including those who refuse to participate in the Demonstration (with the exception of those in the WTC target group, where in some cases a non-response form is completed for those who refuse). Ideally the BIF would be completed electronically by Jobcentre staff and WTC Recruitment Officers. However, to avoid disruptions to usual office business, a paper-based system

may be used instead. If the BIF is a paper document, the completed forms from each office would need to be collated and sent to the evaluation controller's office for keying. Each Intake Clerk or WTC Recruitment Officer will have a set of BIFs with pre-printed serial numbers.

As the BIF represents the original sampling frame, it is very important that it is completed in full for all participants. The Technical Advisers (see Annex 4) will pay particular attention to this. Another method designed to encourage the Intake Clerks/WTC Recruitment Officers to complete the form in full before contacting the Database Controller, is for the latter to ask for additional items of information from the BIF, selected at random, during the process of initiating random assignment.

The list below illustrates the type of information that will need to be collected.

- **Personal identifiers**

Serial number (including components for area, office and person numbers)

Name

Address

Postcode

Tel No

National Insurance Number (NINO)

Date of birth

LMS office

ES region/district

- **Study Status**

Target group (eligibility)

Whether control/programme
(entered after randomisation)



- Demographic and socio-economic characteristics, for example:

Gender

Marital status

Number of children by age group

Age of youngest child

Ethnicity

Housing tenure

Summary of work history over last two years (for example, number of months working full time or part time)

Current work status

Hours worked and earnings (if working)

Duration of current WTC claim²⁷

Highest educational qualification

Whether the customer has driving licence and access to car

Whether has a longstanding illness that affects ability to work

Barriers to work (such as childcare or health problems)

- Tracing information

Name, address, postcode and telephone number of additional named contacts to be used for tracing purposes for the surveys.

- Consent

Consent to participate in ERA research, agreement to participate in surveys, permission to access IR and DWP administrative records if necessary, and to be randomly assigned, if the participant is in target group on a voluntary basis.

The Random Assignment (RA) procedure

After completing the BIF, the Intake Clerk (in the case of the New Deal eligible groups) or WTC Recruitment Officer (in the case of the WTC group), will contact the ERA Database Controller by telephone and give the customer's serial number, personal identifier information and tracing information as recorded on the BIF. The controller will enter these items onto the ERA database and will check that the person is not already a member of the study²⁸. Checks will also be carried out by the Database Controller on the consistency of the postcode and address information as well as the National Insurance Number (NINO) given by the customer.

It will be important that the database controllers are able to meet the expected volume of calls so that Intake Clerks can get immediate access and response.

Address Information

Correct address information is a key requirement for the survey where the aim is to minimise the number of cases lost because of inaccurate contact information. A similar approach to that used by call centres is suggested. The Intake Clerk completing the BIF will give the customer's postcode to the ERA database controller. S/he will enter this on to the ERA system and automatically access a version of the postcode directory that will display the street name and numbers corresponding to the postcode. Any discrepancies can be sorted out while the customer is present.

²⁷ If not available from IR records

²⁸ Once a person joins the study and is randomly assigned, s/he remains in the control/programme group they were initially assigned to irrespective of whether or not they subsequently qualify for a different target sample. For example, a lone parent in the NDLP sample could enter part-time work and claim WTC. S/he would not qualify for the WTC sample because s/he would have already been randomly assigned as part of the NDLP sample. Any cases should be excluded when there is doubt about whether the person has already been randomly assigned.



National Insurance Number (NINO)

The NINO is essential for matching datasets. The possibility of errors due to incorrect recording duplicate or shared numbers needs to be minimised. It is understood that DWP has a 'fuzzy' matching procedure that could be used to check that the NINO is valid. Ideally, this will be run at the same time as the random assignment process, but this will depend on the logistics. When matching data using NINOs, other identifying information (date of birth and name) will also be used as additional match fields to overcome the problem of duplicates.

Eligibility information

It will be useful for the Intake Clerk to give the Database Controller information to confirm the eligibility of the potential study member (for example: economic status, age, ND programme applied for, hours of work, etc.)

Once the ERA Database Controller is satisfied that the details are correct and that the person is eligible, the random assignment process will be run to determine which group, programme or control, the customer is assigned to. The result of the assignment process will be entered automatically by the system on to the ERA database. The Intake Clerk/WTC Recruitment Officer will record it on the BIF and inform the customer.

For the ND samples, it will be necessary to record whether the customer was assigned to the programme or control group on the LMS system as well.²⁹

Adding data from DWP records to the database for the ND samples

Data from DWP records will be added to the ERA database, both to supplement the BIF data and to check that the sampling procedure is operating correctly.

DWP have created New Deal databases containing records for all ND25plus and NDLP entrants. They also have an Evaluation database (NDED), which is currently being re-designed. These databases are constructed from other DWP databases (for example, LMS) and from databases owned by other organisations (for example, the Joint Unemployment and Vacancies Operating System – JUVOS – set up by ONS). It is envisaged that periodically, say every three months, the relevant databases will be interrogated and extract files produced for all those who had entered the programmes since the last sweep.³⁰

Supplementing BIF data

Some examples of the information that could be extracted from DWP records are listed below. These are additional items to those already collected on the BIF, which are likely to be obtained more reliably from administrative records. They relate to the sample characteristics or circumstance *prior to randomisation* so that they can be considered as additional baseline data. The personal identifiers are on both systems to enable matching. If there was a discrepancy, the case would be investigated.

- Personal identifiers:
 - Name
 - Address
 - Postcode
 - NINO
 - Date of birth
 - LMS office
 - ES region/district

²⁹ Ideally, this would be an extra field on the database, but this is likely to be difficult to set up.

If so, it could be possible that a note be made in the 'conversations' facility.

³⁰ DWP databases will also be interrogated periodically for outcome information, as discussed later in this chapter.



- Other information

LA and ward (for survey interviewers)

Date of starting ND

Disability indicator

Benefit history (for examples, the date of first IS claim, number of IS claims, unemployment duration at ND start, total days claiming JSA from 1/1/97)

Checking the sampling process

The New Deal extract file will be matched against the BIF records on the ERA database to ensure that all the cases are present in both systems. Any discrepancies will be followed up. Thus, for example, it will be possible to detect that an office was failing to complete the BIF for some ND entrants who should be entering the Demonstration.

Adding data from IR records to the WTC sample

Supplementing BIF data

Data will be added (with the customer's permission if necessary) from IR records to supplement the BIF information for the WTC sample.³¹ As with the supplementary data for the ND samples, the purpose is to add items not already collected on the BIF that can be obtained more reliably from administrative records and that relate to the situation prior to randomisation. Such supplementary data could include:

- Personal identifiers:

Name

Address

Postcode

NINO

Date of birth

- Other information

LA and ward

Gross earnings of respondent and partner in 2001³²

Other income (e.g. from savings, excludes maintenance)

Use of qualifying childcare, whether the customer receives childcare tax credits, number of qualifying children, cost of childcare

Duration of award (possibly)³³

Non-response analyses

BIF data and/or supplementary information will be available for all eligible cases including those who choose not to participate in the study. It can, therefore, be used to compare the characteristics of participants and non-participants in the study.

Monitoring the programme

Information about participation in different elements of the ND and ERA programmes is needed for two purposes. First, so that there is a record of all the activities that each control and programme group member has undertaken for process analysis. Second, to estimate costs of the New Deal programmes and ERA initiatives.

Some of the programme monitoring information for the New Deal control groups will be available automatically from DWP databases. Additional information, particularly about the activities of the programme group will need to be provided by purpose-built systems, preferably electronic. The types of information that these systems will need to provide are discussed below. Before the project 'goes live', a more detailed list of the items required will need to be produced and added to any system that is developed.

³¹ Outcome information will also be obtained from all relevant administrative records.

³² Because Tax Credits in 2003/4 will be based on income in 2001/2, the latest earnings data will be for 2001/2.

³³ The new Tax Credits will replace WFTC from April 2003 so all awards will be first awards.

It may be possible to derive the duration of the claim by matching cases with old WFTC records.



Data from DWP databases

The LMS is the system that Jobcentre staff use to record each contact and transaction with a customer. Key data about New Deal entrants are extracted onto the New Deal evaluation databases. These should provide information about all the contacts and activities of the ND control samples, for example:

- number and dates of interviews with New Deal advisers;
- details of education and training programmes undertaken;
- details of other initiatives (such as Basic Skills referrals, Restart courses, Training for Work, and Job Clubs).

The programme group will receive some services that are similar to those received by the control group and some services that are new. Ideally, the former type of services would be recorded on the LMS by Jobcentre Plus staff in the same way as for the control group. Some minor adaptations to the LMS would allow this.

As explained above, a special ERA programme group marker will need to be attached to the records of the programme group on the LMS so that the advisers and caseworkers will know which services customers should receive.

Data from purpose built systems

If it is not possible to record information about the contacts of ERA members on the LMS, a new system will need to be developed, preferably electronic, so that Jobcentre Plus staff can update records in the same way that they do for ND customers. It is essential to ensure that comparable information is available for the control and programme groups when they are receiving similar services.

Tailor-made systems will be needed to record the new services (financial incentives) received by ERA programme groups. The media to be used could consist of electronic spreadsheets or paper documents or both.³⁴ It will be easier to set up a process that makes use of existing systems (for example, EXCEL spreadsheets) than one that requires the installation of new software.

The type of information that will need to be collected by existing or new systems will include:

- record of contacts with caseworkers/advisers;
- details of education and training programmes undertaken;
- details of referrals to jobs and other agencies;
- details of any other services received (e.g. in-work support, counselling);
- receipt of money from the ERA crisis fund;
- receipt of ERA financial incentives;
- destination information – see below.

Measuring outcomes from administrative records

The administrative information available for monitoring outcomes is quite limited, particularly with regard to assessing the impacts of the ERA initiatives on advancement and, to a lesser extent, job retention. However, it has the great advantage that, unlike survey data, it should be available for the entire evaluation sample. Administrative records from DWP and IR are potential sources for measuring outcomes.

³⁴ It is understood that the DWP Action Teams have used a spreadsheet system. One of the lessons learned was that the files needed to be protected to ensure that staff could not alter the format. Therefore, it is suggested that those involved with designing such systems liaise with ASD Information Centre who have experienced the types of problems that are likely to be encountered.

Outcome information from DWP databases

DWP can provide information about the out-of-work benefits claimed by all study members and, by deduction, whether they are no longer claiming any benefit. The benefits data are stored on the LMS and updated by taking 'snapshots' of the separate benefits databases. For those on Jobseeker's Allowance, the information comes from JUVOS and covers all claims. For Income Support (IS) claimants, a snapshot is taken fortnightly and, for Incapacity Benefit (IB) claimants, every six weeks and so some shorter claims for IS and IB will not be captured. However, IB claims tend to be of fairly long duration so this should not be a serious problem.

DWP destination data are less comprehensive, especially for non-JSA claimants. The LMS database, which will cover all New Deal eligible study members (programme and control), contains fields indicating customers' destinations on leaving benefit. These fields are not always completed because they depend on the customer notifying staff of their status: the information is usually entered for JSA leavers but is often missing for lone parents. There is a developing interest at DWP in recording destinations for IS leavers, which may ultimately improve the information for lone parents.

The New Deal databases also contain fields for the destination of customers following New Deal participation. Again, the information depends on the customer reporting their status and so is incomplete and possibly inaccurate. This information will only be available for the New Deal control group members unless it can be adapted to cater for ERA programme group members as well. If not, something similar will need to be built

into the system designed to monitor the programme group's progress through ERA. It is likely that the destination data will be more complete for the programme group than for the controls because the former will have more post-employment contact. This is a potential source of bias that would need to be taken into account in the impact analysis.

The information recorded at Jobcentre Plus offices may be better than elsewhere because there are incentives to encourage staff to record whether a job was obtained (although this will not improve the recording of other outcomes). Staff could be encouraged to complete this information as thoroughly as possible during the course of the study, but this is not entirely in their control and it would be safest to assume that the destinations will be missing for some cases and inaccurate for others.

At regular intervals, the DWP databases will need to be interrogated to identify whether members of the programme and control group samples were claiming any benefits and whether members of the New Deal eligible samples (programme and control) had exited benefit/ND and, if so, their destination. This information will be added to the ERA database.

Outcome information from Inland Revenue databases

Inland Revenue (IR) have earnings and other information on all families with children (working or otherwise) and on a fraction of low-income working households without children, via the new Tax Credits. They will, therefore, be able to track the incomes of all lone parents and some ND25plus sample members (programmes and controls) from this source.³⁵

³⁵ The IS/JSA caseload does not move across to the new Tax Credits until 2004.



The IR also have two data sources for earnings of employees: COP and NIRS2. The Analysis and Research Division of IR has sample surveys based on each. COP feeds the Survey of Personal Incomes (SPI) and the Expenses & Benefits Survey; NIRS2 feeds the National Income Survey (NIS). Both sources have data on income from employment but neither has information on working patterns, hours worked, job descriptions, etc..

Both SPI and NIS can be linked to NINOs. NIS is a one per cent sample (based on those with a particular two digit NINO ending – 14) and IR consider that this would probably be the easier one to match with ERA data. SPI is a stratified sample and, although larger than NIS, may not be as good for matching purposes. Whichever source is used, with the current samples, outcome data would be obtained for only a minority of ERA study members. SPI and NIS should, however, provide sufficient information for monitoring earnings for employees in the general population for comparative purposes. Alternatively, IR can extract data for all members of the ERA sample who are employed, as a special exercise. They will require both the names and NINOs of the ERA cases in order to do this. The extract will be run annually in April for the previous financial year. Thus, for example, the extract in April 2005 would draw off the earnings data for the 04/05 tax year.

IR could do the extraction either via an automatic run or through a manual process. The automatic run is likely to have fewer errors, but could only be run in the April after the financial year end, with the sample specified (NINOs and names) up to nine months in advance—that is, around August of the previous year. To illustrate, assuming the ERA intake starts in October 2003, by August 2004 it would be possible to specify details for all of the WTC stock cases and

most, if not all, of the flow cases. Outcome information would then be available in April/May 2005. The other option would be to draw off earnings information manually. This would require a shorter lead time and the sample would only need to be specified a month or so before the financial year end. The work could begin from late April, but depending on how many people were employed on the task (and the size of the sample), may take a while to complete. With a manual process the error rate is likely to be higher. IR will need to confirm the feasibility of these options.

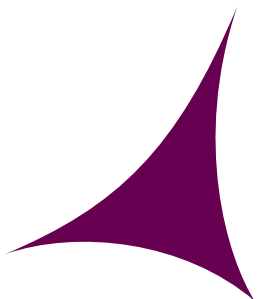
There are two possible options for processing the outcome data. The preferred method is for ERA participants to be asked if they are willing for the ERA evaluators to have access to their earnings data. For those who agree, their data will be added to the ERA database. The drawback of this approach is that, if large numbers do not allow access to their earnings, impacts estimated on these data could be biased. If that happens, another option is for IR to estimate earnings impacts on the full sample and simply communicate the results to the evaluation team. This is less flexible but would not be subject to non-response bias.

Analyses of administrative data

This section summarises the different types of analysis that will be carried out using the administrative data

Outcome measures

- benefits claimed at entry 1-5 years after entry;
- destinations following ERA/New Deal;
- earnings for employees (if necessary, based on aggregate data from IR);
- other IR data (e.g. household income, receipt of tax credit where available).



Descriptive analyses

- descriptive data about sample characteristics;
- descriptive data about ND/ERA services received;
- take-up of various services (to measure the impact of ERA on participation rates).

Non-response analysis

- characteristics of those who do not participate in study;
- characteristics of those who do not respond to the surveys.

Macro information about the sampled areas

- demographic, social and labour market information.³⁶

Validity checking against survey data

- cross checking information about contacts/services received between administrative records and survey data.

The surveys

Surveys will be carried out at 12 and 24 months after the customer has joined the study and been randomly assigned. It is also strongly recommended that a longer-term survey is conducted, at five years after entry, although the detailed design of a five-year follow-up is not considered due to the potential for high attrition rates. However, arrangements for keeping in touch with, and re-contacting, respondents should be incorporated into the survey design so that the practical procedures are in place if a five-year follow-up is considered feasible. The 24-month sample should indicate whether a longer term follow-up will be possible.

This section discusses the sample design for the surveys, the fieldwork, measures to improve response and the information that will be collected.

Sample design

The sample size has been determined on the basis of the funds available for interviews as well as analytical requirements. It is assumed that the total budget for the 12- and 24-month surveys will allow 10,000 achieved interviews over the two interview waves.

The size of the sample needed to detect ERA impacts of varying magnitudes is discussed in Chapter 4 of this report.

Calculating sample sizes requires assumptions about response rates. These are discussed later in this section. Table 12 shows two sets of estimates, the first based on the minimum acceptable response rate (70% of the original sample at 12 months and 65% of the original sample at 24 months) and the target (80% at both waves). Using the minimum rates, the necessary set sample size would be 7,407 people and the estimated number of interviews achieved would be 5,185 at 12 months and 4,815 at 24 months. Using the target rates, the required set sample would be 6,250 people and the estimated number of interviews achieved would be 5,000 at both waves.

There are two alternative designs for sampling at the site level. The choice will depend on a number of factors, including the requirements of the cost-benefit analysis, the degree to which impact estimates are to be computed on the basis of survey data for each site with equal sample power, and the needs of the process study. There will also be practical considerations relating to the administration of the programme.

³⁶ This could include information about the sites from Census, Neighbourhood Statistics or JUVOS.



Equal numbers per site

In theory, this design is preferable for estimating ERA impacts at the site level since it would yield equal numbers across the six sites and across the three target groups. However, the expectation is that only around 270 interviews in each target group at each site will be achieved at the 24-month survey. This is not sufficient to estimate impacts of the modest size expected, although a very large effect for a specific target group, at a given site, would be detected. The design also has a drawback in that the combined data for all sites need to be weighted before the analysis in order to restore uniform selection probabilities in terms of the composition of the population from which the sample was drawn. The ratio between the largest and the smallest weights applied to the site data will be approximately 2:1. This is not a large ratio, but it will still increase the variances for estimates of total sample effects. On the other hand, this design will be easier to manage from a fieldwork point of view, in that it does not require a very large concentration of interviewers in one or two areas.

Sampling with probability proportional to size

The alternative design is to sample with probability proportional to the number of ERA study members at the site. This design results in a variable number of interviews across sites. Impact measures may be possible for the largest sites if the effects are fairly large. No weighting is needed for impact estimates for the total sample, so variances are smaller than for the first design. Such a design is preferable where interest lies primarily in estimating pooled programme impacts. Whichever design is selected, the administrative data will necessarily be the main source for estimating site-specific impacts, reinforcing the importance of ensuring adequate numbers of individuals flow into the demonstration (thereby contributing administrative records) in order for programme impacts to be detectable.

Table 12: Estimated sample sizes for surveys at 12 and 24 months after
Random Assignment

	Minimum response rate (70% Wave 1, 65% Wave 2)	Target response rate (80% Wave 1, 80% Wave 2)
Set sample size	7,407	6,250
<i>Estimated number of achieved interviews:</i>		
Total at 12 months	5,185	5,000
Total at 24 months	4,815	5,000
Per site (24 months)	803	833
Per target group per site (24 months)	268	278
Per programme group per target group per site (24 months) ⁽¹⁾	134	139
Per control group per target group per site (24 months) ⁽¹⁾	134	139

Notes:

(1) Assumes equal numbers per site

Table 13 shows, for each target group, the estimated number of interviews that will be achieved at 24 months for some key subgroups. The figures assume that the subgroups are sampled proportionate to their size. However, there is the potential to

over-sample groups of particular interest, such as ethnic minorities or those aged 50 or over, in order to increase their numbers. Chapter 4 of this report discusses possible booster samples for these groups.

