Net impact evaluation of the Department for Work and Pensions Working Neighbourhoods Pilot

Paul Selby

A report of research carried out by the Department of Economics, University of Sheffield on behalf of the Department for Work and Pensions
Contents

Acknowledgements ........................................................................................................ vii
The Author ................................................................................................................ viii
Abbreviations and glossary ........................................................................................ ix
Abstract ......................................................................................................................1
1 Introduction ..............................................................................................................3
2 Background to the Working Neighbourhoods Pilot .............................................5
   2.1 Policy background .........................................................................................5
   2.2 What the Working Neighbourhood Pilot did differently .............................6
   2.3 Evaluation of the Working Neighbourhoods Pilot so far .............................8
3 Literature review ..................................................................................................9
   3.1 A review of the evaluation methodology literature .....................................9
   3.2 Why Propensity Score Matching is the most appropriate methodology ....13
   3.3 Other issues ..............................................................................................17
4 Data ......................................................................................................................19
   4.1 Data sources ...............................................................................................19
   4.2 Data quality issues ......................................................................................21
   4.3 Sample data ...............................................................................................21
5 Methodology .........................................................................................................25
   5.1 The participation equation .........................................................................25
   5.2 Combined difference-in-differences and Propensity Score Matching ..........26
List of figures

Figure 4.1 Number of starts on WNP by month ...........................................22
Figure 6.1 Overall net impact of WNP on total benefit rate ..........................34
Figure 6.2 Overall net impact of WNP on JSA rate .....................................35
Figure 6.3 Overall net impact of WNP on employment rate ........................36
Figure 6.4 Pattern of benefit claim history for both the treatment and comparison groups both pre- and post-WNP participation ...37
Figure 6.5 Net impact of WNP in EZ/Jobcentre Plus partnership areas on total benefit rate ...........................................................38
Figure 6.6 Net impact of WNP in Jobcentre Plus only areas on total benefit rate ................................................................................39
Figure 6.7 Net impact of WNP in EZ/Jobcentre Plus partnership areas on JSA rate .................................................................40
Figure 6.8 Net impact of WNP in Jobcentre Plus only areas on JSA rate ......40
Figure 6.9 Net impact of WNP in EZ/Jobcentre Plus partnership areas on employment rate ..........................................................41
Figure 6.10 Net impact of WNP in Jobcentre Plus only areas on employment rate ..............................................................................42
Figure 6.11 Pattern of benefit claim history for both the treatment and comparison groups both pre- and post-WNP participation in EZ/Jobcentre Plus partnership areas .....................43
Figure 6.12 Pattern of benefit claim history for both the treatment and comparison groups both pre- and post-WNP participation in Jobcentre Plus only areas ....................................43
Figure 6.13 Net impact of WNP on total benefit rate for those starting WNP while claiming JSA ...................................................45
Figure 6.14 Net impact of WNP on total benefit rate for those starting WNP while claiming IB ...................................................45
Figure 6.15 Net impact of WNP on total benefit rate for those starting WNP while claiming IS ....................................................46
Figure 6.16 Net impact of WNP on employment rate for those starting WNP while claiming JSA ...................................................47
Figure 6.17 Net impact of WNP on employment rate for those starting WNP while claiming IB ...................................................47
Figure 6.18 Net impact of WNP on employment rate for those starting WNP while claiming IS ....................................................48
Figure 6.19 Effect of WNP on total benefit rate, using individuals in selected comparison areas as the comparison group ..........50
Figure 6.20 Effect of WNP on employment rate, using individuals in selected comparison areas as the comparison group ..........51
Figure A.1 Net impact of WNP on total benefit rate, using non-treated individuals in WNP areas as the comparison group ..........58
Figure A.2 Net impact of WNP on employment rate, using non-treated individuals in WNP areas as the comparison group ..................59
Figure A.3 Pattern of benefit claim history for both treatment and comparison groups both pre- and post-participation ................60
Figure B.1 Net impact of WNP on total benefit rate in Glasgow Parkhead (EZ/Jobcentre Plus partnership area) ..................61
Figure B.2 Net impact of WNP on total benefit rate in Glasgow Hutchesontown (EZ/Jobcentre Plus partnership area) .............62
Figure B.3 Net impact of WNP on total benefit rate in Middlesbrough (EZ/Jobcentre Plus partnership area) ..................62
Figure B.4 Net impact of WNP on total benefit rate in Tower Hamlets (EZ/Jobcentre Plus partnership area) ..................63
Figure B.5 Net impact of WNP on total benefit rate in Birmingham (EZ/Jobcentre Plus partnership area) ..................63
Figure B.6 Net impact of WNP on total benefit rate in Knowsley (Jobcentre Plus only area) ..................64
Figure B.7 Net impact of WNP on total benefit rate in Hastings (Jobcentre Plus only area) ..................64
Figure B.8 Net impact of WNP on total benefit rate in Great Yarmouth (Jobcentre Plus only area) ..................65
Figure B.9 Net impact of WNP on total benefit rate in Swansea (Jobcentre Plus only area) ..................65
Figure B.10 Net impact of WNP on total benefit rate in Sheffield (Jobcentre Plus only area) ..................66
Figure B.11 Net impact of WNP on total benefit rate in Wirral (Jobcentre Plus only area) ..................66
Figure B.12 Net impact of WNP on total benefit rate in Newcastle (Jobcentre Plus only area) ..................67
I would like to thank a number of people for their help and advice during the course of carrying out the work described in this paper. Most importantly, I would like to thank Andrew Thomas for giving me sight of the SAS and Stata code that he used for his own evaluation of a different Department for Work and Pensions (DWP) programme. I would also like to thank him for his invaluable and sustained advice throughout. Without it, this project would not have been possible.

Also of great help were the input, guidance and supervision of Professor Andy Dickerson at the University of Sheffield.

Finally, I would like to thank all those at the DWP who provided suggestions and advice at various stages. In particular, I would like to thank Amanda Langdon, Allan Little, Jo Lone, Jayne Middlemass, Paul Mooney, Martin Moran, Tanya Saunders and Dan van der Schans.
The Author

**Paul Selby** has worked as an economist for the Department for Work and Pensions for five years, working on a range of issues, particularly labour market disadvantages such as poor qualifications, drug abuse and homelessness. He studied for his BA(Hons) Economics Degree at the University of Wales, Bangor, and his MSc in Economics at the University of Sheffield, where this working paper was first written as a dissertation.
## Abbreviations and glossary

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIA</td>
<td>Conditional Independence Assumption</td>
</tr>
<tr>
<td>DiD</td>
<td>Difference-in-differences</td>
</tr>
<tr>
<td>DWP</td>
<td>Department for Work and Pensions</td>
</tr>
<tr>
<td>EZ</td>
<td>Employment Zone</td>
</tr>
<tr>
<td>EZ/Jobcentre Plus partnership areas</td>
<td>Working Neighbourhoods Pilot areas run jointly by EZs and Jobcentre Plus</td>
</tr>
<tr>
<td>FDF</td>
<td>Flexible Discretionary Fund</td>
</tr>
<tr>
<td>Jobcentre Plus only areas</td>
<td>Working Neighbourhoods Pilot areas run by Jobcentre Plus only</td>
</tr>
<tr>
<td>JSA</td>
<td>Jobseeker’s Allowance</td>
</tr>
<tr>
<td>HMRC</td>
<td>HM Revenue &amp; Customs</td>
</tr>
<tr>
<td>IB</td>
<td>Incapacity Benefit</td>
</tr>
<tr>
<td>IS</td>
<td>Income Support</td>
</tr>
<tr>
<td>NDYP</td>
<td>New Deal for Young People</td>
</tr>
<tr>
<td>NOMIS</td>
<td>National Online Manpower Information System</td>
</tr>
<tr>
<td>PSM</td>
<td>Propensity Score Matching</td>
</tr>
<tr>
<td>WFI</td>
<td>Work Focused Interview</td>
</tr>
<tr>
<td>WNP</td>
<td>Working Neighbourhoods Pilot</td>
</tr>
<tr>
<td>WPLS</td>
<td>Work and Pensions Longitudinal Study</td>
</tr>
</tbody>
</table>
Abstract

This dissertation carries out a comprehensive net impact evaluation of the Working Neighbourhoods Pilot (WNP). It has done this using a combined propensity score matching and difference-in-differences (DiD) methodology, as is recommended in the literature.

The net impact estimates show that, overall, WNP had a statistically significant positive impact on the total benefit rate, centred around four per cent and persisting over the 39 weeks after participation covered by the analysis. The impacts on employment are insignificant, mainly as a result of the smaller sample sizes and the number of employment observations dropped to make this particular variable reliable.

Sub-group analysis also reveals that WNP areas operated by Employment Zones (EZs) in partnership with Jobcentre Plus, (EZ/Jobcentre Plus partnership areas), had a higher estimated impact than WNP areas operated by Jobcentre Plus only. Some individual EZ/Jobcentre Plus partnership areas had positive impact estimates on the total benefit rate centred around as high as 20 per cent. This is despite the fact that the data appears to show EZ/Jobcentre Plus partnership areas worked with harder to help customers than Jobcentre Plus only areas, on average. It is not entirely clear what drove these higher estimated impacts. However, feedback from staff in the different sites suggests that EZ/Jobcentre Plus partnership areas were better able to use the flexibility allowed in WNP.
1 Introduction

The WNP was introduced in April 2004 and ran for two years in 12 pilot areas. The selection of these areas was based upon the fact that they had very high rates of worklessness and the intention was to help improve the employment rate and therefore, the overall economic standing of these areas. A qualitative evaluation was published in early 2007 but this only included a very basic overall estimate of net impact. As a result, this dissertation aims to produce more comprehensive and reliable estimates for the net impact of WNP. It also intends to carry out sub-analysis to find out whether impact varied by benefit type, delivery mechanism and individual pilot area.

The structure of this dissertation is as follows: Chapter 2 outlines in more detail what WNP was and the background to the pilot. Chapter 3 then discusses the evaluation literature and how reliable programme evaluation methodologies have developed. Chapter 4 moves on to describe the datasets and variables used in the analysis, followed by Chapter 5, which describes the methodology used in detail. Chapter 6 reports the results of the analysis, including the main net impact estimate and the various sub-analysis. It also includes a discussion of these results. Finally, Chapter 7 summarises and concludes the whole dissertation.
2 Background to the Working Neighbourhoods Pilot

2.1 Policy background

Over the past 15 years the UK has experienced a significant increase in employment, with the working age employment rate reaching 74.3 per cent in April 2007, up from a low of 70.3 per cent in April 1993. Policymakers have attributed this success to both a benign macroeconomic climate with less uncertainty, as well as to active labour market policies such as the New Deals, which offer a range of opportunities to encourage jobseekers back into work.\(^1\) However, despite success at an aggregate national level, many local areas still suffer from high levels of worklessness, with its associated deprivation and social problems. This tends to be concentrated in inner city areas, certain isolated coastal and rural areas and areas previously dependent on mining and traditional industries.

Despite these general geographical tendencies, however, many of the specific causes of worklessness vary by location, from problems with seasonal employment to entrenched cultures of benefit dependency. The UK Government, therefore, felt that a new and different approach was needed.

As a result, the WNP began on 26 April 2004 after first being announced by the Chancellor of the Exchequer in late 2002. This project intended to test innovative new approaches that aimed to tackle high levels of worklessness in local areas. The pilots ran for two years in 12 localities at a cost of £42 million.\(^2\) Each locality had a population of between 4,000 and 5,000 people in 2003 and all were specifically selected for their very high levels of worklessness and benefit dependency. In

---

\(^1\) DWP (2007).

\(^2\) The proposed outlay over the two years was initially £77 million but not all of this was spent.
fact, each of the chosen areas had between 35 per cent and 50 per cent of their working age population classed as workless in 2003. They are listed in Table 2.1.

