A Multi-Level Analysis of the Impacts of Services Provided By the UK Employment Retention and Advancement Demonstration

Dorsett, R. and Robins, P.K.

A Multi-Level Analysis of the Impacts of Services Provided By the UK Employment Retention and Advancement Demonstration*

by

Richard Dorsett
National Institute of Economic and Social Research
r.dorsett@niesr.ac.uk

Philip K. Robins
University of Miami
probins@miami.edu

October 2012

Revised March 2013

Revised October 2013

Abstract

The United Kingdom Employment Retention and Advancement (UK ERA) demonstration was the largest and most comprehensive social experiment ever conducted in the UK. It examined the extent to which a combination of post-employment advisory support and financial incentives could help lone parents on welfare to find sustained employment with prospects for advancement. ERA was experimentally tested across more than 50 public employment service offices and, within each office, individuals were randomly assigned to either a program (or treatment) group (eligible for ERA) or a control group (not eligible). This paper presents the results of a multi-level non-experimental analysis that examines the variation in office-level impacts and attempts to understand what services provided in the offices tend to be associated with impacts. The analysis suggests that impacts were greater in offices that emphasized in-work advancement, support while working and financial bonuses for sustained employment, and also in those offices that assigned more caseworkers to ERA participants. Offices that encouraged further education had smaller employment impacts. The methodology also allows the identification of which services are associated with employment and welfare receipt of control families receiving benefits under the traditional New Deal for Lone Parent (NDLP) program.

A Multi-Level Analysis of the Impcts of the UK Employment Retention and Advancement Demonstration

By Richard Dorsett and Philip K. Robins

1. Introduction

The United Kingdom Employment Retention and Advancement (UK ERA) demonstration was the largest and most comprehensive social experiment ever conducted in the United Kingdom. It tested the effectiveness of an innovative method of improving the labor market prospects of low-wage workers and long-term unemployed people. UK ERA took place from October 2003 to October 2007 and offered a distinctive set of 'post-employment' job coaching and financial incentives in addition to the job placement services routinely provided by the UK public employment service (called Jobcentre Plus). This in-work support included up to two years of advice and assistance from a specially-trained Advancement Support Adviser (ASA) to help them remain and advance in work. Those who consistently worked full time could receive substantial cash rewards, called "retention bonuses." Participants could also receive help with tuition costs and cash rewards for completing training courses while employed.

The UK ERA demonstration differed from an extensive set of previous social experiments for low-income families that focused primarily on "pre-employment" (or "work-first") services (see Greenberg and Robins, 2011 and Friedlander, Greenberg, and Robins, 1997 for a summary).[1] Most of these earlier experiments produced modest impacts and it was felt by policymakers and program evaluators that combining pre- and post-employment services and including financial incentives might strengthen the impacts of such programs. UK ERA targeted three groups of disadvantaged people: out of work lone parents receiving welfare benefits (called Income Support in the UK), low-paid lone parents working part-time

---

[1] One exception is an employment retention and advancement demonstration conducted in the US from 2000 to 2003 (see Hendra, et al., 2010). The US ERA was similar in many respects to the UK ERA and served as a prototype for the UK ERA.

and receiving tax subsidies through the Working Tax Credit (WTC), and long-term

unemployed people receiving unemployment insurance (called Jobseeker's Allowance in the

UK). The UK ERA demonstration utilized a random assignment research design, assuring

unbiased estimates of the program's impacts.

The formal evaluation of UK ERA (Hendra et al., 2011) covered five years of

program impacts. Administrative records were used to document impacts on several

outcomes (mainly employment, earnings and benefit receipt) during the five years subsequent

to random assignment. For two of the three target groups (out of work lone parents and WTC

recipients), the impacts were generally quite modest and not statistically significant for most

of the evaluation period.[2] For the other target group (long-term unemployment insurance

recipients), the impacts were statistically significant and sizeable, and persisted into the post-

program period.

Within the six districts in which UK ERA[3] took place, there are more than 50 local

offices. The purpose of this paper is to try to exploit variation in program practices across

these offices in order to determine whether certain features of the local programs' operations

are systematically related to program impacts. Previous studies have shown that program

impacts vary with operational procedures and types of services provided (Bloom, Hill, and

Riccio, 2005, Greenberg and Robins, 2011). Thus, building on these previous studies, we

attempt to get inside the "black box" of ERA implementation practices to see which elements

of the "total package" tended to be associated with stronger impacts on employment and

welfare receipt.[4]

---

[2] The US ERA targeted lone parents and like the UK ERA had generally modest impacts that were mostly not statistically significant. Of the 12 programs formally evaluated in the US ERA, only three produced statistically significant impacts (see Hendra et al., 2010).

[3] Henceforth, we refer to the UK ERA as simply ERA.

[4] For example, some previous studies (such as Hamilton, 2002) have found that programs emphasizing immediate job placement (e.g., job search assistance) generate larger impacts on employment than programs emphasizing human capital development (e.g., placement in education and training). In fact, some studies have found that human capital development programs can lead to short-run reductions in employment. However, a reanalysis of the California GAIN program by Hotz et al. (2006) found that over time the human capital approach can actually generate impacts exceeding those of the work-first approach.

The analysis uses a multi-level statistical model based on the methodology developed by Bryk and Raudenbush (2001) and first applied to the evaluation of social experiments by Bloom et al. (2005). We use both individuals and institutions as the units of analysis, an approach quite appropriate for examining variation in program impacts across offices. Other studies using a somewhat different methodology to exploit variation in office practices to estimate social program impacts include Dehejia (2003) and Galdo (2008).

Implementation practices were not randomized across offices and thus may have been related to client or office characteristics. Because of this, the analysis presented here is non-experimental. We discuss later the assumptions required for the results to be given a causal interpretation and the reader should keep in mind that causal inferences are only valid if these assumptions are satisfied.

The analysis focuses on out of work lone parents receiving welfare.[5] This group is of particular interest because over much of the five-year follow-up period no statistically significant average impacts were detected on most of the outcomes studied (Hendra et al., 2011). Hence, if we are able to identify program features that are associated with inter-office variation in the impacts for this target group we will have added to the knowledge derived from the evaluation of the ERA program.

The remainder of this paper proceeds as follows. In section 2, we describe the ERA demonstration and what it was intended to accomplish. In section 3, we present the hypotheses to be tested in examining cross-office variation in ERA impacts. In section 4, we present the statistical model used to test these hypotheses. In section 5, we discuss the data used to estimate the statistical model. In section 6, we report our estimation results for welfare and emlpoyment outcomes. Results for earnings are provided in section 7. Finally, in section 8, we present our conclusions and policy recommendations.

2. The Policy Setting

---

[5] Insufficient data prevented us from considering the two other groups eligible for ERA.

The ERA demonstration builds on the *New Deal for Lone Parents* (NDLP) policy initiative introduced in the UK in 1998. NDLP's aim was to "*encourage lone parents to improve their prospects and living standards by taking up and increasing paid work, and to improve their job readiness to increase their employment opportunities*" (Department for Work and Pensions, 2002). NDLP participants were assigned a Personal Adviser (PA) through the public employment service office to provide pre-employment job coaching services. PAs could also offer job search assistance and address any barriers participants might have had that challenged their search for work. They also had access to an Adviser Discretion Fund (ADF) that provided money to help participants find employment. Finally, they advised participants on their likely in-work income at differing hours of work and helped them access education or training. NDLP participation was entirely voluntary.

The ERA demonstration project offered services beyond those available under NDLP, mainly in the form of in-work services and financial support. As noted above, these additional services included in-work advice and guidance plus a series of in-work retention bonuses to encourage sustained employment. Support for training was also available; ERA covered tuition costs and offered financial incentives for those in work to train. It also provided an in-work Emergency Discretion Fund (EDF) designed to cover small financial emergencies that otherwise could threaten the individual's continued employment.[6] Importantly, ERA services and financial assistance were available for only thirty-three months.

In order to evaluate the impacts of the multi-dimensional ERA program, a random assignment research design was utilized. NDLP participants who agreed to be included in the experiment were randomly assigned either to a program (or treatment) group that was eligible for the full range of ERA services and financial assistance or to a control group that could

---

[6] The retention bonus consists of up to six payments of ₤400 for each 17-week period during which individuals work 30 or more hours per week for at least 13 weeks. The training bonus consisted of tuition payments up to ₤1,000 for in-work training plus ₤8 for every hour of training completed up to ₤1,000.

only receive standard NDLP services.[7] The randomization process was closely monitored and controlled. The fact that there were no systematic differences between the two groups prior to random assignment (results available from the authors on request) provides some reassurance that the randomization was carried out effectively.

3. Factors Influencing Variation in ERA's Impacts

The simplest measure of the *impact* of ERA is the difference in mean outcomes between the program and control groups over the follow-up period (five years in this paper).[8] The two outcomes examined in this paper are months receiving welfare and months employed. The impact of ERA on months receiving welfare, for example, is the *difference* over the follow-up period between the program and control groups in the average number of months receiving welfare. The follow-up period for ERA is five years, roughly three of which are while the program was operating and two are after the program ended. The results presented later distinguish between these in-program and post-program periods.

ERA impacts can vary over time, across persons, and across geographic areas. Varying impacts over time may have multiple causes including changes in the amount and types of ERA services provided by program administrators, changes in the amount and type of services being provided to the control group under the traditional NDLP program, changes in environmental conditions and changes in the reaction time of participants to the new services being provided. Although we are able to estimate how ERA impacts vary over time, we do not have sufficient data to allow us to identify the precise causes of these varying impacts over time.

---

[7] Goodman and Sianesi (2007) show that 70% of those eligible participated in ERA. Most nonparticipation (86% of cases) was due to (wrongly) not being offered the opportunity to participate. This varied considerably across offices. Participation was higher in areas of higher unemployment. Those already employed at the time of randomization were less likely to participate yet those with substantial prior employment experience were more likely to participate. In the first year after randomization, nonparticipants spent more time in work and less on welfare than participants. It appears, therefore, that offices' tendency to selectively offer the opportunity to participate resulted in the participant sample being made up of individuals with slightly less favorable labor market characteristics than the full eligible population.

[8] In practice, this impact is regression-adjusted so that the influences of variations in individual characteristics at the time of random assignment are taken into account in deriving the impact. This makes the estimated impact more statistically precise.

Varying impacts across persons (sometimes called "subgroup impacts") can arise because certain types of individuals may be more susceptible to program services. For example, those with longer welfare histories or lower levels of education may have been harder to employ and less likely to have been able to use the ERA services effectively than persons with shorter welfare histories or higher levels of education. On the other hand, those with older children may have been more willing to utilize the ERA services than persons with younger children. As will be indicated below, our empirical model allows us to identify subgroup impacts.

Varying impacts across geographic areas may be due to different environmental factors and to different ways ERA was implemented across the various local welfare offices. There are a number of environmental factors that could influence the impact of ERA. For example, persons living in areas with higher unemployment or, generally, in more deprived areas may have found it harder to have made effective use of program services. Our empirical model is specified to allow for the impact to vary with a measure of local area deprivation.

Cross-office variation in impacts can arise due to differences in the overall structure of the individual offices and differences in program implementation practices for both ERA and control group participants. For example, offices with higher caseloads may have been less successful in providing meaningful help to ERA participants, thereby rendering ERA less effective. Or, offices that placed more emphasis on immediate job placement may have had larger impacts than offices that emphasized human capital development. Or, offices that were already providing a rich array of services for control group families may have had smaller impacts than offices that were not. Bloom, Hill and Riccio (2005) find that impacts in several US based welfare-to-work demonstrations vary significantly with differences in program implementation practices across local welfare offices.

