

The impact of employment policy interventions¹

DAVID C MARÉ²

This paper estimates the impact of different categories of employment policy interventions on subsequent outcomes for jobseekers. We generate a range of estimates to help us distinguish programme effects from selection effects. We also examine the robustness of our findings for a range of sub-populations. Referrals to vacancies and job subsidies appear to be most effective in reducing the number of weeks of assistance or contact that jobseekers subsequently have with the public employment agency. The favourable estimated impact of subsidies is not evident until at least a year after the subsidy starts. There are only small differences in the estimated effectiveness across different ethnic groups. Interventions appear to be more effective for males than for females, and to a lesser extent more effective for younger than for older jobseekers. We find evidence to suggest that programme effectiveness is counter-cyclical.

1 Introduction

THIS PAPER EXAMINES THE EFFECT of employment policy interventions (active labour market policies (ALMP), and referrals to vacancies) in influencing the unemployment experience of jobseekers. This is an important issue for both government policy and the large number of individuals who experience unemployment. The government currently spends around \$600 million per year providing assistance to unemployed jobseekers. During 1993,³ over 250,000 distinct individuals registered as unemployed.⁴

The results reported in this chapter contribute to the existing New Zealand evaluation literature on the impacts of active labour market policies. This study is unique in the range of policies considered, and the timeframe over which outcomes are observed.

¹ This paper has benefited from comments, support, and interest from a wide group of people over several years. The Administrative Data Project (ADP) team within the Labour Market Policy Group of the New Zealand Department of Labour provided extensive input (Toby Buscombe, Maria Gobbi, James MacNaughton, David Rea and Andrew Reynolds). I am also grateful for comments from Dean Hyslop, Dan Black, Simon Chapple, Geoff Bascand, Marc DeBoer and participants at seminars at the Department of Labour, University of Auckland, and the 9th Labour Employment and Work Conference. The comments of three anonymous referees are appreciated. Any remaining errors are the author's.

² David Maré is a Senior Fellow with Motu Research Trust (www.motu.org.nz). Most of the work presented in this paper was carried out while David was a Research Adviser with the Labour Market Policy Group of the New Zealand Department of Labour.

³ Most of the paper focuses on 1993 because it is the mid-point of our sample period and, therefore, allows the longest prior and subsequent observation windows.

⁴ Gobbi and Rea (2000).

The administrative dataset used for this project provides complete information on all assistance provided to jobseekers over a nine-year period. It allows us to control for differences in experience prior to receiving assistance, and permits a wide range of outcome measures.

The main drawback of using this dataset for evaluation purposes is a weakness that it shares with any evaluations based on administrative records. The assignment of individuals to different forms of assistance is not random, and observed differences in outcomes for assisted and unassisted jobseekers may therefore reflect differences in the sort of people who are assisted rather than differences due to the assistance.

We use a range of “quasi-experimental” quantitative techniques to estimate the impact of employment policy interventions, and to separate these from other sources of differences in outcomes.

2 Links with previous New Zealand evaluations

There is a strong commitment in New Zealand to evaluate active labour market policies. The Department of Work and Income website contains summaries of 36 evaluations carried out between 1994 and 2000.⁵ As would be expected, the evaluations vary widely in their objectives and approaches. The majority of the evaluations focus on the way that policies were delivered, or on the outcomes experienced by participants. They bring together information from participants, providers, and staff of the public employment service, to evaluate the policy measures. The information collected is often a combination of subjective assessments, stakeholder views about the policy and quantitative measures of outcomes for participants or cost.

A subset of the evaluations, eight in total, address the question that is the main focus of the current paper – “how different were the outcomes for participants compared with the (unobserved) outcomes that they would have experienced had they not participated?”. To answer this question requires the use of some form of control or comparison group. The existing studies have used:

- comparison with earlier cohorts (Job Action – Enhanced (1998));
- comparison with older jobseekers (Youth Action (1996));
- attribute-matched comparison group of non-participants (TPE: Training for Pre-determined Employment (2000); Compass Quantitative (1997); Job Action (1995); TFG – Taskforce Green (1995));
- attribute-matched comparison group combined with multivariate analysis (ECTF – Expanded Community Taskforce and Community Work (2000); CTF – Community Taskforce (1999));
- experimental design (Community Taskforce (1999)).

⁵ The full text of 25 of these studies can be downloaded from the site.

The methods used in this paper offer an alternative to the previously used methods of defining a comparison group. As described below, we use “propensity matching” as well as regression analysis to estimate the effect of treatment on the treated.

We also consider a broader range of interventions than do any of the existing studies, and apply the same method to estimate the effects of training, subsidies, work experience, interviews and referrals to vacancies. The advantage of this approach is that it allows some comparability of findings. The disadvantage is that the study does not use as rich a set of programme-specific information as is taken into consideration in more programme-focused evaluations.

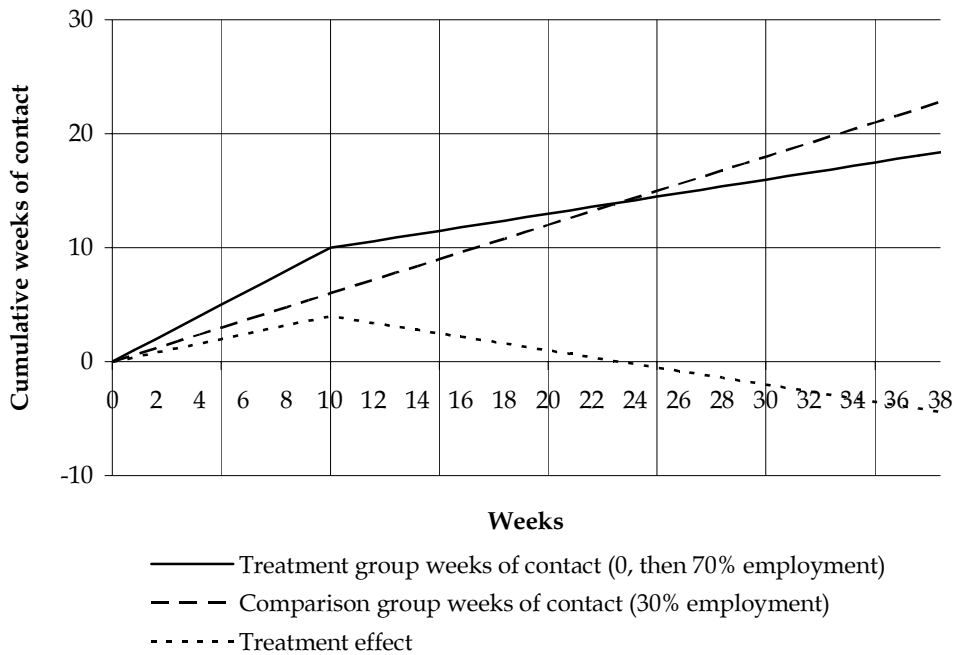
3 The logic of programme effectiveness

In order to be able to compare our results across different forms of assistance, we measure ‘effectiveness’ by means of standard outcome measures. We focus on the effectiveness of interventions in reducing the time that jobseekers are dependent on the public employment service. We do this on the presumption that the *ultimate* objective of active labour market policies is to lead jobseekers into unsubsidised employment.

Of course, some forms of assistance aim to achieve *intermediate* outcomes such as improved self-confidence, motivation, work-skills or ‘job-readiness’. It must, therefore, be borne in mind when interpreting the findings from this study that policies that are found to be ineffective in promoting ultimate objectives may nevertheless be successful in promoting their stated (intermediate) objectives. The outcome measure also takes no account of the quality of job matches, even though this is an important dimension of success for the public employment agency.

Figure 1 illustrates the way that we think about programme effectiveness. The horizontal axis measures calendar time (in weeks). The vertical axis measures the cumulative number of weeks subsequently spent in contact with the New Zealand Employment Service (NZES) for each group. The darkest line represents the experience of the comparison group. For simplicity, they are shown as spending 40 percent of their time in employment. Therefore, after 10 weeks of calendar time, they have had four weeks of employment, and are shown in the graph as having six weeks of contact with NZES. In contrast, the lighter solid line shows the experience of a treatment group that receives 10 weeks of assistance. After 10 weeks of calendar time, they have spent all 10 weeks in contact with NZES.

In order for the intervention to be effective, it must improve the employment chances of the treated group. In the graph, we assume that, following the assistance, the treatment group spend 70 percent of their time in employment. They therefore accumulate contact time more slowly than does the comparison group. After about 23 weeks, the lines cross, indicating that the improved prospects of the treatment group have been sufficient to offset the lack of job-search during the 10 weeks that they were receiving assistance. Beyond the 23-week point, they are, on average, better off than the comparison group.

FIGURE 1: An effective intervention

The dotted line on the graph shows the difference in contact time between treatment and comparison groups. A negative number indicates that the treatment group has experienced fewer weeks of contact, which is a favourable outcome. The main outcome measure used in this paper is the net difference in contact time after three years (156 weeks), although other time periods are considered. We also examine changes in time spent registered as unemployed, which is a subset of the contact time.

4 Methods

The study described in this paper uses non-experimental methods to estimate the impact of ALMPs. In this section, we outline why quasi-experimental methods are needed, and then discuss estimators used in the study.

4.1 Non-experimental methods

As noted in the introduction, the non-experimental nature of the administrative dataset that is used for this project poses some methodological challenges that would not be present in an experimentally designed evaluation.

Ideally (at least from a researcher's perspective), people would be assigned randomly to interventions, so that any difference in outcomes could be attributed to the effects of the intervention. Such an approach to evaluating the effectiveness of interventions is described as 'experimental'. Because of the random selection,

there would be no reason to expect average outcomes to be different for those who received assistance and those who did not, except as a consequence of the assistance. For a range of practical and ethical reasons, there is no history of reliable experimental evaluations in New Zealand, although the approach is widely used in other countries, notably the United States of America.⁶

The approach taken in this paper is to use what are termed “quasi-experimental methods” to control for the fact that expected outcomes for those who receive assistance are likely to be different from expected outcomes for those who do not receive assistance, regardless of the impact of the assistance. The logic of targeting assistance to those most in need, which has been a strong theme in New Zealand labour market policies, makes such differences almost inevitable.