### Table 2.1 Working Neighbourhood Pilot areas

<table>
<thead>
<tr>
<th>Local Authority District (LAD)</th>
<th>Pilot area (2003 ward boundaries)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glasgow*</td>
<td>Parkhead*</td>
</tr>
<tr>
<td>Glasgow*</td>
<td>Hutchesontown*</td>
</tr>
<tr>
<td>Middlesbrough*</td>
<td>Thorntree*</td>
</tr>
<tr>
<td>Birmingham*</td>
<td>Aston*</td>
</tr>
<tr>
<td>Tower Hamlets*</td>
<td>Lansbury*</td>
</tr>
<tr>
<td>Newcastle</td>
<td>Monkchester</td>
</tr>
<tr>
<td>Knowsley</td>
<td>Northwood</td>
</tr>
<tr>
<td>Wirral</td>
<td>Birkenhead</td>
</tr>
<tr>
<td>Sheffield</td>
<td>Manor</td>
</tr>
<tr>
<td>Swansea</td>
<td>Penderry</td>
</tr>
<tr>
<td>Great Yarmouth</td>
<td>Regent (Nelson)</td>
</tr>
<tr>
<td>Hastings</td>
<td>Castle</td>
</tr>
</tbody>
</table>

* = EZ

It is worth noting here that five of the 12 pilot sites were actually located in Jobcentre Plus districts that operated as an EZ. EZ are areas in which certain aspects of Jobcentre Plus provision is private sector led. For the five WNP sites operating in EZ areas, Jobcentre Plus and the private contractor managed the pilots jointly.

### 2.2 What the Working Neighbourhood Pilot did differently

The changes that took place in the pilot areas, which were different to the standard national package of help, were threefold:

- a more intensive intervention regime with more support;
- retention payments to individuals if they gained a job and stayed in continuous employment for certain lengths of time; and
- extra local funding for innovative community-based types of support.

For jobseekers, the standard JSA intervention regime comprises a Work Focused Interview (WFI) with a Jobcentre Plus Personal Adviser at the start of a claim, followed

---

3 In standard EZ areas, individuals claiming Jobseeker’s Allowance (JSA) aged below 25 who have previously experienced spells on the New Deal for Young People (NDYP) and individuals claiming JSA aged 25 or above reaching 18/21 months of unemployment, the EZ providers take over responsibility for these individuals and are incentivised to move them back into work as quickly as possible. Lone parents in EZ districts are also able to voluntarily access EZ provision.
Background to the Working Neighbourhoods Pilot

by fortnightly job search reviews up until they reach six months unemployment. If a jobseeker reaches six months of unemployment, (18 months if aged 25 or over), they are then required to attend a further WFI, before entering the New Deal, and the package of support that entails. In WNP sites, however, the changes introduced meant that jobseekers were required to attend weekly job reviews if and when they reached a claim length of 13 weeks and mandatory New Deal interventions also started at that point. All of these types of intervention, both nationally and in WNP sites, were mandatory. Failure to attend interventions can result in a sanction of benefits withdrawal, if no good cause is given.

For individuals with a new or repeat Incapacity Benefit (IB) claim on or after 26 April 2004, and all Income Support (IS) customers, these groups were required to attend a more regular programme of mandatory WFIs as part of WNP. At these WFIs, individuals were offered additional help and support from WNP, although take up of this extra support was voluntary.

All individuals who gained a job and stayed in that job were entitled to retention payments in two stages. Individuals were entitled to £500 if they stayed in work for 13 weeks and a further £750 if they reached 26 weeks of continuous employment.

Finally, each pilot site was allocated a Flexible Discretionary Fund (FDF) of £1 million in each of the two years of the pilot. Of this £1 million annually, 80 per cent was allocated to the Community FDF, with funds being spent on infrastructure or activities, in order to facilitate the movement of more individuals back into work. The remaining 20 per cent was allocated to an Individual FDF, whose funds were to be used to help with specific customer problems and barriers to work. A full description of the type of activities that the FDF was used to fund can be found in Dewson et al. (2007).

4 It should be noted that the process of weekly signing after week 13 of a JSA claim was rolled out across all districts close to the end of WNP.

5 It is worth noting here that no noticeable difference in non-compliance was observed as a result of the increased interventions associated with WNP.

6 IB customers claiming prior to 26 April 2004 were not required to take part in WNP and any participation was entirely voluntary.

7 At the time, the basic lone parent IS intervention regime required individuals to attend a mandatory WFI at the start of a new claim, again at 26 weeks and 52 weeks and once a year thereafter. For those claiming IB, the basic intervention regime at the time only required individuals to attend a WFI at the very start of their claim.

8 Community FDF was spent on a wide variety of activities and projects, from purchasing/renting new community buildings, through the organisation and sponsorship of local community events, to funding and organising specialist training courses.

9 Examples of Individual FDF use include the purchase of a suit for an interview, travel expenses, debt settlement and individual training.
2.3 Evaluation of the Working Neighbourhoods Pilot so far

Throughout the design of WNP, it was always intended that there would be both qualitative and quantitative evaluation when the pilot ended, in order to assess its net impact and effectiveness, although the specific method of evaluation was never planned. As such, when selecting the pilot areas, comparison sites for each of the pilot areas were also identified. In theory, this would enable researchers to pick individuals in these chosen similar areas, in order to construct a comparison group with which to compare WNP participant outcomes. Lone (2006d) followed this approach by taking the percentage of WNP participants in WNP areas who gained jobs over a year and then compared this to the percentage of all workless benefit claimants who gained jobs over the same year in the chosen comparison sites. This led to a result of job outcomes nine percentage points higher in WNP areas, a figure also reported and published in Dewson et al. (2007). In theory, this is a figure that shows the net impact or additionality of WNP, relative to the standard regime.

This figure is misleading, however. Although Lone (2006c) found no significant differences in overall aggregate characteristics between all individuals in both pilot and comparison sites, the impact estimate in Lone (2006d) failed to control for differences between WNP participants and the group of workless individuals in the selected comparison areas. As such, WNP participants could be more (or less) job ready and this would, therefore, lead to overestimated (or underestimated) effectiveness of the pilot. Indeed, Lone (2006c) reported that, based upon the fact that they spent more time on welfare benefits in the previous four years, WNP participants may actually be further from the labour market than her chosen comparison group.

It is this failure to control for differences in characteristics that this paper aims to explore. Using a combined Propensity Score Matching (PSM) and DiD methodology, it intends to get as close to replicating experimental data as it is possible to achieve with non-experimental approaches. This will enable more robust estimates of the effectiveness of WNP to be computed. The estimates will help to inform the design of future employment policy in local areas.

---

10 Before the pilot areas were finally chosen, a choice of up to three potential areas within each chosen Jobcentre Plus district was given. The comparison areas were selected from those areas not ultimately chosen to be a pilot area.
3 Literature review

3.1 A review of the evaluation methodology literature

The basic problem of programme evaluation is that although the outcomes of participants are observed, it is not possible to observe what their outcomes would have been had they not participated. This is the so called ‘evaluation problem’. This is demonstrated formally in equation (1):

\[ \alpha_T = E(Y^T | d = 1) - E(Y^C | d = 1) \] (1)

Here, \( Y^T \) is the outcome having been treated, while \( Y^C \) is the outcome if untreated, (i.e. the control or comparison group), with \( d = 1 \) indicating that an individual was in the treatment group, and \( d = 0 \) if otherwise. The problem arises because \( E(Y^C | d = 1) \) cannot be observed directly, since participants cannot simultaneously be non-participants.

In a perfect world, all policy evaluation attempting to find a net impact would take place using experimental data derived from the process of random assignment. Here, a randomly chosen group of individuals receives a treatment thought to make an improvement to their lives and the rest of the individuals do not receive treatment (or receive a placebo). This latter group of individuals is called the control group. With a large enough sample size, under these circumstances there should be no reason why the aggregate characteristics, both observed and unobserved, of the treatment and control groups should be any different from each other. In other words, any differences in outcomes that are observed between the two groups after the time of treatment should only be able to be attributed to the treatment. Usefully, randomly assigning a group of potential participants to be untreated enables \( E(Y^C | d = 1) \) in equation (1) to be measured indirectly, assuming that the random assignment was truly random.

However, this is not a perfect world, and for a variety of reasons, experimental methodologies such as random assignment are not used very often. These are described by Stafford et al. (2002).
Firstly, there are ethical issues to consider. Randomly withholding potentially beneficial treatment from individuals is often perceived to be unethical. Secondly, the costs can be prohibitive. To ensure that the random assignment process is truly random, sophisticated procedures and checks often need to be put in place. Thirdly, it can be difficult to obtain as much information about the control group as the treatment group. To have equivalent data, the same questionnaire needs to be given to the control group, even though they did not take part. This can be practically difficult to ensure, since non-participants will be less inclined to answer questions about something they did not actually take part in. Fourthly, there are issues with contamination. Even the most sophisticated procedures can sometimes fail to prevent individuals randomly assigned to the control group from signing up again and being reassigned to the treatment group. Finally, there can also be issues about alternative treatments that the control group receive in the absence of the treatment of interest. This is not necessarily a problem if the result of interest is simply the effectiveness of the treatment compared to alternative treatments already available. However, there can be serious problems with ‘substitution bias’\(^{11}\) and ‘disruption bias’.\(^{12}\)

For a combination of the above reasons, the WNP was not designed to enable the use of an experimental methodology. Indeed, given that the pilots were meant to be testing innovative community-based approaches, it would probably have been counterproductive to exclude a certain proportion of that community from the treatment in order to obtain a control group.

Therefore, to be able to evaluate the effectiveness of WNP, this paper needs to look at non-experimental methodologies, to see if any are relevant or practicable. The literature on this subject has developed quite considerably over the past 20 years, particularly in the labour market intervention arena.

It is important to note that although experimental methodologies may be the ideal, non-experimental methodologies can actually produce comparable results, if the right dataset variables are available and the correct evaluation techniques are used. Important studies in this area use non-experimental techniques on experimental data in order to establish whether the same or similar results are achieved.

Perhaps the most famous and well quoted of these types of studies is by LaLonde (1986). It compared the results of a pilot employment programme that used

\(^{11}\) Substitution bias occurs when the control group receive help from staff to receive similar alternative treatments to the randomised treatment of interest.

\(^{12}\) Disruption bias arises when staff behaviour changes as a result of the treatment existing. For example, they may attempt to manipulate the randomisation process to help favoured individuals originally assigned to the control group, or overenthusiastically help the treatment group more than was proposed by the experiment.
random assignment in the USA, to the results from a variety of non-experimental techniques used on the same pilot programme. Importantly, impact estimates from the non-experimental techniques varied significantly from the ‘true impact’ estimated by the experiment. The author also noted that conventional specification tests on the non-experimental results do not appear to be able to identify methods that get closer to the true impact estimate. The conclusions from this influential paper cast doubt over whether non-experimental methodologies would ever be able to replicate the results from experimental data. However, various studies have revisited the results of LaLonde. For example, Heckman et al. (1997) investigated sources of bias in non-experimental estimators that influence the difference between experimental and non-experimental results. They identified four potential sources:

- differences in unobserved characteristics of the treatment and control groups;
- differences in observed characteristics\(^{13}\);
- differences in outcome and characteristic data collection methods for the treatment and control groups;
- differences in the economic environment experienced by the treatment and control groups.

The authors went through in detail how the problems outlined in the latter three bullet points can be alleviated through particular methods of matching. Importantly, they showed how the ‘selection bias’ associated with bullet point one is a relatively small part of the overall total bias if matching is successful.\(^ {14}\) They did this by decomposing the total bias into three categories and demonstrated the size of each type of bias using their experimental data.