Office variation in impacts according to the way ERA was implemented is the major focus of this paper, although we also examine how impacts vary over time, with individual characteristics, and with environmental characteristics. Introducing office-level variation in

impacts requires a more sophisticated statistical framework than is traditionally used in evaluation research. Specifically, the units of analysis are both the individual and the office and the statistical framework must take this nesting into account. As will be described in greater detail below, multi-level modeling provides a natural framework for analyzing variation in impacts across offices and across individuals within offices. Although the ERA demonstration took place across 58 offices, in practice operations among some of these offices were shared.[9] Where this applies, we have combined the offices, resulting in 37 distinct units of delivery which, for convenience, we continue to refer to as "offices" in the remainder of this paper.[10]

Before proceeding with the specification and estimation of a multi-level statistical model, a fundamental question must be answered. Namely, is there enough variation in the impacts of ERA across offices so that implementation differences can possibly be explained by office-level characteristics? To determine this, we used a multilevel Poisson regression model with program group status as the only regressor in order to construct empirical Bayes estimates of the extent to which program effects on months receiving welfare and months employed varied across the 37 offices in our sample.[11] We estimated separate models for the in-program period (1 to 3 years post randomization) and the post-program period (4 to 5 years post randomization). We conducted formal statistical tests to determine whether the individual office-level impacts were significantly different from the average impact estimated over all offices.

Figures 1A and 1B present the empirical Bayes estimates of office-level effects. Since these are generated by a multilevel Poisson model, they are reported as incidence rate ratios (IRRs). In other words, they are *proportionate* impacts such that a value of 1 indicates

---

[9] For further details, see Dorsett and Robins (2011).

[10] We also performed some analyses using the full 58 office sample, but the results were not as informative as the analyses performed on the combined offices sample. We are grateful to Debra Hevenstone for developing the methodology to combine the 58 offices into the 37 distinct offices.

[11] We discuss the multilevel Poisson model in detail in section 4.

no effect (it implies an increase by a factor of 1). Similarly, an effect of 0.5 implies a reduction of 50 per cent and a factor of 1.5 indicates an increase of 50 per cent.

The welfare impacts (Figure 1A) range from 0.59 to 1.17 for the in-program period and from 0.55 to 4.63 for the post-program period. Although not visible from the chart, these very large impacts for the post-program period correspond to the smallest offices. The overall impact is shown by a vertical line in the figure. The employment impacts are given in Figure 1B. These range from 0.70 to 2.49 for the in-program period and from 0.72 to 1.69 for the post-program period.

For purposes of this paper, the important question is whether the variation across offices in the estimated impacts is statistically significant. We tested this using likelihood ratio tests, comparing our results with restricted results where the impact was not allowed to vary across offices. For both outcomes, the restriction was strongly rejected.[12] Therefore, we conclude that there is sufficient variation in the impacts across offices to warrant a further, more sophisticated, analysis to determine whether part of the variation can be explained by office characteristics.

4. Methodological Framework for Explaining Cross-Office Variation

4.1 Estimation Approach

Our fundamental approach for examining variation in impacts across offices is based on a simple production function framework in which the implementation (or production) of ERA services within a particular office was related to a set of individual, environmental and office factors (or inputs). These factors are based on ERA participant needs and experiences as well as the manner in which ASAs provided the ERA services.

In examining variation in ERA impacts across offices, we focus on ERA services that are consistent with the primary objectives of the demonstration, namely retention and advancement services. Two basic hypotheses will be tested (the specific variables related to each of these hypotheses are described in detail below). First, we hypothesize that the strength (or effectiveness) of ERA's impacts (as opposed to the direction of impacts) will be

---

[12] The test results are available on request from the authors.

systematically related to the *intensity* of ERA services (reflected, perhaps, by the amount of time advisers spend with each ERA participant). Second, we hypothesize that the strength of ERA's impacts will be related to the *types* of ERA services provided (such as help with advancement or help with finding education and training opportunities). Both of these hypotheses are relevant for policy makers. For example, if it is the intensity of services that matters, then hiring additional caseworkers may represent an effective use of public funds. Or, if it is found that particular types of services are associated with greater impacts, then program operators who are not currently emphasizing such services might find it worthwhile to redirect their program delivery activities towards favoring such services.

For both the above hypotheses, the direction of impacts (as opposed to the strength or effectiveness) will depend on the nature of the ERA service. If, for example, the service emphasizes longer-term outcomes beyond the follow-up period (such as encouraging investment in human capital through additional take-up of education and/or training), the impact on months of employment during the follow-up period may be negative and the impact on months receiving welfare may be positive. On the other hand, if the ERA service emphasizes shorter-term outcomes during the follow-up period (such as in-work advice or information about monetary benefits available from ERA) the impact on months of employment during the follow-up period may be positive and the impact on months receiving welfare may be negative. From the policy maker's perspective, negative impacts on employment and positive impacts on welfare receipt during the follow-up period may be viewed as somewhat disappointing, however from the individual's perspective these may lead to better long-term outcomes, beyond the follow-up period.

It is important to keep in mind that when testing hypotheses about the relationship between the intensity and type of ERA services and the impacts of ERA, the control group plays an important role. Previous studies have identified the possibility of "substitution bias" in social experiments (Heckman and Smith, 2005, Heckman et al., 2000). Many control group members received services under the existing NDLP program that were similar to the services received by program group members under ERA. The impact of ERA will be

influenced by the *differential* receipt of services between program and control group members. If control group members receive the same advancement services as program group members, then both might potentially benefit, but the *impact* of ERA would be zero. Thus, when we measure services received by ERA program group members in a particular office, we need to construct them as the *difference* in the receipt of those services between program and control group members, to account for possible substitution bias.[13] The actual level of service receipt of control group members will influence control group (NDLP) outcomes, but not the impacts of ERA.[14]

In addition to individual-level data, our analysis uses office-level variables for both ERA program group members and control group members that allow us to relate the inter-office differences in impacts to the particular characteristics of the offices. Consequently, a multi-level statistical framework is required (see Bryk and Raudenbush, 2001, and Bloom et al., 2005 for details). The basic structure of the multi-level model has the following interpretation. First, the individual-level outcomes (months on welfare and months employed) are allowed to vary across individuals according to program-control group status and individual characteristics. Program impacts are also allowed to vary with individual characteristics, yielding subgroup impacts. In addition, control group outcomes and program impacts are allowed to vary with office characteristics indicating the intensity and type of services received. The framework allows both the control group outcomes and the program impacts to vary with observable and unobservable characteristics.

Because the outcome variables are integer count variables, we use a mixed-effects Poisson model to estimate the parameters. There are two random error terms in the model. The first captures random variation in the average office-level outcome for the control group. The second captures random variation in the average office-level impact for the program

---

[13] Implicit in using the difference between program and control group members in receipt of services is that the services received by the two groups are close substitutes.

[14] Technically, control group members were not supposed to receive ERA services, but in practice control group members may have received similar types of services. In fact, as will be seen, many control group members did receive ERA-like services. This did not apply to the retention bonuses which could only be received by those in those in the program group

group. It is the separate specification of the two error terms and the inclusion of office-level characteristics as explanatory variables that distinguish the multi-level model from the more traditional regression models used in the program evaluation literature.

The multi-level Poisson model described above has the following formal statistical structure:[15]

(1) $\Pr(Y_{ji} = y | \upsilon_j, \mu_j) = \exp(-\theta_{ji}) \, \theta_{ji}^y / y!$

(2) Level 1: $\quad Y_{ji} = \alpha_j + \beta_j P_{ji} + \Sigma_k \delta_k CC_{kji} + \Sigma_k \gamma_k CC_{kji} P_{ji},$

(3) Level 2: $\quad \alpha_j = \alpha_0 + \Sigma_m \zeta_m SI_{mj} + \Sigma_n \eta_n ST_{nj} + \upsilon_j,$

$\qquad\qquad\quad \beta_j = \beta_0 + \Sigma_m \pi_m DSI_{mj} + \Sigma_n \varphi_n DST_{nj} + \mu_j,$

or, combining the equations for levels 1 and 2,

(4) $Y_{ji} = \alpha_0 + \Sigma_m \zeta_m SI_{mj} + \Sigma_n \eta_n ST_{nj} + \beta_0 P_{ji} + \Sigma_m \pi_m DSI_{mj} P_{ji} + \Sigma_n \varphi_n DST_{nj} P_{ji} + \Sigma_k \delta_k CC_{kji} +$
$\Sigma_k \gamma_k CC_{kji} P_{ji} + [\upsilon_j + \mu_j P_{ji}],$

where:

$\theta_{ji} = \exp(Y_{ji}),$

Y_{ji} = outcome for sample member i in office j (number of months receiving welfare or number of months employed),

P_{ji} = a binary indicator of program status for sample member i in office j (1 for program group members, 0 for control group members),

CC_{kji} = the value of individual characteristic k for sample member i in office j (grand-mean-centered),[16]

SI_{mj} = the value of service intensity feature m for control group members in office j (grand mean-centered),

ST_{nj} = the value of service type feature n for control group members in office j (grand mean-centered),

DSI_{mj} = program-control group difference in the value of service intensity feature m for office j (grand mean-centered),

DST_{nj} = program-control group difference in the value of service type feature n for office j (grand mean-centered),

---

[15] For details on the statistical properties of multi-level models, see Bryk and Raudenbush (2001). We use the xtmepoisson procedure in Stata to estimate the models (Stata Corporation, 2009).

[16] As will be described below, included in the set of individual characteristics is an environmental variable (index of local deprivation), which is defined at a finer level than the local offices. Specifically, there are 1,676 distinct values of this variable for the 6,754 individuals in our sample.

α_j = control-group outcome for the average sample member in office j,

β_j = program-group impact for the average sample member in office j,

δ_k = the effect of individual characteristic k on the control group's outcome,

γ_k = the effect of individual characteristic k on the program impact ($\delta_k + \gamma_k$ is the effect of individual characteristic k on the program group's average outcome),

α_0 = the grand mean control group outcome,

ζ_m = the effect of the service intensity feature m on the control group's outcome,

η_n = the effect of the service type n on the control group's outcome,

β_0 = the grand mean impact,

π_m = the effect of the service intensity difference feature m on the program impact,

φ_n = the effect of the service type difference n on the program impact,

υ_j = a random error term for the control group's outcome in office j,

μ_j = a random error term for the program impact in office j.

The basic structure of this model has the following interpretation. First, the outcome Y_{ji} is allowed to vary across individuals according to program-control group status (P) and individual characteristics (CC). Program impacts are also allowed to vary with individual characteristics, yielding subgroup impacts (represented by the γ_k). In addition, average office control group outcomes (the α_j) and the average office program impacts (the β_j) are allowed to vary with office characteristics (SI and ST for control group members and DSI and DST for program group members).[17] Because individual and office characteristics are grand-mean centered, α_j represents the outcome an average control group individual would experience in local office j and α_0 represents the overall value of the outcome for the average control group member of the full sample. Similarly, β_j represents the impact ERA would have on the average program group member in office j, while β_0 represents the average impact of ERA in the sample as a whole.

---

[17] Allowing control group outcomes to vary with office characteristics represents an extension of the model presented in Bloom et al. (2005) which only allows office characteristics to affect program impacts, whereas we allow office characteristics to affect both control group outcomes *and* program impacts.