Table 13: Estimated number of survey interviews in key subgroups

Subgroup	Interviews at 24 months by key subgroup		
	ND25plus	NDLP	WTC/LP
<i>Education⁽¹⁾ (includes vocational quals.)</i>			
Qualifications	1,011	963	1,059
No qualifications	594	642	546
<i>Benefit claim history</i>			
Claim >3 years	498	–	–
Claim <3 years	1,107	–	–
<i>Ethnic group⁽²⁾</i>			
White	1,380	1,461	1,509
Non-white	225	144	96
<i>Partnership status</i>			
Partner	321	–	–
No partner	1,284	–	–
<i>Age</i>			
< 50 years	1,220	–	–
> 50 years	385	–	–
<i>Work history previous 3 years</i>			
Some work	–	1,011	–
No Work	–	594	–
<i>Age of youngest child⁽³⁾</i>			
> 5 years	–	819	465
< 5 years	–	786	1,140
<i>Length of current claim⁽³⁾</i>			
> 24 months	–	706	–
< 24 months	–	899	–
<i>Length of current claim⁽⁴⁾</i>			
> 12 months	–	–	642
< 12 months	–	–	963

Notes

(1) Estimates for WFTC groups come from an analysis of all lone parents claiming FC in 1999, Marsh, McKay, Smith and Stephenson (2001) Table 2.23 p61

(2) Estimates of ethnicity from NDED for NDLP and ND25plus groups. For WFTC group, estimates come from an analysis of all LP FC claims recorded on SOLIF, see Marsh, McKay, Smith and Stephenson (2001) p 42

(3) From the evaluation of the national NDLP, see Lessof C *et al*, 2000, Tables 8.2.4 p 76 & Table 8.3.1 page 77, for WFTC group, estimates come from analysis of all LP FC claims through SOLIF, see Marsh, McKay, Smith & Stephenson (2001), p41

(4) Estimated from WFTC 100% scan at November 2001 – part-time lone parents claiming WFTC only



Sample selection

The sample will be selected by the ERA Database Controller from each month's intake in each target group on each site. The sampling fractions will be calculated according to the total number of people in the target group at a site at the end of the intake period. Because intake will vary by the month, the number sampled each month will also vary.

If it is assumed that intake to the study continues from October 2003 to September 2004, the sample for the first month's interviews in October 2004 will need to be drawn during August so that updates of addresses and telephone numbers can be carried out before the fieldwork starts.

The fieldwork

At entry

As part of their explanation of the ERA study, Jobcentre Plus staff/WTC Recruitment Officers will inform potential participants in the Demonstration that there will be surveys 12, 24 and 60 months after random assignment and that participation in the Demonstration involves taking part in these, including having their name and address passed to a survey organisation. Staff will also explain, if the option is pursued, that survey respondents will be paid for their participation. The survey design outlined here is based on the assumption that survey respondents will be paid for participating in survey interviews and it is this course of action that is recommended. The design team recognises, however, that such payments raise issues beyond those simply relating to survey response in this particular instance, and that, as a result, wider considerations may lead to the decision not to make such payments.

Those who refuse to take part in surveys here do not join the ERA study and so are not part of the sample, although the reasons for refusal to participate should be noted on the BIF whenever possible. Combining agreement to enter the ERA study with participation in the survey should improve co-operation levels, although participation in surveys cannot be enforced.

Interviews at 12 months and 24 months

Prior to interview, the latest addresses and telephone numbers will be obtained from DWP and IR records. An advance letter will be sent from the survey organisation indicating that an interviewer will telephone or visit the sampled individual and mentioning the payment for the interview if this is to be made. The letter will also ask sampled customers to report any new telephone number or address. The first contact will be by telephone and an interview carried out when possible. A face-to-face interviewer will follow up all non-respondents. Attempts to interview each individual will be continued for six months. Refusals will be re-issued to an interviewer with a high success rate at converting such cases.

All sampled people will be approached for the 24-month interview irrespective of whether or not they were interviewed at the 12-month survey.

Keep in touch exercise

The survey organisation will attempt to contact the respondent (but not to interview them) between the survey waves. Prior to contact, the respondent's details will be updated from DWP (for New Deal samples, both control and programme groups) or IR (for WTC sample both control and programme groups) records. The first contact will be by telephone. Non-respondents will be followed up first by letter and then, if necessary, by a face-to-face interviewer.

At this contact, the interviewer will remind the respondent about the ERA study and the survey. S/he will ask about plans for moving and check and update the contact data for relatives and friends. The outcome information from the exercise will be stored on the ERA database.

Pilot test

A pilot test will be carried out before the first wave fieldwork in order to test the questionnaire and fieldwork procedures. If fieldwork for Wave 1 starts in September 2004, the pilot would need to be carried out no later than July 2004 to allow time for amendments to be made. For the pilot study, a supplementary sample could be selected from study members who joined in October 2003 and who had not been selected for the main sample. This would mean that those individuals in the pilot sample would have their interview at about 10 months after entry (rather than 12 months as for the main sample) but this should not affect the findings very much. Alternatively, it is likely that ERA will be pre-tested on a small group of individuals before live-running. The survey could then use the same test cases for the pilot sample. It will not be necessary to formally pilot the Wave 2 questionnaire, as it is likely to be very similar to the Wave 1 questionnaire.

Response

Without any special measures to boost response, it is estimated that interviews would be achieved with about 60% of the original sample at 12 months after entry and with less than 50% of the original sample at 24 months. While these rates may seem low, it should be remembered that in the past, it has been standard practice for benefit claimants to be given a specific opportunity

to withdraw from the sample before their contact details are passed to the survey organisation. However, as discussed below, a different procedure will be used for the ERA surveys. Also, the quality of the contact data tends to be of a poorer standard than for samples drawn from the Postcode Address File (PAF); and, of course, named person samples yield lower response rates than address-based samples, where anyone living at the sampled address can be interviewed.

Surveys in the US generally achieve higher response rates than British surveys; in particular, refusal rates are much lower. There are various reasons for the differences. For example, US surveys often offer financial payments to respondents for taking part, and they use a variety of agencies for tracing respondents. Evaluation studies like ERA typically achieve interviews with over 80% of the original sample several years after intake into the study.

The Great Britain ERA surveys need to obtain high response rates, ideally similar to those achieved in the US. Low response rates would mean that there would be a serious risk of bias in the sample so that the results would not be representative of all ERA study members. The minimum acceptable rates would be 70% at the 12-month interview and 65% at the 24-month interview. If, at the end of the first year's fieldwork, the response rate was below 70%, the research commissioners and the contractors would need to consider whether it was worth continuing.

This relative lack of response in Great Britain suggests that all possible steps to boost response are needed. The measures proposed here, some of which are based on methods used in the US, are discussed below. The study provides a rare opportunity for



assessing the extent to which survey response can be improved by increasing the resources (costs and timescale) beyond those normally available. This has wide-ranging implications for survey research in Great Britain. Since this package has not been used before, it is not possible to quantify the likely effect on the response rate. Clearly the extra steps would need to be extremely effective to raise the response at 24 months from under 50% to 80%. Even with the measures set out below, therefore, there is a risk that the desired rate of 80% will not be achieved.

Measures for improving refusal rates

Reducing the number of stages where opting out is possible

As outlined above, Jobcentre staff and WTC Recruitment Officers would present participation in the surveys as an integral part of joining the study. They would ask for agreement for names and contact details to be passed to the survey organisation at the same time and mention the financial inducements, if these are to be adopted. It will be very important that this is done properly and staff would need training and a script.

If this approach is taken, it would not be necessary for DWP/IR to send an advance letter with an opt out³⁷, nor would permission to recall at the end of each interview be required (although contact information would need to be updated so it would be obvious to the customer that a return interview is intended). There might be a few people who refused to take part in the study because of the survey interview. This might affect the generalisability of the

evaluation but not the control/programme group comparisons, as securing participation should occur prior to random assignment.

Use of a telephone interview

In the ONE survey, which was part of the evaluation of the ONE programme, some benefit claimants were unwilling to have an interviewer call at their home, but a few accepted the offer of a telephone interview instead. Although telephone surveys generally result in lower response rates than face-to-face interviews, they may be a better option for at least some members of this population. The practice in the US for these types of surveys is to use telephone interviews where possible and it is recommended that a telephone interview be offered first, with a face-to-face interview being offered if this is rejected, or if no contact has been made by phone. The telephone interviewers will be briefed to withdraw if they sense that a respondent is going to refuse; a recall at a different, more convenient time may result in an interview.

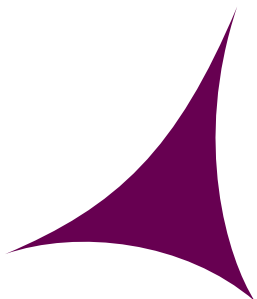
Reissue refusals to the best interviewers

The most successful interviewers will carry out refusal conversions. These will be expensive since some interviewers will have to travel quite long distances.

Offer financial inducements

Experience in the United States suggests that payment for interviews has helped improve survey response rates and it is recommended that this practice be adopted for ERA (though see comments above). The suggested rates are £15 for each interview, and a bonus of £20 for completing both.

³⁷ An advance letter would still be sent from the survey organisation to forewarn of the interviewer's visit, but it would not specifically mention the possibility of opting out.



Measures for improving contact rates

Collect good quality address information at the start of the study

Again this requires the Jobcentre Plus staff to ensure that good quality address and contact information is recorded at the start of the study. There will be an automatic check on the consistency of the street and postcode details. The staff will also provide change of address cards with a reminder about the interview (and that participants will be paid for it, if this is to be the case).

Collect extensive contact information

At the start, Jobcentre Plus staff completing the Basic Information Form (BIF) will ask for details of two or three people (relatives and friends) whom the survey firm could approach if it were unable to contact an individual customer.

Allow interviews to be conducted over a six-month period

ERA study members are being sampled at fixed points post-random assignment. An 18-month field period in total for each wave should be allowed for. The survey firm will continue to seek interviews with people from the one-year anniversary of their entry date for a six-month period. Thus, the last cases to join will be interviewed for Wave 2 at 24-30 months after they entered the study. This will mean that the timing of the interviews could be longer than one and two years after the entry date for some cases.

Obtain regular updates of addresses from DWP/IR before and during fieldwork

This was quite effective on the ONE survey, but is only useful for people who are still claiming benefit. It should also be possible to use IR records for tracing the WTC sample if permission is obtained first.

Movers outside the area

These will be followed up if they have moved anywhere within Great Britain, provided an address or telephone number is obtained for them.

Use of other records for tracing

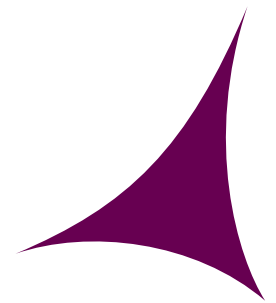
In the US, non-contact rates are a more serious problem than refusals. A variety of records are used there for tracing, including credit ratings and vehicle registration records. It would probably be unacceptable (or even illegal) to use such sources in Great Britain. Even if it were permissible, NINOs are less commonly used on administrative records than US Social Security numbers so the options are more limited.

Questionnaire

The length of the interview at all waves will be no more than 45 minutes as it is judged that this will be acceptable for a telephone interview.

Topics to be covered are listed below, together with some examples of the types of questions to be included. These are intended for illustrative purposes and are not meant to be comprehensive. The aim will be to ask the same questions of programme and control groups members as far as possible.

- Personal identifiers (serial number, NINO, date of birth, etc);
- household composition;
- activity history since entry/previous interview. Time spent in:
 - training and education (type of course, organisation, if respondent paid);
 - paid employment (hours worked, reason for leaving, in-work benefits);



- unpaid work (government schemes/ work experience);
- unemployed and looking for work;
- not working/looking because of illness or injury;
- not working for other reasons
- details of current/most recent job (occupation, pay, conditions);
- job search activities since entry/previous interview (jobclub/programme centre attendance, methods of finding a job, applications, interviews, jobs turned down, type of work sought, minimum acceptable wage);
- contacts with ERA/New Deal personal advisers;
- other ERA/New Deal services received (counselling, in-work support);
- financial incentives received (programme group);
- benefits received by respondent and partner, other sources of income;
- attitudes and barriers to work;
- childcare arrangements (actual/planned).

The questionnaire will be programmed for computer assisted interviewing by telephone (CATI) and face-to-face (CAPI). The CATI/CAPI programme will include consistency and range checks and all the identifier information required for matching purposes.

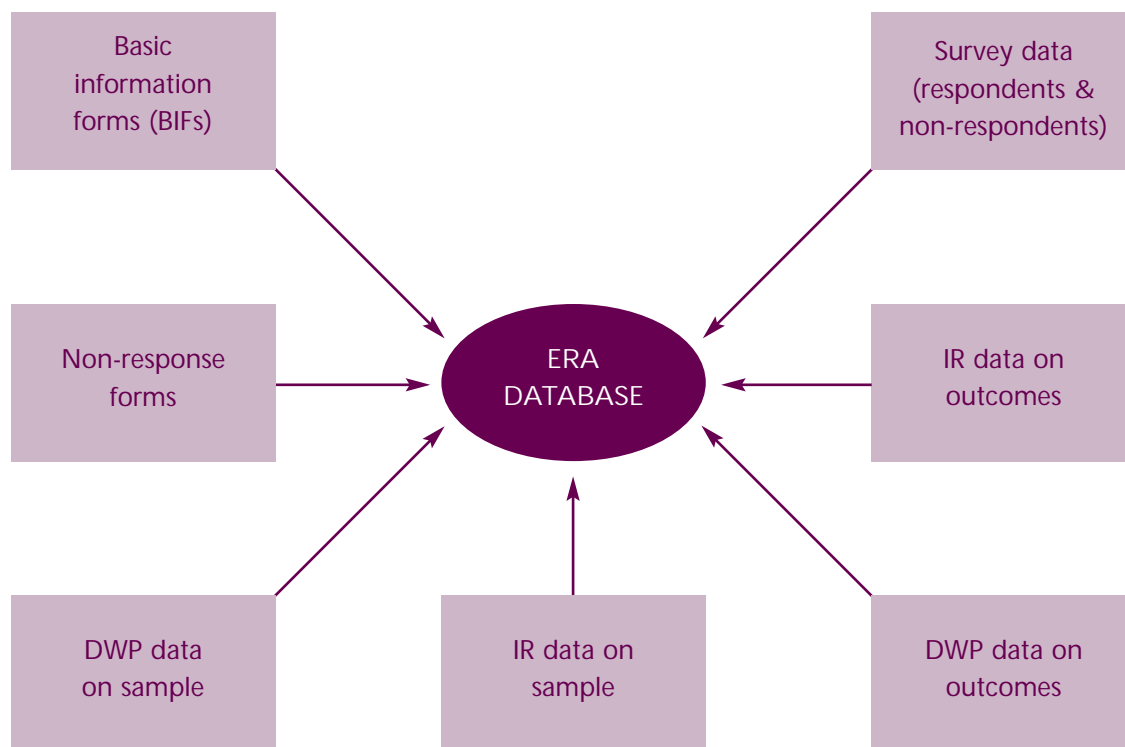
The questionnaire could also incorporate a 'basic question' approach for those who refuse to take part in the surveys. The most likely application of this 'basic question' is to give an additional handle on non-response bias. In a survey of New Deal leavers to

unknown destinations, for example, those refusing to be interviewed were asked a simple question about whether they were currently working and this proved helpful in understanding non-response.

Data processing

After the first month of fieldwork, a test data file will be produced and the frequencies inspected to check that the questionnaire is working as expected. Test data will be provided to the evaluation analysts. At the end of the each wave of fieldwork, an edited and checked data tape will be produced, together with a technical report. The data files will be ready 4-6 weeks after the end of fieldwork. Assuming a start date of October 2004 and an 18-month field period for each wave, the data file for the 12-month survey would be ready for analysis in May 2006. Following a similar timetable, the data file for the 24-month survey would be available in May 2007.

Figure 4: Inputs to the ERA Database



CHAPTER 7 – PROGRAMME COST ESTIMATES

A key element of appraising the ERA project costs is the nature and content of the services to be tested. The costs of the retention bonus and periods in which they arise, for examples, will be determined in part by the amount of the retention bonus that is set and the period over which it can be earned. Once agreed, these design features are known with certainty.

However, and unlike the ex-post evaluation of ERA project costs where the outcomes and costs involved will be known, an ex-ante appraisal of costs also necessarily relies on the use of assumptions. These assumptions are based on forecasts or historic data or, quite often, simple informed best guesses. The numbers in each of the target groups that will participate in ERA, for example, are based on forecasts of the numbers in the target group. Similarly, the entry rates into work of these target groups are based on survey or evaluation data from past comparable projects such as the evaluations of ONE or the New Deal for Lone Parents. Where there are aspects of ERA for which no historical data is available, such as the cost of supporting a customer in work for a two-year period, an informed best guess has necessarily to be made, which could turn out to vary significantly from the outcome.³⁸

Given this, it is inevitable that ex-ante cost estimates will be subject to a large degree of uncertainty. The extent to which ex-post costs vary from the ex-ante costs will depend on the range of variability of the assumptions – where this is large, ex-post costs can be significantly above or below ex-ante costs. Thus the figures presented in this chapter should be viewed with this ‘health warning’ in mind.

Another important point to note, and already discussed in Chapter 5, is that the costs shown are ERA net costs, not the total gross costs that will be incurred by the ERA Demonstration. This is because, in the absence of ERA, most of the target customer groups would still be receiving services and so incurring costs. The New Deal 25plus group, for example, would receive ND25plus services if ERA were not implemented. Thus only the costs arising from ERA services for this group that are over and above those for ND25plus are included. In the case of the WTC group, there would be no such services in the absence of ERA and so the gross costs that arise for this group under ERA are the relevant costs. In short, the costs presented are the *additional* costs incurred through ERA.

The key design features and assumptions built into the ERA costing model are given in the following Table:

³⁸ It can be argued that forecasts or historic data are ‘best estimates’ – the distinction is made here between those estimates based on past trends for which data is available and those estimates for which no such comparable data exists.