The authors went on to criticise the results from LaLonde (1986) on both of the latter two bullet points above: Firstly, LaLonde drew data for his comparison group from national surveys not associated with the pilot of interest. The questionnaires associated with both sets of data, therefore, used different survey techniques. Secondly, the comparison group were from different labour markets, something that Heckman et al. (1997) showed to introduce significant bias. The results from the preferred non-experimental estimator in Heckman et al. (1997) actually produced estimates close to those produced in the experimental results.\(^ {15}\) While the authors stated that the results may not be replicable in other datasets and that some bias remained, the techniques they used have gone a long way to

\(^{13}\) In other words, a lack of ‘common support’. This is discussed in more detail in later sections.

\(^{14}\) The overall total bias is the difference between the impact estimates from experimental and non-experimental methods.

\(^{15}\) Their preferred estimator is a regression adjusted semi-parametric conditional DiD estimator.
demonstrating that non-experimental techniques can produce indicative results at the very least. Dehejia and Wahba (1999) also produced results that suggested the same. They returned to the experiment and dataset used by LaLonde (1986) and carry out propensity score methods which produced much closer impact estimates to the true experimental result than any of the estimates produced by LaLonde.

As an aside, it is worth pointing out here that there is an implicit assumption throughout this section of the literature that the experimental results really do show the ‘true impact’. In fact this may not be the case, especially if substitution or disruption bias occurs. It can be that the researcher is oblivious to the fact that this bias occurred, as the bias can be very subtle. For example, the LaLonde experimental results could have been biased in this way, potentially explaining the difference between the experimental and non-experimental results. However, most researchers would agree that an ‘adequately’ performed experiment is likely to suffer less selection bias than non-experimental estimators that are unable to fully control for unobserved characteristics such as ability and motivation. In the absence of any evidence to the contrary, it seems sensible to assume that the results of an adequately performed experiment are the closest estimate of the true impact.

Perhaps more importantly and most relevant to the analysis carried out later in this current paper, Thomas (2006) used the same datasets and a similar methodology to those proposed later in Chapters 4 and 5. The author used UK administrative data to compare non-experimental estimators of programme impact to the unbiased experimental estimate. Where Thomas differed from LaLonde (1986) is that specification tests are carried out on a whole range of different non-experimental estimates, prior to the estimation of the experimental estimate. This was in order to avoid post-hoc justification for why certain estimates were dropped. Interestingly, most of the more extreme estimates were dropped during the specification tests that Thomas carried out, leaving a more limited range, all of which were quite close to the true programme impact. The author gives advice on how to choose the right estimator, depending on the type of data available to the researcher. His preferred estimator is a propensity score matched panel DiD estimator.\textsuperscript{16} The reason why this is preferred to simple PSM on its own is because the panel techniques remove any remaining fixed differences, (e.g. geographical), between the treatment and comparison groups. By taking differences, any characteristics that are the same in both time periods are ‘differenced out’ or removed. Any changes in the dependent variable between the two time periods can then be attributed to any characteristics that did change over time.\textsuperscript{17}

\textsuperscript{16} This preferred estimator is similar to that preferred by Heckman \textit{et al.} (1997).

\textsuperscript{17} In effect, DiD removes both unobservable fixed individual effects and common macroeconomic trend effects. But this relies on the critically important assumptions of no composition changes within each group and common time effects across groups. The latter assumption is very hard to prove.
Also of interest to this paper is the advice about constructing the comparison group. Thomas (2006) demonstrated that using the largest possible group of non-participants is preferable to selecting the comparison group from labour markets close or similar to the labour market of the treatment group.\(^ {18}\) This is in line with evidence from Heckman et al. (1997), as well as Dehejia and Wahba (1999), who argued that taking sub-samples to select a comparison group should not be used with PSM.

### 3.2 Why Propensity Score Matching is the most appropriate methodology

There are various non-experimental methods used and discussed in the literature, but not all are appropriate, and many have been shown to include strong assumptions. Blundell and Costa Dias (2006) provide a comprehensive outline and discussion of potential methods. Their paper first discusses methods for use on a single cross-section of data, in case that is the only type of data that is available. Since the analysis carried out in this current paper has access to longitudinal data, it makes cross sectional methods inappropriate, as the aforementioned authors argue that panel methods are always preferred.\(^ {19}\) The authors then move on to discuss DiD techniques. This is a relatively simple methodology that requires observations both before and after treatment for both participants and non-participants. Changes in earnings or employment (for example) can then be observed for both groups of individuals, and then compared. Formally this can be presented in equation (2):

\[
\hat{\alpha}_{\text{DID}} = (\bar{Y}_T^T - \bar{Y}_T^C) - (\bar{Y}_T^C - \bar{Y}_T^C)
\]

where \(\bar{Y}_T^T\) and \(\bar{Y}_T^C\) represent mean outcomes for the treated and non-treated respectively, and \(t_1\) and \(t_0\) represent the two different time periods. In effect the Lone (2006d) analysis discussed previously already has observations for time period \(t_1\). This analysis could quite simply have been extended by taking observations in time period \(t_0\) to find an estimate of \(\hat{\alpha}_{\text{DID}}\). An extension of the Lone (2006d) analysis in this way would remove any bias associated with fixed district effects. For example, WNP areas may have been performing worse than other areas prior to the introduction of WNP. This would potentially mean that the nine percentage point difference estimated by Lone was an underestimate of the true impact. However, simple DiD estimators such as the one described above do not control for differences in the types of individual in WNP and other areas.

---

\(^{18}\) This is because the gains from being able to find closer matches from a larger comparison group are greater than the gains through sub sampling from a labour market with similar trends or characteristics. This is especially the case when you include area characteristic variables in the participation equation, as Thomas (2006) does.

\(^{19}\) As is discussed elsewhere in this paper, panel datasets enable researchers to control for fixed unobservable characteristics.
The simple DiD estimator described above can be calculated in an alternate way, using a regression of the outcome on a dummy for before and after, a dummy for treatment and control and an interaction of these two dummies. It is the coefficient on this latter interaction term that gives the same $\hat{\alpha}$ as estimated previously. By adding further controls to the regression equation, it effectively holds these chosen controls constant and creates a more robust ‘regression adjusted’ DiD estimator.

While this is quite a useful estimator, the literature outlines one main criticism that makes this technique less attractive. Although the regression adjusted DiD estimator does control for various chosen characteristics, it cannot account for the fact that the treatment and comparison groups may, in fact, be composed of a very different range of those characteristics. For example, the treatment group may be made up of individuals aged 18 to 35, whereas the comparison group may include individuals aged 30 to 65. Although the age ranges overlap slightly, they are very different, something which basic DiD cannot control for.

This is where matching techniques are preferred. As Bryson et al. (2002) discussed, matching estimators have the advantage of highlighting the above problem of what the literature calls ‘common support’. The area of common support is the area where observed characteristics of the treatment and comparison group overlap. This is important because it ensures that the characteristics of the treatment group are also seen in the comparison group. Where there is little or no overlap, traditional regression-based estimators extrapolate outside the area of common support, based upon the functional form of the regression. This is not always an ideal assumption to make, particularly if the size of the common support is small.

At first glance, it is intuitive to think that if this paper is going to use a matching technique to construct the comparison group, the analysis would want individuals in the comparison group identical in all characteristics to individuals in the treatment group or at least identical in terms of all the observed characteristics. However, this is very difficult to achieve in practice. Attempting to increase the reliability of the match by increasing the number of characteristics matched upon leads to the ‘curse of dimensionality’, in that the chances of finding an identical match reduce. Importantly, an influential paper by Rosenbaum and Rubin (1983) showed that it is possible to match on a single index that reflects the probability of participation, rather than every single characteristic, and still achieve consistent results. It is this index that is termed the ‘propensity score’ and which gives the

---

20 In the specific case of PSM the propensity score defines the area of common support.
21 Results become less robust the more you are forced to extrapolate.
22 The ‘curse of dimensionality’ is a term used to describe the problem caused by the exponential increase in volume associated with adding extra dimensions.
name given to ‘PSM’. This is a powerful result that enables researchers to use the largest number of appropriate and available variables in the matching equation, with much less of a requirement to find a close match.

In PSM, the enforcement of common support is most often carried out by checking whether the distribution of propensity scores for the treatment group is within the distribution of propensity scores for the comparison group. If not, the treatment group observations outside the comparison group distribution are dropped. Trouble arises if large proportions of the sample are eliminated by the enforcement of common support, as it can start to reduce the policy relevance of any results. In other words, just because results show that a policy might have been effective for individuals with common support, it may not have been the case for those eliminated with a lack of common support and vice versa. As Bryson et al. (2002) point out, researchers need to note the proportion of their sample that are eliminated with a lack of common support, the types of individuals within that subgroup and how it might affect their results.

PSM has a further advantage in that it does not make any assumptions about the functional form of the outcome equation. While regression methods often impose functional form restrictions, usually linear, PSM is semi-parametric and, therefore, allows the estimated function to fit the data more precisely.

The main disadvantage is that PSM is very ‘data hungry’. In other words, to make sure that close matches are found, as well as to estimate precise propensity scores, both the number of variables and the number of observations need to be large. The sample size in particular should not be an issue with the datasets being used by this paper. Where there may be an issue is in variants of the main analysis, where the analysis seeks to draw the control group from within the same individual pilot site as treated individuals. The number and closeness of the matches here may be limited and it is a problem that needs to be monitored. However, as noted above,

---

23 Alternate methods include: (1) calliper matching, where treatment observations outside a specified propensity range of any comparison observation are dropped; (2) removing observations where the kernel density is below a certain specified density, (if kernel matching is being used).

24 Heckman et al. (1997) note that the lack of common support is a key source of bias that accounts for differences between experimental and non-experimental estimators.

25 It is worth noting here that PSM relies upon the correct specification being chosen. However, there are no formal tests to check this, unless experimental data is available. And if experimental data was actually available, then PSM is not needed.

26 The WNP database contains records of 23,734 people who started WNP and the preferred comparison group is drawn from a database that initially contains the benefit histories of 13.7 million individuals. The final comparison group will be much smaller, after selecting for relevant dates.
although this paper will be carrying out analysis with these smaller sub-samples for a comparison group, these are unlikely to be the preferred specification. Both Thomas (2006) and Dehejia and Wahba (1999) found that the comparison group should be drawn from the largest possible sample of non-participants. In other words, if using local labour market conditions as a control during the propensity score estimation and panel DiD to control for fixed observed and unobserved heterogeneity, then sub-sampling for ‘similar’ comparison group labour markets is not necessary.

Within the PSM literature, there is still debate about the preferred method of matching individuals in the treatment group to individuals in the comparison group. The most often used non-parametric methods are nearest neighbour matching and kernel density matching. Nearest neighbour matching methods look at the propensity scores for each and every individual in the treatment group, and identifies an individual in the comparison group who is the nearest match, (i.e. has the closest propensity score). Variations within this methodology choose to:

- use averages from a specified number of nearest neighbours (rather than just one nearest neighbour);
- use individuals in the comparison group as a match only once (i.e. without replacement), instead of allowing them to be matched with a number of different individuals in the treated group;
- impose a limit for the closeness of the match (i.e. if no individual in the comparison group falls within a chosen range for closeness of match with a treated individual, the treated individual will remain unmatched.

On the other hand, with kernel matching, the outcome of the treated individual is matched to a weighted average of the outcomes of possibly all the non-treated units, with the weight set by the distribution of the kernel. Those in the comparison group with a very closely matched propensity score to an individual in the treatment group get a higher weighting than those with a poor match. The distribution or shape of the kernel is decided by the data itself. Of high importance here is choosing the size of the bandwidth used to construct the kernel. Choosing a wide bandwidth includes more comparison group observations and so more variation in the estimation. This results in a more smoothly shaped distribution, with less variance and, therefore, more precise estimates. However, the cost of this is increased bias. On the other hand, a narrower bandwidth results in smaller bias and a less smooth distribution, and so is less precisely estimated.