In estimating the parameters of this model, we assume that the office error terms μ_j and υ_j are correlated with each other and are realizations from a bivariate normal distribution with mean 0 and 2x2 variance matrix $\boldsymbol{\Sigma}$. Estimation is performed using maximum likelihood. In all cases, the estimated variances of the error terms are statistically significant and the correlation coefficients of the error terms are negative and statistically significant (full results are available from the authors on request).

4.2 Interpreting the Results

ERA was designed as a randomized control trial and, since randomization was at the level of the individual, office-level impacts estimates are also experimental. However, the analysis in this paper uses non-experimental techniques in order to examine the factors that appear to influence program effectiveness. In view of this, it is appropriate to consider the extent to which the estimation results can be viewed as capturing causal relationships rather than mere associations.

There are two key issues that need to be considered in assessing the causal validity of the results. The first is that the characteristics of individuals may vary across offices in a way that is related to impact. It was explicit in the design of the ERA evaluation that the pilot areas should represent a broad variety of individuals and local economies. We might expect (and indeed our later results confirm this to be the case) that there will be variation across individuals in the effectiveness of ERA. The concern then is that the office-level variation in program effectiveness reflects compositional and other differences across offices. Our analysis controls for the effect of observed individual characteristics on both outcomes (equation 2) and impacts (equation 3). Likewise, we control for variations in area deprivation. There may, of course, be other influences that we do not observe and so cannot be controlled for. Our model assumes that unobserved office-level influences on outcomes are captured by the random error term for control group outcomes (υ_j in equation 3). Similarly, unobserved office-level influences on impacts are captured by the random error term for program impact (μ_j in equation 3). Our model implicitly assumes that, after allowing for the impacts to vary with individual characteristics, the level of local deprivation

15

and unobserved office-level factors, there is no further variation in program effectiveness across subgroups defined by other unobserved characteristics. Given the rich nature of the individual characteristics included in the model and the narrowly defined criteria for inclusion in the experiment (lone parents looking for help re-entering the labor market), this seems a reasonable assumption.

The second concern is that the type of service provided by an office may be endogenous in the sense that it is influenced by characteristics of the individual welfare recipients, local labor market conditions, or other factors that are unobserved. In addition to controlling directly for individual characteristics in the model, the office-level measures of service delivery are constructed in a way that controls for the characteristics of the individuals within that office. This is explained in detail in Section 5.2 (see equation 5) and goes some way towards addressing the potential endogeneity of service type. However, the possibility remains that there are unobserved characteristics that influence both office-level impacts and the implementation strategy adopted by an office. To gain some insight into this, we draw on the qualitative analysis carried out in the course of evaluating ERA and summarized in Hendra et al., (2011). This analysis found little evidence that offices chose strategies to fit around the particular characteristics of the individual welfare recipients. Instead, the intention was very much to deliver a standardized treatment across offices. To achieve this, each district had assigned to it a "Technical Adviser" whose role was to work with caseworkers in that district's offices to ensure that randomization ran smoothly and to advise on delivering in-work support. Furthermore, four of the six districts adopted a centralized approach, thereby limiting the scope for offices to choose their implementation strategies. Other factors do appear to have played a role. Staff shortages were a problem in some areas. In other areas, changes to management policy that were unrelated to ERA had an impact on delivery. For instance, district reorganization meant that some offices were reassigned to a new district, with consequent disruption to delivery, particularly when new district managers did not embrace the ethos of ERA. Overall, the qualitative evidence indicates that variation across offices in the type of support provided is most likely due to exogenous factors.

The strongest basis for achieving causal impact estimates would be if individuals were randomly assigned to offices. This was not feasible, particularly given the large distances between the offices, so we rely instead on a non-experimental approach. However, as with any non-experimental study, there is the possibility that one or more important variables have been omitted. In the discussion of the results, we use causal language, but the reader should remember that those causal statements are only valid when the assumptions of the model are satisfied.

5. Data

To estimate the parameters of equation (4), two kinds of data are required. First, there are the variables measured at the individual level (the outcomes, Y, and the individual characteristics, CC). Second, there are variables measured at the office level (service intensity, SI, and service type, ST). Office variables used in the analysis were derived from staffing forms and the personal interviews conducted during the follow-up period.

5.1 Outcomes

One of the main objectives of the ERA demonstration was employment retention (and hence, a reduction in time spent on welfare). Therefore, the outcomes we examine in this paper are the number of months on welfare and the number of months employed during the five year follow-up period (roughly 2005 to 2009), distinguishing between the in-program and post-program periods. All outcomes were taken from administrative records – the DWP's Work and Pensions Longitudinal Study (WPLS) database. Information on welfare receipt and employment status is available on a monthly basis.[18]

The WPLS contains an identifier that can be used to link to the individuals in the experimental sample. The advantage of this relative to survey data is that there is no attrition in the dataset.

---

[18] For further details on the data sources, see Hendra et al. (2011). Ideally, we would have also considered examining earnings as an outcome. However, both the earnings and log-earnings distributions were highly non-normal implying that a linear specification was not appropriate (indeed, efforts to attempt such a model gave unstable results). We present some alternative estimates of how earnings impacts varied with office characteristics in section 7.

5.2 Individual Characteristics

Individual-level background characteristics were collected as part of the randomization process. Because they were recorded prior to randomization, these background characteristics are exogenous and thus can be included as regressors in the multi-level model without biasing the estimated program impacts. Although, randomization (if implemented properly) guarantees unbiased impact estimates without the need to control for background characteristics, they are included in order to increase precision and to allow subgroup impacts to be estimated (by interacting the background characteristics with the program status indicator, P).

The following individual characteristics (CC) are included in our model:[19]

- Sex,
- Partnership status,
- Highest level of educational qualification,[20]
- Number of children,
- Number of weeks worked in the 3 years prior to random assignment,
- Whether the individual worked in the year prior to random assignment,
- Weekly earnings for the most recent job in the year prior to random assignment,
- Quarter of random assignment,
- Number of months receiving welfare in the 2 years prior to random assignment,
- Age of the youngest child,
- Age,
- Ethnicity.

These individual characteristics were augmented with an official measure of local deprivation.[21] There are 1,676 distinct values of this variable for the 6,784 individuals in our sample. The deprivation measure is intended to control for general economic and social conditions in the areas served by the local offices.

---

[19] A full definition of the variables (including an indication of the reference groups for comparison) is presented in Table 1.

[20] Loosely, A-level qualifications are those typically gained at age 18 while O-level qualification were usually were gained at age 16. "A-level" is used as shorthand for "A-level or higher" and so includes the most highly qualified individuals.

[21] The measure we used is the "Index of Multiple Deprivation," produced by the UK Office of National Statistics. Distinct dimensions of deprivation such as income, employment, education and health are measured and then combined, using appropriate weights, to provide an overall measure of multiple deprivation for each area. Specifically, the areas are "Super Output Areas.". For details, see http://www.neighbourhood.statistics.gov.uk/dissemination/Info.do?page=aboutneighbourhood/geography/superoutputareas/soa-intro.htm.

As will be discussed below, to facilitate interpretation of the estimated coefficients, all individual characteristics were grand-mean-centered (expressed as deviations from the overall mean). In addition, the estimated coefficients from the Poisson model were expressed in monthly equivalents by multiplying the incidence rate ratios minus one by the control group means (that is, percentage effect of each variable times the control group mean for that variable).

Table 1 presents means of the individual characteristics and outcomes used in our analysis, along with their cross-office range. As this table indicates, the sample overwhelmingly comprises female lone parents with generally low levels of educational qualifications. About one-half of these mothers have only one child and in about half of all cases the child is under the age of 6 years. More than 70 percent of the sample did not work in the year prior to random assignment and they received welfare for an average of 17 of the 24 months preceding random assignment.

The median deprivation index in our sample is 27.4, which corresponds to approximately the 71st percentile of deprivation across England.[22] Thus, our sample is somewhat overrepresented by individuals living in relatively disadvantaged areas.

There was considerable inter-office variation in many of the characteristics, including marital status, educational qualifications, number and ages of children, prior work status, age and ethnicity of the individual, and the level of multiple deprivation in the community served by the office.

The average individual in our sample spent about 26 months on welfare during the follow-up period (about 43 percent of the time) and was employed for roughly the same amount of time. Of the two outcomes, average months on welfare showed the greatest inter-office variation, ranging from 14.4 months to 35.3 months. Average months employed ranged from 20.3 months to 33.4.

---

[22] The median deprivation index for England in 2007 was 17.1 (see http://data.gov.uk/dataset/index_of_multiple_deprivation_imd_2007). The 25th percentile in our sample was 16.4, compared to 9.6 for England and the 75th percentile in our sample was 40.3 compared to 30.2 for England.

5.3   Office Characteristics

As indicated above, we classify the office variables into service intensity (individual caseload measures) and service type. For the service-type variables, ERA-control differentials are used to explain variation in program impacts. To explain variation in control group outcomes, control-group values of the service-type variables are used.

The caseload measures were constructed from monthly monitoring forms for the first 17 months of the experiment. All other office-level variables were constructed from individuals' responses to survey interviews carried out 12 and 24 months after random assignment.[23] It is likely that the advice and support offered to individuals were influenced to some extent by their own characteristics. However, more relevant to the analysis is a measure of the extent to which the office emphasized particular elements of ERA (i.e., their philosophical approach to helping persons on welfare achieve self-sufficiency), controlling for differences in the caseload composition. Although office implementation philosophy cannot be observed directly from any of the available data sources, we form proxies for them by adjusting the individual survey measures to control for observable individual characteristics across offices that may have influenced the type of service implemented using the following regression model: [24]

(5) $F_i = \lambda_0 + \Sigma_k \lambda_{1k}O_{ki} + \Sigma_k \lambda_{2k}O_{ki}P_i + \Sigma_l \lambda_{3l}CC_{li} + e_{ji},$

where F_i is the measure of interest for individual i, O_{ki} is a dummy variable indicating whether individual i is in office k, P_i is a dummy variable indicating whether individual i is in the program group or the control group, and CC_{li} is the value of background characteristic l for individual i. The background characteristics include: gender, marital status, the number of children, education, work history and earnings history. Because all variables other than the program group dummy are mean-centered, the coefficient λ_{1k} represents the mean value of F

---

[23] For details on the individual surveys, see Dorsett et al. (2007) and Riccio et al. (2008).

[24] Overall, the adjusted office implementation measures are correlated to some extent with each other (meaning that offices that rank high on one measure have some tendency to rank high on the other), but the correlations are modest at most. Thus, we are able to treat these office implementation measures as separate variables in the statistical analysis.

for control group members in office k, while the corresponding mean value of F for program group members is $\lambda_{1k} + \lambda_{2k}$. The program-control differential is λ_{2k}.

As noted above, the motivation for constructing office-level measures in this way is that it isolates the tendency for offices to vary in the degree to which they emphasize particular aspects of delivery after controlling for the fact that this is driven in part by the between-office variation in caseload composition. A simpler approach would be to use unadjusted measures and rely on the inclusion of individual characteristic variables in the level 1 regression (equation 1) to control for variations across office practices that stem from compositional differences. However, this simpler approach cannot achieve that aim since individual characteristics in the level 1 regression help explain only the variation in the level 1 outcome, not the variation in service intensity or type. A drawback to our approach is that, by subsequently including the λ_{1k} and λ_{2k} terms as regressors in the multilevel model, no account is taken of the fact that they are estimates and subject to error. While this may introduce a specification bias, data limitations prevent us from adopting a better approach.