Table 14: Key assumptions for ERA Demonstration cost estimates

Design feature/assumption	Description/variable
9 months intensive treatment period.	Each target group gets 9 months in which to find full-time work during which they receive intensive treatment under ERA – if they fail to find full-time work in this period, they move on to less intensive treatment for a further 18 months.
24 months of in-work support and opportunity to earn retention bonus.	Each target group gets 24 months in which to earn the retention bonus during which they receive in-work support. If customers fail to find full-time work after the 9-month intensive treatment period, subsequent time spent looking for employment erodes the 24-month period over which the retention bonus can be earned (so, for example, if they take 12 months to find full-time work, they have only 21 months in which to earn the retention bonus).
The refusal to participate in ERA rates are:	ND25plus: 5% NDLP: 5% WTC: 70%
The numbers randomly assigned to the ERA treatment groups are:	ND25plus: 5,600 NDLP: 2,600 WTC: 5,000
The entry rates into work, the additional ERA impact expected, and the total entry rates after 9 months are:	ND25plus: 27%, 5%, 32% NDLP: 40%, 5%, 45% WTC (full-time work): 7%, 5% 12%
The average delays into full-time work are:	ND25plus: 4 months NDLP: 4 months WTC: 4 months
The entry into training rates are:	ND25plus: 8% NDLP: 10% WTC: 12%
The net pre-full-time work costs of an ASA per customer per year during the intensive treatment period are:	ND25plus: £350 NDLP: £250 WTC: £200
The net pre-full-time work programme costs ⁽¹⁾ per customer per year during the intensive treatment period are:	ND25plus: £325 NDLP: £25 WTC: £200
The net pre-full-time work costs per customer per year of an ASA during the less intensive treatment period are:	ND25plus: £70 NDLP: £150 WTC: £50



Table 14: continued	
Design feature/assumption	Description/variable
The net pre-full-time work programme costs per customer per year during the less intensive treatment period are:	ND25plus: £130 NDLP: £25 WTC: £50
The net post-full-time work costs per customer per year of an ASA during the intensive treatment period are:	ND25plus: £280 NDLP: £300 WTC: £200
The net post-full-time work programme costs per customer per year during the intensive treatment period are:	ND25plus: £65 NDLP: £250 WTC: £200
The retention bonus is paid if a customer works full time in any 13 weeks in of a 4-month period. This bonus is paid in the month following the 4-month period.	The amount of the retention bonus is £400 per 4-month period.
The training bonus is paid to those customers in part-time or full-time work who undertake training. It is capped at £1,000.	The training bonus is paid at a rate of £8.00 per hour of training undertaken.

Notes:

1. Programme costs includes such costs as Employment Service programmes that ERA is expected to utilise, childcare costs and subsidies and fares refunds.

Given these assumptions, Table 15 below shows the programme costs that are predicted for ERA by Spending Review Period (SRP). These SRPs run from April to March

starting at April 2003. The costs are shown adjusted by the forecast inflation rate so that they reflect the actual monetary level expected in the relevant SRP.

Table 15: Estimated ERA Demonstration programme costs

	Spending review period (£million)			
	2003 - 2004	2004 - 2005	2005 - 2006	2006 - 2007
Total caseworker costs	£3.5	£8.0	£3.3	£1.7
Cumulative total caseworker costs	£3.5	£11.5	£14.9	£16.5
Total financial incentives costs	£0.1	£3.8	£5.0	£2.4
Cumulative total financial incentives costs	£0.1	£3.9	£8.8	£11.2

Notes:

Numbers may not sum to exact figures due to rounding.

The forecasts in Table 15 have an underlying assumption that there is no 'churning' of customers between full-time work and returning to benefits. It is assumed in the ERA costing model that all customers that find full-time work do not subsequently become unemployed but remain in full-time work for the duration of ERA. The concomitant assumptions to this are that those entering full-time work all receive 24 months of in-work support and that they all earn the maximum six retention bonuses on offer.³⁹ Similarly, it is assumed that every person who does not find full-time work within the intensive treatment period of nine months does not subsequently find full-time work during the remaining 24 months, and hence receive less intensive ERA services during this period.

These are obviously very simplified assumptions, but it would be impossible to try to model the complex churning between work and unemployment that will occur, not least because the unavailability of any detailed data on which to base the model. Instead, however, the option to include a very simplified churning effect is incorporated into the costing model. This takes the observed rates off-work and return to benefits for participants in ND25plus at 12 months of 50 per cent and prorates these attrition rates over 12 months for all three target groups. In other words, half the entrants into ERA who find full-time work are assumed to lose their jobs after 12 months. Thereafter, those retaining work are assumed to stay in work until the end of ERA while those who remain on benefits are assumed to remain out of work for the remaining period.

Table 16, shows the resulting effect of this assumption on the programme costs and compares these costs to the projected costs if no churning is assumed. A primary objective of ERA is to reduce churning by increasing the retention rates of those in the programme group. Thus the table also shows the effect on programme costs if the attrition rate at 12 months is 40% (so that ERA impact on retention is to improve it by 10 percentage points over the observed job-leaving rate). All figures are inflation adjusted to show the actual level of spend in the SRP.

The results in Table 16 indicates, that customers returning to benefits after finding full-time work will lower costs compared to the 'ideal' case where all customers in full-time work remain in full-time work. This is primarily because the costs of financial incentives will be lower if there are fewer customers able to claim them. A secondary factor is that more people go on to receive the less intensive treatment and service, and the programme costs of these are lower than for the intensive pre-work services and the in-work service costs. It is worth noting that if churning increases, there will be higher associated Jobseeker's Allowance and Income Support costs, although these higher associated costs will not affect the costs of ERA.

Table 16 also shows that, very approximately, a 10 percentage point increase in the churning rate from 40% to 50% leads to a fall in programme costs of £0.7 million. This emphasises the points made above, that changes from the expected outcome of key assumptions can have significant impacts on costs and that the figures must be viewed as indicative rather than estimates of the precise costs expected to prevail.

³⁹ Thus the forecast of retention bonus costs sets an upper bound to the costs and the outturn can be expected to be lower than this if the actual entry into full-time work is below the assumed rate. If the entry rate is higher, then this can lead to higher costs even with churning effects. A similar upper bound estimate is made for the training bonus in that all those undertaking training are assumed to earn the maximum training bonus of £1,000 – again, this is an upper bound estimate only if the numbers undertaking training are below the expected levels.



Table 16: Estimated ERA Demonstration costs with and without churning

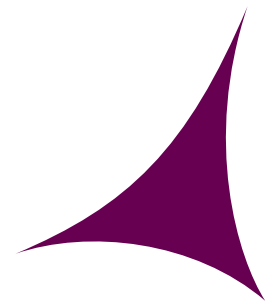
	Spending review period			
	2003 - 2004	2004 - 2005	2005 - 2006	2006 - 2007
			(£million inflation adjusted)	
Cumulative total caseworker costs no churning	£3.5	£11.5	£14.9	£16.5
Cumulative total caseworker costs with 40% returning to benefits after 12 months	£3.5	£11.2	£14.2	£15.8
Cumulative total caseworker costs, 50% returning to benefits after 12 months	£3.5	£11.1	£14.0	£15.6
Cumulative total financial incentives costs, no churning	£0.1	£3.9	£8.8	£11.2
Cumulative total financial incentives costs, 40% returning to benefits after 12 months	£0.1	£3.3	£6.5	£7.9
Cumulative total financial incentives costs, 50% returning to benefits after 12 months	£0.1	£3.1	£6.0	£7.2

Notes

Numbers may not sum to exact figures due to rounding.

CHAPTER 8 - BIBLIOGRAPHY

- Allison, P.D. (1995), *Survival analysis using the SAS system: A practical guide*, Cary, NC: SAS Institute.
- Arulampalam, W. (2002), *State dependence in unemployment incidence: Evidence for British men revisited*, Discussion Paper No.60, Bonn: Institute for the Study of Labor (IZA)
- Arulampalam, W. (2001), 'Is unemployment really scarring? Effects of unemployment experiences on wages', *The Economic Journal*, 111(475), pages 585-606.
- Arulampalam, W. (2000), *Unemployment - non-employment dynamics: Evidence for British men*, Working Paper, Warwick: Department of Economics, University of Warwick.
- Arulampalam, W. and Booth, A. (1998), *Training and labour market flexibility: Is there a trade-off?*, Working Paper 13, Colchester: Institute for Labour Research, University of Essex.
- Arulampalam, W., Booth A. and Taylor, M. (2000), 'Unemployment persistence', *Oxford Economic Papers*, vol. 52, pages 24-50.
- Arulampalam, W., Gregg, P. and Gregory, M. (2001), 'Unemployment scarring', *The Economic Journal*, 111(475), pages 577-584.
- Ashworth, K. and Liu, W. (2001), *Jobseeker's Allowance: Transitions to work and early return to JSA*, In-house Report 80, London: Department for Work and Pensions.
- Ashworth, K. and Young, R. (2000), *Prospects for part-time work: The impact of the Back to Work Bonus*, Department of Social Security Research Report No. 15, Leeds: Corporate Document Services.
- Atkinson, J. and Williams, M. (forthcoming 2003), *Employers perspectives on the recruitment, retention and advancement of low paid, low status employees*, GCSRO Occasional Papers Series, London: Cabinet Office.
- Atkinson, J. (2001), *Evaluation of the New Deal 50 plus: Summary Report*, Employment Service, Research and Development Report 103.
- Bardasi, E. and Gornick, J. (2000), *Women and part-time employment: Workers' choices and wage penalties in five industrialised countries*, Working Paper 2000-11, Colchester: Institute for Economic and Social Research, University of Essex.
- Becker, G. (1975), *Human capital: a theoretical and empirical analysis with special reference to education*, 2nd edition, New York: National Bureau of Economic Research with Columbia University Press.
- Berlin, G., Bancroft, W., Card, D., Lin, W. and Robbins, P.K. (1998), *Do work incentives have unintended consequences? Measuring 'Entry Effects' in the self-sufficiency project*, Ottawa: Social Research Demonstration Corporation.



- Blanchard, O. and Diamond, P. (1994), Ranking employment duration and wages, *Review of Economic Studies*, 34(4), pages 851-886.
- Bloom, H.S., Hill, C.J. and Riccio, J. (2001), Modelling the performance of Welfare-to-Work Programs: The effects of program management and services, economic environment, and client characteristics, New York: Manpower Demonstration Research Corporation.
- Bloom, H.S. (1995), Minimum detectable effects: A simple way to report the statistical power of experimental designs, *Evaluation Review*, 19(5), pages 547-556.
- Blundell, R., Costa Dias, M., Meghir, C. and Van Reenen, J. (unpublished 2002), Evaluating the employment impact of a mandatory job search program: The New Deal for Young People in the UK, London: University College London and Institute for Fiscal Studies.
- Blundell, R. and Costa Dias, M. (2000), Evaluation methods for non-experimental data, *Fiscal Studies*, 4 (21), pages 427-468
- Boardman, A.E., Greenberg, D.H., Vining, A.R. and Weimer, D.L. (2001), Cost-benefit analysis: Concepts and practice, (Second Edition), New Jersey, US: Prentice Hall.
- Boheim, R. and Taylor, M. (2000a), Unemployment duration and exit states in Britain, Working Paper 2000-1, Colchester: Institute for Economic and Social Research, University of Essex.
- Boheim, R. and Taylor, M.P. (2000b), The search for success: Do the unemployed find stable employment?, Working Paper 2000-5, Colchester: Institute for Economic and Social Research, University of Essex.
- Bonjour, D., Dorsett, R., Knight, G., Lissenburgh, S., Mukherjee, A., Payne, J., Range, M., Urwin, P. and White, M. (2001), New Deal for Young People: National survey of participants: Stage 2, Employment Service, Research and Development Report 67.
- Booth, A. (2001), A Theory of part-time/full-time wage differentials with on-the-job training and labour market frictions, Colchester: University of Essex
- Booth, A., Francesconi, M. and Frank, J. (2000), Temporary jobs: Who gets them, what are they worth and do they lead anywhere?, Working Paper 2000-13, Colchester: Institute for Economic and Social Research, University of Essex.
- Bower, C. (2001), 'Trends in female employment', *Labour Market Trends*, 109 (2), pages 107-118.
- Bryk, A.S. and Raudenbush, S.W. (1992), Hierarchical linear models: Applications and data analysis methods, Newbury Park, CA: Sage Publications.
- Bryson, A., Knight, G. and White, M. (2000), New Deal for Young People: National survey of participants: stage 2, Employment Service, Research and Development Report 44.
- Burdett, K. (1978), 'A theory of employee job search and quit rates', *American Economic Review*, 68, pages 212-220.
- Campbell, D. and Green, F. (2002), The long-term pay-off from working longer hours, Department of Economics Discussion Paper 0205, University of Kent at Canterbury.
- Cancian, M. and Meyer, D. (unpublished 1998), Work after welfare: Women's work effort, occupation and economic well-being, New York: MDRC

Card, D., Michalopoulos, C. and Robins, P.K. (2001), *The limits to wage growth: Measuring the growth rate of wages for recent welfare leavers*, National Bureau of Economic Research Working Paper Series 8444, Cambridge, MA: NBER.

Card, D. and Robbins, P. (1999), *Measuring wage progression among former welfare recipients*, Working Paper, University of Berkeley.

Card, D. and Robbins, P. (1998), 'Do financial incentives encourage welfare recipients to work?', *Research in Labor Economics*, 17, pages 1-56.

Card, D. and Robbins, P. (1996), *Do financial incentives encourage welfare recipients to work? Initial 18-month findings from the self-sufficiency project*, Vancouver: Social Research and Demonstration Corporation.

Chang, F. (unpublished 1996), *Evaluating the impact of mandatory work programs on two-parent welfare caseloads*, Doctoral dissertation, Baltimore: University of Maryland Baltimore County.

Dickens, R. (unpublished 2000a), *The low pay - no pay cycle: A report for the Welfare-to-Work team*, HM Treasury, London: Centre for Economic Performance, London School of Economics.

Dickens, R. (2000b), 'Caught in a trap? Wage mobility in Great Britain: 1975-1994', *Economica*, (67), pages 477-497.

Disney, W. and Szyszczak, S. (1984), *Protective legislation and part-time employment in Britain*, *British Journal of Industrial Relations*, 22, pages 78-100.

Edin, K. and Lein, L. (1997), *Making ends meet*, CA: Sage Publications, Thousand Oaks.

Ermisch, J., Francesconi, M. and Pevalin, D., 2001, *Outcomes for Children*, Department for Work and Pensions Research Report No. 158, Leeds: Corporate Document Services.

Friedlander, D., Greenberg, D.H. and Robins, P.K. (1997), 'Evaluating government training programs for the economically disadvantaged', *Journal of Economic Literature*, December, pages 1809-1855.

Garfinkel, I., Manski, C.F. and Michalopoulos, C. (1992), 'Micro experiments and macro effects', in Manski, C.F. and Garfinkel, I. (eds.), *Evaluating welfare and training programs*, Cambridge, MA: Harvard University Press, pages. 253-276.

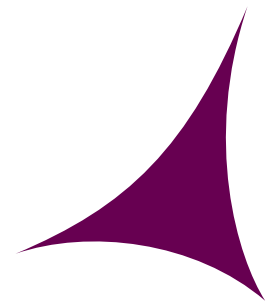
Green, F. (1999), 'Training the workers', in Gregg, P. and Wadsworth, J. (eds.), *The state of working Britain*, Manchester: Manchester University Press, pages 127-146.

Gregg, P. (2001), 'The impact of youth unemployment on adult unemployment in the NCDS', *The Economic Journal*, 111(475), pages 626-653.

Greenberg, D.H. (1997), 'The leisure bias in cost-benefit analysis of employment and training programs', *Journal of Human Resources*, 32, pages 413-439

Greenberg, D.H., Ashworth, K., Cebulla, A. and Walker, R. (forthcoming 2003), 'Do welfare-to-work programmes work for long?' *Fiscal Studies*

Greenberg, D.H., Michalopoulos, C. and Robbins, P.K. (forthcoming 2003), 'What happens to the effects of government-funded training programs over time?' *Journal of Human Resources*.



- Gregory, M. and Jukes, R. (2001), 'Unemployment and subsequent earnings: Estimating scarring among British men', *The Economic Journal*, 111(475), pages 607-635
- Hales, J., Roth, W., Barnes, M., Millar, J., Lessof, C., Gloyer, M. and Shaw, A. (2000), *Evaluation of the New Deal for Lone Parents: Early lessons from the Phase One Prototype - findings of surveys*, Department of Social Security Research Report No. 109, Leeds: Corporate Document Services.
- Ham, J. and Lalonde, R. (1996), 'The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training', *Econometrica*, 64, pages 175-206.
- Heckman, J.J. (1992), 'Randomization and social policy evaluation' in, Manski, C.F. and Garfinkel, I. (eds.) *Evaluating welfare and training programs*, Cambridge MA: Harvard University Press, pages. 201-30.
- Heckman, J.J. (1979), 'Sample selection bias as a specification error', *Econometrica*, 47, pages 153-161
- Heckman, J.J., Ichimura, H. and Todd, P. (1997), 'Matching as an econometric evaluation estimator', *Review of Economic Studies*, 64, pages 605-654
- Heckman, J.J., and Robb, R. (1985), 'Alternative methods for evaluating the impact of interventions', in *Longitudinal analysis of labor market data*, New York: Wiley.
- Heckman, J.J., and Smith, J.A. (1995), 'Assessing the case for social experiments', *Journal of Economic Perspectives*, 9(2), pages 85-110.
- Hedges, L.V. (1994), 'Fixed effects models', in Cooper H and Hedges L V (eds.), *The handbook of research synthesis*, New York: Russell Sage Foundation.
- Hedges, L.V. (1984), 'Advances in statistical methods for meta-analysis', in Yeats, W.H. and Wortman, P.M. (eds.), *Issues in data synthesis*, San Francisco: Jossey-Bass, pages. 25-42.
- HM Treasury (2002), *The Child and Working Tax Credits: The modernisation of Britain's tax and benefit system*, London: HM Treasury.
- Hobcraft, J. (1998), *Intergenerational and life-course transmission of social exclusion: Influences of childhood poverty, family disruption, and contact with the police*, CASE paper 15, London: London School of Economics
- Inland Revenue (2001), *Working Families' Tax Credit statistics: Quarterly enquiry - United Kingdom May 2001*, London: Inland Revenue.
- Johnson, A. (2002), *Job retention and advancement in employment: review of research evidence*, In-house Report 98, London: Department for Work and Pensions.
- Johnson, G. (1979), 'The labor market displacement effects in the analysis of the net impact of manpower training programs', *Research in Labor Economics*, Supplement 1, pages. 227-54.
- Johnson, P. and Reed, H. (1996), 'Intergeneration mobility among the rich and the poor: Results from the NCDS', *Oxford Review of Economic Policy*, 12(1)
- Johnson, T.R., Klepinger, D.H. and Dong, F.B. (unpublished 1990), *Preliminary evidence from the Oregon Welfare Reform Demonstration*.

Johnson, W. (1978), 'A theory of job shopping', *Quarterly Journal of Economics*, 92, pages 261-277.

Jovanovic, B. (1979a), 'Job matching and the theory of turnover', *Journal of Political Economy*, 87, pages 972-990.

Jovanovic, B. (1979b), 'Firm-specific capital and turnover', *Journal of Political Economy*, 87, pages 1246-1259.

Kellard, K., Adelman, L., Cebulla, A. and Heaver, C. (2002), *From job seekers to job keepers: Job retention, advancement and the role of in-work support programmes*, Department for Work and Pensions Research Report 170, Leeds: Corporate Document Services.

Kellard, K., Walker, R., Ashworth, K., Howard, M. and Liu, W. (2001), *Staying in work: Thinking about a new policy agenda*, Department for Education and Employment Research Report RR264, Norwich: HMSO

Kingsmill, D. (2001), *Kingsmill Review of women's pay and employment*, London: Department for Education and Skills.

Knox, V., Miller, C. and Gennetian, L. (2000), *Reforming welfare and rewarding work: Final report on the Minnesota Family Investment Program*, New York: Manpower Demonstration Research Corporation.

Lessof, C., Hales, J., Phillips, M., Pickering, K., Purdon, S. and Miller, M. (2000), *New Deal for Lone Parents evaluation: A quantitative survey of lone parents on Income Support*, Employment Service, Research and Development Report 101.