The major problem that we need to deal with is that of selection bias. Individuals differ in their likelihood of prolonged unemployment. Assistance is often targeted at those most at risk, so that we would expect to observe poorer outcomes for the targeted group than for other jobseekers, even if the assistance had no effect on their fortunes. An intervention that is effective at improving outcomes for targeted individuals will narrow the gap in outcomes between those who are assisted and those who are not, but may not completely eliminate it. What we would observe in this case is that assisted individuals would still have poorer subsequent outcomes than would those who are not assisted. The challenge in evaluating the impact of interventions is to separate out the inherent differences in outcomes (selection bias) from the contribution of the intervention (programme effect).⁷

Our aim in this paper is to evaluate the impact of a programme in terms of the “effect of the treatment on the treated” – that is, compare the outcome for an individual who received the treatment with the outcome that *that individual* would have experienced had they not received the treatment.

Note, that estimating the effect of treatment on the treated (TT) is only one of the possible treatment measures of potential interest. Others include the ‘Average Treatment Effect’ (ATE), the ‘Local Average Treatment Effect’ (LATE), and the ‘Marginal Treatment Effect’ (MTE).⁸ The ATE is the estimated effect of treatment for someone randomly chosen to be assigned to treatment. Treatment on the treated estimates the effect for someone chosen randomly from those who received treatment. Both the LATE and MTE capture the effect for those who are at the

⁶ See, for instance various studies by the Manpower Demonstration Research Corporation, documented at www.mdrc.org.

⁷ For useful discussions of methods and concepts, see Heckman, Lalonde and Smith (1999), Angrist and Krueger (1999), and Heckman (2001).

⁸ See Heckman and Vytlacil (2001) for a description of these different methods, and Heckman, Tobias and Vytlacil (2001) for details of implementation within a common framework.

margin of participating. These four treatment effects are the same only under very specific conditions. Heckman and Vytlacil (2001, p 107) state that “If the effect of treatment is the same for everyone with the same observables, or if it varies among people with the same observables but enrolment decisions are not based on this variation, then all of the mean treatment parameters are the same, and there is a single effect of the intervention”.

Given our focus on the effect of treatment on the treated, we face the problem that we do not observe the same individual both receiving and *not* receiving treatment, and this is the central identification problem for the evaluation. The different approaches discussed in this paper are all different ways of ‘solving’ this problem.

Table 1 illustrates the identification problem. The shaded areas in Table 1 show the outcomes that we do *not* observe directly. The first is the outcome that would have occurred in the absence of the treatment, for those who receive the treatment. The second is the outcome that would occur following treatment, for those who *do not* receive the treatment. The true measure of the effect of treatment on the treated is the difference between the two cells in the first row:

$$\alpha = E[Y_{i1} - Y_{i0} | D_i = 1] = E[Y_{i1} | D_i = 1] - E[Y_{i0} | D_i = 1] \quad (1)$$

The problem is that the individuals who receive the treatment may be quite different from those who do not, and the difference in observed outcomes $E[Y_{i1} | D_i = 1] - E[Y_{i0} | D_i = 0]$ may be due to these differences rather than to the effect of treatment. The term “selection bias” is used to describe the difference between the true α (see equation 1) and the difference in observed outcomes.

A naïve estimator

A naïve estimator is simply the difference in observed outcomes between programme participants and non-participants. This corresponds to comparing the two unshaded boxes in Table 1.

$$\begin{aligned} \alpha_n &= E[Y_{i1} | D_i = 1] - E[Y_{i0} | D_i = 0] = \{E[Y_{i0} | D_i = 1] - E[Y_{i0} | D_i = 0]\} \\ &= \alpha + bias \end{aligned} \quad (2)$$

The bias in this estimator is the difference in outcomes that participants and non-participants would have experienced if neither had received any assistance. The alternative methods described in the sections that follow are attempts to derive estimators that are free from bias.

Random assignment to treatment

If the allocation of people into treatment were random, the expected outcomes that would occur in the absence of treatment (Y_{i0}) would be no different for those that do and those that do not receive treatment, because outcomes are independent of getting the treatment ($Y_{i0} \perp D_i$). Because $E[Y_{i0} | D_i = 0] = E[Y_{i0} | D_i = 1]$, the bias in equation 2 is equal to zero, due to the random allocation. The estimator

used to analyse the treatment effect where assignment is random is the naïve estimator (α_n) but it is, in this case, unbiased.

Matching on observables/regression

In the absence of random assignment, we may still identify the programme effect if outcomes for those that do and those that do not receive treatment depend on observed characteristics X_i in the same way. By comparing outcomes (Y_{i1} and Y_{i0}) for individuals that have the same observed characteristics, we can estimate the outcomes that the treatment group would have had in the absence of treatment. In this case, $E[Y_{i0} | X_i, D_i = 1] = E[Y_{i0} | X_i, D_i = 0]$. Substituting this into a version of equation 1 that is conditional on values of X_i yields:

$$\alpha_m = E[Y_{i1} | X_i, D_i = 1] - E[Y_{i0} | X_i, D_i = 0] \quad (3)$$

This is the difference between outcomes from the two observed cells in Table 1, where the comparison is between sets of individuals with the same observable characteristics.

TABLE 1: Matrix of possible outcomes

	Outcome if treatment is not received (Y_{i0})	Outcome if treatment is received (Y_{i1})
Individuals who receive treatment ($D_i = 1$)	$E[Y_{i0} D_i = 1]$	$E[Y_{i1} D_i = 1]$
Individuals who do not receive treatment ($D_i = 0$)	$E[Y_{i0} D_i = 0]$	$E[Y_{i1} D_i = 0]$

Notation: i denotes an individual

$E[A | B]$ denotes the expected value of variable A given that event B occurs.

In practice, this matching on X_i is usually done on a relatively small set of discrete measures such as age, gender, ethnicity, or location.⁹ Both the treatment and control groups are subdivided into cells, each of which represents a unique combination of values of X 's. There will be as many different estimates as there are distinct cells. These different estimates are weighted according to the number of treated individuals in each cell, so that the weighted average represents the average effect for the population of treated individuals.

A variant on the matching technique, which is common in the econometric literature, is to condition on values of X_i within a regression model. Instead of comparing only individuals with exactly the same X_i values, we assume that the X_i 's are linearly related to outcomes, so that $Y_i = \beta_0 + \beta_1 X_i$. We then compare

⁹ As noted above, this has been the most common method of creating a comparison group in the recent New Zealand ALMP evaluation literature.

individuals with the same expected outcome, which is calculated as a linear function of the X_i 's. In practice, an estimate of α is obtained as the coefficient on D_i in a regression of Y_i on X_i and D_i .

If regression matching is applied to the population of participants and non-participants, α_m will estimate the ATE. If applied to the treatment group and a *matched* comparison group, it estimates the TT effect.

Matching on propensity score

In the matching techniques just described, the identifying assumption is that, in the absence of treatment, the *outcomes* for the treatment and comparison groups are the same, except for the differences due to different X_i 's. An alternative assumption is that the *probability* (or 'propensity') of being assigned to either of the groups is the same (random assignment) except for the effect of X_i 's on the probability. Once we have conditioned on these probabilities [$P(X_i)$], we can treat the treatment and comparison groups as randomly selected, so that:

$$\alpha_p = E[Y_{i1} | P(X_i), D_i = 1] - E[Y_{i0} | P(X_i), D_i = 0] \quad (4)$$

where $P(X_i)$ is the probability of being in the treatment group, as a function of the X_i variables. $P(X_i)$ is conventionally referred to as the propensity score. Instead of comparing individuals with the same X_i values (as is the case for matching on observables), we compare individuals with the same propensity score.¹⁰

Having identified a propensity-matched comparison group, we may then use regression methods to allow for the possibility that differences in covariates (X_i) within a group with the same propensity score may explain some of the difference in outcomes. The resulting estimator differs from α_r and from α_p in that it controls for both X and $P(X)$.

$$\alpha_{pr} = E[Y_{i1} | P(X_i), X_i, D_i = 1] - E[Y_{i0} | P(X_i), X_i, D_i = 0] \quad (5)$$

The covariates can affect outcomes both directly, and through their impact on selection. In a regression framework, the two separate effects are identifiable only by functional form, unless there are covariates that affect participation that do not affect outcomes, in which case we would have $P(X_i, Z_i)$. In the absence of such an instrument, the range of variation in covariates within a propensity matched group will be small, and we would not expect this regression matching to change our propensity estimates substantially.

4.2 Methods used for this study

Rather than rely on any single estimate of programme effectiveness, the approach taken in this study is to derive a range of estimators. Our starting point is the

¹⁰ The use of propensity scores was introduced by Rosenbaum and Rubin (1983, 1985).

naïve estimator – the observed difference in outcomes between participants and non-participants.

In order to reduce the bias that we suspect is present in the naïve estimates, we derive a series of regression estimates, using progressively larger sets of covariates. The covariates chosen are described in the results section below. Finally, we derive and present propensity-matched estimates, based on a full set of covariates.

The propensity-matched estimates are the preferred estimates because they more directly model the selection process, and effect matching without the use of information on outcomes. They also allow for convenient diagnostics and sensitivity testing, and have been shown to produce estimates close to experimental estimates.¹¹

Regression matched estimates provide a measure of the effect of treatment on a jobseeker with average characteristics, which may be very different from the characteristics of those who receive treatment. Instrumental variable methods are not used in this paper because of the lack of a plausible instrument – something that is related to participation but does not have an independent effect on outcomes.

We consider the effects of five different types of intervention, and a range of outcome measures. We also examine the sensitivity of our findings to different sub-populations and time periods.

5 Data

The data used in this paper are a subset of the administrative data collected by the NZES between 1 October 1988 and 31 December 1997. The New Zealand Employment Service was the public employment agency, and was also the delivery agency for a wide range of ALMPs during this period. Anyone receiving the unemployment benefit was required to register with NZES. In addition, about 15 percent of jobseekers registered with NZES but were not in receipt of a benefit. The NZES administrative records contain information on each episode of unemployment experienced by registered jobseekers, as well as information about referrals to and participation in various forms of ALMP assistance.