The specific office variables used in this study are as follows:[25]

- Caseload per advisor,
- Proportion of advisers working with ERA participants,
- Proportion of Individuals Advised to Think Long-Term,
- Proportion of Individuals Helped Finding an Education or Training Course,
- Proportion of individuals whose advisers discussed in-work advancement,
- Proportion of individuals given a lot of support while working,
- Proportion of ERA participants aware of the work retention bonus.[26]

All of the office-level variables (both control group measures and program group measures) were grand-mean-centered. The control values of the grand-mean-centered variables were included in the level 2 equation determining α_j (the control mean outcome) while the difference in the values of the grand-mean-centered variables between the program

---

[25] A fuller description of the office variables, their construction, and their expected effects on control group outcomes and ERA program impacts is given in Dorsett and Robins (2011).

[26] A possible concern with the bonus variable is that awareness of the retention bonus may have been affected by individuals' employment status. However, a key aim of ERA was to inform every participant, employed or not, of the availability of the bonus. Overall, awareness was very high among both workers and nonworkers.

and the control groups were interacted with the program group dummy variable (P_{ji}) and included in the level 2 equation determining β_j (the program impact).

5.4   Summary Statistics for the Office Variables

Table 2 present the means and the cross-office range of the (regression-adjusted) office variables used in the multi-level analysis. The caseload averages about 29 individuals per adviser and about 42 percent of these advisers, on average, work with ERA participants. There is significant variation in the caseload across offices (from about 3 individuals per adviser to 110 individuals per adviser) and in the proportion of advisers working with ERA participants (from about 20 percent to 94 percent).

For each of the service type measures, Table 2 presents the mean proportion for the control (NDLP) group, the mean proportion for the program (ERA) group, and the mean ERA-control group difference in the proportion. The first and third of these (control group value and ERA-control group difference) are used as variables in the multi-level model. The second (ERA value) is not directly included in the multi-level model (except for the retention bonus awareness variable) and is shown for informational purposes only.

On average, for every service type, the ERA group had a higher proportion receiving that service than the control group. This is as would be expected, however the differential is not always that great. In some offices, a greater proportion of the control group received the services, as reflected in the negative minimum values of the differential in the cross-office ranges.[27] In no office were less than three-quarters of the ERA participants aware of the retention bonuses and in some offices all of the ERA participants surveyed were aware of the bonuses.

The considerable amount of services received by control group members may have contributed to the fact that there were few significant overall impacts in the ERA evaluation and highlights the importance of the type of model presented in this paper that attempts to

---

[27] Specifically, there were 4 offices in which the proportion of individuals advised to think long-term was higher among the control group than the program group; 7 offices where the proportion of individuals receiving help finding an education or training course was higher; 6 offices where the proportion receiving help with in-work advancement was higher; and 7 offices where the proportion receiving support while working was higher.

control for possible substitution bias in estimating impacts of particular program features across offices. As will be indicated later in section 6.5, by empirically taking into account the possibility of substitution bias, the estimated coefficients on the office-level program-control group differences represent the impacts assuming no substitution bias (that is the impact assuming all program group members receive the particular feature in question and no control group members receive it). We describe how these coefficients need to be interpreted to reflect the actual substitution biases present in the data.

Table 3 presents a correlation matrix of the office variables for the control group and the ERA program group. For both groups, the correlations between the non-caseload variables are all positive, suggesting that retention and advancement services were being delivered together, although not perfectly. For the ERA group these positive correlations are consistent with the goals of the demonstration. From a statistical standpoint, the fact that the correlations are modest implies that it is theoretically possible to estimate the contribution of each element separately.

6. Results

We present the results of estimating the multi-level Poisson model in Tables 4-7. As was done for the empirical Bayes estimates in Figures 1A and 1B, we present separate estimates for the in-program period (years 1 to 3) and the post-program period (years 4 and 5). Recall that the Poisson coefficients are presented in monthly terms to facilitate interpretation of the results.

Table 4 shows the effects of the individual characteristics on the five-year control group outcomes. Table 5 shows how these individual characteristics affect the program impact (subgroup impacts). Table 6 shows how the office characteristics affect the control group outcomes and Table 7 shows how the office characteristics affect the program impacts.

6.1   Effects of Individual Characteristics on Outcomes

Table 4 presents the effects of the individual characteristics on months receiving welfare and months employed for control group members during the three-year in-program and two-year post program periods. For comparison purposes, the grand mean control group

23

outcome (α_0) is also presented in the table. The coefficients of the individual characteristic variables in Table 4 represent the *deviation* from the mean outcome of the omitted reference group (see Table 1). Thus, for example, the coefficient of 2.62 for individuals with A-level qualification on months employed in years 1-3 is their additional months employed compared to individuals with no qualifications.

The average control group member spent 17.9 months on welfare and 13.8 months employed during the in-program period and 7.3 months on welfare and 10.2 months employed during the post-program period. As would be expected, many of the individual characteristics are significantly related to the outcomes in both periods. Individuals who are younger (below age 30), less educated (qualifications below O-level), have less previous work experience (worked 12 or fewer months in the past three years), are non-white, and live in more deprived areas spent longer periods of time on welfare and had less time employed than their counterparts (who are aged at least 30, qualified at O-level or higher, worked more than 12 months in the three years before random assignment, white, and living in less deprived areas). Individuals who were not previously partnered also spent more time on welfare than those who were previously partnered, but did not spend less time employed during the in-program period, although they spent less time employed during the post-program period.

Interestingly, time spent on welfare declines systematically during the in-program period according to the calendar time of random assignment (the later the time of random assignment, the fewer the months spent on welfare). At first sight, this seems somewhat surprising given the onset of recession in the second quarter of 2008 will have affected labor market outcomes of those randomized earlier less than those randomized later. However, there are two countervailing factors. First, a feature of the recent recession is that, up until the second quarter of 2010 (the latest period for which outcomes are considered in this analysis) the reduction in the overall employment rate was driven almost entirely by the fall in the proportion of men in work. As we have already seen, the NDLP group is predominantly female and women's employment remained comparatively stable. Second, policy

24

developments in the UK have increased the conditions placed on lone parents. For example, those in receipt of welfare have had to attend an increasing number of work-focused interviews and, since 2005, to agree an action plan with their adviser to prepare themselves for work (Finn and Gloster, 2010). As another example, since 2008, lone parents with a youngest child aged 12 or over are no longer entitled to welfare solely on the grounds of being a lone parent (DWP, 2007). Those randomly assigned more recently will have been subject to the new regulations for a greater proportion of their follow-up period than those randomly assigned earlier.[28]

6.2 Effects of Individual Characteristics on Program Impacts

Table 5 presents the effects of the individual characteristics on program impacts over the three year in-program and two-year post-program periods. For comparison purposes, the grand mean impact of ERA (β_0) is included in the table. The coefficients represent deviations from the impacts of the omitted reference groups (see Table 1). Thus, for example, the coefficient of 2.66 for individuals with A-level qualification on months employed during the in-program period is their *additional* impact compared to individuals with no qualifications. Note that the impacts for individuals in the reference groups (those with no qualifications in this example) are not shown in Table 5. All that the table shows are *deviations* in impacts from the reference group, and not the impacts themselves for either group.

The average response to ERA (the grand mean impact in Table 5) is statistically significant for both outcomes during the in-program period, but is not statistically significant during the post-program period. During the in-program period, months on welfare declined by about one and a half months (8.5 percent) and months employed increased by about three-quarters of a month (5.5 percent).

Several of the impacts vary significantly across subgroups. One notable finding has to do with educational qualifications. It appears that individuals with O- and A-level qualifications had stronger responses to ERA over the full five-year follow-up period than

---

[28] A fuller discussion of policy developments in the UK during the years ERA was conducted is presented in Hendra et al. (2011).

individuals with no qualifications. They had larger reductions in the number of months on welfare, and larger increases in the number of months employed than individuals with no qualifications. Another notable result is that during the in-program period, ERA seems to have had its biggest impacts on individuals who had the least amount of employment during the three years prior to random assignment.   Specifically, months on welfare fell by more and months employed increased by more for individuals who had been employed for a year or less in the three years prior to random assignment. These impacts did not carry over into the post-program period—in fact, months on welfare actually rose for these individuals during the post-program period. Still another notable result is that the impacts on months receiving welfare and months employed seem to have varied with the degree of local area deprivation, particularly during the in-program period. Specifically, ERA participants living in more deprived areas had larger reductions in months on welfare and larger increases in months employed than ERA participants living in less deprived areas. Thus, ERA appears to have been more effective in more deprived areas. Finally, ERA seems to have caused larger reductions in months on welfare and greater increases in months employed for older individuals (aged 30 years and above) and minority individuals.

6.3   Effects of Office Characteristics on Office Control Group Outcomes

Table 6 shows how the office characteristics affect office control group outcomes. In other words, the results in Table 6 provide an indication of whether office characteristics are systematically related to office outcomes for standard NDLP participants. For comparison purposes, the grand mean control group outcome (α_0) is also shown. In addition to presenting the coefficient estimates, we also present the interquartile range of the outcome across offices. The interquartile range is the predicted outcome from the 25th percentile of the office characteristic to the 75th percentile.   The interquartile range provides an indication of how the control group outcome varies across offices possessing the middle 50 percent range of values of a particular characteristic.

As Table 6 indicates (and as already shown in Table 4), the average control group member spent about 18 months (50 per cent of the time) on welfare during the in-program

26

period, about 7 months (30 per cent of the time) on welfare during the post-program period, about 14 months (39 perent of the time) employed during the in-program period, and 10.2 months (43 percent of the time) employed during the post-program period. The interquartile range across offices is modest, from 17 to 19 months on welfare during the in-program period, 7 to 9 months on welfare during the post-program period, 13 to 15 months employed during the in-program period, and 10 to 11 months employed during the post-program period.

As Table 6 further indicates, both the intensity and type of service appears to matter for the control group over the entire five-year follow-up period. Offices with larger adviser caseloads have control individuals that spent more months on welfare and fewer months employed during the follow-up period. The effects are slightly larger during the in-program period, but not significantly different from the effects in the post-program period. The results imply that for the average office with almost 30 individuals per adviser (see Table 2), a doubling of the caseload during the in-program period would result in about 1.5 more months on welfare (.05x30) and almost .6 fewer months of employment (-.02x30) for control (NDLP) individuals. The interquartile range across offices of the effects of caseload size during the in-program period ranges from 17 to 18 months for welfare and 13.8 to 14.2 months for employment. While these are small differences across offices, they do illustrate that the size of adviser caseloads matters for effectively administering services under the standard NDLP program.[29]

The results also suggest that in offices where all NDLP recipients receive help in finding education courses, the amount of time spent on welfare is increased by 11 months during the in-program period and 6 months during the post-program period and the amount of time spent in work is reduced by 6 months during the in-program period nd 3 months during the post-program period relative to in offices where no recipients receive such help. These are sizeable effects. While they imply greater dependence on welfare during the five-year follow-up period, they may imply greater self-sufficiency in the long-run if the education

---

[29] Note that while the inter-quartile range provides some indication of the variation across offices, the office-level variables considered tend to have "long tails" such that a comparison of the full (minimum to maximum) range would give the impression of substantially more variation.

eventually leads to upgraded skills and higher employment and earnings. However, inspection of Table 6 reveals that the effects of prolonging welfare and reducing employment are stronger during the in-program period and gradually weaken after that. As was the case for the effects of caseload size, the variation across offices in the proportion of recipients receiving help in finding education courses is not great.