Light, A. and McGarry, K. (1998), 'Job change patterns and the wages of young men', *The Review of Economics and Statistics*, 80(2), pages 276-286.

Lipsey, M.W. and Wilson, D.B. (2001), *Practical meta-analysis*, Applied social research methods series volume 49, Thousand Oaks, CA: Sage Publications.

Lin, W., Robins, P., Card, D., Harknett, K. and Lui-Gurr, S. (1998), *When financial incentives encourage work: Complete 18-month findings from the Self-sufficiency Project*, Vancouver: Social Research and Demonstration Corporation.

Lissenburgh, S. (2001), *The New Deal for the Long-term Unemployed Pilots: Quantitative evaluation using Stage 2 Survey*, Employment Service, Research and Development Report 81.

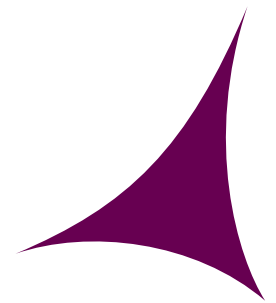
Lissenburgh, S. (2000), *The New Deal for the Long-term Unemployed Pilots: Baseline report*, Employment Service, Research and Development Report 59.

Lockwood, B. (1991), 'Information externalities in the labour market and the duration of unemployment', *Review of Economic Studies*, 58(4), pages 733-753.

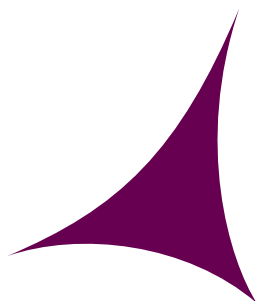
Long, D. and Knox, V. (1985), *Documentation of the data sources and analytical methods used in the benefit-cost analysis of the EPP/EWEP program in San Diego*, New York: MDRC.

Machin, S. (1999), 'Wage inequality in the 1970s, 1980s and 1990s', in Gregg, P. and Wadsworth, J. (eds), *The state of working Britain*, Manchester: Manchester University Press,.

Machin, S. (1998), 'Childhood disadvantage and intergenerational transmissions of economic status', in Atkinson, A.B. and Hills, J. (eds.) *Exclusion, Employment and Opportunity*, London: Centre for Analysis of Social Exclusion, London School of Economics, CASE paper 4.



- Manning, A. (2000), 'Pretty vacant: Recruitment in low wage labour markets', *Oxford Bulletin of Economics and Statistics*, 62, 51, pages 747-770.
- Manski, C.F. (1995), Learning about social programs from experiments with random assignment of treatments, Institute for Research on Poverty Discussion Paper No. 1061-95, Madison: University of Wisconsin.
- Manski, C.F. (1993) What do controlled experiments reveal about outcomes when treatments vary? Institute for Research on Poverty Discussion Paper No. 1005-93, Madison: University of Wisconsin
- Manski, C.F. and Garfinkel, I. (1992), 'Introduction', in Manski, C.F. and Garfinkel, I. (eds.), *Evaluating welfare and training programs*, Cambridge, MA: Harvard University Press, pages. 1-22.
- Marsh, A. (unpublished 2001), *Employment stability among British Lone Parents: Paper II*, London: Department for Work and Pensions
- Marsh, A. (2001a), 'Employment stability among British lone parents', in K. Kellard (ed.) *Lone parents and sustainable employment*, Proceedings from a seminar held on 1st May 2001, Loughborough: Centre for Research in Social Policy.
- Marsh, A., McKay, S., Smith, A. and Stephenson, A. (2001), *Low-income families in Britain: Work, welfare and social security in 1999*, Department of Social Security Research Report No. 138, Leeds: Corporate Document Services.
- Michalopoulos, C., Card, D., Gennetian, L., Harknett, K. and Robins, P. (2000), *The Self-Sufficiency Project at 36 months: Effects of a financial work incentive on employment and income*, Ottawa: Social Research and Demonstration Corporation.
- Mincer, J. (1974), *Schooling, experience and earnings*, New York: National Bureau of Economic Research with Columbia University Press.
- Moffitt, R.A. (1996), 'The effect of employment and training programs on entry and exit from the welfare caseload', *Journal of Policy Analysis and Management*, 15(1), pages. 32-50.
- Moffitt, R.A. (1992), 'Evaluation methods for program entry effects in evaluating welfare and training programs', in Manski, C.F. and Garfinkel, I. (eds.), *Evaluating welfare and training programs*, Cambridge, MA: Harvard University Press, pages. 231-52.
- Montgomery, M. and Cosgrave, J. (1997), 'The effect of employee benefits on the demand for part-time workers', *Industrial and Labor Relations Review*, 47, pages 87-98.
- Narendranathan, W. and Elias, P. (1993), 'Influences of past history on the incidence of youth unemployment: Empirical findings for the UK', *Oxford Bulletin of Economics and Statistics*, (55, 2), pages 161-185.
- Nathan, M. (2001), *Getting attached: New routes to full employment*, London: Fabian Society.
- Nickell, S., Jones, P. and Quintini, G. (2002), 'A picture of job insecurity facing British men', *The Economic Journal*, (112), pages 1-27
- Orr, L. (1999), *Social experiments: Evaluating public programs with experimental methods*, CA, Thousand Oaks: Sage Publications.
- Paull, G., Taylor, J. and Duncan, A. (2002), *Mothers' employment and childcare use in Britain*, London: Institute for Fiscal Studies.



Phillips, E.H. (unpublished 1993), The effect of mandatory work and training programs on welfare entry: The case of GAIN in California, Doctoral dissertation. Madison: University of Wisconsin.

Pissarides, C. (1992), 'Loss of skill during unemployment and the persistence of employment shocks', *Quarterly Journal of Economics*, 107, pages 1371-1391.

Rangarajan, A. (1996), Taking the first steps: Helping welfare recipients who get jobs keep them, Princeton, NJ: Mathematica Policy Research Inc.

Rangarajan, A., Schochet, P. and Chu, D. (1998), Employment experiences of welfare recipients who find jobs: Is targeting possible?, Princeton, NJ: Mathematica Policy Research Inc.

Riccio, J., Bloom, H.S. and Hill, C.J. (2000), 'Management, organizational characteristics, and performance: The case of Welfare-to-Work Programs', in J. Heinrich and E. Lynn Jr., (eds.), *Governance and performance: New perspectives*. Washington, DC: Georgetown University Press.

Schiller, B.R. and Brasher, C.N. (1993), 'Effects of welfare saturation on AFDC caseloads', *Contemporary Policy Issues*, 11(1), pages. 39-49.

Shropshire, J. and Middleton, S. (1999), *Small expectations: Learning to be poor?* York: Joseph Rowntree Foundation.

Slaughter, E., Whiteneck, G. and Baumheier, E.C. (1982), Postplacement services to WIN clients: Final report of a Denver WIN research laboratory project, Denver, CO: University of Denver.

Stewart, M. and Swaffield, J. (1999), 'Low pay dynamics and transition probabilities', *Economica*, 66, pages 23-42.

Strawn, J. and Martinson, K. (2000), *Steady work and better jobs: How to help low-income parents sustain employment and advance in the workplace*, New York: MDRC.

Sweeney, K. (1996), 'Destination of leavers from claimant unemployment', *Labour Market Trends*, October.

Wissoker, D.A. and Watts, H.W. (1994), *The impact of FIP on AFDC caseloads*, Washington, D.C.: The Urban Institute.

Young, M. (2001), People leaving employment: Characteristics and flows, *Labour Market Trends*, 109(11), pages 515-522.

ANNEX 1 – THE JUSTIFICATION AND EVIDENCE FOR INTERVENTION

Overview

Implicit in any welfare-to-work project such as ERA is the view that some feature of the labour market is undesirable and needs rectifying – this is the justification for intervening. The reasons why action may be warranted can vary:

- Often, the motivation is to correct or compensate for market failure that occurs because the conditions needed for efficient functioning do not hold. Most commonly, imperfect information or barriers to entry into the market cause these market failures.
- Distribution or inequality concerns and concern over the level of wellbeing of certain groups are also frequent grounds for intervention. The observed outcome, for instance, whether due to market failure or not, might be such that some individuals or groups are at or below some defined level of poverty; the goal of intervention here would be to raise the living standard of such groups.
- Similarly, a policy encouraging work and economic self-sufficiency might be implemented with the aim of improving the government's budgetary position by reducing transfer payments and increasing tax receipts.

- Justification can also be based on some normative, social value judgement such as dependency on welfare or social exclusion being intrinsically a 'bad' thing and work being a 'good' thing. Such judgements often have a further underlying and usually unexpressed rationale – for example, that welfare dependency and the absence of work is detrimental because it traps individuals in poverty.

The above rationales are neither mutually exclusive nor necessarily consistent with each other. A policy to mitigate market failure by providing more training to low-wage workers, for instance, could also have the goal of raising the living standards of these individuals. As an example of inconsistency, an intervention might achieve its aim of encouraging work but at a net cost to taxpayers.

These factors need taking into account if the justification for intervention is to be convincing. This Annex examines the evidence on retention and advancement in Great Britain, focusing on the problems of individual or circumstantial barriers to work, unemployment and wage scarring, and the low pay/no pay cycle that are the main justifications for intervention. It looks at how various economic theories explain these problems and what these theories imply for any remedial action for individuals prone to such problems.

The problem of retention and advancement in Britain

Viewing the UK working population as a whole, at first sight there does not appear to be much of a job retention problem. In an analysis of the Labour Force Survey (LFS), Young (2001) shows that, quarter to quarter, only a very small proportion of those in employment leave work for non-employment. Over a year, 92 per cent of men and 87 per cent of women who were in work at the beginning of the period were in employment in all four quarters (Young, 2001).⁴⁰

Within this overall picture, however, there is evidence of a retention and advancement problem among certain types of individuals. These individuals are more likely to find it difficult to retain work. If they do retain jobs, these are overwhelmingly jobs in the low wage sector; moreover, they involve a higher proportion of part-time, temporary, or casual work than is the case for the working population as a whole.

Once in work, such persons do not advance in terms of improvements in pay and terms and conditions. The result is that a significant minority of the workforce become welfare dependent or trapped in low-wage work.

Evidence on employment retention and advancement among low-wage workers is quite limited, particularly in Britain. Nevertheless, drawing together information from the few studies that are relevant illustrates the scope and severity of the problems faced.

The evidence on job retention

Evaluations of the various New Deals help shed light on the extent to which new entrants to employment, in other words those leaving the New Deals for work, are able to retain their jobs. These studies support the view that, relative to the working population as a whole, new entrants face greater problems retaining work:

- Analyses of data from the New Deal for Lone Parents Prototype evaluation found that 20 per cent of lone parents who had left Income Support (IS) at the time of the baseline survey had returned to Income Support some ten months later (Hales, *et. al*, 2000).⁴¹
- Survey data from the evaluation of the New Deal for Long Term Unemployed People (ND25plus) shows that 38 per cent of those who entered employment within six months of the start of the programme were not in work a year later.⁴²
- Around a quarter (24 per cent) of young people leaving New Deal for Young People (NDYP) for subsidised employment returned to benefit within 13 weeks (Johnson, 2002). This figure rises to 50 per cent by six months (Nathan, 2001).
- The New Deal for 50 plus (ND50+) evaluation shows that of those entering full-time work under the programme, 19 per cent were no longer working six months later while six per cent were working part-time. Among those who entered part-time work, 68 per cent were still working part-time six months later, two per cent were working full-time, and six per cent were self-employed. The remaining 24 per cent were out of work (Atkinson, 2001).⁴³

⁴⁰ Women tend to leave employment and become economically inactive rather than unemployed – this tends to be because they are looking after their family or home. Men have traditionally left employment for unemployment rather than economic inactivity, though analysis of the most recent quarter to quarter transitions between work and non-employment, based on the LFS, shows that for the first time more men left employment for inactivity than unemployment (Young 2001: 517, Figure 2).

⁴¹ It cannot be assumed that all lone parents leaving IS entered employment as some may have re-partnered.

⁴² This figure comes from a re-analysis of survey data from the evaluation of the ND25plus specifically undertaken for the ERA Demonstration design project.

⁴³ ND50+ involves the payment of an Employment Credit to those taking a full-time job under the programme and paid less than £15,000 per annum. This incentive is paid for a period of a year. The rates of job retention recorded in Atkinson's (2001) study are those prevailing while the vast majority of sample members were in receipt of the incentive (Atkinson, 2001: 35).



Findings from a broad range of studies, which do not focus on New Deal participants, reveal a similar pattern of relatively high exit rates from work into non-employment for certain groups:

- Findings from an analysis of the first two waves of the Survey of Low Income Families show that between interviews held in 1999 and 2000, 17 per cent of lone parents in employment left for either unemployment or inactivity (Marsh 2001). The percentage of lone parent entrants who had entered work of 16 hours or more between the survey interviews and had left employment by the 2000 interview was nearly double, at 32 per cent.
- Sweeney (1996) finds that among recipients of the Jobseeker's Allowance (JSA) who move into work, 25 per cent return to JSA within just 13 weeks and 40 per cent return within six months. Ashworth and Lui (2001) find that 12 per cent of those who leave JSA for permanent work return to claim the benefit within three months whilst the figure for those who leave for temporary work⁴⁴ is 38 per cent.⁴⁵
- More generally, an analysis of the British Household Panel Survey (BHPS) finds that over the period 1991-1997, workers who entered a job from unemployment or inactivity were three times more likely to return to unemployment than those entering a new job from a prior job (Boheim and Taylor, 2000b).
- Using quarterly LFS panel data between 1996-98, Dickens (2000a) finds that of those in entry jobs, which are likely to be low paid, some 27 per cent leave for non-employment after nine months compared to five per cent for the stock of all employees.

In summary, the evidence indicates that there are groups of people in Great Britain facing uncertain and unstable employment prospects. Depending on the group in question, between 15 to 30 per cent or more of new entrants to employment, will leave work and return to unemployment or inactivity within a year. Additionally, those entering a job from inactivity are more likely to return to inactivity than those entering from another job.

Personal barriers to work

The characteristics of workers and job seekers in the labour market vary. This heterogeneity may mean that those experiencing repeated unemployment and/or continual low wages do so because they have different characteristics from those not repeatedly unemployed. These characteristics can range from ill-health or poor literacy and numeracy skills, to more intangible and hard-to-measure traits such as a lack of confidence or poor work motivation, and they can form barriers to getting and holding a job.

Individuals also vary in the situations and circumstances they face, such as the level of local labour demand where they live or the available childcare provision, or they may suffer from a lack of permanent accommodation. Together or in isolation these can be obstacles to obtaining or retaining work. Like individual characteristics, such circumstances need not be mutually exclusive and often a person can face multiple barriers to work due to any combination of circumstances and/or characteristics.

⁴⁴ Of those entering work from JSA, 48 per cent understood their job to be permanent, as opposed to being temporary in some way (Ashworth and Lui, 2001: 10).

⁴⁵ The data set used 'pooling' cohorts of claimants sampled before and after the introduction of JSA. The pooling was done in order to generate a sufficient sample size but as a result, the data, although referred to as describing JSA leavers, in fact includes observations on those leaving the old Unemployment Benefit/IS regimes

A good illustration of such personal barriers is given by Table 1, derived from the baseline report on the New Deal for the Long-term Unemployed pilots (Lissenburgh, 2000). Table 1 shows the types of problems experienced by participants in finding or keeping a job:

An evaluation of the New Deal for Lone Parents (NDLP) (Hales, *et al.*, 2000) shows that almost all the barriers mentioned in the New Deal for the long-term unemployed pilots report are also cited in the NDLP survey, as illustrated in Table 2. Here, the

numbers of multiple barriers faced by the respondents are reported and, as is shown, are considerable: only five per cent of lone parents who were 'work ready'⁴⁶ had no barriers, while 61 per cent faced three or more barriers.

The NDLP survey participants were also asked another question about 'things that would make work more difficult for you' – the responses to this question show the nature of the more intangible barriers to work that can arise.

Table 1: Barriers to working identified in the New Deal for the long-term unemployed pilots¹

Types of problems	Identified barrier ² %	Discussed with NDPA ³ %
No jobs nearby	34	45
Considered too old	31	41
Lack of personal transport	29	51
Own ill health or disability	26	58
Lack of qualifications	24	49
Lack of public transport	17	50
Lack of references from previous employer	13	31
Debt or money problems	10	30
Difficulties with reading or writing	7	65
Illness of another member of the family	6	58
Problems with the law or previous record	6	41
Lack of childcare or affordable childcare	2	18
No permanent place to live	2	33
Problems with drugs or alcohol	2	14
Mortgage problems	2	22
Considered too young	1	33
No problems	13	–

Notes:

1. Based on sample of 942 persons weighted base, 936 unweighted base

2. These were barriers to work identified by the respondent

3. These were barriers to work discussed by the respondent with their New Deal Personal Adviser

⁴⁶ This was a somewhat subjective categorisation and included not only those who were unemployed and looking for work and those unemployed, not looking for work but would like to work but also those employed for less than 16 hours (Hales, *et al.*, 2000: 74).



Table 2: Barriers to work for NDLP sample Lone Parents ¹		
	Work ready %	Postpone work %
<i>Nature of barrier</i>		
Problems with finding or arranging childcare	58	62
Lack of qualifications	46	44
Lack of work experience	39	35
No local job opportunities	30	21
Transport costs	26	25
Attitudes of employers towards lone parents	25	20
Availability of transport	14	12
No appropriate clothes for job interviews	15	16
Debts	15	13
Other	8	10
None of these apply	5	10
<i>Number of barriers</i>		
None	5	11
One or two	34	42
Three to five	45	40
Six or more	16	7
<i>Things that make work difficult</i>		
Need to be more flexible about the hours worked	55	48
Children need parent around	39	58
Nervousness about job interviews	27	25
Need free time to run errands for family	29	45
Confidence about working is low	25	27
Working makes too many demands on time	8	13
None of these apply	18	12

Notes:

1. Based on sample size of 988 for work-ready persons and 919 for postpone work individuals. Some respondents indicate more than one response

This problem of multiple barriers is also noted in a national survey of participants in the New Deal for Young People (NDYP) that finds that 36 per cent of respondents had two or more barriers to work. As the survey report comments:

"There is increasing awareness that some of the unemployed face multiple disadvantages in entering and holding onto work. Some have gone further and argued that these disadvantages can result in deprivation and social exclusion. There is evidence that multiple disadvantage reduces subsequent employment chances." (Bryson et al., 2000: 24)

A follow-up survey on the NDYP participants, conducted a year on from the initial survey, also indicates that obstacles to work are not static for individuals, though they can persist for many people. Table 3 shows, for the eleven major barriers identified in the initial survey, how many people persisted in the problem they identified and how many had 'entered' or 'exited' such problems by the time of the follow-up survey.

The follow-up survey report notes:

"A notable point ... is that new entries were generally a larger part of the barriers reported at the follow-up interview than were the cases of persistence across the two surveys. This suggests the potential importance for this group of having continuing access to personal support and assistance, from which help can be sought as new problems arise."
(Bonjour et al., 2001: 59)

Barriers caused by circumstances or by traits intrinsic to an individual have important implications for projects such as the ERA Demonstration. They suggest that if these barriers can be identified and support given to overcome them, then job retention can be achieved. So, for example, if a lone parent finds it difficult to hold a job due to undependable childcare, help in finding childcare can greatly enhance the chances of the lone parent remaining in work. Similarly, a person suffering from low confidence or poor motivation may be assisted through a caseworker fostering and boosting their confidence or motivation.