We select treatment and comparison groups on a combination of calendar time and the timing of intervention spells, and refer to the selected group as an ‘intervention cohort’.

¹¹ Dehejia and Wahba (1999) demonstrate the ability of propensity-matched estimates to approximate experimental estimates for the evaluation of a US training programme – the National Supported Work Demonstration, although the strength of this finding has been questioned by Smith and Todd (2001). Heckman *et al* (1999, p 1955) discusses the conditions under which the propensity matched estimates will be identical to experimental estimates.

The treatment group from an intervention cohort contains all jobseekers who started a spell of assistance in a specified three-month period. Prior and subsequent experience is measured from the reference date, which is the date that the intervention spell commences.

Comparison group observations are for unemployment register spells experienced by people who did *not* receive any treatment during the quarter. We use only the subset of register spells that are current at the middle of the quarter. Clearly, the reference date for the comparison group cannot be the date that the intervention spell commences, since they are chosen because they do not commence an intervention spell. We assign the middle of the period as the reference date for the comparison group.¹² This is notionally the date that they would have received treatment.

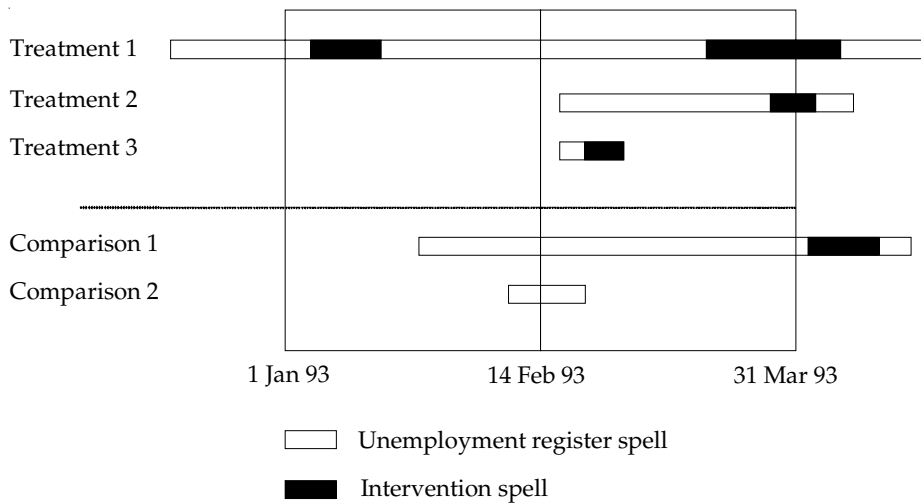
For simplicity, we use the same comparison group for all forms of treatment. Thus, members of the comparison group do not commence a spell of any form of assistance during the three month sample window.¹³ For ease of computation, a 50 percent sample of the common comparison group was used for the results presented in this paper. Other than the expected increase in standard errors, none of the main results changed as a result of this sampling.

Figure 2 illustrates the selection method. Each horizontal bar represents a single jobseeker's experience over time – including spells of unemployment (unshaded bars) or intervention assistance (shaded bar). The length of the bar represents calendar time, and the dates shown are the dates used for selecting the 1993Q1 intervention cohort. The three treatment cases are included because they each contain an intervention that starts within the quarter. In the case of 'Treatment 1', this jobseeker contributes twice to the treatment group (if the interventions were of the same type), or to two different treatment groups (if the interventions are of different types). The two comparison cases are included because they are unemployed on 14 February 1993 – the midpoint of the quarter.

The treatment and comparison groups chosen in this way will differ not only because they have different characteristics, but also because they have different

¹² An alternative approach would be to assign comparison group members a notional reference date, chosen at random from the intervention starting dates for the treatment group. The distribution of reference dates for the treatment and comparison groups would therefore be similar, reducing the potential bias arising from changes over time in unobserved factors. In the current study, the range of reference dates is relatively small – within a three-month period – so we have assigned a common reference date of mid-quarter for the comparison group.

¹³ It would be possible to select a different comparison group for each type of intervention, excluding only those who had not started a spell of that particular type of intervention. This was done in earlier versions of the paper. The results are barely altered by using a common comparison group.

FIGURE 2: Method of selecting intervention cohort

durations of unemployment. By design, selected jobseekers have intervention starting dates/reference dates that fall within the same quarter. They will, however, have different unemployment durations at the time of intervention, due to their different enrolment dates. Rather than force the duration structure of the treatment and comparison groups to be the same, by selecting from a cohort who commenced their unemployment spell at the same time, we control for duration differences by including duration at the reference date as a control in our analyses.

One feature of our method of choosing the treatment comparison group is that the comparison group is 'stock-sampled' (sampled from all jobseekers unemployed at a particular point in time) whereas the treatment groups are 'flow-sampled' (sampled from jobseekers starting an intervention during a period). Stock sampling produces a sample with higher average durations than does flow sampling. Given that most interventions are targeted on the basis of longer durations, we expect this choice of sampling method to reduce some of the selection bias that would otherwise arise.

The reference quarter for most of the analysis is the first quarter of 1993. This period is chosen because it is in the centre of our sample period, and thus provides the longest period of observation prior to *and* subsequent to enrolment. Analysis is also presented for different time periods, to check the sensitivity of results to time-varying factors such as business cycle conditions.

5.1 Range of interventions considered

We analyse five broad types of intervention, and derive estimates of the effectiveness of each of the five types. The categories are:

TABLE 2: List of grouped interventions

<i>Intervention group</i>	<i>Interventions included (with proportion of the 1993Q1 intervention cohort treatment group, where applicable)</i>	
Vacancies:	<ul style="list-style-type: none"> • Successful referrals to unsubsidised jobs (26 percent) 	<ul style="list-style-type: none"> • Unsuccessful referrals to unsubsidised jobs (74 percent)
Interviews:	<ul style="list-style-type: none"> • Work focus interviews (95 percent) • Job Action interviews (4 percent) • Job Action Case Management (0 percent) • Job Action Focus (0 percent) 	<ul style="list-style-type: none"> • Youth action interview (0 percent) • Job Action post-interview
Seminars: (none in 1993Q1 sample)	<ul style="list-style-type: none"> • Job Action • Job Wise • Job Search seminar • Job Club • CV seminar 	<ul style="list-style-type: none"> • Careers Advice • Careers Counselling • Careers Guidance • Enrolment seminar • Miscellaneous seminars
Subsidies:	<ul style="list-style-type: none"> • Job Plus (66 percent) • Taskforce Green (25 percent) • Enterprise Allowance (9 percent) • Enterprise allowance w/ capitalisation (0 percent) • Job Connection (0 percent) • Restart • Special Groups • Development • Self-development 	<ul style="list-style-type: none"> • Wage Subsidy • Project Employment Programme • Contract Work Scheme • Tertiary Employment Assistance • Student Employment Assistance Scheme • Job Opportunities Scheme
Training:	<ul style="list-style-type: none"> • Training Opportunity Programme (93 percent) • Other Training (7 percent) • Job Plus Training (0 percent) • Wahine Pakari (0 percent) 	<ul style="list-style-type: none"> • Youth Action commissioned training • Training in Employment • Referral to Access or TOP
Work experience:	<ul style="list-style-type: none"> • Community Taskforce (99 percent) • Job Link (1 percent) • Conservation Corps (0 percent) 	<ul style="list-style-type: none"> • Youth Services Corps • Job Introduction
Work confidence (none in 1993Q1 sample)	<ul style="list-style-type: none"> • Wahine Ahuru • Limited Service Volunteers • Stepping Stones 	<ul style="list-style-type: none"> • Hikoi ki paerangi • Tane Atawhai • Job Action workshop
Miscellaneous interventions	<ul style="list-style-type: none"> • Referral to external agency (100 percent) 	<ul style="list-style-type: none"> • Māori youth

- vacancies: referrals to unsubsidised job vacancies;
- interviews: interviews with the jobseeker, carried out by NZES staff;
- subsidies: payment of a subsidy to employers to hire unemployed jobseekers;
- training: provision of training to improve jobseekers' chances of gaining unsubsidised employment;
- work experience: providing jobseekers with work experience.

Although referrals to vacancies are not generally included under the heading of ALMP, we will, for convenience, refer to them as such in this paper. Table 2 provides a full list of the interventions from the study period that fall under each of the five headings. The Table also shows the proportion of the observed interventions from each category that relate to each individual intervention. The proportions are based on the sample used in most of the analyses (from the 1993Q1 intervention cohort). Note that there are no seminars or work confidence courses observed in the sample, because of the way that these interventions were (not) recorded in the administrative dataset, or because interventions had not been introduced in the first quarter of 1993. We also exclude from our analysis the group 'Miscellaneous interventions', which, for the 1993 cohort, comprise only referrals to other agencies.¹⁴

5.2 *Intervention cohort summary statistics*

Table 3 summarises the characteristics of people in the intervention cohort. We present summaries for the comparison group and for each of the treatment groups. Four sets of characteristics are presented: first, a range of demographic characteristics; second, summary measures of prior experience of unemployment; third, summary measures of subsequent unemployment and, finally, information about prior and subsequent receipt of assistance.

It is clear from the characteristics of the various treatment groups that interventions are targeted in notably different ways. Those being referred to vacancies are more likely to be Pākehā, female, highly qualified, have few identified barriers to employment, and have fewer weeks of prior unemployment or contact than those not being referred. Those receiving training are particularly young and poorly qualified. Māori are over-represented in the treatment groups for training and work experience.

¹⁴ There were only 511 such interventions recorded in the 1993 cohort. Those referred to other agencies spent less time subsequently in contact with NZES than did the comparison group – a finding evident in naïve, regression and propensity-matched estimates. Because of the small size of the treatment group, estimates were volatile and mostly statistically insignificant.