In contrast to education services, in offices where all recipients receive help with in-work advancement, there is a statistically significant effect on months employed during the in-program period, but not during the post-program period nor on months receiving welfare at any time during the full five-year follow-up period. The employment effect is sizeable, but doesn't vary much across offices.

Finally, in offices where individuals receive support while working, months on welfare decline and months employed increase during both the in-program and post-program periods, although the post-program effect on welfare is not statistically significant, but as in the case of help with in-work advancement, there is little variation in this effect across offices.

Taken together, these results suggest that certain services matter for traditional NDLP recipients, particularly those that target education and employment activities. However, those that target education tended to prolong welfare receipt and delay employment during the five-year follow-up period while those that target employment tended to have the opposite effect, reducing time spent on welfare and increasing time employed during the five-year follow-up period. Because we do not have data beyond the five-year follow-up period, we are unable to determine whether the additional education help received during the five-year follow-up period eventually leads to lower receipt of welfare and greater employment over the longer run.

6.4   Effects of Office Characteristics on ERA Program Impacts

Table 7 shows how the office characteristics are related to ERA program impacts. Recall that these results are based on a non-experimental analysis and can only be given a causal interpretation if the assumptions of the model are satisfied. Also recall that for the ERA input types available to control group members (advice for thinking long-term, help in

finding education courses, help with in-work advancement, and support while working), the office characteristics included in the multi-level model are measured as *differences* in the proportions receiving such services between the ERA program group and the control group (see Table 2). The other two office characteristics included in the multi-level model (the proportion of advisers working with ERA participants and the proportion of ERA participants aware of the employment retention bonus), apply only to ERA program group members and, hence, are simply measured as the proportion for ERA program group members.

As indicated in Table 7, there are statistically significant impacts of ERA on welfare receipt and employment during the in-program period, but not during the post-program period. During the in-program period, welfare receipt is reduced by 1.5 months (an 8 percent impact) and employment is increased by .8 months (a 6 percent impact).

During the in-program period, five of the six office characteristics are estimated to be significantly related to ERA program impacts. First, in offices where all of the advisers were working with ERA participants, the average program group member spent 3 fewer months on welfare, but was not employed significantly longer, than in offices where no advisers were working with ERA participants. To put it another way, an individual in an office with a 10 percentage point higher proportion of advisers working with ERA participants will have .3 fewer months on welfare than an individual in an office where the same proportion of advisers worked with ERA participants and control group members (NDLP recipients). The information on interquartile ranges is very important because, in practice, few of the program-control group differences in receiving this kind of help were very large, so the effect translates to only about a .6 month interquartile range across offices in the impact of the advisers on welfare receipt.

Second, in offices where all ERA participants were given help finding education courses but control group members were not, the average program group member spent almost 4 more months receiving welfare and 4 fewer months employed, although the welfare effect is not statistically significant. Again, the information on interquartile ranges is very important because, in practice, few of the differences in receiving this kind of help were very

29

large, so the effect translates to only about a 1.1 month interquartile range across offices in the impact on welfare receipt and about a 1.2 month interquartile range across offices in the impact of this service on months employed.

Third, in offices where all ERA participants received help with in-work advancement, but control group members did not, the average program group member spent almost 8 more months employed but not a statistically significant shorter time on welfare. Again, few of the differences in receiving this kind of help were very large across offices, so the effect translates to only about a 1.2 month interquartile range across offices in the impact of this service on months employed.

Fourth, in offices where all ERA participants received support while working, the average program group member spent 3.5 fewer months on welfare and was employed for 3.2 more months. The interquartile range of impacts was about 1.1 months for welfare and 1.0 months for employment.

Finally, in offices where all ERA participants were aware of the bonus, the coefficient implies that they would have spent 9.4 fewer months on welfare than in offices where no ERA participants were aware of the bonus. There is also a sizeable coefficient of 8.4 months for employment, but it is not statistically significant. In practice, almost all ERA participants were aware of the bonus (no office had fewer than 75 per cent aware), so while the bonus was apparently an important part of the ERA program design, it translated into a moderately small (about 1 month) interquartile range of ERA program impacts across offices.

Virtually all of the services that had a statistically significant impact during the in-program period retain their statistical significance during the post-program period. The one exception is for the impact of help with in-work advancement on employment which is no longer statistically significant in the post-program period. However, the impact of this service remains positive. For all of the services, as was the case during the in-program period, the interquartile ranges of impacts were modest because of mostly small program-control group differences in receipt of these services.

7. An Alternative Specification to Examine Earnings

As indicated earlier, the chief objective of ERA was to encourage employment retention and so our main outcomes of interest were time spent employed and time spent on welfare. However, ERA also aimed to promote advancement in employment. Pay progression is one possible manifestation of advancement, so it is of interest to consider earnings as an outcome.

Unfortunately, as noted in section 5, it was not possible to estimate a multilevel models for earnings. In order to have some sense of how earnings impacts vary with program-control differences in office characteristics, we present in this section supplementary results using a "reduced from" estimation approach, similar to the one used by Somers et al.(2010) in examining how impacts on student grades vary with program implementation conditions in a demonstration of supplemental literary courses for struggling ninth graders.

Methodologically, we use a linear regression model, but cluster the standard errors in order to allow for within-office correlation of errors. This approach implies a simplified version of equation (4) as follows:

$$(1) \quad Y_{ji} = \alpha_0 + \beta_0 P_{ji} + \Sigma_m \pi_m DSI_{mj} P_{ji} + \Sigma_n \varphi_n DST_{nj} P_{ji} + \Sigma_k \delta_k CC_{kji} + \Sigma_k \gamma_k CC_{kji} P_{ji} + \varepsilon_j + u_{ji}.$$

It is helpful to highlight the differences between this specification and the multilevel model. First, to control for variations between offices in the level of earnings, an office-specific error term, ε_j, has replaced the random effect υ_j. A consequence of this is that variables that do not vary within offices cannot be be included, so the $\Sigma_m \zeta_m SI_{mj}$ and $\Sigma_n \eta_n ST_{nj}$ terms from equation (4) are no longer present and therefore variation in control group outcomes with office characterisitics cannot be estimated. Second, this specification does not involve the interaction term $\mu_j P_{ji}$. This amounts to an assumption that the the office-level error term in equation (4) is zero. In other words, all variation in program impacts is assumed to be explained by the program-control differences in services. Third, an individual-level error term, u_{ji}, has been introduced since we are now estimating a linear regression model rather than a Poisson model.

31

The results provided by this model are of interest both in themselves and also because they represent a more common estimation approach seen in the literature. We preface them by noting that, for the welfare and employment outcomes, the estimated variances and correlation coefficients of the office-level error terms are statistically significant, so our expectation might be that this would also apply when considering earnings. In view of this, the results in this section may be based on a mis-specified model.

With this caveat in mind, the results are presented in Table 8. Earnings data are only available on a financial year basis – from 2005/6 to 2008/9 – so, unlike the welfare and employment outcomes (available monthly), it is not possible to relate earnings outcomes directly to the timing of randomization. We present the results for each financial year but note that ERA eligibility would be ongoing for all program group individuals throughout 2005/6 and would have expired for all before 2008/9.

With regard to the overall impact of ERA, this was statistically significant in 2005/6, increasing annual earnings by an estimated £309. There was no significant impact in later years. This is consistent with the welfare and employment impacts which showed significant impacts during the in-program period but not the post-program period. Under this specification of the model there is no variation in program impacts other than that associated with program-control differences in services. Consequently, Table 8 does not report an interquartile range around the grand mean impact.

Program impacts did not vary with the proportion of advisors working with ERA participants except in 2008/9, where the reported positive coefficient translates into an interquartile range of nearly £300 across offices in the impact of advisers. This is consistent with the reported results for time spent on welfare, which also showed a variation that became more statistically significant in the post-program period.Higher earnings impacts in 2005/6 were also seen in offices where the proportion of ERA participants advised to think long-term was higher. The interquartile range in this case was just over £500. However, this variation was not statistically significant in later years. For welfare and employment outcomes, there was no significant variation in any year.

There is evidence that the earnings impacts were lower in offices that provided more help with finding education courses. This was consistent across all years, although only statistically significant in 2005/6 and (especially) 2008/9. The interquartile range in 2008/9 is £769. It is perhaps of some concern that these longer-term outcomes are not suggestive of emphasis on education being rewarded with positive returns. It is of course possible that this finding could be reversed with even longer-term outcomes. We note that these results are consistent with those reported for employment impacts.

Emphasizing help with in-work advancement on the other hand is associated with stronger earnings impacts in all years. Beginning 2006/7, these variations are statistically significant, and the interquartile range is quite stable at £375, £511 and £461 in this and the successive two years respectively. The employment impacts showed similar variation during the in-program period but not during the post-program period.

There was no significant impact variation in any year with the proportion of ERA participants receiving support while working. This is in contrast to the welfare and employment impacts, for both of which this appeared to be a key factor along which impacts varied. Nor was there any variation associated with awareness of the retention bonus, something that had been shown to correlated with program effectiveness when considering exits from welfare. However, the bonus awareness coefficients are positive and large for all four years and in three of the years the coefficients are not too far from being statistically significant.

Overall, this summary of the earnings results has revealed a general consistency with the welfare and employment outcomes, but there also some differences. The reasons for the differences are not clear but could simply be the result of the different estimation techniques followed. In view of this, and of our preference for the multi-level specification, we do not attempt to interpret these differences.

8. Conclusions and Policy Implications

For out of work lone parents, the ERA demonstration had statistically significant impacts on welfare receipt and employment during the in-program period (years 1 to 3) and

these impacts varied significantly across the offices that participated in the demonstration

The main purpose of this study has been to examine how program impacts varied with differences across the offices in the way the ERA program was implemented. Secondary objectives of this study have been to determine whether office characteristics can help explain cross-office variation in the control environment (under the standard NDLP program) and whether the impacts of ERA vary with certain personal characteristics of the ERA participants (subgroup impacts).

In interpreting the results of this study, it is important to understand that while certain office characteristics may be quite important in explaining outcomes and impacts, lack of variation in these characteristics across offices may lead to only a small estimated variation in these outcomes and impacts across offices. Thus, for example, while our results indicate the importance of conveying information about the financial rewards available to lone parent ERA participants who maintain employment (given by the estimated coefficients in Table 7), there was not much variation in the actual conveying of this information across offices, so it is associated with only modest variation in program impacts across offices.

Our results indicate that ERA was especially effective at reducing welfare receipt and increasing employment for lone parents with O- and A-level qualifications, those living in more deprived areas, and those aged 30 or over. Subgroup variation was not, though, the primary focus of this analysis. Our main results concern impact variation with office characteristics. Several such characteristics were found to be related to the control environment (outcomes of control group members under the standard NDLP program). Offices with higher adviser caseloads had control group lone parents that spent more months on welfare and fewer months employed over the five-year follow-up period. Where offices prioritized help in finding education courses, control group individuals had more months on welfare and fewer months employed. Finally, where offices emphasized in-work advancement, control group individuals had more months employed.

Most importantly, we find that several office characteristics were associated with greater *impacts* of the ERA demonstration on lone parents. It is important to keep in mind

that these results are non-experimental and such associations can be given a causal interpretation only if the assumptions of the model are valid. ERA's main design feature was to extend the NDLP program by providing help *after* employment was obtained. We have estimated that such retention services can lead to additional impacts beyond those obtained under the New Deal program and can help individuals achieve economic self-sufficiency (by spending fewer months on welfare and more months employed). Offices that assigned more caseworkers to ERA participants tended to be more successful in reducing time spent on welfare. Offices that emphasized in-work advancement and in-work support more generally tended to deliver stronger effects of ERA, as did those offices where awareness levels of the employment retention bonus were higher. On the other hand, offices that emphasized human capital investment tended to have weaker employment impacts.