Table 3: Persistence, entries, and exits to problems affecting NYDP survey participants¹

Type of problem	Number in each category		
	Persistence	Entries	Exits
Own ill-health	163	223	168
Ill-health of other family member	21	65	99
Childcare	9	39	13
Lack of public transport	72	169	202
Lack of personal transport	250	275	375
Lack of jobs in locality	300	329	443
Lack of employer references	121	183	283
Debts or money problems	84	162	196
Lack of permanent accommodation	20	57	86
Prosecution or criminal record	68	60	96
Drugs or alcohol problems	15	30	30

Notes:

1. Based on sample size of 2,347 persons



The evidence on scarring effects

Labour market theory suggests that leaving a job can have positive as well as negative impacts on an individual's subsequent earnings and employment. Job search theory, for example, proposes that termination of employment is largely due to a poor 'match' between the individual and the type of job they have. Under such circumstances, a process of intensive job search undertaken while in non-employment can lead to a better job match that will be reflected in higher productivity and higher wages.

Under this theory, if there is an intervention to encourage individuals to remain in work or in a particular job, this may actually harm their longer-term economic prospects by preventing them from acquiring a better match. This would be especially true for young people who are inexperienced in the labour market and might need to make a number of moves between jobs to find a good match (a phenomenon often referred to as 'job shopping').

The evidence does support this view to a limited extent: Ashworth and Liu (2001) note that the high rates of return to JSA from entry jobs, particularly the number of resignations, suggest job mismatches. Similarly, Gregory and Jukes (2001) find that a spell of unemployment for young men only negatively impact on wages when the spell duration is prolonged, implying that, for these individuals, briefer spells of unemployment can lead to better job matches and no loss in subsequent wages.

Moves into and out of work may, therefore, simply reflect the workings of an efficient labour market with job separations leading eventually to better jobs. To the extent that

unemployment allows an improved 'sorting' of workers among jobs, the expectation is of higher earnings and a greater attachment to work in the longer run.

What would not be anticipated, is that those who experience a single or a small number of job interruptions, continue to experience further spells of unemployment or persistent wage penalties⁴⁷ or both as a direct consequence, *ceteris paribus*, of such interruptions. Any such 'scarring' effect, whether through unemployment or wages or both, would indicate some sort of market failure.

The evidence is that there are unemployment scarring effects in Great Britain so that workers' previous unemployment experience has implications for their future employment experience. Arulampalam, Booth and Taylor (2000), for instance, use BHPS data for the period 1991-95 to model persistence in unemployment occurrence for men. They find evidence that strong unemployment scarring effects exist with respect to previous unemployment incidence, especially for men aged 25 and over. Arulampalam (2002) revisits the BHPS data set and extends the analysis in Arulampalam, Booth and Taylor (2000) concluding that:

"The results confirm the earlier finding that strong state dependence effects do exist with respect to previous unemployment. This finding is consistent with the 'scarring' theory of unemployment – an individual's previous unemployment experience has implications for his future labour market behaviour, perhaps because of depreciation of human capital, or because employers use an individual's previous labour market history as a screening device about his productivity" (Arulampalam, 2002: 22)

⁴⁷ This phenomenon is referred to in the economic literature as state or structural dependence. Arulampalam *et al* (2000) give the definition "True state dependence – or scarring – is where there is a causal link between past unemployment and current unemployment, so that an individual who does not experience unemployment now will behave differently in the future to an otherwise identical individual currently experiencing unemployment." (Arulampalam, Booth, and Taylor, 2000: 25). Throughout this paper, 'scarring' will be taken to mean such state or structural dependence.

This further study once again confirms that scarring effects are much greater for men aged 25 and over than for those aged less than 25.

Much of the empirical evidence for scarring is based on studies typically involving time spans of between three to six years (see Arulampalam, 2000, Boheim and Taylor, 2000b, Arulampalam, Booth, and Taylor, 2000, and Narendranathan and Elias, 1993). One notable exception, is the analysis by Gregg (2001), who examines scarring over a much longer time span, using the National Child Development Study, a survey cohort of individuals born in the same week in March 1958. His study shows that:

- Unemployment experience is concentrated on the same minority of the workforce over extended periods.
- There is strong evidence of scarring induced by early experience of unemployment for men but only minor effects for women.

One issue raised in this analysis, and common to other studies (see Gregory and Jukes, 2001, Arulampalam, 2000, Stewart and Swaffield, 1999, and Light and McGarry, 1998), is the extent to which findings of concentrated unemployment spells amongst a minority of the workforce is really due to scarring or to unobserved heterogeneity. If the cause is heterogeneity, what appears to be the result of scarring could well be related instead to individual characteristics similar to those discussed above.

Gregg attempts to control for this heterogeneity and finds that factors such as low educational attainment, depressed local labour market, and coming from a disadvantaged family background all raise a person's susceptibility to unemployment. Despite this, the study finds that the

structural dependence relationship exists independently of these other characteristics, although Gregg concedes that the method used to control for heterogeneity is problematic.

To examine wage scarring, Arulampalam (2001) used BHPS data on the wages of men over the period 1991-97. This study finds that:

- A single spell of unemployment for men leads to a wage on employment re-entry some six per cent lower than would be the case where no job interruption had been experienced. Over a three-year period of continuous employment, this 'wage penalty' rises to around 14 per cent before declining.
- The first spell of interruption carries the largest penalty, with subsequent spells of unemployment carrying a less pronounced wage scar.
- Once the effects due to the incidence of unemployment are taken into account, there are no additional detectable scarring effects due to unemployment duration.

Gregory and Jukes (2001) also examine the effects of male unemployment on wages, combining New Earnings Survey (NES) and Joint Unemployment and Vacancies Operating System (JUVOS) data for the period 1984-94. They split the effect of unemployment into two components, that attributable to the job interruption, or the incidence, and that due to the duration of the unemployment spell. Their findings are⁴⁸:

- Unemployment incidence does give rise to a wage penalty in the first year but that this is only a temporary effect with the penalty largely dissipating after two years of continuous employment.

⁴⁸ The findings in this analysis are contrary to those in Arulampalam's regarding the relative importance of unemployment duration and incidence of scarring. The disagreement may be attributable to the much larger sample size in the Gregory and Jukes study, differences in the labour market during the different time periods covered by the studies, and to differences in the definition of unemployment used in the two studies (see Arulampalam, Gregg, and Gregory, 2001: 581, and Arulampalam, 2001: 601).



- Conversely, the effect of unemployment duration is permanent and proportional to the length of the spell. This duration effect includes the cumulative duration built up through a number of repeat spells.
- The impact of spell duration on future earnings is greatest for older workers and for more skilled workers, and least for low-paid and younger workers.⁴⁹
- Although the incidence of unemployment causes a reduction in wages, continuous subsequent employment brings a substantial recovery. The irrecoverable dimension arises from the duration of the unemployment spell.

Nickell, Jones, and Quintini (2002) examine three aspects of job insecurity facing British men between the period 1982-97, also using NES and JUVOS data. They find that:

- On average, there has been no systematic increase over time in the chances of becoming unemployed but that there is some indication that low-level manual workers have faced a small but steady rise in their chances of becoming unemployed.
- Despite this stability in the chances of becoming unemployed, there has been a strong tendency for the cost of unemployment in terms of wage losses after unemployment spells to increase, particularly for older age groups and higher skill groups.
- For both men in continuous employment and those who change jobs, there is a clear and significant increase in the chances of a substantial year-on-year decline in real hourly wages over the period, particularly for men in the lower skill groups. This increase in earnings insecurity is across all sectors and not just manufacturing sectors subject to structural decline.

The pattern of wage losses resulting from unemployment is similar to that found in the Gregory and Jukes study, where the young and the low-paid suffer the smallest effect. As these latter groups are most likely to be affected by unemployment however, the relatively reduced impact of scarring partially re-balances proneness to unemployment, perhaps implying that the problem is diminished in dimension.

Nonetheless, three considerations suggest that scarring effects are important for young or low-paid workers. First, all the studies quoted do find a persistent scarring effect for the young and the low-waged. Second, it is plausible to argue that the marginal disutility to these groups of a loss of earnings is greater than for groups with higher income. Lastly, if they are prone to a greater incidence of unemployment, this continual interruption to their recovery of wages will magnify the impact of their earnings losses. The Nickell *et al* (2002), study finds that on average, 17 per cent of the low skill group suffered one or more spells of unemployment compared to 11 per cent for the high skill group over the period examined.

The existence of scarring has important ramifications for any intervention. It suggests that a policy to prevent repeated unemployment spells as well as limit the duration of those spells, particularly among more mature workers, may independently act to improve job retention and employment advancement. Such a policy might, for instance, inform individuals of the existence of scarring and its consequences or tackle underlying causes that make certain persons prone to repeated and extended unemployment spells, or a mixture of both. The nature and emphasis of the policy will be determined by what economic theories have to say about the causes of poor retention and advancement. These theories are examined next.

⁴⁹ This is true provided their unemployment spells are not extended in duration.

Underlying economic theories

As discussed above, unemployment as part of the better job match process can lead to greater economic efficiency. Yet these efficiency gains do not rule out unfavourable impacts, either temporary or permanent, on individuals from the reallocation process or from intervening unemployment spells. Job search theory, for example, implies that whenever a good job match, reflected by high productivity and high wages, is disrupted, a subsequent wage loss is to be expected unless an equivalent or better match is found.

Job search theory is one variant of a number of job mobility theories.⁵⁰ The ‘search good’ model of job matching explains mobility in terms of voluntary moves to more productive employment where the productivity of the new job is known ex ante, making jobs ‘search goods’ (Jovanovic, 1979b and Burdett, 1978). This model predicts that workers move to increasingly better quality matches and hence their mobility slows over time. Here, mobility per se does not affect wages as the latter is only affected by match quality, which is a time invariant characteristic⁵¹ – that is, mobility has no effect on wages after its relationship with time-invariant job attributes are accounted for – a good match could be found quickly, or may take some period of time to secure.

Scarring under this model would occur if the effort invested to find a good match decayed with the duration of unemployment. There would also be scarring effects if the initial job separation were involuntary and the worker needed to take a lesser quality job for reasons such as labour market conditions, loss of benefits, diminishing utility from leisure, or liquidity constraints. In these circumstances,

intervention would focus on support and assistance to get the best job match possible, achieved through the provision of information and through developing the person’s skills and abilities.

Another variant is the ‘experience good’ model of job matching (Jovanovic, 1979a, and Johnson, 1978), so called because productivity in the new job is not known ex ante but revealed over time through working at the job. If the match turns out to be worse than initially perceived, wages will accordingly adjust downwards because the worker is less productive than that employer anticipated. This will lead to separation if wages fall below those available to the workers in other jobs. Alternatively, the employer might simply dismiss the employee. Although true match quality again is time-invariant in this model, mobility is driven by time varying perceptions of job quality and thus will be correlated with wages even when the relationship between wages and unobserved time-invariant personal and job effects are controlled for. A study by Light and McGarry (1998) finds evidence that workers who undergo persistent mobility have lower wage paths than less mobile workers, independent of unobservable individual and job-specific traits.

This experienced good model allows for the possibility that workers can experience a sequence of bad matches with resultant persistent wage losses. In this situation, any action would concentrate on reducing persistent mobility by again seeking to obtain the best job match, with the emphasis on better information about the jobs on offer. There could also be an effort to inform the employer as fully as possible about job candidates’ capabilities.

⁵⁰ The oldest of these is the ‘mover-stayer’ model where underlying individual characteristics determine high productivity workers who avoid job turnover and low productivity workers who undergo persistent mobility. In this model, movers’ mobility is time invariant and any observed association between job turnover and lower wages is due to the correlation between mobility and the unobserved heterogeneous characteristics that determine productivity. According to this theory, if this relationship could be taken into account, there would be no relationship between mobility and wages. The findings of the studies cited above, therefore, would be explained under this model as merely due to the heterogeneity problem discussed previously.

⁵¹ In other words, the chances of obtaining a good match do not vary with time. If a worker’s chances of a good match did vary with time due to, for example, the accumulation of human capital, this would challenge the conclusions of the theory.



Other ‘non-job mobility’ economic theories can also explain why scarring occurs. Human capital theory (Pissarides, 1992, Becker, 1975, and Mincer, 1974) for example, argues that, through work experience and training, workers accumulate skills that are rewarded through their wages as evidenced by the association between wages and job tenure. These skills have two components: the job- or firm- or occupation-specific skills and more general transferable skills.

To the extent that job-specific skills are non-transferable, their contribution to a worker’s productivity, and hence wage, will be permanently lost when employment in that job is ended. Moreover, it is contended that even general transferable skills can depreciate with unemployment, with this depreciation accelerating as the unemployment duration lengthens. Again, the effect will be to lower productivity and thus a lower entry wage in future employment.

The implication of the loss of job-specific skills is that the incidence of unemployment is important and duration should not be of influence. Conversely, the loss of general skills should be linked directly to duration and not incidence. A further inference is that job-specific skills will start to accrue again with tenure whilst the decline in general skills will stop, and possibly reverse, on re-employment. Thus the earnings setback on re-entry into employment should be followed by an upward resumption in earnings as human capital increases and productivity rises.

The studies previously cited do support these inferences to some extent: they all find that the impact of scarring diminishes with continued tenure in a subsequent job. Gregory and Jukes (2001) and Nickell *et al* (2002) find that employment interruption for young workers brings only a relatively small wage penalty but that the penalty rises

progressively higher with age – this is consistent with human capital theory. The young are likely to have built up less human capital than older workers and so suffer a proportionately smaller loss of skills from unemployment.

Similarly, Gregory and Jukes (2001) explain their finding that the cumulative duration of unemployment is significant in explaining wage scarring by stating that:

“Repeat spells not only build up cumulated duration but interrupt the recovery process from the previous dislocation.” (Gregory and Jukes, 2001: 622).

Market or institutional failures can perpetuate the loss of human capital, as can an initial deficit in human capital. Low-income persons, for instance, may not have the capital resources to invest in education to compensate for the loss of, or insufficient, human capital. A lack of collateral and a high risk of default may limit their access to private financing.

The human capital model suggests that a range of policies may be needed to tackle different aspects of the problem. To compensate for the loss of job-specific skills following unemployment, training could be encouraged in subsequent re-employment to rapidly build up these skills in the new job. If the problem is seen more as one of the decay of general skills over time, individuals could be encouraged back into employment as soon as possible even if the job is not the best match that could be achieved, or education and training could be provided during the period of inactivity. Similarly, those individuals facing financial or other obstacles to acquiring or improving human capital could be given assistance to overcome these obstacles.

A third relevant economic theory is that of asymmetric information or, more specifically, signalling theory (Blanchard and Diamond, 1994, and Lockwood, 1991). As in the 'experienced good' search model, there is limited prior knowledge of the new worker's productivity, this time on the part of the employer. The employer will thus look for signals that convey information on the worker. The prospective employee's history of unemployment in terms of incidence and duration will be one key signal. A record of numerous spells or long durations of unemployment will, according to the theory, provide a negative signal to employers.

In looking at who employers choose to interview for low-skilled entry-level jobs, Manning (2000) shows that they select against the unemployed and those with no relevant job experience. Similarly, a 1996 survey carried out by Atkinson, Giles, and Meager, cited in Atkinson and Williams (2003) found that previous experience and a continuous job record were important in the selection criteria of UK employers.

Atkinson and Williams (2003) also cite a 1987 study by Meager and Metcalf that found that, in at least half of the jobs examined, the long-term unemployed were at risk of rejection simply because they were long-term unemployed. Taking these two studies together, the observation is made that:

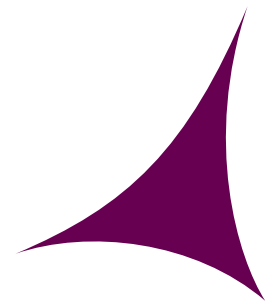
"Where both sets of evidence concur is that for longer durations of unemployment, attitudes harden and willingness to recruit diminishes. They both indicate that both extended, unbroken spells, and repeated discontinuous ones, seem to be matters likely to be taken seriously and widely into consideration by recruiters." (Atkinson and Williams, 2003: 14).

The history of wages can be another key signal for employers. Past low wages could be viewed by employers as a negative indicator of an individual's productivity, or as a signal of a high turnover propensity, and thereby discourage job offers, leading to the 'low pay/no pay' cycle discussed below.

Signalling theory makes two predictions on the impact of unemployment on subsequent earnings. The first is that any initial wage penalty because of incomplete information about a worker's productivity should be quickly eroded if productivity proves to be higher than the employer inferred from the individual's employment history or wages record. The second, confirmed by Arulampalam (2001), is that unemployment due to redundancy should give less of a negative signal than unemployment due to dismissal. As a result, redundant workers should experience less wage scarring than dismissed workers.

Imperfect information is also applicable to workers. Some persons may have a high preference for leisure over work and thus be less committed to retaining work. This preference may be derived from an erroneous under-assessment of the value of work to them and the longer-term costs to them of being unemployed. In other words, individuals may not have sufficient information about the consequences of further spells out of work in terms of their future job prospects and wages.

If they are unaware, for instance, of the way employers assess applicants, they may place a lower premium on retaining work than they would do had they access to full information. If scarring were a feature of the labour market, a continual return to benefits without knowledge of this factored into their decision would be an economically inefficient choice on their part. Similarly, some persons may make insufficient investments in their own



human capital because of imperfect foresight, inadequate information about the returns to education and training, or because they put a much greater value on current gratification than future rewards.

The obvious remedial intervention in situations, where asymmetric information is thought to hold, is to improve the quality of information so that employers and employees make better-informed judgements and decisions. This could include persuading employers that existing signals do not necessarily convey wholly accurate information and that alternative signals would be better. In addition, workers should be advised of the full consequences of additional spells of unemployment, especially if these spells result from dismissals.

The low pay/no pay cycle

Further empirical evidence on retention comes from research into what is termed the 'low pay/no pay' cycle. The 'low pay/no pay' cycle exists where a group of workers persistently cycle between short-lived, low-paid jobs and unemployment or inactivity. Stewart and Swaffield (1999) find, for example, that not only are the low paid more likely to exit from work into non-employment than higher-waged workers, but that they are also more likely to enter low-wage jobs upon reattachment to the labour market. Additionally, their analysis finds evidence of considerable persistence in low-paid work, a conclusion confirmed by Dickens (2000b).

This low pay/no pay cycle has a number of implications for scarring under each of the economic theories discussed. As alluded to, if employers see low wages as some sort of indicator of a worker's abilities or propensities, then this can affect employment chances. Alternatively, under the human capital approach, such a cycle can result because there is little opportunity for human

capital accumulation in low-paid jobs, or because human capital depreciates during the 'no pay' phase, thereby keeping an individual's productivity low and reducing the probability of emerging from low pay in the future.

Under 'job search' theory, continuous spells of low pay/no pay could influence workers' perceptions of their market value, thereby discouraging them from applying for better paid employment. It could make them more likely to accept poorer quality jobs that have more chance of being destroyed, thereby increasing their chances of future unemployment. Additionally, if poorer quality, low-waged jobs are easy to find, the cost of searching for them will be low. Job search theory then implies that it will be relatively costless to leave one job because the next will be easily obtained.