TABLE 3: Summary statistics - 1993Q1 intervention cohort

<i>Characteristic</i>	<i>Comparison group</i>	<i>Vacancy treatment group</i>	<i>Interview treatment group</i>	<i>Subsidy treatment group</i>	<i>Training treatment group</i>	<i>Work experience treatment group</i>
<i>Number of observations</i>	20,644	59,767	32,285	7,988	12,576	1,786
<i>Demographics</i>						
<i>Percentage male</i>	67	56	68	74	53	61
<i>Percentage Māori</i>	31	20	30	26	39	34
<i>Percentage Pacific peoples</i>	8	6	8	4	10	6
<i>Age</i>	31.3	30.0	31.8	29.6	26.0	30.4
<i>Percentage no qualifications</i>	51	34	51	45	62	48
<i>Percentage tertiary qualifications</i>	5	6	4	5	1	6
<i>Barriers</i>	5	4	8	6	6	7
<i>Prior outcomes</i>						
<i>Unemployment duration at reference date</i>	62.6	22.1	38.1	26.3	14.9	48.2
<i>Unemployment in prior 12 months (weeks)</i>	34.1	25.7	40.6	38.8	17.2	37.6
<i>Unemployment in prior 36 months (weeks)</i>	76.0	48.7	72.4	76.8	41.4	77.1
<i>Contact in prior 12 months (weeks)</i>	35.8	27.9	42.1	42.4	22.2	40.2
<i>Contact in prior 36 months (weeks)</i>	79.2	52.5	76.4	83.5	48.3	83.0
<i>Subsequent outcomes</i>						
<i>Unemployment in next 12 months</i>	33.6	23.3	34.9	12.5	21.5	37.5
<i>Unemployment in next 36 months</i>	69.6	44.9	72.9	38.2	66.2	78.3
<i>Contact in next 12 months</i>	35.6	27.0	39.2	35.2	39.4	43.7
<i>Contact in next 36 months</i>	76.0	51.9	82.0	65.1	95.0	92.1

continued

TABLE 3: continued

<i>Characteristic</i>	<i>Comparison group</i>	<i>Vacancy treatment group</i>	<i>Interview treatment group</i>	<i>Subsidy treatment group</i>	<i>Training treatment group</i>	<i>Work experience treatment group</i>
<i>Percentage with prior assistance (in previous 36 months)</i>						
<i>Vacancies</i>	27	67	39	54	33	45
<i>Interviews</i>	51	38	47	68	50	63
<i>Subsidies</i>	9	13	11	22	8	16
<i>Training</i>	10	13	14	17	36	18
<i>Work experience</i>	3	4	3	7	4	22
<i>Percentage with subsequent assistance</i>						
<i>Vacancies</i>	35	100	48	40	53	61
<i>Interviews</i>	43	37	100	31	54	57
<i>Subsidies</i>	19	24	26	99	25	40
<i>Training</i>	12	10	17	7	94	16
<i>Work experience</i>	4	5	6	4	7	99

Source: See the description in the text of the selection criteria for the comparison and intervention groups.

The differences in outcomes that are evident in the ‘subsequent assistance’ block of the table are the basis of naïve estimates, as described above. These are reproduced in Table 4, which is discussed below.

The relative sizes of the treatment groups reflect the nature of active labour market policies in New Zealand in 1993.¹⁵ Referrals to vacancies are the most numerous form of assistance (59,767 occurrences). The number of interviews is also large (32,285). These were mainly ‘Work-focus Interviews’, which were scheduled on the basis of unemployment duration for almost all long-term jobseekers. Training spells (12,576) were slightly more numerous than were subsidy spells (7,988). The number of spells of Work Experience assistance commenced in the quarter (1,786) is the smallest group, and comprised almost entirely Community Task Force participants.

¹⁵ Martin (2000, Table 5) shows the substantial change in the focus of New Zealand’s active labour market policy expenditures between 1985 and 1995 – the percentage due to direct public sector job creation dropping from 65 percent to 4 percent (compared with a change in the Organisation for Economic Cooperation and Development (OECD) average from 17 percent to 14 percent) and the percentage going to labour market training rising from 16 percent to 46 percent (compared with the OECD average rise from 23 to 27 percent).

5.3 Variable definitions

Outcome measures

A significant weakness in the outcome measures available for the study is that we do not observe whether someone who is absent from the administrative records is absent because they have a job, or because they are no longer actively seeking employment. Absence from the records is treated as a favourable outcome, and continued active contact as unfavourable.¹⁶

We consider two different measures of subsequent experience to capture outcomes that are potentially influenced by the interventions. First, we count the number of weeks that the jobseeker is registered as unemployed during the 36 months following the reference date. Second, we count the number of weeks that the jobseeker is either registered as unemployed *or* in receipt of assistance through NZES over the same period (referred to as 'any contact'). Some forms of assistance (for example, receipt of a wage subsidy) involve the jobseeker leaving the unemployment register. By our first measure, time spent in receipt of a wage subsidy shows as a favourable outcome. In the second, it does not.

For the 'any contact' measure, we also consider outcomes over the subsequent six months and 12 months, rather than just over 36 months.

For a small number of intervention spells, we do not observe end-dates. For instance, there are missing or clearly wrong data on the actual durations of around 3.5 percent of Training Opportunity Programme (TOP) courses. New Zealand Employment Service data contain information about referrals, but the training courses were administered by a different agency, so NZES records do not contain full course information.¹⁷ Where intervention end-dates are missing, we have used average durations for each specific intervention to impute an end-date.¹⁸ This imputation does not significantly affect the registered unemployment outcome measure, although it will enter directly into the contact weeks outcome measures. There should be no bias resulting from the use of an average duration measure but the precision of the impact estimates will be reduced.

¹⁶ This would be an appropriate approach if we were interested only in a narrow fiscal approach. Our choice of measure is a consequence of the unavailability of a better measure, not an indication that we are solely interested in fiscal considerations.

¹⁷ Subsequent to this paper being prepared, the Department of Work and Income has incorporated additional information from Skill New Zealand into its data to rectify this gap.

¹⁸ We have also adjusted end-dates where we have additional information. For instance, where we observe a job-seeker commencing a different spell before the imputed duration, we set the intervention end-date to the observed spell start-date. Negative durations are set to zero. Durations that are more than two standard deviations above median duration for the specific intervention are set to the average duration. For the 3.5 percent of TOP spells with imputed durations, one-third have missing end-dates, and two-thirds have excessively long recorded durations.

TABLE 4: Summary of interventions - 1993Q1 intervention cohort

	<i>Vacancies</i>	<i>Interviews</i>	<i>Subsidies</i>	<i>Training</i>	<i>Work experience</i>
Sample size					
Treatment group (#)	59,767	32,285	7,988	12,576	1,786
Comparison group (#)	20,644	20,644	20,644	20,644	20,644
Outcome = weeks of contact in subsequent three years					
Treatment group	51.9	82.0	65.1	95.0	92.1
Comparison group	76.0	76.0	76.0	76.0	76.0
Raw difference	-24.2	6.0	-10.9	19.0	16.1
Outcome = weeks of registered unemployment in subsequent three years					
Treatment group	44.9	72.9	38.2	66.2	78.3
Comparison group	69.6	69.6	69.6	69.6	69.6
Raw difference	-24.7	3.3	-31.4	3.3	8.7

Notes: 'Contact' time is time spent registered as unemployed *or* in receipt of some form of active labour market assistance.

We are also aware that there are instances where a wage subsidy intervention is recorded for an individual, but where no subsidy payments were made. This may be due to an employer not claiming the subsidy, or to the individual never starting the subsidised position. The intervention should therefore be thought of as a *referral* to a subsidised position. To the extent that jobseekers do not start the position, and therefore return to the register, the estimated impact of wage subsidies will appear less effective.¹⁹

Table 4 shows the actual three-year outcomes, and the raw differences in these outcomes for treatment and control groups for each intervention type. The observed raw differences are the 'naïve estimates' referred to above. They make no allowance for differences in characteristics of those receiving treatment and those not receiving treatment. For instance, jobseekers who are referred to vacancies experience 24.2 fewer weeks of contact (and 24.7 fewer weeks of unemployment) in the three years following the referral. The comparison group spent 76.0 of the 156 weeks, or almost 50 percent of their time with some sort of contact or assistance. The vacancy treatment group experienced only 51.8 weeks, or a third of their time.

¹⁹ An indication of the extent of this problem is evident in the final (Subsequent Assistance) block of Table 3. For subsidies, training and work experience forms of assistance, the summary statistics show the proportion of the respective samples with spells of at least one day. Only 99 percent of the subsidy and work experience treatment groups, and only 94 percent of the training treatment group had intervention spells of at least one day.

Covariates

The summary statistics in Table 3 summarise the main sources of variation used to create covariates. The specific form of covariates is described more precisely in this section.

Demographics

Dummy (0–1) variables are included for sex, six age categories, seven education categories, five ethnicity categories, 14 locations, and 10 occupational groupings. We also include information from the administrative records of the barriers faced by jobseekers. This is included in the form of seven dummy variables.²⁰

Prior experience

The jobseeker's prior unemployment experience is included as a spline of weeks of unemployment. The spline is piecewise linear over the ranges zero to six months, six to 12 months, 12 to 18 months, 18 to 24 months and 24 to 36 months. We also include unemployment duration at the reference date. This is included as a fourth order polynomial.²¹

To capture prior intervention experience, we include a set of six dummy variables indicating whether the jobseeker received each form of assistance at any time in the previous three years.

²⁰ *Age*: [15–19; 20–24; 25–29; 30–39; 40–49; 50+; n/a]; *Education*: [No formal school qualifications; less than three School Certificate subjects; three or more School Certificate subjects; Sixth Form Certificate/University Entrance; other school qualifications; post-secondary or trade qualifications; degree or professional qualifications; n/a]; *Ethnicity*: [New Zealand European/ Pākehā; Sole Māori; Mixed Māori; Pacific Islands peoples; Other; n/a]; *Location*: [Northland; Auckland North; Auckland Central; Auckland South; Waikato; East Coast; Bay of Plenty; Central; Taranaki; Wellington; Nelson; Canterbury; Southern; n/a]; *Occupation*: NZES occupation codings for preferred occupation; *Barriers*: Psychiatric disability; physical disability; intellectual disability; Education/learning/literacy/English for Speakers of Other Languages (ESOL) barrier; Alcohol and drugs barrier; Multiple disability; no barrier. One dummy variable is omitted for each set of characteristics, to avoid perfect collinearity.