The results of this study are also interesting in another regard. While the overall impact of ERA on welfare and employment 4 to 5 years post-randomization was not statistically significant (see Table 7), we find that this masks significant variation of impacts across offices, some being positive and some negative. This suggests that, in addition to focusing on overall impacts, which is typically done in employment and training demonstrations such as the one examined here, policy evaluation should, where possible, pay attention to implementation procedures across offices where the program is being conducted. Rather than concluding a policy to be ineffective, the type of approach presented in this paper may offer a means of learning from those with positive impacts in order to refine policy and, in time, raise overall effectiveness.

Although we were unable to estimate a mult-level model of earnings due to statistical convergence problems, we were able to estimate a simpler, more restrictive, earnings model that has been used in other studies to examine variation in program impacts with program implementation practices. The earnings model estimates are roughly consistent with the multi-level welfare and employment models, but there are also some differences, primarily in statistical significance rather than direction of effects.

In conclusion, it is relevant to mention that, as with any long-term study, the economic and policy environment changes. Most obviously, the results relate to a period marked by severe recession and associated increases in unemployment. Equally relevant though is the fact that the last few years have seen a number of policies introduced that directly affect lone parents in the UK. Lone parents have been increasingly required to attend work-focused interviews and those with a youngest child aged 7 or over now have to actively seek work. Furthermore, In-Work Credit was introduced in 2008, providing weekly subsidies to lone parents entering work of 16 or more hours per week. The effect of such policy developments is to reduce the contrast between the service available to the ERA group and that available to the control group and has an important bearing on how to view the overall effect of ERA. However, despite these policy changes and despite that fact that our analysis is non-experimental, we have obtained plausible results identifying those particular implementation features that tended to be linked to stronger impacts of ERA.

References

Bloom, Howard, Carolyn J. Hill, and James A. Riccio. 2005. "Modeling Cross-Site Experimental Differences to Find Out Why Program Effectiveness Varies," in Howard S. Bloom, ed. *Learning More From Social Experiments*, New York, Russell Sage Foundation, pp. 37-74.

Bryk, Anthony S. and Stephen W. Raudenbush. 2001. *Hierarchical Linear Models: Applications and Data Analysis Methods*, second edition, London, Sage Publications.

Dehejia, Rajeev H. 2003. "Was There a Riverside Miracle? A Hierarchical Framework for Evaluating Programs With Grouped Data," *Journal of Business and Economic Statistics*, Vol. 21, No. 1, January, pp. 1-11.

Department for Work and Pensions. 2002. *Lone Parent Brief*, unpublished, Sheffield, UK.

Department for Work and Pensions. 2007. *In Work, Better Off: Next Steps to Full Employment*, London: The Stationery Office.

Dorsett, R., Campbell-Barr, V., Hamilton, G., Hoggart, L., Marsh, A., Miller, C., Phillips, J., Ray, K., Riccio, J., Rich, S. and Vegeris, S. 2007. *Implementation and First Year Impacts of the UK Employment Retention and Advancement (ERA) Demonstration*, Sheffield, UK, Department for Work and Pensions Research Report No. 412.

Dorsett, Richard and Philip K. Robins. 2010. *The Employment Retention and Advancement (ERA) Demonstration: Effects on Employment and Benefit in the 4 Years Post RA and Earnings in 2005/6, 2006/7, and 2007/8 Tax Years*, Unpublished Memorandum submitted to the Department for Work and Pensions, Sheffield, UK, October.

Dorsett, Richard and Philip K. Robins. 2011. *In-Work Support for Lone Parents: Using the UK ERA Demonstration to Examine Cross-Office Variation in Effectiveness*, Working Paper Number 105, Department for Work and Pensions, Sheffield, UK, November.

Finn, Dan and Rosie Gloster. 2010. *Lone Parent Obligations: A Review of Recent Evidence on the Work-Related Requirements Within the Benefit Systems of Different Countries*, Sheffield, UK, Department for Work and Pensions Research Report No. 632.

Friedlander, Daniel, David H. Greenberg, and Philip K. Robins. 1997. "Evaluating Government Training Programs for the Economically Disadvantaged," *Journal of Economic Literature*, Vol. 35, No. 4, December, pp. 1,809-1,855.

Galdo, Jose. 2008. *Treatment Effects for Profiling Unemployment Insurance Programs: Semiparametric Estimation of Matching Models with Fixed Effect*, Carleton University, unpublished working paper.

Goodman, Alissa and Barbara Sianesi. 2007. *Non-Participation in the Employment Retention and Advancement Study: A Quantitative Descriptive Analysis*, Sheffield, UK, Department for Work and Pensions Working Paper 39.

Greenberg, David, Johanna Walter, and Genevieve Knight. 2009. *The Cost of Services and Incentives in the UK Employment Retention and Advancement (ERA) Demonstration: Preliminary Analysis*, Sheffield, UK, Department for Work and Pensions Working Paper No. 64.

Greenberg, David and Philip K. Robins. 2011. "Have Welfare-to-Work Programs Improved Over Time in Putting Welfare Recipients to Work?" *Industrial and Labor Relations Review*, Vol. 64, No. 5, October, pp. 910-920.

Hamilton, Gayle. 2002. *Moving People from Welfare to Work: Lessons from the National Evaluation of Welfare-to-Work Strategies*, New York, MDRC, July.

Heckman, James J. and Jeffrey Smith. 1995. "Assessing the Case for Social Experiments," *The Journal of Economic Perspectives*, Vol. 9, No. 2, Spring, pp. 85-110.

Heckman, James, Neil Hohmann, Jeffrey Smith, and Michael Khoo. 2000. "Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment," *Quarterly Journal of Economics*, Vol. 115, No. 2, May, pp. 651-694.

Hendra, Richard, Keri-Nicole Dillman, Gayle Hamilton, Erika Lundquist, Karin Martinson, and Melissa Wavelet. 2010. *How Effective Are Different Approaches Aiming to Increase Employment Retention and Advancement? Final Impacts for Twelve Models*, New York, MDRC, April.

Hendra, Richard, James A. Riccio, Richard Dorsett, David H. Greenberg, Genevieve Knight, Joan Phillips, Philip K. Robins, Sandra Vegeris, and Johanna Walter, with Aaron Hill, Kathryn Ray, and Jared Smith. 2011. *Breaking the Low-Pay, No-Pay Cycle: Final Evidence from the UK Employment and Advancement (ERA) Demonstration*, Sheffield, UK, Department for Work and Pensions Research Report No. 765.

Hotz, V. Joseph, Guido W. Imbens and Jacob A. Klerman. "Evaluating The Differential Effects Of Alternative Welfare-to-Work Training Components: A Reanalysis Of The California GAIN Program," *Journal of Labor Economics*, 2006, Vol. 24, No. 3, July, pp. 521-566.

Riccio, James A., Helen Bewley, Verity Campbell-Barr, Richard Dorsett, Gayle Hamilton, Lesley Hoggart, Alan Marsh, Cynthia Miller, Kathryn Ray, and Sandra Vegeris. 2008. *Implementation and Second-Year Impacts for Lone Parents in the UK Employment Retention and Advancement (ERA) Demonstration*, Sheffield, UK, Department for Work and Pensions Research Report No. 489.

Somers, Marie-Andrée, William Corrin, Susan Sepanik, Terry Salinger, Jesse Levin, Courtney Zmach, and with Edmond Wong. 2010. *The Enhanced Reading Opportunities Study Final Report*, NCEE 2010-4021, Washington, D.C., U.S. Department of Education, July.

Stata Corporation. 2009. *Longitudinal-Data/Panel-Data Reference Manual, Release 11*, College Station, TX, Stata Press, 2009.

Figure 1A: Proportionate impacts on welfare, by office

Months receiving welfare years 1-3

Months receiving welfare years 4-5
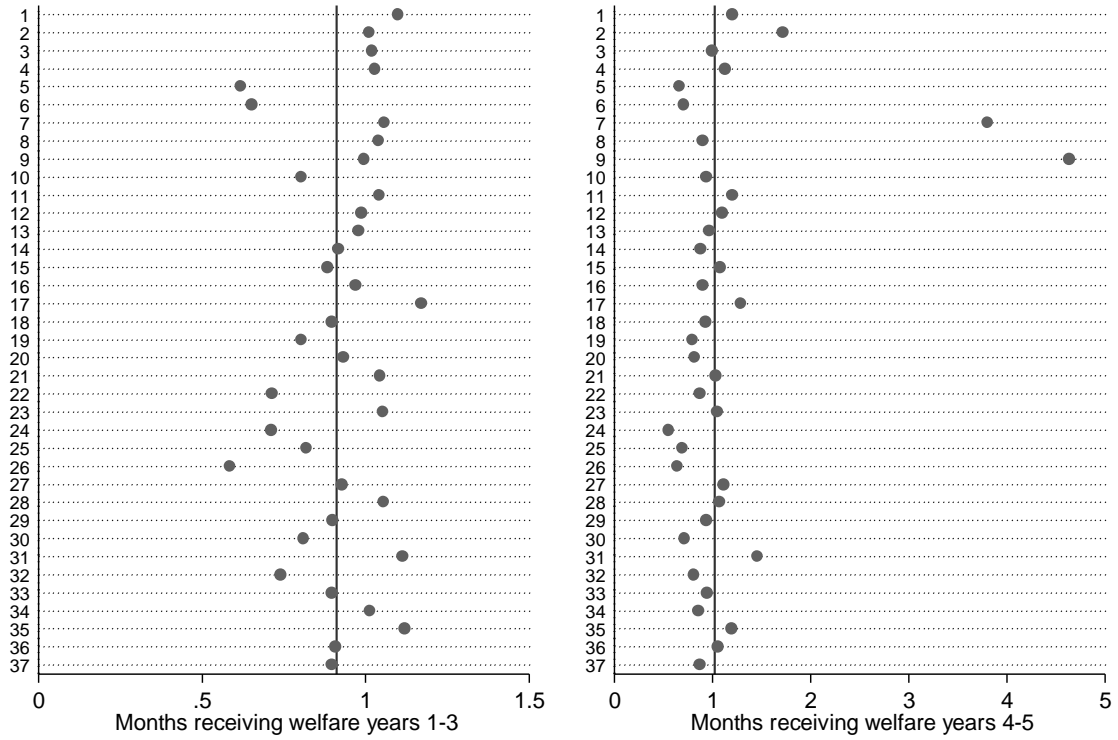
Figure 1B: Proportionate impacts on employment, by office