As is noted by Bryson et al. (2002), the use of PSM does not necessarily alleviate the problem of being able to find a match. For example, it is still possible that there will not be any individual in the comparison group with a ‘similar’ propensity score to individuals in the treatment group.
3.3 Other issues

A further issue to be aware of that is not specific to PSM methodologies is the substitution effect. This is an important general equilibrium effect. One of the major assumptions made by most labour market evaluations, both experimental and non-experimental, is that any success achieved by the treatment in getting more of the treated into work does not displace other individuals who may have got that specific job otherwise. However, very few studies even mention substitution, let alone try to control for it. There is no evidence that WNP had substitution effects, either reported in Dewson et al. (2007) or anecdotally. It is, therefore, likely that the substitution effects are small and can be discounted.

It is also important to be aware of selection bias. This is when providers of the treatment intentionally or unintentionally pick individuals who are easiest to help. This would not necessarily be a problem if it was mandatory for all of the eligible population to receive the treatment but this is clearly not the case for those individuals claiming IS or IB in WNP areas. For claimants of these two welfare benefits, most pilot activities other than the mandatory WFIs were voluntary for them. In fact, Dewson et al. (2007) showed that even where WNP was mandatory for claimants of JSA, only 60 per cent recorded a start on WNP, across all pilot sites. This is potentially a serious problem for the analysis, as it could be that there are some unobserved rules or individual characteristics that staff are using to select participants from non-participants, within the JSA population. If selection was occurring in this way, controlling for observable characteristics such as benefit history, can potentially remove some of this bias. Even so, there will be unobserved residual effects that continue to bias results. Thankfully, analysis shows that most of the 40 per cent non-participation amongst the JSA population came at the end of the two years of the pilot. This is likely to have occurred for two reasons: Firstly, Jobcentre Plus district managers and staff correctly anticipated that funding for the pilot was due to end and reduced their focus and attention on enrolling new participants. Secondly, Jobcentre Plus nationally were suffering a large reduction in staff numbers throughout the last few months of the pilot. All Jobcentre Plus services were under pressure during this time, not just WNP. As a result of these two points, the last three months of starts on WNP have been excluded from the analysis presented in this paper. Also, because all pilot schemes take a while to bed in and settle down, the first three months of starts on WNP have been excluded from the analysis as well. In summary, the fact that nearly all relevant characteristics have been included in the participation equation and the panel data allows the removal of any remaining fixed differences, should mean that selection bias is minimised.

As has been alluded to above, it is very important to identify the correct functional form as well as the variables that specify the participation equation, which itself

---

28 This will be the case if we assume that benefit history is correlated with individual selection rules.
determines estimates of the programme effect. For example, Heckman et al. (1997) showed that omitting important variables such as recent individual labour market history and local labour market performance could seriously bias results.

Bryson et al. (2002) explained why this might be the case. They noted how the use of pre-programme unemployment duration, which they suggest is correlated with unobserved motivation, could help to capture some of the motivation effect, and help to minimise the bias associated with unobservable factors. Indeed, in his analysis of the reliability of PSM, Thomas (2006) demonstrated that the use of previous unemployment history brings the estimate of programme impact much closer to the estimate from experimental data than estimates that exclude previous unemployment history.

Related to all of the above is what Blundell and Costa Dias (2006) called the most important assumption of PSM, the Conditional Independence Assumption (CIA). This assumption states that the outcomes of the non-treated are independent of their participation status, after controlling for observables. This allows the counterfactual outcome for the treatment group to be inferred. But to be credible, a rich dataset is needed, with the added assumption that the unobservables have no effect. This is a strong assumption, but given what was discussed earlier about labour market history reflecting aspects of motivation, the bias associated with unobserved characteristics may be limited.

Another important point that the literature makes is about the so-called ‘Ashenfelter’s dip’.29 This description is given to the observation of reduced employment or earnings for the treatment group in the time period just before treatment. It occurs because participants anticipate the upcoming treatment and reduce their jobsearch as a result. If unaccounted for, it can lead to overestimating the impact of treatment. Similarly, some individuals eligible for treatment may refuse treatment if they are expecting a job offer in the near future, biasing results in the other direction. Most researchers avoid these two effects by ignoring the time periods immediately before and after treatment when presenting their results.30

---

29 This comes from findings observed in Ashenfelter (1978).
30 Often, this is a matter of judgement, based upon observation of any pickup or drop-off in pre- and post-event outcomes, shown graphically.
4 Data

4.1 Data sources

Most of the data used in this analysis is drawn from DWP and Jobcentre Plus databases. These databases include individual client characteristic information, mainly drawn from adviser records during client interviews. Included in this data are specific variables taken from the WNP database. The databases include variables such as start and end dates on WNP, as well as variables recording the WNP site. The databases also include every individual’s benefit history going back to 28 June 1999.

Using data taken from National Online Manpower Information System (NOMIS), this paper constructs variables for population density and local labour market conditions.\(^{31}\) These hold self-explanatory information for each individual based upon the local conditions when each individual participated (or non-participated) in WNP.

Each individual’s recorded employment history is drawn from the Work and Pensions Longitudinal Study (WPLS). This dataset is originally sourced from HM Revenue & Customs (HMRC) tax system data, and includes employment records for all individuals who have ever claimed any welfare benefit since 28 June 1999.\(^{32}\)

The specific list of variables used and how they were constructed is as follows:

- age, (and age squared);\(^{33}\)
- gender;
- ethnicity;
- marital status;

---

\(^{31}\) NOMIS at http://www.nomisweb.co.uk/

\(^{32}\) This information is received by HMRC when employers send them completed P45 and P46 forms.

\(^{33}\) Age squared is included as the literature extensively shows a non-linear relationship between employment and age.
• disability;
• labour market history\textsuperscript{34};
• local labour market conditions\textsuperscript{35};
• population density;
• lone parenthood;
• month of claim start/participation start on WNP.

All this data is then merged together based upon each individual’s unique identifier. Using the merged data enables the replication of the basic table of gross job outcomes for those starting WNP, by district. This is shown in Table 4.1.

### Table 4.1  Starts and job outcomes through WNP, by district

<table>
<thead>
<tr>
<th>Site</th>
<th>Starts</th>
<th>Jobs gained</th>
<th>Jobs as a % of starts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birmingham</td>
<td>1,190</td>
<td>400</td>
<td>34</td>
</tr>
<tr>
<td>Great Yarmouth</td>
<td>2,790</td>
<td>940</td>
<td>34</td>
</tr>
<tr>
<td>Hastings</td>
<td>1,730</td>
<td>490</td>
<td>28</td>
</tr>
<tr>
<td>Glasgow (Hutchesontown)</td>
<td>2,110</td>
<td>720</td>
<td>34</td>
</tr>
<tr>
<td>Knowsley</td>
<td>2,310</td>
<td>870</td>
<td>38</td>
</tr>
<tr>
<td>Middlesbrough</td>
<td>2,310</td>
<td>700</td>
<td>30</td>
</tr>
<tr>
<td>Newcastle</td>
<td>1,940</td>
<td>770</td>
<td>39</td>
</tr>
<tr>
<td>Glasgow (Parkhead)</td>
<td>1,840</td>
<td>760</td>
<td>41</td>
</tr>
<tr>
<td>Sheffield</td>
<td>1,700</td>
<td>700</td>
<td>41</td>
</tr>
<tr>
<td>Swansea</td>
<td>2,150</td>
<td>580</td>
<td>27</td>
</tr>
<tr>
<td>Tower Hamlets</td>
<td>1,570</td>
<td>500</td>
<td>32</td>
</tr>
<tr>
<td>Wirral</td>
<td>2,090</td>
<td>790</td>
<td>38</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>23,730</strong></td>
<td><strong>8,220</strong></td>
<td><strong>35</strong></td>
</tr>
</tbody>
</table>

NB. Table replicated from Dewson et al. (2007), numbers rounded to the nearest 10.

\textsuperscript{34} This is constructed as a string variable, as is used by much of the literature, including Thomas (2006). For example, if an individual spent a majority of a three-month period claiming welfare benefits, this is recorded as 1 or 0 if otherwise. The string variable is 12 characters long, reflecting three years of benefit and employment history.

\textsuperscript{35} Both the JSA claimant count unemployment rate at the start of a benefit claim and the change in the JSA claimant unemployment rate over the two months prior to the start of a claim were used.
4.2 Data quality issues

There are well documented issues regarding problems with the WPLS data. For example, a small proportion of employment spells have arbitrary start and end dates coinciding with the tax year, rather than the actual dates. There can also be problems when merging the employment and benefit spells data, as the spells can sometimes overlap. These, and a few other minor issues, are outlined in detail by Arrowsmith (2006), which demonstrated analytical solutions to overcome these problems. This paper’s analysis has followed such advice. In any case, there is no reason why the problem should be any different for the treatment and comparison groups.

Another potential problem with the administrative data used is that advisers do not always fill in some characteristic fields during client interviews. For example, this is the case with lone parenthood, disability and ethnicity and leads to missing values for these variables in the datasets used by this paper. Missing values can potentially be a large problem, since dropping observations with missing values can lead to much reduced sample sizes. However, in this case the problem is not a result of selection bias caused by jobseekers, since it is advisers who fill in the characteristic fields. It is, therefore, assumed that missing values do not differ systematically across jobseekers. Indeed analysis of the data shows that there is no significant difference in the number of missing values between the treatment and comparison groups for any of the characteristics variables. Therefore, in the analysis, any missing values are treated as a valid category.

4.3 Sample data

All the datasets discussed in the previous section are merged together and cleaned to remove duplicate records. The amount of data is also cut down in three ways in order to reduce the size of the dataset and, therefore, the amount of computer processing and time it takes for the analysis to be carried out: Firstly, only individuals who started their benefit claim between the dates of 1 August 2004 and 31 December 2005 are kept. The reason why benefit start date was used instead of WNP start date for the sampling frame is because there is no way of constructing a pseudo start date for a programme they did not actually start. The above dates exclude the first and last three months of WNP. The first three months have been excluded because all pilot projects take time to get up to speed and settle down and this is evidenced in Dewson et al. (2007). The last three months have also been excluded, as monitoring and evaluation data suggests that advisers and managers working in pilot areas correctly anticipated that the pilot would not be rolled out nationally and, in some cases, shifted effort to other services during the latter stages of the pilot. This is also demonstrated in Figure 4.1, which shows the number of starts on WNP in each month of the pilot. The start and end dates chosen apply to both the treatment and comparison groups.
Secondly, for all individuals, employment and benefit records are only kept for the 156 weeks prior to their benefit start date and the 39 weeks after their benefit start date.\footnote{The 156-week (three-year) labour market history before the benefit start date is used to compare and match participants with non-participants in the participation equation of the PSM. The 39 weeks post-participation takes us to the maximum length of time that can be observed for individuals that started WNP in its final week of operation. This length of 39 weeks post-participation is also imposed on individuals who started the pilot earlier, in order to achieve consistency.} Thirdly, for every individual recorded, only the data associated with their first benefit start date within the two-year period outlined above is kept. This is in order to remove any bias associated with outcomes from the same individual being included multiple times.