²¹ Initially, we tested an unrestricted functional form, allowing for a different intercept for each single week of duration up to six months, and for each four-week period thereafter. This entailed including 58 dummy variables. Inspection of the coefficients on these dummies suggested that the relationship between our chosen outcome variables and current duration was close to linear. The fourth order polynomial was chosen as a (perhaps excessively) flexible functional form to capture the relationship.

6 Results

6.1 Estimates of the impact of interventions

Controlling for observed characteristics (regression estimates)

The first method of reducing the influence of selection bias is to use regression, as outlined above. This removes that part of the selection bias that is (linearly) related to observed characteristics. Table 5 reports estimates from regressions of the form:

$$Y_i = \delta + \alpha_j D_{ij} + X_i \beta_j + \varepsilon_i \quad (6)$$

where i = individual

j = intervention type

D_{ij} = dummy variable indicating that individual i is in the treatment group for treatment j

X_i = characteristics of the individual (unemployment and intervention history, age, qualifications, location, occupation employment barriers).

A separate regression is run for each choice of intervention type (j). Each of these regressions contains observations on the same comparison group, and on the treatment group for the selected intervention type. The j subscripts on the α and β coefficients reflect these sample differences. The main coefficient of interest is α that shows the effects of being ‘treated’ with assistance of type j . Coefficient α can be interpreted as the average difference in outcomes between those who received assistance and those who did not, controlling for the effect of other covariates (the ATE).

The regression estimates calculate the difference that receiving assistance makes by comparing outcomes for the treatment group with outcomes for the comparison group, controlling for the average relationship between observed characteristics and outcomes (averaged across all observations in the regression). This reduces the influence of selection bias, to the extent that differences between outcomes of participants and non-participants depend on differences in the characteristics that are included as covariates. Selection bias will still be present if:

- it depends on unobserved characteristics; or
- the relationship between covariates and outcomes is different for participants and non-participants.

The first row of Table 5 shows the number of observations in each treatment group. As noted earlier, the same comparison group of 20,664 observations is used for all treatments. Each reported coefficient in Table 5 is from a separate regression. Each column reports results for a particular choice of treatment and comparison group. The first column, for instance, relates to a regression sample containing the comparison group and the treatment group for the vacancy treatment. Each row relates to a particular choice of covariates, starting with the naïve estimate (no covariates) in the first row. Subsequent rows progressively add

TABLE 5: Estimated impacts of interventions – any contact

Dependent variable: Weeks of *any contact* in the 36 months after the reference date

Estimate (standard error)

[Adj R² for regression]

	Vacancies	Interviews	Subsidies	Training	Work experience
Number in Comparison group = 20,664					
Number in Treatment group	59,767	32,285	7,988	12,576	1,786
Naïve estimate (raw difference)	-24.2 (0.4) [4.9]	6.0 (0.4) [0.3]	-10.9 (0.6) [1.0]	19.0 (0.6) [3.4]	16.1 (1.3) [0.7]
Regression Estimate I (with prior interventions as covariates)	-20.3 (0.4) [15.0]	7.2 (0.4) [5.8]	-15.6 (0.7) [8.3]	17.6 (0.6) [8.4]	11.2 (1.2) [10.0]
Regression Estimate II (after adding prior register spells and current duration)	-9.9 (0.4) [21.8]	9.0 (0.5) [11.4]	-6.5 (0.7) [16.5]	26.2 (0.6) [13.7]	15.4 (1.2) [17.8]
Regression Estimate III (adding demographics, location and labour market)	-7.9 (0.4) [26.0]	8.6 (0.5) [16.5]	-6.2 (0.7) [21.6]	20.1 (0.7) [18.9]	15.1 (1.2) [22.8]
Propensity Estimate	-11.7 (0.4) [1.2]	6.3 (0.4) [0.4]	-13.6 (0.6) [1.7]	19.4 (0.5) [3.9]	14.6 (1.2) [0.6]
Propensity Estimate (with regression adjustment)	-7.5 (0.3) [21.4]	8.8 (0.4) [14.3]	-8.9 (0.6) [25.6]	21.4 (0.5) [17.6]	15.6 (1.1) [23.2]

Notes: Each cell in this table is from a separate regression. Cells in the same column relate to outcomes for the intervention specified at the head of the column. Cells in the same row are derived using the same estimation technique and specification.

'Contact' time is time spent registered as unemployed *or* in receipt of some form of active labour market assistance.

a richer set of covariates. The second row adds prior interventions, the third adds prior register spells and current unemployment spell duration. The fourth row adds demographic, location and labour market variables. Each coefficient is presented with its estimated standard error, and the adjusted R² from the regression from, which it is taken.

Table 6 provides a comparable analysis using as the outcome variable the number of weeks that the jobseeker was registered as unemployed during the three years following the reference date (as defined above).

Referrals to vacancies

The first coefficient in Table 5 (-24.2) indicates that jobseekers who were referred to vacancies during the first quarter of 1993 had, on average, 24.2 fewer weeks of subsequent contact with NZES than did those who were not referred to a vacancy. We suspect that those who are referred to vacancies are not a random sample of jobseekers, and that some of this better subsequent outcome is due to the better prospects that the treatment group would have had even if they had not received the treatment. If this were true, we would expect that adding covariates would diminish the estimated impact of the treatment, to the extent that the covariates are correlated with the factors that explain the inherently better prospects of the treatment group.

The second cell in the first column shows that differences between the treatment and comparison groups in the receipt of assistance over the previous three years can explain some of the raw difference in outcomes. The estimated impact is reduced to 20.3 weeks. Adding prior register spells and current duration reduces the estimate significantly - to 9.9 fewer weeks of contact. Finally, adding the full set of covariates reduces the estimate further, to 7.9 weeks. Selection on observable characteristics can account for about 70 percent of the raw difference in outcomes, although the final estimate of 7.9 fewer weeks of contact still represents a favourable impact of referral to vacancies.

Interviews

As noted above, jobseekers were required to attend interviews when they reached a particular unemployment duration. The selection bias is therefore not expected to be large, except to the extent that the duration profile of the treatment and comparison groups differs.

The estimated impact of interviews does not appear to be as strongly influenced by selection on observables. The naïve estimator is that those receiving assistance in the form of interviews experience 6.0 *more* weeks of contact than do those in the comparison group. This rises slightly to 8.6 weeks even when a full set of covariates is added, suggesting that those who attend interviews have observable characteristics that are associated with better outcomes. Controlling for differences in observables removes this discrepancy, making interviews look less favourable. The implied effect of interviews is that they lead to poorer outcomes.

Subsidies

Subsidies appear to improve subsequent outcomes. The naïve estimate of 10.9 fewer weeks of contact for the treatment group than for the comparison group increases to 15.6 fewer weeks when we control for previous assistance, suggesting that the jobseekers receiving subsidy assistance have intervention histories that are associated with poorer outcomes. The naïve estimator looks less favourable

because it reflects the effects of unfavourable intervention histories. As in the case of vacancies, controlling for prior register spells and current duration reduces significantly the estimated impact of subsidies, in this case to 6.2 fewer weeks of contact. Table 3 shows that the average current duration for the subsidy treatment group is relatively low. To the extent that shorter prior duration is associated with better outcomes, the naïve estimator will be biased, showing an excessively favourable impact of subsidy assistance. The 'full-specification' estimate of 6.2 fewer weeks nevertheless suggests that subsidies have a favourable impact on subsequent outcomes.

This finding is consistent with overseas research into active labour market policies that finds that subsidies improve the labour market prospects of the jobseekers receiving subsidies. Note that our measure of outcomes does not take into account the high dollar cost of providing subsidies, or the likelihood that subsidised jobs take the place of unsubsidised jobs.²²

Training

The naïvely estimated impact of training assistance is that it is associated with 19.0 *more* weeks of contact in the subsequent three years. This estimate is not greatly changed by the addition of covariates – the final regression estimate of 20.1 more weeks of contact experienced by the treatment group still shows much poorer outcomes for jobseekers who receive training than for those who do not.

Work experience

The picture is similar for work experience assistance. The raw difference of 16.1 weeks more contact for the treatment group than for the comparison group is reduced to only 15.1 weeks more contact when we use the full set of covariates.

Impact using registered unemployment as the outcome measure

Table 6 repeats the same analysis, but uses weeks of registered unemployment rather than contact. The results are similar to those obtained in Table 5, with the notable exceptions of the subsidy and training interventions. Both these forms of intervention entail jobseekers leaving the unemployment register while in receipt of assistance. The effects of this fact are clearly evident in Table 6. Whereas subsidies are estimated to reduce weeks of contact by 6.2, the reduction in weeks of unemployment is about 20 weeks greater, at 26.3 weeks. Similarly, for training forms of assistance, the 20.1 more weeks of contact are clearly not weeks spent

²² Martin (2000) reports that most evaluations show that subsidies have both large deadweight effects (that is, employers use the subsidy to hire workers they would otherwise have hired anyway) and displacement effects (many subsidised hires displace others who would have been hired in the absence of the subsidy).

TABLE 6: Estimated impacts of interventions – registered unemployment

Dependent variable: Weeks of *registered unemployment* in the 36 months after the reference date

Estimate (standard error) [Adj R ² for regression]	Vacancies	Interviews	Subsidies	Training	Work experience
Number in Comparison group = 20,664					
Number in Treatment group	59,767	32,285	7,988	12,576	1,786
Naïve estimate (raw difference)	-24.7 (0.3) [6.0]	3.3 (0.4) [0.1]	-31.4 (0.6) [8.3]	-3.3 (0.5) [0.1]	8.7 (1.2) [0.2]
Regression Estimate I (with prior interventions as covariates)	-20.8 (0.4) [14.3]	4.8 (0.4) [5.2]	-35.6 (0.6) [14.7]	-4.3 (0.5) [5.8]	4.8 (1.2) [9.3]
Regression Estimate II (after adding prior register spells and current duration)	-10.9 (0.4) [21.5]	5.6 (0.5) [11.4]	-26.8 (0.7) [22.7]	5.9 (0.6) [12.0]	9.0 (1.1) [17.6]
Regression Estimate III (adding demographics, location and labour market)	-9.1 (0.4) [25.5]	5.0 (0.5) [16.5]	-26.3 (0.7) [27.4]	2.1 (0.6) [16.3]	8.7 (1.1) [22.5]
Propensity Estimate	-12.7 (0.3) [1.8]	4.0 (0.4) [0.2]	-32.5 (0.6) [10.0]	-1.9 (0.5) [0.04]	8.2 (1.2) [0.2]
Propensity Estimate (with regression adjustment)	-9.3 (0.3) [20.2]	6.0 (0.4) [14.1]	-28.8 (0.5) [30.6]	0.9 (0.5) [21.8]	9.0 (1.0) [21.8]

Notes: Each cell in this table is from a separate regression. Cells in the same column relate to outcomes for the intervention specified at the head of the column. Cells in the same row are derived using the same estimation technique and specification.