Months employed years 1-3

Months employed years 4-5
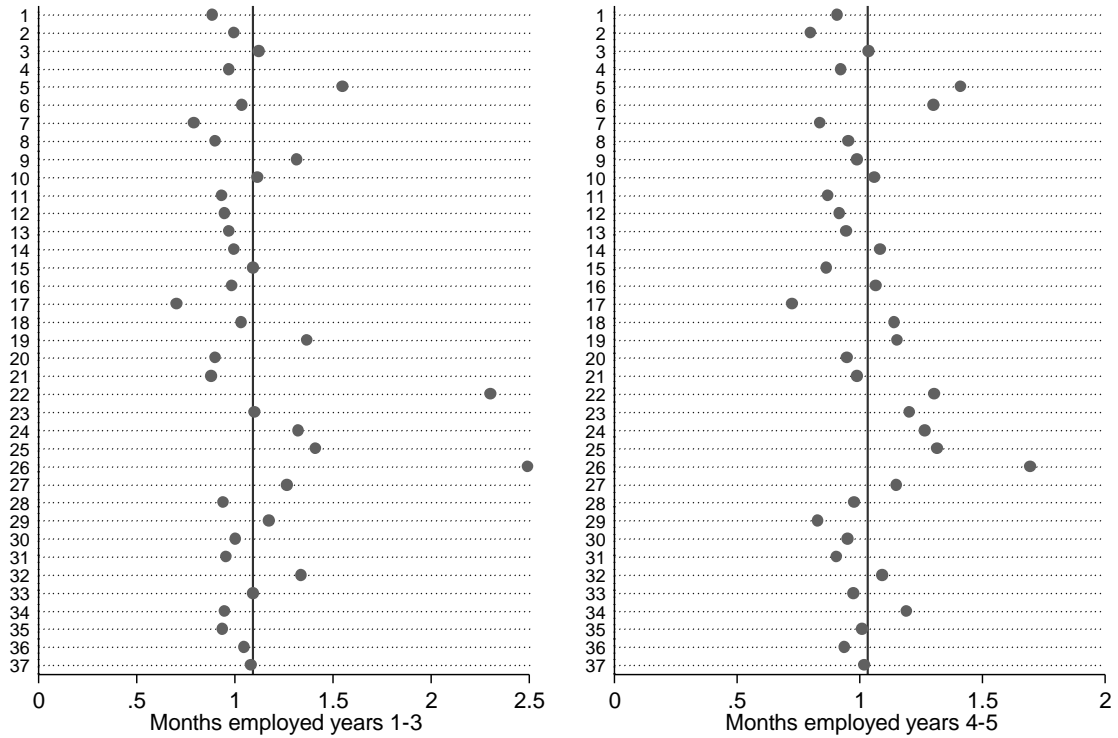
Table 1			
Customer Characteristics			
(Measured at Random Assignment (RA) Unless Otherwise Noted)			
		Cross Office Range	
	Mean	**Minimum**	**Maximum**
Customer Charactertistic			
1=Female (reference category is "male")	0.95	0.86	0.98
1=Never partnered (reference category is "previously partnered")	0.72	0.37	0.86
Qualification level (reference category is "no qualifications")			
1=O level qualification	0.48	0.24	0.63
1=A level qualification	0.22	0.09	0.64
1=Other education qualification	0.07	0.00	0.16
1=More than one child in family (reference category is "one child")	0.44	0.30	0.68
Weeks worked in past 3 years (reference category is 25-36)			
1=Worked for 12 or fewer months in the past 3 years	0.73	0.50	0.93
1=Worked for 13-24 months in the past 3 years	0.13	0.00	0.32
1=worked in the past year (reference category is "no work in year prior to RA")	0.29	0.10	0.64
Weekly earnings in most recent job in the year before random assignment (£)	27.17	12.65	61.09
Quarter of random assignment (reference category is July-September 2004)			
1=Randomly Assigned between Oct 03-Dec 03	0.10	0.00	0.29
1=Randomly Assigned between Jan 04-Mar 04	0.30	0.10	0.63
1=Randomly Assigned between Apr 04-Jun 04	0.21	0.07	0.52
Number of months receiving IS in the two years prior to random assignment	17.31	13.95	19.42
1=Youngest child in household less than 6 at random assignment (reference category is "youngest child older than 6")	0.49	0.14	0.69
Customer Age (reference category is "less than 30")			
1=Aged 30-39 years old	0.40	0.26	0.59
1=Aged 40 plus years old	0.19	0.00	0.53
1=Non-white (reference category is "white")	0.15	0.00	0.63
Index of multiple deprivation	33.87	15.43	62.28
Outcome (measured over the five-year follow-up)			
Months receiving Income Support	26.19	14.43	35.28
Months of Employment	25.29	20.34	33.40
Sample Size			
Program group = 3,348			
Control group = 3,406			
Total = 6,754			

Office Variable	Mean	Cross-Office Range	
		Minimum	Maximum
Caseload (per adviser)	29.47	3.27	109.99
Proportion of advisers working with ERA participants	0.42	0.20	0.94
Proportion of individuals advised to think long-term			
Control Group	0.30	0.00	0.74
ERA Group	0.49	0.02	1.00
ERA-Control Differential	0.19	-0.49	0.93
Proportion of individuals receiving help finding education course			
Control Group	0.43	0.00	1.00
ERA Group	0.64	0.14	0.99
ERA-Control Differential	0.21	-0.17	0.54
Proportion of individuals receiving help with in-work advancement			
Control Group	0.20	0.00	0.67
ERA Group	0.40	0.00	1.00
ERA-Control Differential	0.20	-0.69	1.00
Proportion of individuals receiving support while working			
Control Group	0.45	0.00	1.00
ERA Group	0.65	0.21	1.00
ERA-Control Differential	0.21	-0.49	1.00
Proportion of ERA participants aware of retention bonuses	0.87	0.75	1.00
Sample Size	37	37	37

Table 2
Office Characteristics
(Measured During the Follow-Up Period)

Note: Caseload variables taken from staffing form. All other variables taken from individual surveys. The office variables have been regression adjusted to control for differences in individual characteristics across offices.

		Table 3					
		Correlation Matrix of Office Characteristics					
Office Variable							
Control Group							
		1	2	3	4	5	
Caseload (per advisor)	1	1					
Proportion of individuals advised to think long-term	2	-0.0617	1				
Proportion of individuals receiving help finding education course	3	0.0047	0.3383	1			
Proportion of individuals receiving help with in-work advancement	4	-0.1565	0.4858	0.2967	1		
Proportion of individuals receiving support while working	5	0.0142	0.4318	0.2504	0.1095	1	
ERA Group							
		1	2	3	4	5	6
Proportion of advisers working with ERA participants	1	1					
Proportion of individuals advised to think long-term	2	-0.0060	1				
Proportion of individuals receiving help finding education course	3	0.0930	0.6456	1			
Proportion of individuals receiving help with in-work advancement	4	-0.1728	0.6717	0.3664	1		
Proportion of individuals receiving support while working	5	-0.1617	0.4295	0.4475	0.2758	1	
Proportion of ERA participants aware of retention bonuses	6	-0.2520	0.4375	0.2953	0.3784	0.2527	1

Note: Caseload variables taken from staffing form. All other variables taken from individual surveys. The office variables have been regression adjusted to control for differences in individual characteristics across offices.

		Years 1-3			Years 4-5	
		Coefficient	Standard Error		Coefficient	Standard Error
Table 4						
Effects of Individual Characteristics on Control Group Outcomes Over Five Years						

		Years 1-3			Years 4-5	
		Coefficient	**Standard Error**		**Coefficient**	**Standard Error**
Months on welfare						
Control group grand mean outcome (α_0)		17.91 ***	0.43		7.30 ***	0.40
Effect on control group outcome (deviation from reference group)						
1=Female		0.62 *	0.38		0.38	0.25
1=Never partnered		1.25 ***	0.20		1.27 ***	0.14
1=O level qualification		-1.87 ***	0.16		-1.12 ***	0.09
1=A level qualification		-3.46 ***	0.18		-2.00 ***	0.10
1=Other education qualification		-1.51 ***	0.28		-0.22	0.17
1=More than one child in family		-0.14	0.15		-0.39 ***	0.09
1=Worked for 12 or fewer months in the past 3 years		4.45 ***	0.36		2.11 ***	0.24
1=Worked for 13-24 months in the past 3 years		2.23 ***	0.37		0.13	0.22
1=worked in the past year		-2.52 ***	0.25		-1.07 ***	0.15
Weekly earnings in most recent job in the year before random assignment (£)		0.01 ***	0.00		0.00 **	0.00
1=Randomly Assigned between Oct 03-Dec 03		4.05 ***	0.31		1.66 ***	0.19
1=Randomly Assigned between Jan 04-Mar 04		2.25 ***	0.20		0.80 ***	0.12
1=Randomly Assigned between Apr 04-Jun 04		1.18 ***	0.21		0.82 ***	0.13
Number of months receiving IS in the two years prior to random assignment		0.29 ***	0.01		0.18 ***	0.01
1=Youngest child in household less than 6 at random assignment		1.83 ***	0.20		2.20 ***	0.14
1=Aged 30-39 years old		-1.37 ***	0.18		-0.78 ***	0.10
1=Aged 40 plus years old		-1.99 ***	0.23		-1.69 ***	0.13
1=Non-white		2.62 ***	0.27		0.76 ***	0.16
Index of multiple deprivation		0.06 ***	0.00		0.03 ***	0.00
Months employed						
Control group grand mean outcome (α_0)		13.82 ***	0.30		10.20 ***	0.22
Effect on control group outcome (deviation from reference group)						
1=Female		5.32 ***	0.48		2.11 ***	0.34
1=Never partnered		0.12	0.16		-0.58 ***	0.12
1=O level qualification		1.49 ***	0.18		1.28 ***	0.16
1=A level qualification		2.62 ***	0.23		1.18 ***	0.19
1=Other education qualification		0.14	0.28		-0.73 ***	0.22
1=More than one child in family		0.52 ***	0.14		0.41 ***	0.12
1=Worked for 12 or fewer months in the past 3 years		-3.04 ***	0.16		-1.45 ***	0.15
1=Worked for 13-24 months in the past 3 years		-1.42 ***	0.20		0.49 **	0.20
1=worked in the past year		4.01 ***	0.27		1.81 ***	0.22
Weekly earnings in most recent job in the year before random assignment (£)		-0.01 ***	0.00		-0.01 ***	0.00
1=Randomly Assigned between Oct 03-Dec 03		-1.86 ***	0.20		-1.28 ***	0.18
1=Randomly Assigned between Jan 04-Mar 04		-0.17	0.15		-0.11	0.13
1=Randomly Assigned between Apr 04-Jun 04		-0.50 ***	0.17		-0.50 ***	0.14
Number of months receiving IS in the two years prior to random assignment		-0.04 ***	0.01		-0.07 ***	0.01
1=Youngest child in household less than 6 at random assignment		-0.54 ***	0.15		-0.87 ***	0.12
1=Aged 30-39 years old		1.02 ***	0.18		0.64 ***	0.16
1=Aged 40 plus years old		2.27 ***	0.25		0.97 ***	0.21
1=Non-white		-1.81 ***	0.20		-0.72 ***	0.18
Index of multiple deprivation		-0.03 ***	0.00		-0.02 ***	0.00

Coefficent on characteristic represents *deviation* from mean outcome of omitted reference group (see Table 1).
  Thus, the coefficient of 1.25 for individuals who were never partnered on months on welfare in years 1 to 3
  implies they spent 1.25 months longer on welfare than customers who were previously partnered (not shown in table).
*Significant at 10 percent level; **Significant at 5 percent level; ***Significant at 1 percent level