Doing so leaves us with the final dataset, which includes all of the outcome, participation, characteristic and labour market variables. This is this dataset that is used to carry out the analysis discussed in the next chapter. Table 4.2 presents, in full, how the final sample for the participation and comparison groups were arrived at.
Table 4.2  Final preferred sample selection

<table>
<thead>
<tr>
<th>Participation group</th>
<th>Comparison group</th>
</tr>
</thead>
<tbody>
<tr>
<td>WNP database: All recorded starts</td>
<td>National Benefits Database: All recorded spells</td>
</tr>
<tr>
<td>Basic selection:</td>
<td>Basic selection:</td>
</tr>
<tr>
<td>• Selection of only the first WNP start if an individual has multiple starts</td>
<td>• Selection only if benefit claims were for JSA, IS or IB</td>
</tr>
<tr>
<td>• Selection if WNP start has a valid open benefit claim at time of start</td>
<td>• Removal of those without a valid identification code</td>
</tr>
<tr>
<td>• Removal of those without a valid identification code</td>
<td></td>
</tr>
<tr>
<td>Final sample of participants:</td>
<td>Final sample of comparison group:</td>
</tr>
<tr>
<td>• Selection if the benefit start date is between 1 August 2004 and 31 December 2005</td>
<td>• Removal of individuals in WNP areas</td>
</tr>
<tr>
<td></td>
<td>• Selection if the benefit start date is between 1 August 2004 and 31 December 2005</td>
</tr>
<tr>
<td></td>
<td>• Selection of only the first benefit start if an individual has multiple starts</td>
</tr>
<tr>
<td></td>
<td>• Random selection of 2.5 per cent of starts (to reduce dataset size)</td>
</tr>
</tbody>
</table>

| 23,734 | 65,644,119 |
| 22,739 | 38,209,154 |
| 9,182  | 79,192     |

It is worth noting that an unexpectedly large number of treatment observations were dropped in the final stage of the sample selection. This is a result of there being long periods of time between starting a benefit claim and starting on WNP for many individuals.
5 Methodology

Having discussed the literature and outlined similar work in Chapter 3, this chapter outlines the proposed methodology. As was mentioned previously, the data available to this paper is longitudinal, enabling us to ignore cross-sectional methodologies. As was also noted in Chapter 3, the analysis and advice provided by Thomas (2006) is important. It demonstrated that using a propensity score matched panel DiD technique, while controlling for labour market history, impact estimates can be produced that are very close to the ‘true impact’ that might be estimated when using experimental data. Not only that but Thomas also used the same datasets as proposed by this paper. As a result, it shall attempt to follow Thomas’ advice as closely as is possible. Therefore, in terms of the specific methodology, this paper shall also use the matched panel DiD estimator, as Thomas recommends.

5.1 The participation equation

In terms of the specification of the participation equation that is used to construct the propensity score, this paper will use a probit model, as is used in most of the PSM literature. The dependent variable in this equation is the likelihood of participating in WNP. The independent variables used to explain participation are limited by the data available to us. For example, educational qualifications are likely to be important in explaining both participation and outcomes, and are potentially observable, but are not recorded in Jobcentre Plus administrative datasets. Also, as has been explained previously, the motivation levels of individuals are likely to be important but again, are not observed or available to us. It is worth repeating what the independent variables used were:

- age, (and age squared);
- gender;
- ethnicity;
- marital status;
- disability;
- labour market history;
• local labour market conditions;
• population density;
• lone parenthood;
• month of claim start/participation start on WNP.

The participation equation is estimated using all observations for individuals in both the treatment group and the comparison group. Fitted propensity scores are then calculated for each individual using this estimated equation and their own individual characteristics. The results of the probit regression are presented in Appendix C for interest.

5.2 Combined difference-in-differences and Propensity Score Matching

Once the likelihood of participation (propensity score) for each individual has been calculated, the PSM can be carried out. This is done by matching participants with non-participants, based upon the closeness of the propensity score. As mentioned in Chapter 3, the matching can be carried out using a variety of non-parametric methods, the most popular being nearest neighbour and kernel density matching. As was preferred in Thomas (2006) and used by Speckesser and Bewley (2006), this paper carries out kernel density matching, using an Epanechnikov kernel and using a bandwidth of 0.001.37

It is important at this stage to check the success of the match. As was mentioned in Chapter 3, the CIA is very important. However, this is impossible to test without experimental data. Checking the success of the matching variables used, therefore, relies upon comparing the mean observed characteristics of the treatment and matched comparison groups. Ideally there should be no significant difference in any of the characteristics between the two groups. The analysis will also need to make sure as small a number of treated individuals as possible are dropped under the enforcement of common support. Also important is the need to compare pre-treatment outcomes of the treatment and comparison group, to ensure that the two groups are as similar as possible.

37 In the literature, Epanechnikov kernels appear to be used most often, followed by Gaussian kernels. The popularity of these two kernel types is likely to be self-reinforcing as well as related to their default use in statistical packages such as Stata. It is also important to note that the choice of bandwidth is much more important than the choice of kernel type. While this paper would have preferred to try alternate specifications using both wider and narrower bandwidths to the chosen 0.001, time limitations prevented this.
The expectation is to find examples of Ashenfelter’s dip in the weeks immediately before participation, as has been described earlier. Of more importance, however, is that ideally there needs to be little systematic difference in the longer term pre-treatment outcomes. Key here is the use of panel DiD techniques, which help to remove any remaining observed and unobserved fixed heterogeneity after matching.

The general matched DiD estimator as presented in Thomas (2006) is replicated here as equation (3):

$$\hat{\alpha}_{MDiD} = \sum_{i \in (d_t = 1)} \left[ \left( Y_{it} - \sum_{j \in (d_t = 1)} w_{ij} Y_{jt} \right) \right] - \left[ \left( \sum_{j \in (d_t = 1)} w_{ij} Y_{jt} - \sum_{j \in (d_t = 0)} w_{ij}^C Y_{jt} \right) \right] w_i$$  

Here \( t' \) represents the first period of time, and \( t \) represents the second period of time. \( d_t = 1 \) indicates whether an individual was treated and indicates otherwise. \( Y_i \) represents outcomes for individuals who were treated and \( Y_j \) represents outcomes for those in the comparison group. Finally, \( w_{ij}^T \) is the weight attributed to observation \( j \) at time \( t' \) in forming the counterfactual for observation \( i \) from the treated group at time \( t \).

More specifically, as the author explained, the panel DiD estimator used later in the analysis in this current paper imposes two restrictions on the above equation: Firstly, \( w_{ij}^T = 1 \) since the same treated individual is used to construct the pre-programme counterfactual. Secondly, \( w_{ij}^C = w_{ij}^C \) because the same comparison observation is used in the pre- and post-programme periods.

When the difference is actually taken can be crucial to the final results. It makes no sense to take the difference a week before participation, as this is when Ashenfelter’s dip occurs. It is actually a matter of judgement by the researcher, based upon observation of pre-programme benefit histories. In terms of the research carried out by this paper, the difference is always taken 32 weeks prior to participation, well before any changes in benefit claim rate occur for the treatment group.

Once the above has been completed, the analysis can then move on to look for differences in post-treatment outcomes. Any significant differences observed are the estimate of the net impact of WNP. On the issue of significance, the calculation of standard errors is important. An issue arises as a result of each individual’s propensity scores being estimated rather than being observed. This introduces variation not accounted for in the later DiD regression equation. The solution used by authors such as Thomas (2006) and Bewley and Speckesser (2006), is to use a
procedure called bootstrapping. As a result of the considerably longer amount of time it takes for statistical packages to carry out bootstrapping procedures, this paper has chosen not to attempt bootstrapping. This potential criticism will be revisited later.

5.3 Adapting the model for different research questions

The above description of the proposed methodology is a general description of the techniques that this paper proposes to use. However, it intends to use these techniques to produce results using a number of different comparison groups, as well as to produce impact estimates for a number of different sub-groups. These are discussed below.

5.3.1 Comparison group selection

One of the points made in Heckman et al. (1997), and demonstrated in Thomas (2006), is that it can be useful to produce a number of different impact estimates using differently constructed comparison groups. The reason for this is because it is unlikely that any chosen comparison group will be identical in terms of observed and unobserved characteristics to the treatment group. While the literature does suggest ways to resolve problems with comparison group selection, there is not always an ideal way to do this. It is, therefore, worth selecting a number of different comparison groups so as to be able to compare the resulting estimates and in order to state a range for the ‘true’ programme impact. The analysis carried out in this current paper proposes to use three differently constructed comparison groups:

1. using all possible non-participants across the country as the comparison group;
2. using non-participants in the selected comparison areas only;
3. using non-participants in WNP areas only.

As discussed in the previous chapter, the largest possible comparison group is usually preferred for PSM, so of the three above, the most reliable estimates are likely to come from the first specification. The second specification is included because these comparison areas were selected prior to the introduction of WNP

---

38 Bootstrapping effectively takes a resample with replacement of ‘N’ observations from the original total sample population. It then uses the coefficients obtained from the estimated regression equation to obtain fitted outcomes and, therefore, the overall estimated impact. This bootstrapping procedure is repeated ‘B’ times. The estimated standard error is the standard deviation of the B separate estimates. The choice of both N and B is arbitrary and generally decided by time and computational considerations.
for their close similarity to the labour market conditions of the pilot areas. These comparison areas were also the areas used to construct a comparison group by Lone (2006d), so it will be interesting to compare this paper’s alternate impact estimate with Lone’s estimate.

The third specification of comparison group has been included for interest only and the results will only be presented in the Appendix A. The reason for using only non-participants in WNP areas as a comparison group is because they are based in the same labour market. However, there are two serious problems with this: Firstly, there may be reasons why non-participants chose not to participate in WNP, particularly for IS and IB claimants. This introduces unobserved selection bias to the estimates. Secondly, by design, WNP was meant to be a community-based pilot that had wider community effects. In particular, the use of the Community FDF to fund local events and buy premises for use by the whole community were meant to help all individuals in WNP areas, not just participants. This means that, if WNP worked as intended, a comparison group of non-participants in WNP areas may actually also have benefited from WNP and as such, analysis including this contamination will underestimate the impact of WNP.

5.3.2 Impact estimates

For each of the three comparison group specifications described above, this paper will produce a number of different net impact estimates. The most important will of course be the aggregate figure across all WNP sites, as well as a district breakdown. However, this paper also intends to produce estimates for a number of different sub-groups:

- a comparison of impact estimates between Jobcentre Plus only and EZ partnership WNP pilot areas;
- a comparison of impact estimates between individuals starting WNP claiming different types of benefit.

By carrying out these different sub-group estimates, the analysis aims to find out whether WNP was better carried out by the private sector in partnership with Jobcentre Plus and whether those claiming a particular type of benefit were helped more than others. It is worth remembering that this sub-sampling will be using smaller sample sizes and will result in larger standard errors and, therefore, less precise estimates.

39 Often the comparison areas are adjacent to the WNP areas within the same LAD.
6 Results

This chapter presents the results of the analysis. In turn, it goes through estimates for the overall impact of the WNP, as well as breakdowns by delivery method, benefit type and WNP site. It also presents estimates using alternate comparison groups. Finally there is a discussion of the results and what they might mean in the whole.

Throughout this section, the results are presented with the same style of graph, with time (in weeks) presented on the horizontal axis and net impact (i.e. the difference between the treatment and comparison groups) on the vertical axis. This net impact is the percentage difference, so a figure of 0.2 on the vertical axis means a 20 per cent impact. The varying line going from left to right through each graph measures impact. Ninety-five per cent confidence intervals surround either side of this central line, indicating the range of the estimates. The 52 weeks prior to benefit start date are shown effectively as a specification test, so as to get an idea of how closely matched the treatment and comparison groups are in terms of benefit/employment history. Ideally, if the treatment and comparison groups are closely matched, the horizontal axis should be within the confidence interval in most of the 52 weeks prior to the benefit start date, although Ashenfelter’s dip is to be expected immediately prior to the benefit start date. The 39 weeks after the benefit start date are the furthest the analysis can go when looking at potential impact. If WNP is to have the desired positive impact, the upper confidence interval line should be below the horizontal axis when looking at benefit rates (i.e. showing a reduction in the likelihood of claiming benefits). Similarly, the lower confidence interval line should be above the horizontal axis when looking at employment rates, (i.e. showing an increase in the likelihood of being in work).