'Contact' time is time spent registered as unemployed *or* in receipt of some form of active labour market assistance.

unemployed. Training has a small effect on subsequent unemployment experience (2.1 more weeks). The imputation of duration, which was described above, affects training interventions but not subsidy interventions. For training, the average difference between unemployment weeks and contact weeks reflects the imputation, although we do not expect this to have caused any bias.

For subsidy, training and work experience interventions, the difference between the 'any contact' results and the 'registered unemployment' results is in

part because of the time spent receiving the assistance. This is counted as contact time but not as time registered as unemployed. For instance, those receiving assistance in the form of subsidies spent on average 21.1 weeks in the subsidised placement. This is almost exactly the difference between the (naïve or regression) estimated impacts in Tables 5 and 6 (20.5 weeks or 20.1 weeks respectively). Note that the raw 10.9 fewer weeks of contact that the subsidy treatment group received (relative to the comparison group) occurred despite the fact that the average duration on subsidy was 21.1 weeks.

In contrast, the average length of a training spell was 10.4 weeks. This is too short to account for the difference in contact time between treatment and control groups (19.0 weeks) or the difference between contact and register outcome measures (22.3 weeks). The results imply that those who received training assistance subsequently spent about the same amount of time registered as unemployed as did the comparison group, but more time receiving other forms of assistance that removed them from the register.

Finally, for work experience assistance, the raw difference between treatment and control of 16.1 more weeks of contact was around the same as the average length of assistance (16.5 weeks). However, the treatment group spent around nine weeks more time registered as unemployed.

In much of the subsequent analysis, we focus on weeks of contact as the more meaningful measure of outcome. The unemployment measure is presented here because it is a commonly used outcome measure, and is therefore useful as a comparison. It is also of interest to anyone doing programme costings, because the effect of interventions on the fiscal cost of unemployment (unemployment benefit payments) can then be counted separately from the direct programme costs.

Choosing a comparable sub-group from the intervention cohort (propensity matching)

All of the estimates presented so far have controlled only for regression matching on observables. In the final two rows of Tables 5 and 6, we present propensity-matched estimates, as an alternative way of controlling for selection bias. As with the regression matching, it controls only for differences in observables, but uses the information in a different way.

We will discuss the results for the first intervention type, vacancies, in more detail, to clarify the method, and then summarise results for other forms of intervention.

The first step in deriving a propensity-matched estimate is to calculate the propensity score ($P(X_i)$ in the notation above). Using the same definition of treatment and comparison groups as used for the regression matching, we estimate the probability that each observation is in the treatment group. To do this, we use a logit regression, with the same full set of covariates as used for regression matching, again entered as main effects (that is, without interactions). We can then compare outcomes for those in the treatment group with outcomes

for those in the comparison group, giving most weight to those who were most likely to receive assistance (but did not). The specification for the logit regression is:

$$P[D_j = 1] = \frac{e^{X_i\beta}}{1 + e^{X_i\beta}} \quad (7)$$

where the X_i are the same as those used in the regression model outlined in the previous equation.

Figure 3 helps to illustrate the method. Having obtained a predicted probability of treatment for each observation, we can rank jobseekers according to this prediction. The horizontal axis is the predicted probability, or propensity score. The two lower lines (using the right axis) show the densities for the treatment and comparison samples – the proportion of each sample with a particular propensity score.²³ The solid lower line shows that the treatment group is concentrated at the upper end of the graph – with high estimated probabilities of being referred to a vacancy. The comparison group is more evenly spread across the propensity range. The treatment group accounts for 74.3 percent of the full sample, so the regression-matching coefficients are already weighted towards the comparison group. The weighting for the propensity-matched estimate is based on the density of the treatment group. The most weight is therefore given to outcome differences between those in the treatment and comparison groups with high treatment propensities.

The formula for the propensity-matched estimate is thus a weighted average of the form:

$$\hat{\alpha}_p = \sum_i \Delta^i \lambda_T^i \quad (8)$$

where i indexes the propensity scores;

Δ^i = the difference in outcomes between treatment and comparison groups with propensity score i ;

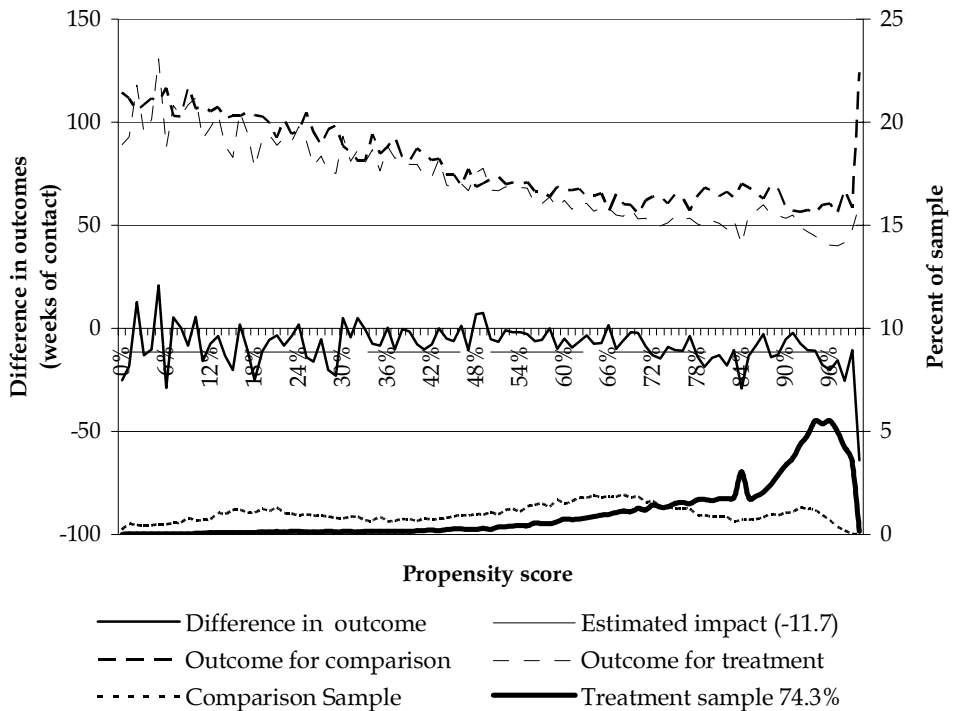
λ_T^i = the proportion of the treatment group that has propensity score i .

The upper two lines in Figure 3 show, for each propensity score, the average outcome for the treatment group (solid line) and the comparison group (dotted line). Both lines slope downwards, implying that jobseekers with a high estimated propensity to be referred to vacancies had lower subsequent contact time, whether or not they were referred to a vacancy.

The solid line in the centre of the graph, labelled ‘difference in treatment’ is the difference between outcomes for treatment and comparison groups with the

²³ The predicted probability is rounded to the nearest percentage point, so there is a maximum of 100 possible distinct values. The approach thus uses non-overlapping rectangular kernels of width 0.01, centred on multiples of 0.01. There are many alternative and more refined ways in the literature of smoothing estimates across the propensity domain.

FIGURE 3: Propensity score estimation for vacancies



same propensity score. This is a measure of impact (D^i). Note that, although the line fluctuates, there is no clear slope, implying that referrals to vacancies appear to have the same favourable effect on subsequent outcomes regardless of how likely a jobseeker is to be referred to vacancies. There appears to be little evidence that those who are referred to vacancies are those who are likely to benefit most.

The light dashed horizontal line that passes through the impact line is our estimate of (α_p) – the weighted average of the differences, using the density of the treatment group as weights. The estimate is -11.7 weeks. This estimate is shown in Table 5, in the block labelled ‘Propensity estimate’. This is significantly different from the naïve estimate of -25.8, but very similar to the regression-matched estimate of -7.8 shown in the row above. Note that the adjusted R-squared for the propensity estimates is very low. Most of the variation in outcomes occurs within groups of jobseekers with the same propensity score and the same treatment status (treatment or comparison group). This is not surprising given that there are only (up to) 101 propensity groups, and between about 22,000 and 80,000 observations. What is more significant is that the treatment effects are estimated with a level of precision comparable with that of the regression estimates.

In practice, our estimation of α_p is implemented by carrying out a weighted regression of outcomes on a constant and a dummy variable denoting member-

ship of the treatment group. Each observation for the treatment group has a weight of one. The weight for a comparison group observation with propensity score i is $W_i^C = P_i^T / P_i^C$ where P_i^T is the proportion of treatment group members with the score i , and P_i^C is the comparable proportion for the comparison group.

Propensity matching with regression matching

The propensity-matched estimates differ from the regression-matched estimates in two respects. First, the regression-matched estimates control for a linear relationship between covariates and outcomes, which is assumed to be constant across treatment and comparison groups, whereas the propensity-matched estimates control for a relationship between covariates and the probability of being in the treatment group, with the assumption that pre-intervention outcomes are equal for jobseekers with the same propensity score. As noted in Section 4.1, the conditions under which each of these provides an unbiased estimate of the treatment effect differ.

Second, the estimates differ in the population for which the treatment effect is being estimated. The regression-matched estimate is an ATE estimator for a population with the characteristics of the selected sample,²⁴ whereas the propensity-matched estimate is weighted to provide an estimate for a population with the characteristics of the treatment group, and is thus a TT estimator.