		Years 1-3		Years 4-5	
		Coefficient	Standard Error	Coefficient	Standard Error
Months on welfare					
Grand mean impact of ERA (β_0)		-1.52 ***	0.35	0.04	0.43
Subgroup Impact (deviation from reference group)					
1=Female		-2.37 ***	0.40	-1.06 ***	0.28
1=Never partnered		0.38	0.25	-0.45 ***	0.16
1=O level qualification		-0.74 ***	0.23	-0.46 ***	0.14
1=A level qualification		-1.61 ***	0.27	-0.91 ***	0.18
1=Other education qualification		-0.68 *	0.38	-1.64 ***	0.21
1=More than one child in family		0.32	0.21	0.33 **	0.14
1=Worked for 12 or fewer months in the past 3 years		-1.08 ***	0.35	0.77 **	0.30
1=Worked for 13-24 months in the past 3 years		-0.21	0.43	1.93 ***	0.39
1=worked in the past year		0.67 *	0.38	0.06	0.25
Weekly earnings in most recent job in the year before random assignment (£)		-0.01 ***	0.00	0.00	0.00
1=Randomly Assigned between Oct 03-Dec 03		-1.26 ***	0.31	-0.61 ***	0.21
1=Randomly Assigned between Jan 04-Mar 04		0.45 *	0.24	0.26	0.16
1=Randomly Assigned between Apr 04-Jun 04		0.12	0.26	-0.67 ***	0.16
Number of months receiving IS in the two years prior to random assignment		0.03 **	0.02	-0.02 *	0.01
1=Youngest child in household less than 6 at random assignment		0.06	0.23	-0.53 ***	0.14
1=Aged 30-39 years old		-0.90 ***	0.24	-0.68 ***	0.15
1=Aged 40 plus years old		-2.26 ***	0.30	-1.11 ***	0.20
1=Non-white		-0.52 *	0.29	-0.05	0.20
Index of multiple deprivation		-0.03 ***	0.01	-0.01 ***	0.00
Months employed					
Grand mean impact of ERA (β_0)		0.76 **	0.39	0.12	0.22
Subgroup Impact (deviation from reference group)					
1=Female		0.93 *	0.54	1.94 ***	0.48
1=Never partnered		-1.03 ***	0.21	0.30	0.20
1=O level qualification		1.20 ***	0.28	0.45 **	0.22
1=A level qualification		2.66 ***	0.35	2.10 ***	0.30
1=Other education qualification		1.74 ***	0.46	4.02 ***	0.48
1=More than one child in family		-0.63 ***	0.19	-0.32 *	0.16
1=Worked for 12 or fewer months in the past 3 years		0.71 **	0.31	-0.10	0.25
1=Worked for 13-24 months in the past 3 years		0.46	0.34	-1.14 ***	0.25
1=worked in the past year		-1.53 ***	0.28	-1.16 ***	0.23
Weekly earnings in most recent job in the year before random assignment (£)		0.01 ***	0.00	0.01 ***	0.00
1=Randomly Assigned between Oct 03-Dec 03		0.12	0.35	0.33	0.30
1=Randomly Assigned between Jan 04-Mar 04		-1.08 ***	0.21	-0.60 ***	0.18
1=Randomly Assigned between Apr 04-Jun 04		0.01	0.26	0.35	0.22
Number of months receiving IS in the two years prior to random assignment		0.01	0.01	0.00	0.01
1=Youngest child in household less than 6 at random assignment		-0.38 *	0.22	0.21	0.19
1=Aged 30-39 years old		0.85 ***	0.27	0.50 **	0.22
1=Aged 40 plus years old		0.93 ***	0.34	1.15 ***	0.30
1=Non-white		1.74 ***	0.38	0.34	0.28
Index of multiple deprivation		0.05 ***	0.01	0.00	0.00

Notes: All characteristics are interacted with a dummy variable for being an ERA customer.

Coefficients represent deviations from mean impacts of omitted reference group (see Table 1).

Thus, the coefficient of 0.38 for individuals who were never partnered on months on welfare in years 1 to 3 is their additional impact compared to customers who were previously partnered (not shown in table).

*Significant at 10 percent level; **Significant at 5 percent level; ***Significant at 1 percent level

	Table 6						
Effects of Office Characteristics on Office Control Group Outcomes Over Five Years							
			Standard		**Interquartile Range**		
	Coefficient		**Error**		**Across Offices**		
Months on IS, years 1 to 3							
Control group grand mean (α_0)	17.91	***	0.43		16.75	to	18.74
Caseload (per adviser)	0.05	***	0.02		16.70	to	18.00
Proportion of customers advised to think long-term	0.63		3.51		17.85	to	17.95
Proportion of customers receiving help finding education course	10.62	***	3.52		16.94	to	18.99
Proportion of customers receiving help with in-work advancement	-3.06		2.91		18.21	to	17.69
Proportion of customers receiving support while working	-3.32	**	1.19		18.30	to	17.48
Months on IS, years 4 to 5							
Control group grand mean (α_0)	7.30	***	0.40		6.68	to	8.60
Caseload (per adviser)	0.03	***	0.01		6.69	to	7.34
Proportion of customers advised to think long-term	-1.94		1.86		7.50	to	7.18
Proportion of customers receiving help finding education course	5.51	***	2.77		6.79	to	7.85
Proportion of customers receiving help with in-work advancement	0.23		2.65		7.28	to	7.32
Proportion of customers receiving support while working	-0.96		0.89		7.41	to	7.17
Months Employed, years 1 to 3							
Control group grand mean (α_0)	13.82	***	0.30		12.84	to	15.13
Caseload (per adviser)	-0.02	*	0.01		14.24	to	13.79
Proportion of customers advised to think long-term	-1.40		2.32		13.97	to	13.73
Proportion of customers receiving help finding education course	-5.84	***	1.00		14.36	to	13.23
Proportion of customers receiving help with in-work advancement	10.07	***	4.22		12.85	to	14.56
Proportion of customers receiving support while working	2.57	**	1.24		13.52	to	14.15
Months Employed, years 4 to 5							
Control group grand mean (α_0)	10.20	***	0.22		9.53	to	10.89
Caseload (per adviser)	-0.01		0.01		10.37	to	10.19
Proportion of customers advised to think long-term	2.29		2.35		9.96	to	10.35
Proportion of customers receiving help finding education course	-2.89	***	0.88		10.47	to	9.91
Proportion of customers receiving help with in-work advancement	1.77		2.18		10.03	to	10.33
Proportion of customers receiving support while working	1.59	*	0.90		10.02	to	10.41
Interquartile range is the predicted outcome from the 25th percentile of the office characteristic							
to the 75th percentile.							
*Significant at 10 percent level; **Significant at 5 percent level; ***Significant at 1 percent level							

	Coefficient		Standard Error	Interquartile Range Across Offices		
Table 7						
Effects of Office Characterisitics on Office Program Impacts Over Five Years						
Months on IS, years 1 to 3						
Grand mean impact of ERA (β_0)	-1.52	***	0.35	-2.20	to	-0.24
Proportion of advisers working with ERA customers	-3.03	*	1.62	0.30	to	-0.25
Proportion of customers advised to think long-term	0.17		2.12	-0.02	to	0.02
Proportion of customers receiving help finding education course	3.62		2.27	-0.68	to	0.42
Proportion of customers receiving help with in-work advancement	-2.15		1.70	0.17	to	-0.18
Proportion of customers receiving support while working	-3.49	***	1.02	0.60	to	-0.51
Proportion of ERA customers aware of retention bonuses	-9.42	***	2.03	0.55	to	-0.50
Months on IS, years 4 to 5						
Grand mean impact of ERA (β_0)	0.04		0.06	-0.92	to	0.84
Proportion of advisers working with ERA customers	-2.50	**	1.05	0.25	to	-0.21
Proportion of customers advised to think long-term	-0.23		1.83	0.03	to	-0.03
Proportion of customers receiving help finding education course	1.96		2.02	-0.37	to	0.23
Proportion of customers receiving help with in-work advancement	0.43		1.89	-0.03	to	0.04
Proportion of customers receiving support while working	-1.64	*	0.90	0.28	to	-0.24
Proportion of ERA customers aware of retention bonuses	-6.43	***	0.50	0.38	to	-0.34
Months Employed, years 1 to 3						
Grand mean impact of ERA (β_0)	0.76	**	0.03	-0.63	to	2.12
Proportion of advisers working with ERA customers	0.53		2.29	-0.05	to	0.04
Proportion of customers advised to think long-term	-0.11		2.49	0.01	to	-0.01
Proportion of customers receiving help finding education course	-4.06	**	1.71	0.76	to	-0.47
Proportion of customers receiving help with in-work advancement	7.68	**	3.25	-0.60	to	0.63
Proportion of customers receiving support while working	3.23	*	1.75	-0.56	to	0.47
Proportion of ERA customers aware of retention bonuses	8.41		7.58	-0.49	to	0.45
Months Employed, years 4 to 5						
Grand mean impact of ERA (β_0)	0.12		0.02	-0.71	to	0.73
Proportion of advisers working with ERA customers	1.79		1.56	-0.18	to	0.15
Proportion of customers advised to think long-term	1.09		1.67	-0.13	to	0.12
Proportion of customers receiving help finding education course	-2.33	**	1.06	0.44	to	-0.27
Proportion of customers receiving help with in-work advancement	1.67		1.52	-0.13	to	0.14
Proportion of customers receiving support while working	3.73	***	1.17	-0.64	to	0.54
Proportion of ERA customers aware of retention bonuses	4.87		4.49	-0.28	to	0.26

Interquartile range is the predicted impact from the 25th percentile of the office characteristic to the 75th percentile.

*Significant at 10 percent level; **Significant at 5 percent level; ***Significant at 1 percent level

		Standard	Interquartile Range		
	Coefficient	Error	Across Offices		
Earnings 2005/6					
Grand mean impact of ERA (β_0)	309 **	144	-	-	-
Proportion of advisers working with ERA customers	-458	959	46	to	-39
Proportion of customers advised to think long-term	2184 *	1142	-258	to	246
Proportion of customers receiving help finding education course	-1447 *	844	271	to	-167
Proportion of customers receiving help with in-work advancement	1355	1168	-105	to	112
Proportion of customers receiving support while working	-35	579	6	to	-5
Proportion of ERA customers aware of retention bonuses	745	2308	-43	to	40
Earnings 2006/7					
Grand mean impact of ERA (β_0)	164	120	-	-	-
Proportion of advisers working with ERA customers	626	1041	-63	to	53
Proportion of customers advised to think long-term	-26	1197	3	to	-3
Proportion of customers receiving help finding education course	-517	860	97	to	-60
Proportion of customers receiving help with in-work advancement	2340 *	1269	-182	to	193
Proportion of customers receiving support while working	128	583	-22	to	19
Proportion of ERA customers aware of retention bonuses	4141	2591	-242	to	221
Earnings 2007/8					
Grand mean impact of ERA (β_0)	129	118	-	-	-
Proportion of advisers working with ERA customers	1081	965	-109	to	91
Proportion of customers advised to think long-term	1106	1176	-131	to	125
Proportion of customers receiving help finding education course	-1287	941	241	to	-149
Proportion of customers receiving help with in-work advancement	3193 **	1277	-249	to	263
Proportion of customers receiving support while working	155	455	-27	to	22
Proportion of ERA customers aware of retention bonuses	3749	2555	-219	to	200
Earnings 2008/9					
Grand mean impact of ERA (β_0)	-10	121	-	-	-
Proportion of advisers working with ERA customers	1605 *	834	-161	to	135
Proportion of customers advised to think long-term	1432	1553	-169	to	161
Proportion of customers receiving help finding education course	-2541 **	1003	476	to	-294
Proportion of customers receiving help with in-work advancement	2881 *	1698	-224	to	237
Proportion of customers receiving support while working	5	562	-1	to	1
Proportion of ERA customers aware of retention bonuses	4609	3037	-269	to	246

Interquartile range is the predicted impact from the 25th percentile of the office characteristic
 to the 75th percentile.
*Significant at 10 percent level; **Significant at 5 percent level; ***Significant at 1 percent level