6.1 Overall impact of the Working Neighbourhoods Pilot

6.1.1 Success of the match

Before presenting estimated impacts, it is important to present and discuss indicators of the success of the match for the variables matched upon. Table 6.1 shows the differences in mean characteristics between the treatment and comparison groups before and after matching.
Table 6.1  Average characteristics of treatment and comparison
group before and after matching

| Variable           | Sample   | Treated | Control | % bias | % reduction in bias | t value | p>|t| |
|--------------------|----------|---------|---------|--------|---------------------|---------|------|
| Male               | Unmatched| 0.67789 | 0.6001  | 16.3   | 11.29               | 0       | 0    |
|                    | Matched  | 0.67309 | 0.67782 | -1     | 93.9                | -0.52   | 0.6  |
| Disabled           | Unmatched| 0.20398 | 0.1702  | 8.7    | 6.22                | 99.9    | 0    |
|                    | Matched  | 0.20089 | 0.22272 | -5.6   | 35.4                | -2.78   | 0.006|
| Age                | Unmatched| 31.604  | 34.619  | -24.6  | -16.69              | 0       | 0    |
|                    | Matched  | 31.726  | 31.58   | 1.2    | 95.2                | 0.66    | 0.507|
| Ethnicity – white  | Unmatched| 0.09698 | 0.22785 | -36.1  | -23.31              | 0       | 0    |
|                    | Matched  | 0.09989 | 0.09312 | 1.9    | 94.8                | 1.19    | 0.233|
| Ethnicity – black  | Unmatched| 0.81404 | 0.67495 | 32.3   | 21.66               | 0       | 0    |
|                    | Matched  | 0.81801 | 0.84052 | -5.2   | 38.3                | -3.11   | 0.002|
| Ethnicity – South Asian | Unmatched| 0.03337 | 0.03272 | 0.4    | 0.26                | 0.797   | 0    |
|                    | Matched  | 0.02838 | 0.0218  | 3.7    | -904.5              | 0.18    | 0.03 |
| Ethnicity – mixed  | Unmatched| 0.02444 | 0.0386  | -8.1   | -5.39               | 0       | 0    |
|                    | Matched  | 0.02409 | 0.01819 | 3.4    | 58.3                | 2.13    | 0.033|
| Ethnicity – Chinese| Unmatched| 0.01097 | 0.00837 | 2.7    | 1.94                | 0.053   | 0    |
|                    | Matched  | 0.01038 | 0.01018 | 0.2    | 92.5                | 0.1     | 0.921|
| Ethnicity – other  | Unmatched| 0.00219 | 0.00195 | 0.5    | 0.39                | 0.697   | 0    |
|                    | Matched  | 0.00204 | 0.00136 | 1.5    | -174                | 0.86    | 0.393|
| Ethnicity – missing| Unmatched| 0.01802 | 0.01557 | 1.9    | 1.37                | 0.171   | 0    |
|                    | Matched  | 0.01723 | 0.01483 | 1.9    | 1.8                 | 0.99    | 0.32 |
| Marital – not known| Unmatched| 0.08852 | 0.24545 | -43    | -27.46              | 0       | 0    |
|                    | Matched  | 0.09526 | 0.09788 | -0.7   | 98.3                | -0.46   | 0.644|
| Marital – single   | Unmatched| 0.67946 | 0.45715 | 46.1   | 31.84               | 0       | 0    |
|                    | Matched  | 0.67161 | 0.67979 | -1.7   | 96.3                | -0.91   | 0.364|
| Marital – married  | Unmatched| 0.07489 | 0.16341 | -27.6  | -17.86              | 0       | 0    |
|                    | Matched  | 0.07784 | 0.07561 | 0.7    | 97.5                | 0.43    | 0.665|
| Marital – widowed  | Unmatched| 0.00407 | 0.00471 | -1     | -0.66               | 0.509   | 0    |
|                    | Matched  | 0.00426 | 0.00467 | -0.6   | 35.9                | -0.32   | 0.752|
| Marital – divorced | Unmatched| 0.047   | 0.04412 | 1.4    | 0.98                | 0.327   | 0    |
|                    | Matched  | 0.048   | 0.04799 | 0      | 99.7                | 0       | 0.998|
| Marital – separated| Unmatched| 0.04387 | 0.03145 | 6.5    | 4.81                | 0       | 0    |
|                    | Matched  | 0.04059 | 0.03677 | 2      | 69.3                | 1.03    | 0.305|
| Marital – cohabiting| Unmatched| 0.0622  | 0.05372 | 3.6    | 2.6                 | 0.009   | 0    |
|                    | Matched  | 0.06245 | 0.05728 | 2.2    | 39                  | 1.13    | 0.257|
It can clearly be seen that matching helps to remove most of the differences in characteristics between the two groups. In fact only the disability dummy and three of the ethnicity dummies have remaining significant differences between the two groups, as shown by the t tests in the second column to the right. Given the small number of people in each of the ethnic groups, the remaining differences are not too concerning. The remaining difference in the proportion who are disabled is unexpected, although in reality the difference is still only just over two percentage points, unlikely to be enough to significantly bias the results.

Table 6.2 shows in more detail the success of the match. The chi squared test shows how prior to the match, there was approximately a zero probability that the treatment and comparison group had the same set of characteristics. After matching there is effectively no difference between the two groups in terms of the variables used in the matching.

### Table 6.2 Output showing overall success of the match

<table>
<thead>
<tr>
<th>Sample</th>
<th>Pseudo R2</th>
<th>LR chi²</th>
<th>p&gt;chi²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmatched</td>
<td>0.378</td>
<td>11465.52</td>
<td>0.000</td>
</tr>
<tr>
<td>Matched</td>
<td>0.041</td>
<td>615.07</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Of more concern is the fact that 26 per cent of the treatment observations were dropped through lack of common support. As has been discussed in earlier chapters, net impact estimates that are based upon samples where a large proportion of observations have been dropped with a lack of common support, can be biased away from the ‘true impact’ that might have been estimated using the full sample and experimental data, if they had been available. This is because observations dropped are usually at the two tails of the distribution of propensity scores, i.e. the least or most likely to participate. If the observations that were dropped benefited more (or less) than those used in the propensity score estimation, then the net impact estimate will be biased downwards (or upwards). Inspection of the two distributions of propensity scores reveals that most of the dropped treatment observations were those with a higher propensity to participate. WNP was intended for more disadvantaged benefit claimants, although we have no idea whether the interventions actually used by WNP were more or less appropriate for those with the extreme worst characteristics. Therefore, there is no indication as to whether any bias would be upwards or downwards, if any bias exists at all. It is important that this is kept in mind when interpreting the results that follow.

### 6.1.2 Impact estimates

The first impact estimate for WNP is shown in Figure 6.1. This shows the overall average impact of WNP across all WNP areas on the total benefit rate.
Reassuringly, in the majority of the 52 weeks prior to participation, there is no significant difference in the proportion of individuals in the treatment and comparison groups claiming benefit. However, there is an unusual pick up in the proportion of people claiming benefits from around 20 weeks prior to participation. Normally, Ashenfelter’s dip shows the opposite, a drop off in the proportion claiming benefits just prior to participation. An example of this can be seen for the comparison group in Figure 6.4. Indeed, this differing pattern for the treatment and comparison groups is repeated in all of the analysis this paper carries out and it is not obvious why this might be. The issue will be returned to in the later discussion of results.

Post-participation, Figure 6.1 shows that the impact centres around four percentage points less of the treatment group being on benefits than the comparison group, with the 95 per cent confidence interval showing that the range of estimation lies between two per cent and seven per cent.

It is also worth considering impact on JSA only, as opposed to all the main welfare benefits. This is shown in Figure 6.2.
There actually appears to be a marginally negative impact on JSA claims, although not always statistically significant, especially in the first 20 weeks after starting WNP. Given the impact on all benefits has been estimated as positive, it would seem to indicate that WNP is encouraging some inactive benefit recipients to move closer to the labour market, by claiming JSA. This is consistent with recent evaluation findings for Pathways to Work, another pilot active labour market programme for claimants of IB. Looking at the impact of WNP from a different viewpoint, Figure 6.3 shows the estimated overall impact of WNP on employment.
As can be seen, the confidence interval is much wider, and no statistically significant impact is detected. There are indications that there may have been a positive impact, with the central estimate being above the horizontal axis, but this is not enough for impact to be certain.

It is worth investigating in more detail what is driving the net impact estimate. Figure 6.4 shows the pattern of benefit claim history for both the treatment and comparison group both pre- and post-WNP participation. It is the difference between the two lines that is used to derive Figure 6.1. The estimated central impact line for the treatment group is shown with the 95 per cent confidence interval surrounding it, while the comparison group is shown as a single line.
As can clearly be seen, post-participation, both the treatment and comparison group have a near identical drop off in the proportion claiming benefits. Indeed, the comparison group impact line is within the confidence interval of the treatment group, indicating no significant difference. However, there is a significant difference in the proportions claiming benefits in the 52 weeks prior to participation. To be clear, this indicates that there is a difference even after carrying out the PSM. This is where the DiD methodology becomes of use, as it takes into account any of the remaining differences, before the final net impact is estimated in graphs such as Figure 6.1. Of importance here is that the net impact estimate for WNP is actually being driven by differences in benefit claim rates prior to participation, rather than differences in off flow post-participation. It could, therefore, be argued that WNP is not actually doing anything fundamentally different to drive higher off flows. Rather, it could instead be argued that the more intensive intervention regime in WNP areas is helping the harder to help claimants in those areas to ‘catch up’ with those nationally who appear to be less hard to help. This will be discussed further later in this chapter.
6.2 Impact of the Working Neighbourhoods Pilot, by delivery method

It is also useful to investigate whether the impact of WNP varied by the type of delivery mechanism, in this case whether WNP sites run by EZs in partnership with Jobcentre Plus performed better or worse than sites run solely by Jobcentre Plus. Figures 6.5 and 6.6 show the impact of WNP on total benefit rate in EZ/Jobcentre Plus partnership areas and Jobcentre Plus only areas respectively.

**Figure 6.5  Net impact of WNP in EZ/Jobcentre Plus partnership areas on total benefit rate**
Figure 6.6  Net impact of WNP in Jobcentre Plus only areas on total benefit rate

It can clearly be seen that the overall impact on total benefit rate is much higher in EZ/Jobcentre Plus partnership areas, centred around 18 per cent just after start of claim and gradually falling to 12 per cent in week 39, at the end of the period of monitoring. This compares to an impact centred around eight per cent, falling to five per cent, in Jobcentre Plus only areas.

It is also worth considering impact on JSA only, as opposed to all welfare benefits. This is shown in Figures 6.7 and 6.8 for EZ/Jobcentre Plus partnership areas and Jobcentre Plus only areas respectively.
Results

Figure 6.7 Net impact of WNP in EZ/Jobcentre Plus partnership areas on JSA rate

Figure 6.8 Net impact of WNP in Jobcentre Plus only areas on JSA rate
Both show a smaller impact of roughly the same pattern as for the impact on all welfare benefits. The main difference is that Jobcentre Plus only areas have a generally insignificant impact on the JSA rate, indicating that these areas’ main impact was on inactive benefit recipients, whereas EZ/Jobcentre Plus partnership areas appear to have impacted on all benefit recipients. In terms of impact on employment, Figures 6.9 and 6.10 show the impact in EZ/Jobcentre Plus partnership areas and Jobcentre Plus only areas respectively.

**Figure 6.9** Net impact of WNP in EZ/Jobcentre Plus partnership areas on employment rate
Figure 6.10 Net impact of WNP in Jobcentre Plus only areas on employment rate

Both have wide confidence intervals and there is, therefore, no discernable significant impact on employment in either type of area. What can be seen, however, is differing impact patterns post-participation. EZ/Jobcentre Plus partnership areas do appear to cause a step change in terms of an increased central estimate of employment immediately after starting WNP, followed by a gradual decline towards zero. In Jobcentre Plus only areas there is much less of a step change in the central estimate of the impact on employment, with more variation afterwards.