All the economic theories discussed above imply that low pay/no pay cycles tend to perpetuate themselves. Once caught in such a cycle, it will be difficult to escape without outside intervention. This is in contrast to the pure heterogeneity case, where the individual characteristics that affect the chances of low pay are not affected by experience of low pay. Stewart and Swaffield's (1999) study indicates that, even after allowing for heterogeneity, the contribution of scarring is considerable: being low paid in one period in itself increases the probability of being low paid in the following period.

Whatever economic theory is thought to underlie low pay/no pay cycles will again determine the main response to this phenomenon. Any intervention will be centred on improving human capital through education and training if this is thought the key element, whereas signalling theory would suggest the provision of better information so that the signals are improved or the use of other signals encouraged. Job search

theory would emphasise the importance of boosting individuals' morale and perceptions of themselves so that they have the confidence to try to find better paid work.

Implications for interventions to address retention problems

As seen above, the remedial action that is appropriate in response to scarring or to the fact that some workers are trapped in a low pay/no pay cycle, differs according to the theory one accepts. Yet the competing theories make various predictions that are not always clear-cut or mutually consistent. Moreover, there are interactions between the theories. Atkinson and Williams, for example, find that:

"...there is clear evidence in this research that most employers believe that motivation, behaviour and skills (in that order) deteriorate during unemployment, and would thus be looking for indications of this (one way or another) among longer term unemployed." (Atkinson and Williams, 2003: 14).

In other words, many employers believe that human capital deteriorates with unemployment but, because of imperfect information, use employment history signals to assess the extent of deterioration. Would a policy intervention here focus on improving human capital or on persuading employers to reconsider how they evaluate the unemployed, or both?

It is no surprise that there is no consensus as to which is the 'right' model. It is also unsurprising that none of the available theories offer a complete explanation of all the empirical evidence. Interventions will be based on addressing a range of factors which different labour market theories suggest are important. Many of the studies cited place greatest emphasis on human capital aspects, but asymmetric information and job search theory have also been used to explain findings.

What the available evidence does suggest, is that tackling scarring and the low pay/no pay cycle through retention policies will have important long-run effects. As Arulampalam *et al* note:

"...if there is ... (scarring) ..., then policies reducing short run unemployment incidence will have longer run effects by reducing the natural rate of unemployment." (Arulampalam, Booth, and Taylor, 2000: 25)

Another implication is that improving retention through tackling scarring effects and the low pay/no pay cycle will, of itself, lead to advancement for many workers, as it is these phenomena that retard progress into better paid jobs.

Summary on retention

To summarise, the evidence shows that there is a problem of job retention for certain individuals and that being unable to retain work has negative, longer-term implications for future labour market attachment, job stability, and wages. These negative phenomena are related to poor job retention in and of itself, as well as to work disincentives and individual job and personal characteristics, and arise from market failures that lead to unemployment and wage scarring.

Not all unemployment, therefore, appears to be about improved job matches. Consequently, there seems to be scope for intervention to address the problems caused by poor retention. The nature of the intervention will be determined by the mechanisms through which the problems manifest themselves, whether through inadequate or deteriorating human capital, information asymmetry, or uninformed job search. The conclusion, on the basis of the empirical evidence, is that there is scope for intervention to help individuals at risk of losing work to remain attached to work, avert future spells of unemployment, and accrue improvements in lifetime earnings.



Table 4 summarises the type of policy interventions that the different economic theories suggest should be implemented for various categories of workers. Table 4, also considers the policies necessary if there are

personal barriers such as childcare, transportation, and health problems that limit work and other work disincentives that result from high marginal tax rates.

Table 4: Retention policies for possible target groups

	Human capital	Job matching	Information asymmetry	Personal barriers	Work disincentives
New workforce entrants	Guide workers to jobs that offer opportunities to acquire specific human capital, as they are less likely to leave a job.	Obtain as good a job match as possible first time so individual continues working.	Provide employer accurate information about individual so the individual is not fired after being hired. Emphasise importance to worker of initial job paying good wages so as not to send negative signals to potential future employers.	Provide needed services directly or help finance needed services so that once job is found, individuals continue working.	Not applicable.
Long-term unemployed	Provide training or subsidised employment while not working so skills do not deteriorate. Find job as quickly as possible to stop human capital decay.	Ensure that the next job is not worse than last job. Ensure effort invested in looking for work does not deteriorate with time. Provide information and assistance to overcome reasons that necessitate taking on poorer quality jobs.	Provide employee with accurate information about job prospects so the individual is not fired after being hired. Find job as quickly as possible to minimise negative signal. Provide individual with better information about the costs of spells out of work. Advise how to improve signals.	Provide needed services directly or help finance needed services, including those relating to more 'intangible' barriers such as poor motivation so that once job is found, individuals continue working.	Ensure deduction rates not punitive so that benefits loss compensated for by wages.
Intermittent or part-year workers	Provide incentives to stay in job so specific human capital not lost. Provide training during non-work periods.	Obtain or improve good job match so that the individual continues working. Provide information and assistance to overcome reasons that necessitate taking on non full-time jobs.	Inform individuals of the negative signal resulting from job switches, especially if they result from dismissals. Advise how to improve signals.	Provide needed services directly or help finance needed services if barriers cause irregular work patterns.	Provide financial incentives to work a greater fraction of the year.
Part-time workers	Retention not necessarily a problem for part-time workers.				

The evidence on advancement

If the evidence on retention is relatively sparse, there is an even greater paucity of evidence for advancement, not least because data spanning longer periods is required. A recent review of the research evidence for job retention and advancement notes there is very little evidence on the problems faced by benefit leavers in job advancement (Johnson, 2002).

To a considerable extent, the issues of retention and advancement are intertwined. It has already been suggested that promoting job retention can actually be detrimental to future advancement if it prevents better job matching. On the other hand, it appears difficult to promote job advancement without a background of job stability. The reported link between tenure and wages seems to demonstrate this. As noted above, improving retention by reducing scarring effects can lead to advancement for some workers. The evidence already cited on retention problems, therefore, can also be used in a discussion of advancement.

Concern over advancement stems in part from the fact that wage inequality has risen sharply over the last two decades (Machin, 1999). In addition to distributional concerns, there is the fear that those at the lowest wage levels may be trapped in in-work poverty.

This would not be of such concern if there were a substantial degree of mobility within the wage distribution so that low-paid workers could progress to higher earnings in the near future, with the low-paid position acting as a springboard to rising lifetime earnings. Such evidence as there is, however, suggests that mobility is limited at best and that those earning the lowest wages are more likely to enter unemployment or

inactivity than improve their earnings, at least in the short run (Dickens, 2000b). Moreover, Boheim and Taylor (2000a) suggest that previous labour market status is important in determining the probability of upward career mobility, although they also find that duration of the previous unemployment spell apparently has no effect.

Furthermore, there is evidence that wage mobility has actually declined at the same time as wage inequality has grown (Dickens, 2000b, and Stewart and Swaffield, 1999). As Dickens summarises:

"It appears that individuals find it harder now to better their position in the wage distribution than they did 20 years ago. This has occurred against a backdrop of a huge rise in cross-sectional wage dispersion. Not only are the differences in wages between individuals in a given year larger than they were, but the possibility of moving up the distribution over the next year has now become more remote. So the low paid are worse off both in terms of the relative wage they receive and in terms of their opportunity to progress out of the low-pay trap." (Dickens, 2000b: 496)

As already seen, prospects for advancement in terms of improvements in pay and conditions are reduced because of scarring for those entering employment after a spell, or spells of unemployment. Poorer prospects for advancement are also associated with the nature of entry-level jobs that are taken on. In the context of the low pay/no pay cycle, Dickens (2000a) examines the characteristics of entry jobs⁵² against that of the total stock of jobs and finds:

- Some 54 per cent of all entry jobs are part-time compared to 23 per cent of the stock of jobs.

⁵² That is, the types of jobs that individuals entering work are taking. Although not further defined, this presumably means first time entrants and entrants from unemployment or inactivity.



- Moreover, 36 per cent of entry jobs are temporary jobs as against six per cent of all jobs. These temporary jobs are much more likely to be seasonal, casual, or agency work than existing temporary jobs.
- Entry jobs are less likely to be in the public sector or in larger firms compared to existing jobs – 18 per cent against 28 per cent and 49 per cent against 68 per cent respectively.
- In terms of occupations, entry jobs are more likely to be junior positions – 51 per cent are in junior, non-manual, personal services, or unskilled manual positions, compared to 30 per cent of existing jobs.

The tendency in recent years towards part-time work and other forms of 'flexible' employment such as temporary jobs, particularly seasonal and casual work, is confirmed in a number of studies (Ashworth and Liu, 2001, Boheim and Taylor, 2000b, and Booth, Francesconi, and Frank, 2000 among others). The issue of part-time or other forms of flexible employment is not pertinent to advancement if such work is undertaken on a voluntary basis and on the basis of complete information. There is nothing intrinsically wrong with such types of employment per se and, in some circumstances, it would be rational and preferred for both workers and employers. Parents who work part time, for instance, may choose to do so in order to have more hours available for childcare. Similarly, workers for whom leisure has a greater marginal utility than earnings from work may elect to be employed intermittently in order to have more time during the year for other activities. From an employer's perspective, such flexible employment could be the best response to market demand patterns. Atkinson and Williams (2003) make the point that:

"It must be remembered that the major expansion of part-time work in the past decades has been in sectors where total labour inputs are not limited to less than full time, but are rather where employers find advantage in making up total hours from several part-time inputs, which can then be varied to match customer flows." (Atkinson and Williams, 2003: 19).

The problem with flexible working arises when it is involuntary or based on imperfect information. Part-time and intermittent employment might not be in the long-run economic interest of some persons who, because of inadequate foresight or insufficient information, may not realise this or be aware of the potential gains that can accrue with respect to adding to their human capital through full-time work.

Flexible employment on an involuntary basis reflects barriers to entry into the market place since individuals with particular characteristics, constraints, or employment histories take such employment because they are unable to obtain full-time work. Taking the example of childcare again, it may well be that lone parents are in part-time work not because they wish to be, but because they lack sufficient resources or contacts to arrange for childcare if they increase their working hours. A study on low-income families in Britain (Marsh *et al*, 2001), reports that, of lone parents working less than 16 hours a week who were asked if there were any particular reason that prevented them looking for a job of 16 hours or more, some 45 per cent cited problems related to childcare availability or affordability.

There are other problems with flexible working arrangements. Various pieces of research have connected temporary work with jobs of poor quality, shorter duration and, in some cases, lower pay and fewer prospects (Ashworth and Liu, 2001, Boheim and Taylor, 2000b, and Booth, Francesconi, and Frank, 2000). Booth, Francesconi, and Frank (2000) show that, despite evidence that some temporary work does lead to permanent employment, temporary workers are, on average, less happy with their jobs, particularly with their promotion prospects and job security.

Moreover, those in temporary work receive less training⁵³ and are, on the whole, less well paid; they also suffer a greater variation in hourly wages, and so greater income instability, than those in permanent employment (Booth, Francesconi, and Frank, 2000, Arulampalam and Booth, 1998). Booth, Francesconi, and Frank (2000) find that, over the period 1991-97, wages are lower for men in all forms of temporary work compared to those in permanent work, though this is tempered for temporary workers with fixed-term contracts. These results hold for women in casual and seasonal jobs, but not for women in fixed-term contracts.

Part-time work is also associated with the types of jobs and the types of workers for whom advancement is more problematic. A third of women in permanent employment work part-time (Bower, 2001) especially women with young children. The wages received by these part-time workers tend to be lower than for full-time female workers (Bardasi and Gornick, 2000).

The types of jobs females working part time are in and the fact that they tend to possess lower levels of human capital, largely account for this wage differential. In addition, Manning and Robinson (2000) suggest that part-time women workers are paid less than those working full time because they are more likely to have experienced job interruptions of some form, usually associated with child birth.

Part-time workers are also less likely to receive work-based training (Arulampalam and Booth, 1998). To some extent this is not surprising, as it is known that part-time employees have lower levels of human capital and that those with lower levels of human capital receive less training (Green, 1999). Furthermore, the returns to training for employers are likely to be less substantial in the case of part-time workers, simply because there are fewer working hours in which the investment can be recouped.

The 1997 DfEE skills survey shows that 28 per cent of part-time workers had received training in the last 12 months compared to 46 per cent of full-time workers (Kingsmill, 2001). Based on the first five waves of the BHPS covering the period 1991-95, Arulampalam and Booth (1998) look at the incidence of training and education undertaken in the workplace, particularly training to improve skills directly related to a current job. They examine the rates of participation in training across all jobs and part-time jobs, for both men and women, and find:

- Around 36 per cent of all male employees received training (Arulampalam and Booth, 1998). For men in part-time work, around a third (31 per cent) received work-focused training.

⁵³ Particularly men and women in seasonal or casual employment compared to those in permanent work (Booth, Francesconi, and Frank, 2000: 15).



- For all women employees, the proportion stood at 33 per cent, whilst for women in part-time work 23 per cent had the opportunity to increase or improve their work-based skills.
- Taking into account job, employer, and personal characteristics, part-time male workers are seven percentage points, and part-time females nine percentage-points less likely to receive work-related training than their full-time counterparts (Arulampalam and Booth, 1998).

The evidence that exists concerning the relative opportunities for job training in part-time and full-time employment pertains *to formal* job training. However, informal opportunities for learning on the job may be more important, especially in positions that pay above entry wages. In many such jobs it may take some time after being hired before a worker reaches his or her full productive potential. This would be the case, for example, in positions where the skills to operate highly specialised equipment must be acquired or reputations need to be established or contacts developed.

In these cases, the firm's investment cost is not a direct cash outlay for training, but the worker's reduced productivity until the learning is completed. It seems likely that firms would prefer to hire full-time, rather than part-time, workers into positions where informal learning is important simply because the learning will be completed more quickly and they will have more hours over which to recoup their investment. Thus, individuals who prefer part-time to full-time work may find many such jobs preclude them.

Low initial human capital and a lack of training opportunities need not be the only reason why part-time workers face a disadvantage when compared to full-time

workers. One reason that employers may pay a lower wage or provide fewer opportunities to part-time workers, than to full-time workers, are the so-called 'quasi-fixed costs of employment'; these are the costs of employing workers that do not increase proportionally with hours worked. The costs to employers of non-wage benefits, job security, recruitment, and training are all examples of quasi-fixed costs because their costs per hour diminish as a greater number of hours are worked.

There are two ways in which employers can respond to the existence of such costs. First, they can hire fewer part-time workers and more full-time workers, thereby reducing the relative demand for part-time workers (Disney and Szyszczak, 1984, and Montgomery and Cosgrave, 1997). Second, they can offer a less expensive compensation package to part-time workers (for example, a lower hourly wage or fewer non-wage benefits) than to full-time workers and invest in less job training for part-time workers. The important implication of this, in the context of advancement, is that a worker who chooses to work part-time, rather than full-time, is likely to receive less remuneration and have fewer opportunities for advancement.

To a large extent, the ability on the part of employers to offer differential employment packages to full-time and part-time workers will depend on market conditions and the ability of workers through unions to resist such differentials and negotiate better terms. Booth (2001) develops a theoretical model of the behaviour of firms and workers in a fictional labour market and this shows that part-time/full-time differentials will be determined by differences between part-time/full-time workers in terms of, *inter alia*, employers' monopsony power and variation in union coverage.

These two reasons are of particular significance in that they are not specific to, and wholly outside the control of, individual workers. Because of market imperfections such as the non-wage costs of changing jobs (particularly for women), oligopolistic product markets, or asymmetric information about worker quality, firms may have some monopsony power, creating a wedge between wages and marginal product. Both part-time and full-time wages are affected by firms' monopsony power but the magnitude of the wedge will typically differ across part-time and full-time employment (Booth, 2001). Similarly, union coverage can lead to negotiated benefits, especially non-pecuniary benefits, for full-time workers not available to part-time workers because union coverage is smaller in the part-time sector.

If these two factors are of any importance in reality (and the evidence is extremely sparse although Arulampalam and Booth (1998) show, in the context of work-related training that, among other factors, a union presence is associated with more training), the implication for policy is that a switch from part-time to full-time work will assist advancement.

Such a conclusion is further reinforced by the evidence that there appears to be a long-term pay-off from working longer hours for both men and women (Campbell and Green, 2002). Using BHPS data, Campbell and Green (2002) estimate the impact of working longer hours over 1991 to 1995 on 1996 wages. The study finds that:

"...working longer hours in Britain is positively related to future labour market earnings... Unsurprisingly, investment in hours has

diminishing returns. There are no long-term incentives for working very long hours, either for men or for women... For women, the most substantial incentives are to work a normal full-time week, rather than part-time; for men, most of whom work at least 40 hours, the future loss from working only 35 hours instead of the average 45 hours is substantial... In general, the marginal incentives are somewhat greater for most women, because of their lower average hours." (Campbell and Green, 2002: 14).

Campbell and Green go on to observe:

"Although this study has detected a significant link between past hours of work and current labour market earnings, the relationship is consistent with more than one theoretical interpretation. One possibility... is that individuals work longer hours in order to signal a higher level of commitment. This may increase their chances of securing promotion or a better job, and moving further along the wage distribution associated with their firm or industrial grouping... It is equally possible, however, that individuals choose to work longer hours in order to enhance their work skills, thereby raising their earnings capacity. The relationship between past hours and current earnings could therefore be consistent with either a human capital or signalling model." (Campbell and Green, 2002: 15-16).

Whatever the causal relationship, the clear implication of this study is that, particularly for women, working part time rather than full time has a long-term cost.

To summarise, those employed on a temporary basis, particularly in casual or seasonal work, and those in part-time work, particularly women, appear to suffer lower wages and poorer access to non-pecuniary



benefits. They tend to earn a lower hourly wage, in a large part due to the types of jobs in which flexible work is concentrated and as a result of low initial human capital endowments. They are also less likely to acquire job-specific human capital.⁵⁴ In addition, female part-time employees are more likely to have experienced episodes of job interruption that can be expected to have consequences for their current and future earnings; this is also the case if they work part time.⁵⁵

The extent to which persons who take on flexible work do so because of barriers to full-time work, insufficient knowledge of the longer-term consequences, or other factors is not known. What evidence there is suggests that, for many people, intermittent or part-time employment is probably involuntary – that is, it is due to the characteristics of the workers themselves, labour demand factors, and the social context within which individuals make decisions about work. A number of British and American studies point to a variety of reasons that contribute to problems of low employment retention and advancements (Kellard *et al*, 2001, Cancian and Meyer, 1998, Rangarajan *et al.*, 1998, Edin and Lein, 1997, Rangarajan, 1996, Slaughter *et al.*, 1982, and Strawn and Martinson, 2000) and it appears likely that a significant minority of workers take on flexible employment because of these reasons.

Summary on advancement

The empirical evidence of advancement is very sparse. What there is suggests that, against a background of increasing wage disparity over the last two decades, there has been little improvement in wages among those employed at the lower end of the labour market. In fact, wage mobility has actually declined in recent years.

This problem is compounded by the type of low-wage entry-level jobs that are available, which can be characterised as junior and low skilled, and often part time or temporary, with fewer prospects for training. The evidence suggests that intervention can help raise the advancement prospects for those who work in such jobs involuntarily, or for those who elect to do such work because of imperfect information about the consequences.

Table 5 summarises the type of policy intervention, this time for advancement, that the different economic theories suggest should be implemented for various categories of workers, together with the policies necessary if there are personal barriers and other work disincentives.