It is possible to combine elements of the two estimators to obtain a propensity-matched (TT) estimate that controls for the relationship between covariates and outcomes. The final block of Table 5 contains propensity-score estimates that have been adjusted by carrying out a propensity-weighted regression. This is a weighted regression, with each observation weighted according to the propensity-score distribution of the treatment group. The regression also allows for a separate intercept for each propensity group, so that the estimated relationship between covariates and outcomes is based on within-propensity group variation.

The regression equation is an augmented and weighted version of equation 6, and is of the form:

$$Y_i = \delta + \alpha_j D_{ij} + X_i \beta_j + \sum_k \pi_k P_k + \varepsilon_i \quad (9)$$

where k = index of propensity-score value

P_k = dummy variable indicating that the individual has an estimated propensity score of i .

Weights are those used for the propensity-matched estimates.

Because of the absence of a convincing instrument (a variable that influences participation but does not affect outcomes) the same covariates are used in the

²⁴ The regression-matched estimator provides an estimate of the average treatment effect for a somewhat arbitrary population, namely one where the proportion of treatment and comparison group members is the same as that of our selected sample (with selection as described at the beginning of Section Five and shown in Figure 2).

regression as were used in the logistic regression that produced the propensity scores. Therefore, there is limited variation in the covariates for observations with the same propensity score.

The estimates from the 'propensity and regression' method are generally fairly close to the regression estimates, suggesting that outcomes, and the relationship between outcomes and covariates, do not differ greatly between different propensity groups. The picture of the effectiveness of different forms of assistance is remarkably similar, whether we use regression, propensity, or propensity-regression estimates. In the interests of conciseness, subsequent tables will focus primarily on a single estimator – the propensity-matched estimator, although none of the key inferences in the paper would be altered if we were to rely on regression or propensity-regression estimates.

The graphs in Figures 3 to 7 summarise the patterns of outcome differences and sample densities that generate the propensity estimates. In no case is there a significant gradient that would suggest that interventions are more effective for the sort of jobseekers who receive assistance. In the case of subsidies, there is a slightly more favourable impact for those with a high probability of receiving a subsidy, but the pattern is not strong.

FIGURE 4: Propensity score estimation for interviews

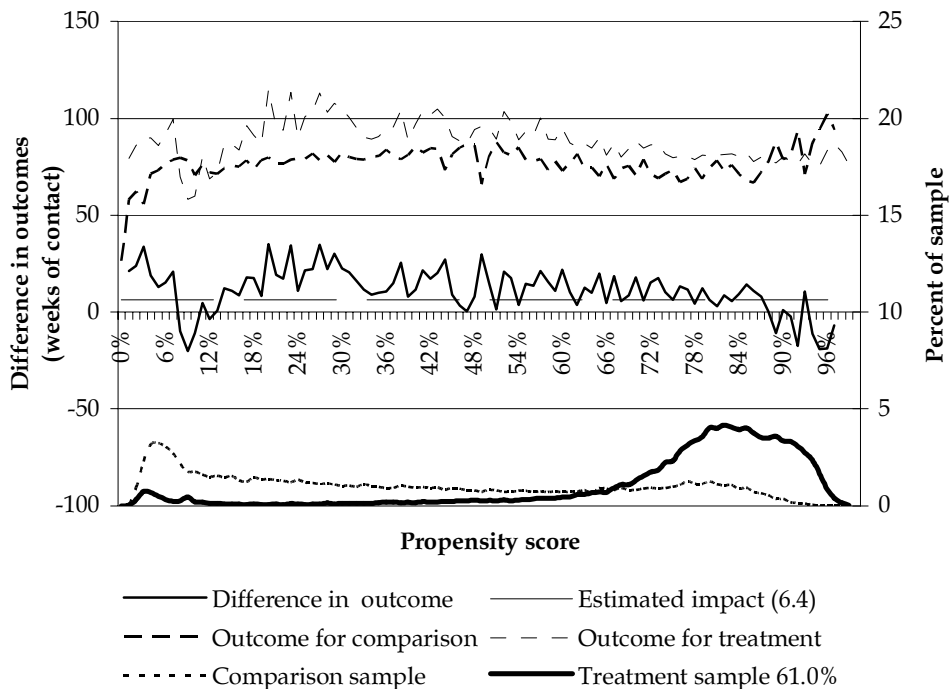


FIGURE 5: Propensity score estimation for subsidies

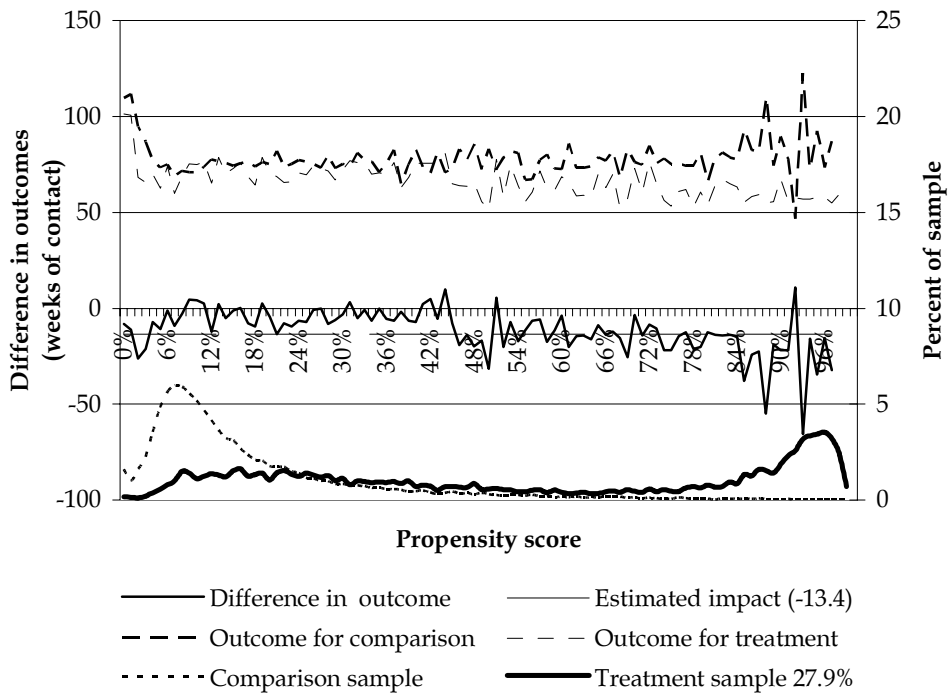


FIGURE 6: Propensity score estimation for training

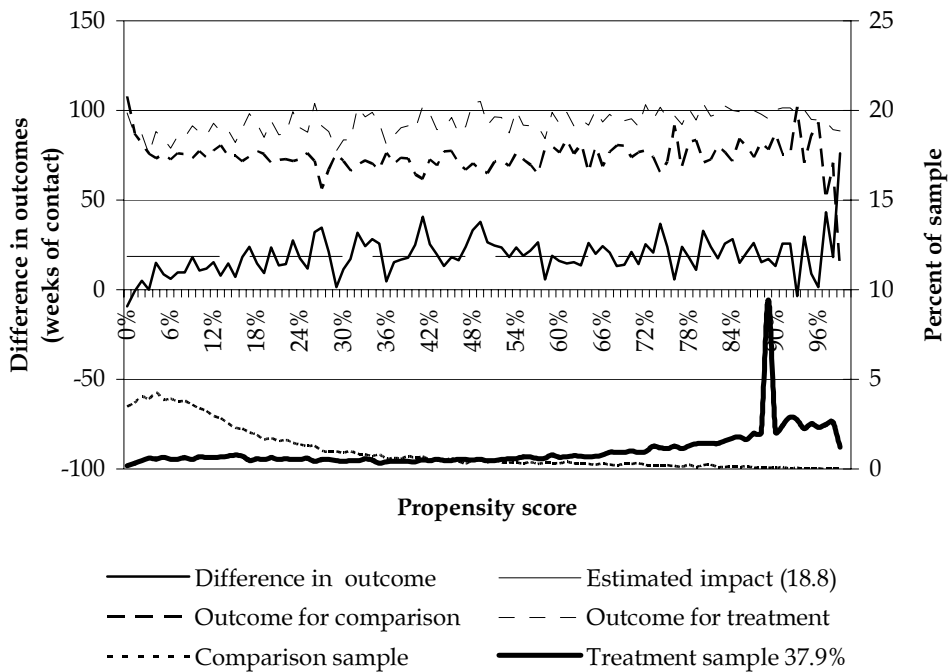
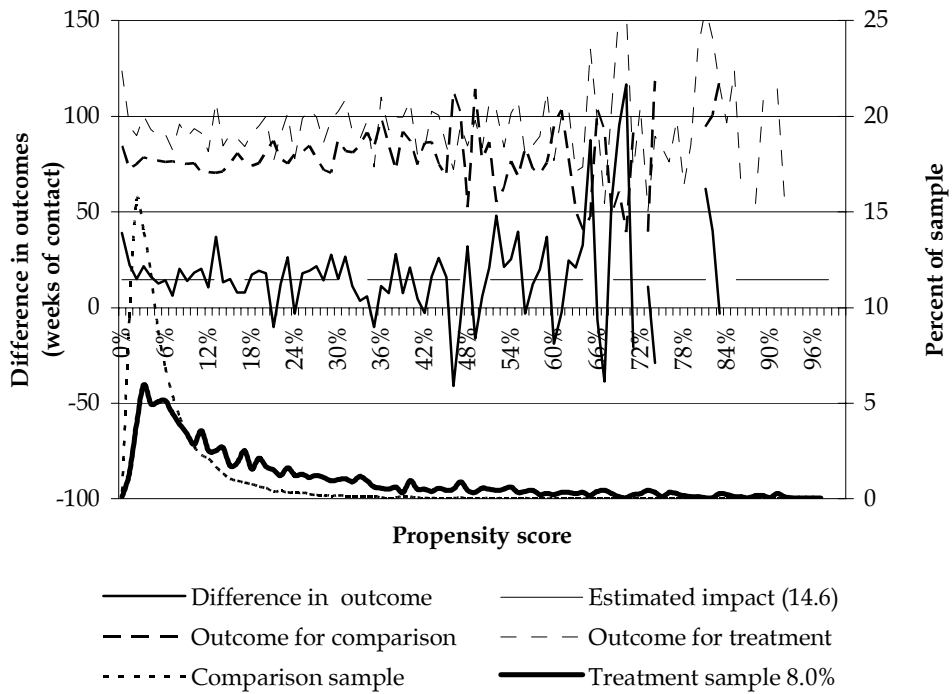


FIGURE 7: Propensity score estimation for work experience



6.2 Stability of results

In this section, we examine whether the general inferences reported in Section 6.1 hold for sub-populations, other samples, other outcome measures and using different covariates.