It is potentially worth uncovering where the higher performance for EZ/Jobcentre Plus partnership areas came from. This can be done by looking at Figures 6.11 and 6.12. They show the pattern of benefit claim history for both the treatment and comparison groups in EZ/Jobcentre Plus partnership areas and Jobcentre Plus only areas, respectively. Again, the treatment group line is shown with the 95 per cent confidence interval surrounding it, while the comparison group is shown as a single line.
In the months prior to participation, it can be seen that treated individuals in EZ/JCP partnership areas are ten percentage points more likely to be on benefits than those in JCP only areas. This would indicate that WNP participants in EZ/JCP partnership areas were potentially harder to help on average than WNP participants in JCP only areas. Interestingly however, despite the potentially harder to help claimants, post participation, EZ/JCP partnership areas actually appear to perform better than JCP only areas in terms of benefit off flows. In simple terms, it would appear that EZ/JCP partnership areas may have achieved higher outcomes despite working with harder to help claimants. This is potentially a very significant finding, as there was no difference in funding between the two types of WNP area.

It is important to note here that for both EZ partnership areas and JCP only areas, quite a large number of observations were dropped through a
In the months prior to participation, it can be seen that treated individuals in EZ/Jobcentre Plus partnership areas are ten percentage points more likely to be on benefits than those in Jobcentre Plus only areas. This would indicate that WNP participants in EZ/Jobcentre Plus partnership areas were potentially harder to help, on average, than WNP participants in Jobcentre Plus only areas. Interestingly, however, despite the potentially harder to help claimants, post-participation, EZ/Jobcentre Plus partnership areas actually appear to perform better than Jobcentre Plus only areas in terms of benefit off flows. In simple terms, it would appear that EZ/Jobcentre Plus partnership areas may have achieved higher outcomes despite working with harder to help claimants. This is potentially a very significant finding, as there was no difference in funding between the two types of WNP area.

It is important to note here that for both EZ partnership areas and Jobcentre Plus only areas, quite a large number of observations were dropped through a lack of common support. This means that the impacts reported above are only for a specific subsection of participants and are not for the population of participants as a whole. All the same, the estimates are comparable to each other and, therefore, indicative.

6.3 Impact of Working Neighbourhoods Pilot by benefit type

Another important distinction to look at is the varying impact of WNP on those claiming the three different primary welfare benefits prior to starting on WNP. As was outlined in Chapter 2 of this paper, the WNP intervention changes varied by benefit type. Most importantly, while WNP interventions were mandatory for JSA claimants, the majority of WNP interventions for IB and IS claimants were only voluntary. Figures 6.13, 6.14, and 6.15 show the impact of WNP on total benefit rate for those claiming JSA, IB and IS respectively, when they started WNP.

---

40 Twenty-four per cent of observations were dropped for EZ areas and 12 per cent for Jobcentre Plus only areas.
Figure 6.13 Net impact of WNP on total benefit rate for those starting WNP while claiming JSA

Figure 6.14 Net impact of WNP on total benefit rate for those starting WNP while claiming IB

Results
The largest impact appears to have occurred for IB claimants, with an estimated impact centred around 25 per cent immediately after starting WNP, falling to around 15 per cent at 39 weeks. The impact on JSA claimants is more stable, centred around six per cent for most of the 39 weeks observed. There is a larger confidence interval around the IS claimants impact estimate, although it remains significant throughout, centred around ten per cent.

Moving on to the impact of WNP on employment, Figures 6.16, 6.17 and 6.18 show the impact of WNP for those claiming JSA, IB, and IS respectively, at the time they started WNP.
The largest impact appears to have occurred for IB claimants, with an estimated impact centred around twenty five percent immediately after starting WNP, falling to around fifteen percent at 39 weeks. The impact on JSA claimants is more stable, centred around six percent for most of the 39 weeks observed. There is a larger confidence interval around the IS claimants impact estimate, although it remains significant throughout, centred around ten percent.

Moving on to the impact of WNP on employment, Figures 6.16, 6.17, and 6.18 show the impact of WNP for those claiming JSA, IB, and IS respectively, at the time they started WNP.

**Figure 6.16 Net impact of WNP on employment rate for those starting WNP while claiming JSA**

**Figure 6.17 Net impact of WNP on employment rate for those starting WNP while claiming IB**

**Figure 6.18 Net impact of WNP on employment rate for those starting WNP while claiming IS**
Figure 6.18 Net impact of WNP on employment rate for those starting WNP while claiming IS

Sadly, the confidence intervals are too large to be able to see a significant impact on employment for any of the three benefit types. If anything, there is a more noticeable step change in the impact on IB claimants but the confidence intervals are largest here, so nothing conclusive can be stated.

6.4 Impact of the Working Neighbourhoods Pilot, by individual pilot location

It is also worth investigating whether, and how much, the impact of WNP varied by pilot location, so as to be able to identify whether some areas performed better than others.\textsuperscript{41} Given the limited number of observations available, the confidence intervals, when looking at impact on employment, are very wide and so these impact estimates are not reported. However, impact estimates on total benefit rate are presented in Appendix B, with 12 graphs, one for each pilot district. The general trend is for EZ/Jobcentre Plus partnership areas to have larger and significant impact estimates, compared to smaller and insignificant impact estimates in Jobcentre Plus only sites. This confirms the aggregate evidence shown in Section 6.2 earlier. The largest EZ impacts centre around 20 per cent in Tower Hamlets.

\textsuperscript{41} It is worth remembering that WNP was never set up to be competitive or to compare one site with another. WNP was really meant to test a more flexible approach and to compare pilot sites to their chosen comparison areas.
and in both of the two Glasgow pilot sites. On the other hand, Birmingham has too wide a confidence interval to discern impact, while Middlesbrough has the smallest estimated EZ impact, centred around five to ten per cent, and becoming insignificant after week 20.42

The largest impact estimates from a Jobcentre Plus only WNP area come from both Hastings, with a consistent impact centred around 15 per cent, and Wirral, with a short-term impact estimate of 20 per cent. The worst performing Jobcentre Plus only sites are Sheffield and Great Yarmouth, where even ignoring the confidence intervals, there is no evidence shown in their respective graphs to indicate WNP was at all effective in reducing benefit rates.

Interestingly, these impact estimates by individual WNP area reflect the anecdotal evidence reported at the time. For example, we do know from the qualitative evaluation that both of the Glasgow sites are reported to have had excellent relationships between the two partner organisations, which appear to be reflected by higher impact estimates. Also, the Sheffield Jobcentre Plus only WNP site had staff resource issues, which is again potentially demonstrated by the lack of any noticeable impact.

Again, it is worth pointing out that for all the estimates for individual WNP sites, around 25 per cent of observations were dropped through a lack of common support. While this means estimates will differ from the ‘true impact’ for reasons already described above, the estimates are still comparable to each other.

6.5 Sensitivity analysis

As discussed in earlier chapters, it is worth using an alternative comparison group to see if net impact estimates differ from the main specification. The literature suggests that the final matched comparison group should be drawn from the largest possible group of people, controlling for different labour market conditions in the participation equation, and with any remaining fixed differences controlled for when using the panel DiD techniques. That is what the main analysis in this paper is actually carried out. As an alternative specification, this paper specifically selects the comparison group from more closely matched labour markets, even before matching takes place, using the selected neighbouring areas to WNP areas to draw the alternate comparison group.

In reality, like the main specification, a large proportion of the sample participants, 25 per cent, are dropped through a lack of common support. The first estimates can be seen in Figure 6.19, measuring the impact of WNP on claiming all welfare benefits.

42 In some respects these results are not surprising. The take-up rate of WNP in Birmingham was very low and it was anecdotally reported that Middlesbrough was not following the mandatory intervention regime.
Reassuringly, once again there is no significant difference between the treatment and comparison groups in the majority of the year before participation. In the weeks after participation, the impact slowly increases, stabilising at around ten per cent less people on benefits in WNP areas when compared with the selected comparison sites. On the flip side, Figure 6.20 measures the impact of WNP on employment.
Again, there is evidence of Ashenfelter’s dip, but the main point is that there is an immediate and increasing significant impact on employment after participation, reaching around five per cent by week 39 after participation. This impact appears to be just about stabilising but it would be interesting to monitor the weeks after week 39 to see if the impact is maintained.

Overall, it would appear that using an alternate comparison group does not seriously alter the impact estimates. Comparing Figure 6.1 with Figure 6.19, although the central impact estimates are different, the confidence intervals just about overlap, so there is no real significant difference between the two estimates. This is quite reassuring.

Appendix A also shows a further alternative specification using non-participants within the WNP areas as the comparison group. There are various reasons why using this comparison group should not really be reported. These are described in more detail in Appendix A. Interestingly, despite concerns about using this specification of comparison group, the results are relatively similar to both the main and alternative specifications. Also of interest is the fact that after matching there is no pre-programme difference in the benefit claim rate of the treatment group and comparison group, as may well be expected given both groups live in exactly the same area. This is not the case in the main specification and proves the validity of using DiD after PSM, in order to remove remaining differences in the main comparison group specifications.
6.6 Discussion of results

Overall, the results presented in the previous few sections paint a generally positive view of the estimated net impact of WNP. The main specification shows an estimated positive impact on the total benefit rate centred around four per cent (within a confidence interval of two to seven per cent) and persisting over the full 39 weeks covered by the analysis. It is this figure that is the main preferred estimate to be reported for WNP, as it uses the biggest sample size and, therefore, takes into account area variation.

While less importance is given to the results using alternative comparison groups, their estimates are encouraging, as they are generally within the confidence intervals of the main specification.

It is worth returning to the issue briefly discussed in Section 6.1.2. It highlighted the fact that the impact estimates for WNP were being driven by pre-programme rather than post-programme differences, with a larger proportion of the treatment group claiming benefits than the comparison group prior to participation and little difference post-participation. It was suggested that WNP might not actually do anything fundamentally different to drive higher off flows, and that maybe the more intensive intervention regime in WNP areas is helping the harder to help claimants in those areas to ‘catch up’ with those nationally, who appear to be less hard to help. On an overall level this would appear to be the most sensible suggestion, although it is not obvious how this could be tested. Interestingly though, the results for delivery method do show that EZ/Jobcentre Plus partnership areas appear to both deal with harder to help claimants on average (in terms of the proportion claiming benefit in the year prior to participation) and have higher off flows post-participation. In other words, unlike Jobcentre Plus only areas, EZ/Jobcentre Plus partnership areas may actually have done something different to drive their higher benefit off flows.

As a result, impact estimates are generally larger in EZ/Jobcentre Plus partnership areas and in particular in Tower Hamlets and the two Glasgow WNP sites, where the estimated impact on the total benefit rate centres around 20 per cent. This is reflected in both the individual pilot site impact estimates presented in Appendix B and the estimates presented in Section 6.2. What is driving this much higher estimated impact is difficult to say. The qualitative evaluation of WNP carried out by Dewson et al. (2007) does not identify any particular specific differences between the two types of delivery method. All there is to go on is feedback from members of staff in the various pilot sites. This suggests that the flexibility allowed by WNP was fully utilised in EZ/Jobcentre Plus partnership areas, whereas attempts to do so in Jobcentre Plus only areas were stymied by rules, regulations and general bureaucracy.

It is worth returning to the fact that results for employment impact are insignificant throughout the analysis, despite the significant impact on benefit rates. This is a result of both less individuals recording employment spells and the amount of
employment records removed in the data cleaning and results in wider confidence intervals. It means that the impact estimates for employment and benefit rate do not at first appear to tally, especially if only the central estimates are looked at. However, the confidence intervals do actually overlap for the main overall WNP impact estimate, meaning no significant difference between the two measures. Interestingly though, there is a significant difference between the EZ/Jobcentre Plus partnership employment and benefit impact estimates. It could be that EZ/Jobcentre Plus partnership areas successfully move people off welfare benefits but do not actually get them into work. While this would appear to be good for the Exchequer, at least in the short-term, it is to be debated whether it is good for the individuals involved.