⁵⁴ Green (1999) shows that in-work training opportunities tend to accrue disproportionately to those who already possess higher levels of educational qualifications and human capital. Given the fact that those entering part-time work tend to have fewer qualifications and lower levels of human capital formation, this finding would tend to re-enforce the view that part-time workers enjoy fewer opportunities to add to their skills and develop them. Looking at the LFS for Spring 1997, Green shows that around 12 per cent of part-time workers had received training over a four week period, compared to 15 per cent of those working full-time. Green did not look at the differences between men and women.

⁵⁵ The evidence of the relationship between job interruption and the duration of that interruption on re-entry and future wage rates discussed in the context of retention was based studies that largely focused on males. From theory, however, there is no reason to expect that women will not also suffer a similar wage penalty.

Table 5: Advancement policies for possible target groups

	Human capital	Job matching	Information asymmetry	Personal barriers	Work disincentives
New workforce entrants	Guide workers to jobs that offer opportunities to acquire specific human capital.	Help worker find job that offers opportunities for advancement.	Inform individual of importance of finding a good first job so that a positive signal will be sent to future employers.	Policies in connection with retention will also aid advancement.	Not applicable.
Long-term unemployed	Guide unemployed workers to jobs that offer opportunities to acquire human capital.	Help unemployed worker find job that offers opportunities for advancement.	Not applicable.	Policies in connection with retention will also aid advancement.	Not applicable.
Intermittent or part-year workers	Provide training during non-work periods.	Help workers find jobs they are less likely to leave or be dismissed from.	Inform workers of the long-term costs to them of an irregular work history.	Policies in connection with retention will also aid advancement.	Provide financial incentives to work a greater fraction of the year.
Part-time workers	Provide training while working. Incentives to work full time	Help worker find full-time job if involuntarily employed at part-time job.	Inform worker of limited opportunities for advancement at part-time jobs.	Policies in connection with retention will also aid advancement.	Provide financial incentives to work full time.

Overall conclusions on the evidence on retention and advancement

Most people leaving benefits for work in Britain tend to find employment in low-wage jobs. Some of these individuals have great difficulty in retaining their jobs and advancing to higher wages or better positions; in particular, they appear to be more likely to exit work prematurely compared to those moving job-to-job. The jobs they take are often temporary or part-time and typically poorly paid. Persons who move into the labour market directly from a spell of unemployment also tend to experience lower wages compared to those who start work having moved job-to-job.

Low wages are of less concern if people only face them for a short period and are able to progress up the earnings ladder and out of in-work poverty. The incidence and duration

of unemployment, however, appears to be related to lower future earnings and poorer subsequent employment prospects, and there is evidence that wage mobility for the low-paid has diminished. The evidence also suggests that some experience a low pay/no pay cycle, where they move from one low-paid job into unemployment and re-enter work in another low paid job. Because of relatively high levels of wage immobility, people can remain trapped on in-work state support and in in-work poverty for long periods of time.

The empirical evidence is neither as comprehensive nor as conclusive as would be desired, particularly for advancement. A project such as ERA, therefore, will be important in obtaining more robust evidence on retention and advancement and will prove of great value in informing future policy decisions in this area.

ANNEX 2 – KEY ISSUES IN ASSIGNING INDIVIDUALS AT RANDOM

Approaches to random assignment

In this annex we discuss four alternative approaches to random assigning individuals:

- the simple random draw;
- random assignment based on individual identifiers;
- batch random assignment;
- block random assignment.

Simple random draw

Given our research design, it would appear that the best way to generate a programme and control group would be to assign individuals completely at random – or at least as close to random as possible. There are random number generators built into most programming languages, and these work well in avoiding statistical patterns occurring between programme and control group members. Using a simple random draw can best be thought of as equivalent to drawing names from hat; on average, groups drawn in this way will be equivalent, at least in a statistical sense.

Random assignment using individual identifiers

Another approach would be to employ an algorithm based on an identifier (ID) that is available for everyone in the sample, for example the US Social Security Number, or in Great Britain, the National Insurance Number. Since most identifiers, however, have some kind of pattern associated with

them (being assigned in sequence, for example) they are not truly random, but for most purposes can be treated as such. Some random assignment evaluations have, for example, assigned individuals with an even number in the last digit of their identifier to the programme group and those with an odd number to the control group. Or, if a non-50:50 ratio is needed (for example 2:1), one could take two or three digits of the ID and determine how its value relates to a predetermined threshold. In the case of a 2:1 assignment ratio for example, one could string three digits from the ID together and specify that if that number were less in value than 667 the individual would go into the programme group and if it were above 667, into the control group. Such an approach has one advantage over the random number approach – the managers of random assignment can check that the process is running as planned and that no tampering with the data has occurred.

Where an individual is assigned on the basis of a simple random draw, using a random number generator, it is unlikely that any tampering or error could be detected. On the other hand, if the ID ‘algorithm’ is known to an outsider, it is possible for someone interested in predetermining a result to modify the ID – or discourage (or encourage) individuals with certain ID number combinations from participating in the study. This could have the effect of unbalancing the two samples.



Both of the ‘random’ methods described so far have a weakness that can affect programme operations. A ‘bad draw’ can happen in any random (or quasi-random) process. It is possible, but very unlikely in large samples, that the final sizes of the two groups may be statistically different from what had been intended. The ‘Law of Large Numbers’ helps keep such quirks in check. Localised bad draws in a small subsample, for example in the case of subgroups or at particular experimental sites can, and will, occur. Some programmes are unaffected by such vagaries, while others can be seriously affected. If programme staff expect to assign 50 people in a given week to the programme, but instead are faced with processing 75, staff may experience an overload affecting the services delivered and/or their future cooperation. Similarly, too few people assigned to the programme group can result in wasting staff resources. Moreover, localised bad draws can give people the misguided idea that the process is not random, that there is a bias somewhere. In fact, quite the opposite is true. If a ‘bad draw’ *doesn’t* occur at some point in a lengthy build-up process, that can be seen as evidence that the draw is *not* random.

Strictly speaking, we frequently don’t want a truly random process because ‘bad draws’ can adversely affect the operation of the programme being tested. We prefer more control. We want unpredictability for an outside observer so the outcomes of the assignment process cannot be tampered with or anticipated, but we also need to accommodate the pragmatic concerns of service providers who may require a smooth flow of clients into the programme. We also want to avoid the perception by staff and clients that the process is not ‘random’ – they need to feel comfortable that the assignments are fair. From the point of view of the

evaluation, it is helpful to be able to check how the process is proceeding and know that at any time what is happening can be monitored with relative ease.

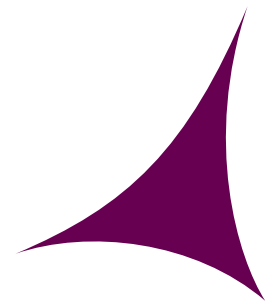
Batch random assignment

There are a number of ways assignments can be made that satisfy these concerns. The safest is to wait until a large number (a ‘batch’) of potential sample members are identified, order them randomly, and then specify that those in the first half of the list, for example, are put into the programme group, and the rest into the control group. It’s even possible – if baseline characteristics are available at the point of random assignment – to match pairs of individuals on the basis of key variables and then randomly assign one individual from each pair to the programme group and the other to the control group. The methodology for pairing individuals can be complicated to develop, but the effort can be worthwhile since the end result is to help guarantee that the two groups are similar in key characteristics.

Block random assignment

Unfortunately, the practicalities of creating research groups often demands that we cannot wait for a number of eligible clients to be assembled before random assignment is carried out. In this case, the most common approach is to carry out random assignment as eligible sample members come through the door. If so, it may be helpful to ensure that the proportion of programme and control group members stays very close to the intended level, even in a small subgroup.

One easy solution is to assign people alternately as individuals present themselves. This method, of course, has the disadvantage of being easily detected and manipulated. Once the first assignment is known, the rest



can be predicted. Moving back a little, one could take pairs of sequential candidates and randomly determine that the first goes to one group and the second to another. In this scenario, an outside observer would only be able to successfully predict (and hence manipulate) half of the assignments. Stepping back a bit farther, one could randomly block four assignments (PPCC, PCPC, CPPC, CPCP, CCPP, PCCP, etc. where 'P' represents an assignment to the programme group and 'C' assignment to the control group) and assign from such blocks in sequence as people come through the door, but here, too, care is required.

First of all, an outside observer could detect the pattern here and would know for certain once the third of four assignments in a block is revealed what the next assignment will be. Moreover, it is not necessary to wait until the fourth assignment is reached to have an advantage. Once the first assignment is revealed, the observer knows that, instead of a 50:50 chance of predicting the next assignment, there is a 67:33 chance of making that prediction successfully. For example, if the first of a block of four is a 'P', then you know you are twice as likely to get a 'C' as a 'P' on the next assignment. In other words, information has been conveyed and can be used to circumvent the required unpredictability of the assignment process.

The logic here continues as you lengthen the size of the block. You always know that an outsider can manipulate the last assignment in a block. Moreover, the outsider can 'gamble' intelligently on most of the other assignments as well. It is possible to construct some degree of control by lengthening the size of the block more and more, but we then also lose our ability to avoid localised bad draws. The implication is clear: by adding information into the process, you add predictability.

However, there is a way to take some information out of the process of randomising within blocks, and that is to randomise the length of the blocks. For example, we could have at random blocks of size four, six, two, four, four and then eight etc. Clearly, any advantage the outside observer might have in predicting future assignments would be greatly reduced. This approach makes it very difficult, expensive and impractical to manipulate the process.

A randomised block sequence can be generated well in advance of the start of random assignment. It can be reviewed for peculiarities, such as too many 'Ps' or 'Cs' in a row. For example, if the largest block used is of size six, it is possible to end up with two adjacent blocks of six where the last three assignments from the first blocks is 'CCC' and the first three assignments from the next block is 'CCC'. So, six 'Cs' in a row (or six 'Ps' in a row) is possible. If that occurs, one can choose to tolerate it; or break it up by trading places with Ps nearby in the sequence; or determine in advance that a block size of six is too dangerous and decide that the largest block size should be four. In addition, in a 50:50 design, one is not restricted to even-numbered block sizes. One could, for example, have two blocks of five: one with three Ps & two Cs and the other with two Ps and three Cs. An additional advantage of odd sized blocks is that they add to the lack of a perceivable pattern by allowing a wider range of block sizes. On the down side, they increase the danger of a localised bad draw, but only very slightly.

One approach to using randomised blocks is to develop a programmed utility that allows the user to specify a range of permissible block lengths. So, for example, a block of four with a 50:50 ratio, the block could be 'PPCC'. The application then builds a

sequence by randomly picking one of the block sizes, scrambling the contents, and adding the result to the end of the sequence. The process ends when the length of the sequence matches or exceeds a specified number.

Generating a sequence in such a way allows it to be 'tested'. A random position in the sequence is picked and the process calculates how far into the sequence one would have to go from that starting point before getting to within a preset tolerance level. Thus, for example, for a 50:50 ratio one might want to get back to within a range of 48%–52%. If such a calculation is repeated over and over, a statistic can be calculated that reflects how many assignments are required to get arbitrarily close to a target ratio. This measure is used to determine the likelihood of getting a bad draw; the more assignments it takes to get back to 50:50, the more likely the sequence is to exhibit a 'bad draw'. In this way a check can be made to ensure that an 'outlier' sequence does not occur.

Once the sequence has been constructed, each location or site can be started at a random position in that sequence. The position is then advanced as each random assignment is made. It is important not to start each site at the same place because knowing the series of random assignment at one site would be informative to another site. When a site's pointer gets to the end of the sequence, the sequence is then started again at the beginning. (Note: stratification complicates this process – see below).

One other consideration that favours the use of sequence files (randomised block or otherwise) is that retaining the pointer numbers, along with the database of all random assignment data, provides a way of auditing the process. Breaks in sequence are

very useful. Moreover, if anything goes wrong with a transaction (as when a computer network access is interrupted in the middle of a transaction) or for some other reason a data record becomes corrupted, the pointers can be quite informative in helping resolve the problem.

Monitoring for bias

The primary task of a random assignment process is to create a programme group and a control group of individuals who, in aggregate, share characteristics in common. So, it is important to build in a process to check that the groups are reasonably similar in baseline characteristics. Such checks are usually carried out after enough of the sample has been accumulated.

Because so many statistical checks are possible, there is always the danger of coming across some patterns that raise concerns of differences between the two groups. Since statistical significance can be reached even for random data in five per cent of all tests, marginally significant results sometime occur, suggesting that the groups might be dissimilar. Such results can be very disturbing for those responsible for the random assignment process, but they do happen occasionally.

One good way of checking whether the samples were created randomly is to look for patterns that, together, tell a similar story. For example, if the programme group is shown to have statistically fewer children than the control group (making it easier for members of that group to enter work), one should look for other characteristics that might be seen as beneficial to the programme group. Are those children also significantly older? Do other characteristics seem biased in favour of the programme group? If there is a consistent pattern of



advantage for one research group over another, then there is reason for concern that something is biasing the assignment process. If the significant differences that are detected seem haphazard, they can be written-off as being the result of running too many tests.

Stratification

Stratification is a method to insure that certain key subgroups are fully represented in the final sample. For example, in the case of the ERA Demonstration, it may be necessary to insure that 'Ps' and 'Cs' are equally distributed among the three target groups at each of the six sites.

Issues to consider in stratifying for each of the approaches to random assignment described above are:

Random individual assignment
or ID-based assignment:

If the random assignment ratios for the stratified groups are the same, nothing needs to be (or can be) done. All groups are assigned randomly and bad draws are always possible. There is no point to stratifying. If the ratios differ, then each individual assignment is made with the probability appropriate for that group, but it is still not possible to avoid a bad draw.

Batch random assignment

Each subgroup is assembled into its own batch. The ratio for each subgroup's batch can be independently set.

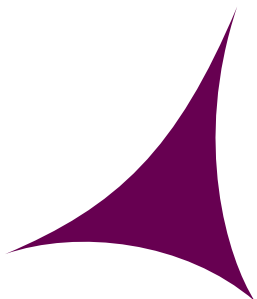
Block random assignment

Each subgroup uses its own randomised block sequence or has an independent pointer for a commonly used sequence. But note, if randomised block designs are being

used to control the local P/C pattern and to avoid even a temporary bad draw, as previously seen, the more stratified the design the less overall control one will have over the local pattern. For example, in a straightforward model where there is a 50:50 ratio with block sizes as high as six, six Cs in a row can occur through two blocks of length six being adjacent to one another in the sequence. In a stratified model – for example, stratifying on long-term and short-term benefit receipt – one will be picking independently from two sequences; hence, it is then possible to get 12 Cs in a row. Complex stratified designs are not really compatible with a block-randomised approach where the goal is to maintain local control of the sequence of 'Ps' and 'Cs'.

Adjusting sample sizes

Planning a random assignment experiment frequently requires certain assumptions about the flow into the random assignment process. For the purposes of discussion, imagine a total sample size of 6,000 is required and that this sample builds up over one year. In other words, there are an average of 500 assignments a month. It is often found that the vagaries of intake process and the capacity of the programme to provide services, makes it necessary to change random assignment ratios (the proportion of the programme group size to control group size). In the example above, 500 assignments a month at a 50:50 ratio means 250 new participants in the programme each month. After a number of months, the service provider may be overwhelmed and ask that the workload be reduced. When this occurs it is not uncommon for evaluators to be asked to change the random assignment ratio to cut down the flow into the programme. Going to a 3:1 ratio (three controls for every programme group member) would mean



cutting down the programme intake load from 250 per month to 125 per month, effectively halving the workload.

Such a change in random assignment ratio is possible once. If it is done any more often than that, the results become intractable to analyse; for example, cohort effects cannot be adjusted for. Making one change in the random assignment ratio means that there are no degrees of freedom for future changes; consequently, such an option has to be employed with care. In addition, if the random assignment ratio is changed, statistical power might decline. All other things being equal, statistical power is maximised by a design in which the proportion of eligible individuals assigned to the programme is 0.5. If some other proportion is used, the easiest way to recapture lost statistical power is to increase the sample size through extending the intake period.

There is one technique that can be used to make life easier when the size of the flow into random assignment is larger than required for measuring programme impacts. In our example, rather than changing the ratio from 1:1 to 3:1, we can simply randomly pick some proportion of the intake not to go into the study at all. We refer to such sub-samples as a 'non-research' group. Although we might monitor their baseline data if we got it, we would not collect their outcome data.

In many cases it is advantageous to build in non-research groups when developing random assignment procedures. If possible, we should plan how many random assignments are needed through the expected intake period and start with an N:P:C ratio that results in just as many Ps and Cs as necessary to reach that target. When the current flow is known, the size of the non-research group can be readily projected

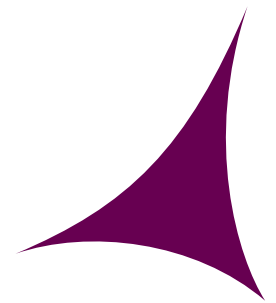
and that number would be incorporated into the design. As long as all of the assumptions hold true, things can be left alone.

In circumstances where the service provider was overloaded, more sample would be drawn-off into the non-research group. If the previous month was unusually slow and build-up is low, then the proportion drawn off into the non-research group can be decreased, thereby raising the number going into the two research groups. None of these sorts of changes decreases the statisticians' ability to deal with the data; no cohort adjustments are needed. Moreover, the one-degree of freedom to make a real change in the random assignment ratio remains.

Monitoring the process

There are many ways that random assignment and its associated sample functions need to be monitored:

1. In randomised block and identifier-based random assignment, it is useful to monitor that the procedures are working as planned. These approaches can be audited. Batch assignment procedures can also be audited, but are so fully under programme control that it may not be necessary to do so.
2. Baseline data quality should be monitored to ensure that its quality and completeness are maintained. Sometimes it is possible to prevent data problems by detecting such issues and re-training staff in collecting and entering baseline data.



3. It is essential that no individual can come back through the assignment process and get a different assignment. Detecting 'repeaters' sometimes requires a pattern-perception algorithm since identifiers may not be reported the second time exactly as they were reported the first time. Any algorithm employed in this way needs to be checked on occasion. When designing the process for detecting repeaters, evaluators would be well served to avoid very demanding algorithms. When an algorithm requires too much similarity in the identifiers, the likelihood of identifying duplicates is reduced; hence, when a true duplicate comes along, a rigid algorithm might decide that the duplicate is not a repeater and subsequently the record would be randomly assigned. If two different assignments are made for the 'same' person, the repeater's impact data has to be included with both the programme and control group's outcome measures, reducing the measurable impact. The cost of creating a more relaxed algorithm comes when a non-repeater is identified as a repeater. The record for such a case is not randomly assigned and so the individual's outcome measures are left out of the analysis entirely. A potential sample member has been lost, but, for the most part losing a sample member is much less costly than having an individual in both programme and control conditions.
4. Another arena where matching is an issue is when an algorithm has to be used to link random assignment and baseline records, since they come in separately and may have some discrepancies in identifiers. These algorithms are usually more straightforward than those used in detecting repeaters, but their performance should be monitored. For example, if the algorithm is too rigid, some BIF records

will end up unattached to any random assignment records (i.e. 'orphan' BIFs) and some random assignment records will be missing a link to a required BIF record. A manual review of these cases is sometimes helpful.