It is also worth returning to the issue raised earlier in Section 6.1.2 about the noticeable difference in pattern for treatment and comparison groups in the weeks before participation. As discussed previously the treatment group lack the normal pattern of an Ashenfelter’s dip. Interestingly, this odd pattern is the same in each of Figures 6.13, 6.14 and 6.15, for the different types of benefit. This would appear to indicate that the pattern is not a result of different intervention regimes, as had first been suspected. Careful inspection of the statistical coding used to construct the samples of treatment and comparison group individuals also did not reveal any errors that might have led to this oddity. There, therefore, appears to be no obvious explanation for the unusual pattern. Thankfully, the patterns of benefit claim rates for the treatment and comparison groups are remarkably stable between weeks 52 and 20 prior to participation, and given that the difference is taken in week 32 prior to participation, comfort can be taken that impacts estimated post-participation are reliable.

6.7 Limitations

It is important to note some of the limitations of the analysis performed in this paper: Firstly, on a technical note, and as previously mentioned in Chapter 5, no attempt was made to bootstrap for standard errors, as is now common in the literature. The reason this paper chose not to do so was a result of time limitations, in that bootstrapping takes a lot longer for the computer to process. Its implication for the results presented earlier is that the confidence intervals are slightly smaller than they would be if using bootstrapping, as a result of standard errors being smaller than they should be. This means that some of the impact estimates may be presented as statistically significant, when in fact they might not have been, had bootstrapping been used. However, many of the estimated impacts, including the main overall impact of WNP are significant enough to make it unlikely that they will be strongly affected by this.

It is also worth returning to the issue that in most of the presented results, 25 per cent of treatment observations were dropped with a lack of common support. As has been pointed out previously, this does not change impacts estimated for the observations still included in each piece of analysis. What it does mean, however,
is that the estimated impact may be different from the ‘true impact’ that could be estimated, had the whole sample of participant data been available for analysis. The size and direction of the bias is not clear, however.

Finally, the analysis carried out in this paper cannot identify what caused positive impacts where they were identified. WNP included more intensive intervention regimes, retention payments, and significant amounts of extra money being spent in the pilot areas. It is not clear whether one, two, or all three of these differences contributed to any positive impact, simply that, as a whole, they appear to have done so, particularly in the EZ/Jobcentre Plus partnership areas.
7 Conclusions

The purpose of this dissertation has been to carry out a comprehensive net impact evaluation of the WNP. It has done this using a combined PSM and DiD methodology, as is recommended in the literature.

The net impact estimates show that, overall, WNP had a statistically significant positive impact on the total benefit rate, centred around four per cent, and persisting over the 39 weeks after participation covered by the analysis. The impacts on employment are insignificant, mainly as a result of the smaller sample sizes and the number of employment observations dropped to make this particular variable reliable.

Sub-group analysis also reveals that WNP areas operated by EZ in partnership with Jobcentre Plus (EZ/Jobcentre Plus partnership areas) had a higher estimated impact than WNP areas operated by Jobcentre Plus only. Some individual EZ/Jobcentre Plus partnership areas had positive impact estimates on the total benefit rate centred around as high as 20 per cent. This is despite the fact that the data appears to show EZ/Jobcentre Plus partnership areas worked with harder to help customers than Jobcentre Plus only areas, on average. It is not entirely clear what drove these higher estimated impacts. However, feedback from staff in the different sites suggests that EZ/Jobcentre Plus partnership areas were better able to use the flexibility allowed in WNP.
Appendix A
Results using non-participants in Working Neighbourhoods Pilot areas as an alternative comparison group

As was discussed in the main body of this paper, there are likely to be two serious problems in using non-participants within WNP areas as an alternative comparison group: Firstly, there may be reasons why non-participants chose not to participate in WNP, particularly for IS and IB claimants. This introduces unobserved selection bias to the estimates. Secondly, by design, WNP was meant to be a community-based pilot that had wider community effects. In particular, the use of the Community FDF to fund local events and the building of premises for use by the whole community were meant to help all individuals in WNP areas, not just participants. This means that, if WNP worked as intended, a comparison group of non-participants in WNP areas may actually also have benefited from WNP and, as such, analysis on this basis will underestimate the impact of WNP. The only really good reason why non-participants in WNP areas might be chosen as a comparison group is because they are, by definition, living in the same labour market as the treatment group and so subject to the same labour market conditions. But as has been shown, the methodology used in the main body of this paper can control for different labour market conditions in the participation equation, with any remaining fixed differences controlled for using panel DiD techniques.
In any case, 30 per cent of treated individuals dropped through a lack of common support. This is a large proportion of the sample of participants to drop and as such, is likely to significantly bias the impact estimates away from the ‘true impact’ that might have been measured if experimental data existed.\(^{43}\)

Figure A.1 shows the net difference between the treatment group and comparison group in terms of claiming any type of welfare benefit. The impact slowly increases to ten per cent less participants being on benefit 39 weeks after participation, compared to non-participants in the same areas.

**Figure A.1 Net impact of WNP on total benefit rate, using non-treated individuals in WNP areas as the comparison group**

On the other hand, in terms of employment, Figure A.2 shows that the treatment group were increasingly likely to be in employment in the weeks after participation. The net impact had reached between four per cent and nine per cent 39 weeks after participation and looks to be on a continued upward trend.

---

\(^{43}\) This is the point made by Bryson et al. (2002). Just because an impact may be found for those with common support, does not mean that there was an impact for those without common support. On the other hand, observations without common support are not dropped when using experiment data, and will, therefore, lead to differing impact estimates.
Figure A.2 Net impact of WNP on employment rate, using non-treated individuals in WNP areas as the comparison group

One final finding of interest is shown in Figure A.3. It shows the pattern of benefit claim history for both treatment and comparison groups. As can be seen, in the majority of the 52 weeks prior to participation, there is no significant difference in their benefit claim rates, after matching. This may be expected given that both groups are drawn from the same pilot area. However, as has been shown in the main results section, this is not the case when using the preferred comparison group taken from a national sample, where differences between the treatment and comparison groups remain. It, therefore, justifies the need for use of DiD after the PSM has been carried out.
Figure A.3  Pattern of benefit claim history for both treatment and comparison groups both pre- and post-participation
Appendix B
Results by individual Working Neighbourhoods Pilot site

All figures shown in this appendix look at WNP’s impact on total benefit rate only. This is because the sample sizes are too small and confidence intervals too large to make looking at impact on employment worthwhile.

Figure B.1 Net impact of WNP on total benefit rate in Glasgow Parkhead (EZ/Jobcentre Plus partnership area)
Figure B.2  Net impact of WNP on total benefit rate in Glasgow Hutchesontown (EZ/Jobcentre Plus partnership area)

Figure B.3  Net impact of WNP on total benefit rate in Middlesbrough (EZ/Jobcentre Plus partnership area)
Figure B.4  Net impact of WNP on total benefit rate in Tower Hamlets (EZ/Jobcentre Plus partnership area)

Figure B.5  Net impact of WNP on total benefit rate in Birmingham (EZ/Jobcentre Plus partnership area)
Figure B.6 Net impact of WNP on total benefit rate in Knowsley (Jobcentre Plus only area)

Figure B.7 Net impact of WNP on total benefit rate in Hastings (Jobcentre Plus only area)
Figure B.8  Net impact of WNP on total benefit rate in Great Yarmouth (Jobcentre Plus only area)

Figure B.9  Net impact of WNP on total benefit rate in Swansea (Jobcentre Plus only area)
Figure B.10 Net impact of WNP on total benefit rate in Sheffield (Jobcentre Plus only area)

Figure B.11 Net impact of WNP on total benefit rate in Wirral (Jobcentre Plus only area)
Figure B.12 Net impact of WNP on total benefit rate in Newcastle (Jobcentre Plus only area)
Appendix C
Results of the propensity score probit regression

| Variable             | Coefficient | Std. Error | z   | P>|z| | 95 per cent confidence interval |
|----------------------|-------------|------------|-----|------|-----------------------------|
| Male                 | 0.03801     | 0.0236127  | 1.61| 0.107| -0.0082697 - 0.0842904     |
| Disabled             | 0.162088    | 0.0276725  | 5.86| 0    | 0.1078505 - 0.2163249      |
| Age                  | 0.033477    | 0.0058252  | 5.75| 0    | 0.0220596 - 0.044894       |
| Started WNP September 2004 | -20.3194    | 1.837989   | -11.06| 0    | -23.92174 - -16.71695     |
| Started WNP October 2004  | -20.068     | 1.837458   | -10.92| 0    | -23.66931 - -16.46661     |
| Started WNP November 2004 | -19.8281    | 1.836913   | -10.79| 0    | -23.42836 - -16.22779     |
| Started WNP December 2004 | -19.6509    | 1.838202   | -10.69| 0    | -23.25369 - -16.04807     |
| Started WNP January 2005  | -19.6465    | 1.83658    | -10.7 | 0    | -23.24609 - -16.04683     |
| Started WNP March 2005   | -19.6527    | 1.837442   | -10.7 | 0    | -23.25401 - -16.05137     |

Continued
Table C.1  Continued

| Variable                        | Coefficient | Std. Error | z     | P>|z| | 95 per cent confidence interval |
|---------------------------------|-------------|------------|-------|-----|-----------------------------|
| Started WNP April 2005          | -19.7354    | 1.837603   | -10.74| 0    | -23.33703 -16.13376         |
| Started WNP May 2005            | -19.6459    | 1.837593   | -10.69| 0    | -23.24748 -16.04425         |
| Started WNP June 2005           | -19.5623    | 1.836762   | -10.65| 0    | -23.16231 -15.96234         |
| Age^2                           | -0.00053    | 0.00008    | -6.62 | 0    | -0.0006858 -0.0003724       |
| Ethnicity – white               | -0.48238    | 0.2369546  | -2.04 | 0.042| -0.9468038 -0.0179587       |
| Ethnicity – black               | -0.02623    | 0.2353978  | -0.11 | 0.911| -0.4875972 0.4351453        |
| Ethnicity – South Asian         | -0.59671    | 0.2431843  | -2.45 | 0.014| -1.073344 -0.1200789        |
| Ethnicity – mixed               | -0.75525    | 0.2439454  | -3.1  | 0.002| -1.233373 -0.2771242        |
| Ethnicity – Chinese             | -0.17802    | 0.2580776  | -0.69 | 0.49 | -0.6838473 0.3277983        |
| Ethnicity – other               | -0.25589    | 0.2486271  | -1.03 | 0.303| -0.7431873 0.2314131        |
| Marital – single                | -0.36008    | 0.0620927  | -5.8  | 0    | -0.4817774 -0.2383785       |
| Marital – married               | -0.03139    | 0.0571672  | -0.55 | 0.583| -0.1434323 0.0806589        |
| Marital – widowed               | -0.46395    | 0.0632844  | -7.33 | 0    | -0.5879862 -0.3399158       |
| Marital – divorced              | -0.0587     | 0.1703866  | -0.34 | 0.73 | -0.3926526 0.2752506        |
| Marital – separated             | -0.01818    | 0.0726359  | -0.25 | 0.802| -0.1605397 0.1241879        |
| Marital – cohabiting            | -0.17314    | 0.0704088  | -2.46 | 0.014| -0.3111344 -0.0351372       |
References


DWP (2007), In work, better off: Next steps to full employment, Department for Work and Pensions Policy Document.


Thomas, A. (2006), Evaluating the econometric evaluations of active labour market programmes using administrative data: Evidence from the Jobseeker's Allowance pilots, MSc Thesis from Nottingham University.