The ERA Demonstration

The previous discussion raises a number of issues for the design of a random assignment procedure for the ERA Demonstration programme. As indicated in the main text, the ERA programme will be tested in six experimental sites and will be targeted at three groups in each site. Ideally, we would want our sample to be balanced, in terms of numbers assigned to programme and control groups, across target groups and by site. It is also important to bear in mind that random assignment will take place centrally and be under the control of the ERA Database Controller.

Several of the approaches to random assignment that are described above are probably inappropriate for use in the ERA Demonstration evaluation. For example, using a simple random draw to assign individuals to programme and control groups in this instance would leave open the possibility of 'bad draws'. In other words, in a given target group, at a particular site, the balance between the numbers in the control and programme group might be quite different from the 50:50 target. An alternative, involving random assignment on the basis of an individual identifier, for example NINO, makes manipulation of the assignment possible. For example, those with a particular number combination who are known to be assigned to the control group, might be dissuaded from joining the programme because they or an administrator know in advance that they have zero chance of being assigned to the programme group.

Given the need to inform eligible individuals quickly of their random assignment status, a random assignment procedure constructed to assign individuals on a batch basis would also appear to be ruled out.

This leaves the option of conducting random assignment using the block approach. Because we wish to effectively stratify the random assignment, however, the question is whether there should be a single sequence of blocks for each site or one sequence for each target group at each site? The problem with having a sequence for each target group (or stratum) at each site is that this would in effect increase the probability of ‘bad draws’. It is therefore suggested that a single sequence of blocks, of a random length, be set up for all sites but with a different starting point for each site determined at random. Given the ‘Law of Large Numbers’, at the end of the 12-month intake period, it is highly likely that the numbers assigned to programme and control groups within in each target group at each site will be broadly in balance, although an exact 50:50 ratio is unlikely to be achieved.

One issue remains – how do we allow the programme administrators to adjust the flow of individuals into the programme group if it is greater than anticipated? If the projected inflows (see Chapter 2) prove to be under-estimates for any experimental site, then all that will happen is that individuals will enter a queue to wait to see their Advancement Support Adviser (ASA). However, if the length of time between being assigned and seeing an ASA become very large, because of volume of individuals assigned to the programme group is larger than expected, this could cause individuals to drop out or lose interest in the programme.⁵⁶ For this reason, the potential for establishing

a non-research group, as explained above, should be built into the random assignment process. The non-research group can then be activated for an individual experimental site, if the inflow of clients proves unmanageable. Moreover, with a non-research group in place, it is also possible to adjust to conditions where the service provider might prefer an increase in flow. The proportion of eligible clients assigned to the non-research group could simply be reduced. Such an approach may have implications, however, for the length of time the intake process runs for at a specific site, in order that sample power requirements for estimating impacts be met.

⁵⁶ This is obviously of greater concern where participation is voluntary.

ANNEX 3 – MEASURING NON-EXPERIMENTAL IMPACTS

The aim of this Annex is to examine the earnings progression experienced by those individuals who receive the ERA programme and to evaluate the impact of the ERA programme on earnings progression. The WTC and New Deal samples pose rather different issues for this analysis primarily because in the WTC case the randomisation occurs for individuals already in employment, whereas in the New Deal case individuals will be looking for employment when they are assigned.

There are some distinct advantages for measuring earnings growth using the WTC sample relative to the New Deal case. In the New Deal case, the individual necessarily has to find employment before earnings can be observed and the comparison of those in employment between the programme group and control group at any point after they enter the programme is no longer a randomised comparison. This is not the case for those in the WTC group who are observed in work at the beginning of the programme.

To examine earnings progression, one needs at least two observations on an individual's earnings. One early on in the programme and one after the main effects of the programme can be expected to have taken place. In the case of the New Deal sample, the employment rate at any point in time is likely to differ systematically between the programme group and the control group. Consequently, the measurement of earnings

growth and of the impact of the ERA programme on earnings growth is quite complex. In the case of the WTC sample, which is discussed first, systematic differences only occur after the programme has been in place for some time. That still means that non-experimental methods are required; but, because the comparison is experimental at the employment baseline, this provides a simpler and natural starting point for the discussion.⁵⁷

In the WTC-ERA design, the randomisation occurs once individuals are already in employment and receiving WTC. For comparisons of employment durations, for example, this provides a valid experimental comparison. That is, a comparison of mean employment durations between the programme and the control group will provide an unbiased estimate of the programme impact. However, since earnings growth requires a measurement at the beginning period, say period 1, and at a later period, say period 2, a problem still remains. Those who receive treatment are likely to have a higher probability of employment in period 2. Consequently, the comparison of observed earnings between programmes and controls in the second period will no longer be a randomised comparison.

Using the subset of individuals who remain in work may suffer from selection bias if, as might reasonably be expected, those with lower levels of wage growth tend to leave employment. If individuals in the control and

⁵⁷ This Annex draws upon a small and related literature (see Ham and Lalonde (1996) and Card, Michalopoulos and Robins (2001), for example).

programme groups fall out of employment with the same propensity, then comparing their wage growth experimentally would give an unbiased estimate of the impact of the ERA. However, the chances are that this propensity will not be the same. Indeed, if the ERA is successful, it should retain more individuals in employment. The characteristics, observed and unobserved, of those in employment in the programme and control groups will be different.

So what can be done? There are essentially two general approaches that can be followed. Both of which adapt the programme and control comparison using non-observational methods.⁵⁸ To begin with, imagine that there are just two points of measurement: $t=1$ and $t=2$. The first point of observation is when the programme begins and the second is some follow-up period.⁵⁹ Suppose also that there are a set of characteristics 'x' that are measured at time '1' for each individual in the programme and the control group. To correct for the selection bias one can either assume selection on the observables, 'x', and use a *matching method*⁶⁰ to adjust the relative growth rate in wages between controls and programmes or use a *selection bias correction approach*⁶¹ and assume a subset of the observed variables drive selection but not wage growth directly.

To consider the problem and the alternative solutions, suppose wage growth can be written as the sum of one term that depends on observable characteristics 'x' and another that is unobservable. For the control group, wage growth between periods 1 and 2 can be written as:

$$\Delta w_i = a_c(x_i) + \varepsilon_{ci} \forall i \in C \text{ with } E[\varepsilon_{ci} | x_i, C] = 0 \quad [1]$$

and for the programme group as:

$$\Delta w_i = a_T(x_i) + \varepsilon_{Ti} \forall i \in T \text{ with } E[\varepsilon_{Ti} | x_i, T] = 0 \quad [2]$$

Note that this allows the unobservables to differ between control and programme for a given x type of individual. This is because, even though there is randomisation at time 1, by time period 2 the programme and control groups will have experienced different impacts on the wage growth.

There are two potentially important parameters of interest. The first is the differential wage growth between those who received the ERA programme and those who did not. This is given by:

$$b(x_i) = a_T(x_i) - a_c(x_i) \quad [3]$$

Note that this is defined for each 'x' type of individual and measures the net impact of the ERA. The average effect over all those in the programme group is given by:

$$E_{x \sim T}[a_c(x_i) - a_T(x_i)] = \int a_c(x_i) - a_T(x_i) dF_x^T \quad [4]$$

where F_x^T is the distribution of the x covariates among the treated.

Even though (4) looks complex it is easy to compute once (3) is known by taking the simple average of (3) over all individuals in the programme group. The second parameter of interest is the wage growth for those in the programme group. This is given simply by:

$$a_T(x_i) \quad [5]$$

or its corresponding average over the programme group.

These parameters of interest are simple to define but less easy to estimate from the observable data on wage growth. To be more precise let $E_{2i}=1$ occur when individual i is in employment and therefore has an

⁵⁸ Blundell and Costa-Dias (2000) review non-experimental approaches to evaluation.

⁵⁹ In practice, there may be many points of observation after the intervention occurs, but the same arguments will hold.

⁶⁰ See Heckman, Ichimura and Todd (1997).

⁶¹ See Heckman (1979) and Heckman and Robb (1985).



observed wage in period 2. In this case observed wage growth for the control and programme group is given by:

$$\Delta w_i = a_c(x_i) + E\varepsilon_{ci}|E_{2i}=1, x_i] \quad [6]$$

and for the programme group by:

$$\Delta w_i = a_T(x_i) + E\varepsilon_{Ti}|E_{2i}=1, x_i] \quad [7]$$

The second term on the right hand side of each of these wage growth expressions is the 'selection bias' that occurs from only observing wages for those in employment in the second period.

In principle there are three alternative approaches that can be taken with regard to estimation. The first case is to assume *balancing biases*, the second is to assume *selection on observables* and use a *matching estimator*, and the third is a *selection correction or control function* estimator. The balancing biases approach assumes that:

$$E\varepsilon_{Ti}|E_{2i}=1, x_i] = E\varepsilon_{ci}|E_{2i}=1, x_i] \quad [8]$$

in which case the selection bias terms in (6) and (7) cancel out when estimating the first of the parameters of interest. Note that this does not help in the estimation of the second parameter of interest. However, assumption (8) is potentially very strong. One would expect the type of events that happen to individuals who receive the ERA programme would be quite different to those who do not receive the programme.

Another way in which the biases can be eliminated is by using a matching estimator. Under the matching assumption, conditioning on the observable 'x' variables is sufficient to eliminate all the selection bias. Consider the expressions for the wage growth (3) or (4).

Matching assumes that the set of conditioning variables x is sufficient to induce the conditional independence assumption:

$$\varepsilon \perp E_2/x \quad [9]$$

that is the unobservables 'ε' are distributed independently of second period employment conditional on the observables 'x'. In this case provided the 'x' variables are such that (9) is satisfied, the selection bias terms disappear. This is a case of selection on the observables. A slightly weaker condition can be used for the first parameter of interest because all that is required is that:

$$\varepsilon_T - \varepsilon_c \perp E_2/x \quad [10]$$

Formally, this is precisely when the biases resulting from the unobservables balance.

If it is not believed the 'x' variables are such that (9) or (10) are satisfied, then use of the control function estimator should be considered. Suppose there is a set of variables, 'z', that determine employment in period 2, but they do not directly determine wage growth between periods 1 and 2. In this case, one can write:

$$\Delta w_i = a_c(x_i) + S_c(x_i, z_i) \quad [11]$$

where S is the selection bias correction or control function. Similarly for the programme group one can write:

$$\Delta w_i = a_T(x_i) + S_T(x_i, z_i) \quad [12]$$

Provided the selection terms 'S' can be manipulated independently of 'x' through movements in 'z' then all the parameters of interest can be estimated.

Under joint normality this becomes the standard selection bias estimator of Heckman (1979). For example,

$$\Delta w_i = a_c(x_i) + \rho_c \lambda(x_i \pi_x + z_i \pi_z) \quad [13]$$

where ' ρ_c ' relates to the correlation between the wage growth error term and the error term in the probit equation for employment in the second period. In this case ' $\lambda(x_i \pi_x + z_i \pi_z)$ ' is the corresponding employment hazard in period 2.

Estimation of this control function or selectivity model requires a set of 'z' variables that impact on employment in the second period but not directly on wage growth. In the WTC case these could relate to the parameters of the WTC programme, for example, as they differentially affect each individual.

Finally, consider the evaluation of earnings progression in the New Deal group. For this group the analysis is further complicated by the fact that the initial period earnings observation will also suffer from non-random selection. In principle the same approaches for recovering the parameters of interest can be followed. However, now there are two selection issues. Either it must be assumed that the 'x' variables are sufficient to ensure no remaining selectivity bias. Or sufficient omitted 'z' variables must be found to construct a full set of selection correction terms.

ANNEX 4 – DESCRIPTION OF ERA DEMONSTRATION TECHNICAL ADVISER ROLE

The role of the technical adviser is described in this Annex. It also discusses options for the delivery of this key role. It is vital that the technical advice role is in place at an early stage in the implementation phase, so that proposed changes to the policy or evaluation design can be monitored and, if necessary, challenged. Ensuring that robust evaluation results are obtained from this Demonstration Project presents particular implementation challenges. The period of time in which the new services will be provided (a maximum of 33 months) means that the Demonstration will run from October 2003 to at least July 2007 (depending on take-up). It is important that members of the control group are not offered new services throughout this demonstration. Previous British experience, particularly with random assignment based pilots, has been mixed. The lessons from this and US experience highlight the need to ensure that:

- policy is implemented as designed, and consistently over time;
- random assignment processes are adhered to; and
- accurate and timely data are collected throughout the project.

In US Demonstration Projects, 'technical advisers' make a critical contribution to the successful delivery of such projects. These advisers have a number of responsibilities:

- working with front-line staff and their managers to help ensure that the new services are implemented as designed;

- liaising with the evaluation team to ensure that the evaluation is proceeding to plan e.g. that sample numbers are building up as needed;
- ensuring that control group customers do not receive programme group services;
- explaining the rationale for random assignment and rigorous evaluation to local staff and managers;
- acting as a link amongst sites so that consistency among them is maintained.

Although elements of this role are often carried out in British pilots, others are not. Clearly there are benefits in amalgamating these responsibilities so that one person is up to date with operational issues and difficulties and is able to ensure that any corrective action does not subvert the evaluation. The technical adviser must be independent of local management to enable them to avoid being subject to the operational pressures that may lead to:

- diverting ERA resources (e.g. using ASAs for New Deal PA work);
- changes in the processes to accommodate non-project aims;
- 'easements' (dropping elements of the process for operational reasons); or,
- *in extremis*, subversion of the programme (for example, staff influencing the composition of the group put forward for random assignment).

The technical adviser will look for instances in which things are not going according to plan because of unanticipated problems, constraints or events; they will be able to help the site managers make the appropriate adjustments.

Technical advisers will perform four main functions:

- Validation of the ERA processes – ensuring each phase of customer interaction (including that with control group members) conforms to the programme design.
- Monitoring outputs of the process via the collected data.
- Alerting local management, the Steering Group and project team, and the evaluation contractor to risks emerging in sites.
- Managing and resolving (in collaboration with local managers and the project team) conflict arising from competing priorities.

Validation

Validating the ERA processes will be an on going task undertaken in collaboration with the local site manager. It will involve accompanying ERA team members at various stages of customer interaction; talking to ERA team members and managers about their understanding of the processes; on occasions, interviewing programme and control group customers; and observing the processes at first hand (such ensuring full completion of the Basic Information Form or that the Intake Clerk and ERA Database controller perform the random assignment process correctly).

Monitoring

Technical advisers will regularly monitor all data sources and analyse trends to ensure that the programme remains on track. Advice from practitioners in the US is that at each site targets be agreed upon for those outputs that matter most to the programme – customer contact rates, participation rates, etc. These can be negotiated with local management at each site. Included in these may be targets that trigger pre-emptive action if sufficient customers are not entering the programme or ultimately allow an exit strategy if the site cannot generate an adequate sample size or the programme does not engage enough Customers. The technical adviser could monitor these targets in collaboration with local managers and agree to any necessary corrective action.

The monitoring conducted by the technical adviser may include:

- bi-weekly personal (preferably) or phone contact with staff during implementation design phase;
- consultation about ERA staff job specifications and anticipated programme outputs – an exercise to ensure all parties agree on what roles ERA staff should perform and what the programme measures. This, in turn, helps define the differences in approach from current PA roles;
- random assignment training of the Intake Clerk and recruitment officer who will be responsible for the baseline data collection and initiating random assignment;
- early site visits to speak with front line programme staff and managers and to observe customer/staff interaction;



- regular site visits and phone check-ins after early operations site visits to monitor sample build up, baseline data quality, customer contact and engagement rates, and programme activities;
- regular validation of processes (see above);
- six-month formal programme assessment.

At the six-month formal programme assessment, the technical adviser, local managers and the ERA implementation team will review target achievement and to agree any modifications. They will also review whether the ERA processes (including intake, random assignment, pre-work and in-work activity, data collection, etc.) are working as they should and decide on any remedial action they may be needed.

Highlighting emerging risks

During day-to-day operations, the technical adviser should liaise with the demonstration project local manager, who is responsible for the agreed outcomes, to resolve any issues as they arise; this eliminates the confusion over who to contact when something happens. Good communication is critical, and it is vital to have a person on both sides who is responsible for keeping track of the flow of information and requests. The technical adviser will be the first point of contact with the evaluation contractor.

It is envisaged that there will be one technical adviser for each ERA regional site. ERA technical advisers will meet regularly with evaluators to discuss progress, lessons, best practices, etc. Local managers will need to be clear about the nature of the role at each site to ensure that all parties understand that the technical adviser is accountable to the project, rather than to local management. Any proposed changes to the ERA process at sites will have to be formally analysed by the

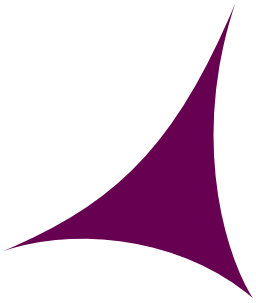
technical advisers and evaluators. For changes with potentially significant impacts, the results of the analysis, with recommendations, will be put to the Steering Group for acceptance or rejection. In effect, the technical advisers and evaluators will perform the function of a central design authority that reports to the Steering Group.

Managing and resolving conflicting priorities

Technical advisers need to be in place early in the implementation planning stage. They need to be able to advise the implementation team on changes to the processes suggested prior to go-live. Technical advisers can be key in the 'dry-run' or pilot phase, ensuring the integrity of the programme design from the outset. The dry run tests the activities that are planned and demonstrates how the collection of baseline information and random assignment affect the process. In the US, the most effective dry runs have tried to mirror the study as much as possible to uncover issues that are not apparent on paper.

The role of the Evaluation Co-ordinator

A beneficial site-specific role that has emerged from the US experience is that of an evaluation co-ordinator who is located at each site, wherever possible. The evaluation co-ordinators would be members of the local Jobcentre Plus ERA team. They would be responsible for the co-ordination and ensuring timely collection and despatch of management and evaluation data. An important aspect of this role is that it forms the first point of contact for the technical adviser – and the focal point for clarifying problems or rectifying bottlenecks. The task would not be full time and could be combined with other team roles (e.g. the Intake Clerk or recruitment officer).



The Government Chief Social Researcher's Office

Sue Duncan is the Government Chief Social Researcher. The Government Chief Social Researcher's Office (GCSRO) is based in the Prime Minister's Strategy Unit. It provides strategic leadership to the Government Social Research Service and supports it in delivering an effective service. It has a broad role in promoting the use of evidence in strategy, policy and delivery and leads on strategic social research issues and standards for social research in government. It represents GSR and its work within government and in the wider research community. It also provides practical support and advice to departments on the organisation and delivery of the research function and on recruitment, career development and training.

A web version of the research can be found on Policy Hub (<http://www.policyhub.gov.uk>). Policy Hub is a web resource launched in March 2002 that aims to improve the way public policy is shaped and delivered. It provides many examples of initiatives, projects, tools and case studies that support better policy making and delivery and provides extensive guidance on the role of research and evidence in the evaluation of policy.

Strategy Unit, Admiralty Arch, The Mall, London SW1A 2WH

Tel: 020 7276 1881

Email: strategy@cabinet-office.x.gsi.gov.uk

Website: www.strategy.gov.uk

ISBN No. 0 7115 0445 7

The text of this document may be reproduced free of charge in any format or media without requiring specific permission. This is subject to the material not being used in a derogatory or in a misleading context. The source of this material must be acknowledged as Crown copyright and the title of the document must be included when being reproduced as part of another publication or service.

© Crown copyright 2004

This report is printed on recycled paper produced from at least 75% de-inked post consumer waste, and is totally chlorine free.