The discussion and results presented in this section provide some broad indicators of 'what works, and for whom'.

Is the impact different for different sub-groups?

Interventions may be more or less effective for different groups of jobseekers.²⁵ In this section, we consider six broad dimensions of jobseeker characteristics – gender, age, ethnicity, qualifications, prior unemployment duration and location.

By running separate analyses on different sub-samples, we are allowing the relationship between covariates and propensities to be different for different sub-groups, rather than weighted towards the relationship that holds for the average member of the treatment group.

In general, differences across sub-groups in estimated effectiveness for the different forms of assistance are not great. Because estimates for sub-groups are

²⁵ Martin (2000) and Robinson (2000) provide useful summaries of key findings.

based on smaller sample sizes, standard errors are larger, and most differences are not statistically significant.

Separate estimates of effectiveness for males and females are shown in the first two blocks of Table 7. Referrals to vacancies appear to be equally effective for males and females, reducing subsequent contact for each group by around 11 weeks.²⁶ For all other forms of interventions, males appear to benefit more than do females. Each of the estimated impacts is more favourable (smaller positive or larger negative coefficient). The relative effectiveness of the different interventions is similar for males and females, with vacancies and subsidies having the most favourable impact on subsequent contact time.

The second panel of estimates in Table 7 shows differences across three age groups. There is no clear pattern of effectiveness across jobseekers of different ages.

Similarly, the estimates of effectiveness that are shown separately for Māori and Pacific peoples do not reveal any significant differences in effectiveness for different ethnic groups. The estimates are remarkably similar to the aggregate estimates shown in Table 5.

The separate estimates for jobseekers with low qualifications (less than three School Certificate passes) and those with high qualifications (University Entrance or above) suggest that subsidies are most effective for those with low qualifications and that training is most effective for those who already have relatively high levels of qualifications (at least University Entrance).

The panel of estimates by unemployment duration show that referrals to vacancies are most effective for jobseekers with low unemployment durations, and that subsidies and training are most effective for long duration jobseekers.

The final set of results in Table 7 show estimates for the 13 NZES regions. While there are statistically significant differences between different regions in the effectiveness of the various forms of interventions, we have been unable to identify any systematic pattern in the results. We recommend that future work should attempt to tease out the reasons for the regional differences. For instance, are the differences due to different selection patterns of who receives assistance, were there specific labour market conditions that can explain the differences, or did the nature of the different interventions vary across regions? These are all important questions that are beyond the scope of the current study.

²⁶ The pooled results in Table 5 are *not* a weighted average of the gender-specific estimates in Table 7 because the estimation of the Table 5 propensity score constrained the coefficients in the propensity-regression estimates to be constant across genders (apart from a separate constant term) whereas the Table 7 estimates were obtained from separate gender-specific regressions. The same comment applies to other sub-group analyses in Table 7.

TABLE 7: Estimated impacts of interventions - any contact (for selected sub-groups)Dependent variable: Weeks of *any contact* in the 36 months after the reference date

<i>Estimate (standard error)</i>	<i>Vacancies</i>	<i>Interviews</i>	<i>Subsidies</i>	<i>Training</i>	<i>Work experience</i>
<i>Gender</i>					
<i>Male</i>					
N (Comparison) = 13,910					
N (Treatment group)	33,722	22,026	5,871	6,654	1,092
Propensity	-10.7 (0.5)	4.1 (0.5)	-16.2 (0.7)	11.9 (0.7)	13.4 (1.6)
<i>Female</i>					
N (Comparison) = 6,734					
N (Treatment group)	25,348	10,194	2,099	5,088	691
Propensity	-11.4 (0.6)	7.4 (0.8)	-7.4 (1.0)	24.9 (0.9)	18.8 (1.8)
<i>Age</i>					
<i>Less than 26 years of age</i>					
N (Comparison) = 8,052					
N (Treatment group)	26,596	12,516	3,604	8,045	762
Propensity	-11.6 (0.6)	6.2 (0.7)	-12.0 (0.9)	21.0 (0.7)	15.3 (1.8)
<i>Twenty-six to 40 years of age</i>					
N (Comparison) = 8,022					
N (Treatment group)	21,978	12,227	3,048	3,261	669
Propensity	-11.2 (0.6)	7.7 (0.7)	-14.0 (1.0)	18.9 (1.0)	15.4 (2.1)
<i>Forty-one or more years of age</i>					
N (Comparison) = 4,570					
N (Treatment group)	11,193	7,542	1,336	1,270	355
Propensity	-12.6 (0.9)	8.0 (1.0)	-9.6 (1.5)	17.8 (1.5)	13.8 (2.7)
<i>Ethnicity</i>					
<i>Māori</i>					
N (Comparison) = 6,341					
N (Treatment group)	11,674	9,545	2,075	4,838	610
Propensity	-11.3 (0.7)	8.6 (0.8)	-12.6 (1.2)	21.4 (0.9)	13.5 (2.1)
<i>Pacific Peoples</i>					
N (Comparison) = 1,699					
N (Treatment group)	3,707	2,484	320	1,252	102
Propensity	-10.5 (1.3)	12.8 (1.5)	-10.3 (2.5)	19.8 (1.7)	12.3 (4.4)

continued

TABLE 7: continued

<i>Estimate (standard error)</i>	<i>Vacancies</i>	<i>Interviews</i>	<i>Subsidies</i>	<i>Training</i>	<i>Work experience</i>
<i>Qualifications</i>					
Low qualifications (less than three School Certificate passes)					
N (Comparison) = 13,242					
N (Treatment group)	28,871	20,824	4,748	9,777	1,104
Propensity	-12.9 (0.5)	6.3 (0.6)	-17.1 (0.8)	13.1 (0.6)	12.2 (1.6)
High Qualifications (University Entrance or above)					
N (Comparison) = 5,115					
N (Treatment group)	21,326	7,731	2,162	1,108	461
Propensity	-16.9 (0.7)	2.6 (0.9)	-13.4 (1.1)	10.8 (1.5)	18.1 (2.3)
<i>Unemployment duration</i>					
0-13 weeks					
N (Comparison) = 6,878					
N (Treatment group)	34,853	6,625	4,765	9,744	480
Propensity	-12.0 (0.6)	21.7 (1.7)	-9.3 (0.8)	14.4 (0.7)	14.8 (1.8)
14-26 weeks					
N (Comparison) = 3,172					
N (Treatment group)	8,469	83	387	379	317
Propensity	-3.7 (0.9)	2.2 (5.1)	1.3 (2.5)	24.0 (2.6)	12.6 (2.8)
27-52 weeks					
N (Comparison) = 2,531					
N (Treatment group)	8,558	15,227	1,350	1,152	374
Propensity	-1.1 (1.2)	11.9 (1.2)	2.7 (1.5)	14.5 (1.6)	20.0 (2.3)
53-104 weeks					
N (Comparison) = 3,374					
N (Treatment group)	5,486	9,378	979	840	394
Propensity	-4.3 (1.1)	5.2 (1.0)	-7.3 (1.6)	7.1 (1.7)	12.4 (2.3)
105 or more weeks					
N (Comparison) = 4,689					
N (Treatment group)	2,401	972	507	461	221
Propensity	-8.0 (1.2)	-0.8 (1.5)	-14.2 (1.9)	2.4 (2.0)	11.7 (2.8)

continued

TABLE 7: continued

<i>Estimate (standard error)</i>	<i>Vacancies</i>	<i>Interviews</i>	<i>Subsidies</i>	<i>Training</i>	<i>Work experience</i>
<i>Location</i>					
Northland					
N (Comparison) = 1,209					
N (Treatment group)	1,410	860	486	539	140
Propensity	-14.8 (2.0)	7.9 (2.3)	-11.8 (2.7)	15.5 (2.5)	12.6 (4.5)
North Auckland					
N (Comparison) = 2,038					
N (Treatment group)	6,127	2,629	771	767	87
Propensity	-13.5 (1.1)	7.2 (1.4)	-10.4 (1.8)	21.2 (1.9)	16.2 (5.1)
Auckland Central					
N (Comparison) = 1,745					
N (Treatment group)	5,208	2,986	673	708	64
Propensity	-9.1 (1.2)	4.7 (1.5)	-12.2 (2.1)	18.0 (2.0)	13.6 (6.7)
South Auckland					
N (Comparison) = 2,222					
N (Treatment group)	4,419	3,132	667	977	154
Propensity	-11.0 (1.2)	2.3 (1.4)	-9.4 (1.9)	21.5 (1.7)	20.3 (3.8)
Waikato					
N (Comparison) = 1,376					
N (Treatment group)	4,887	2,662	519	593	252
Propensity	-12.2 (1.4)	10.6 (1.7)	-12.8 (2.4)	17.8 (2.3)	15.6 (3.4)
Central					
N (Comparison) = 1,218					
N (Treatment group)	4,028	2,288	576	764	105
Propensity	-7.5 (1.5)	5.1 (1.8)	-18.7 (2.3)	16.9 (2.1)	18.9 (5.3)
Bay of Plenty					
N (Comparison) = 2,037					
N (Treatment group)	5,710	3,038	569	960	170
Propensity	-3.9 (1.2)	9.9 (1.5)	-10.7 (1.9)	26.1 (1.7)	8.5 (3.9)

continued

TABLE 7: continued

<i>Estimate (standard error)</i>	<i>Vacancies</i>	<i>Interviews</i>	<i>Subsidies</i>	<i>Training</i>	<i>Work experience</i>
East Coast					
N (Comparison) = 1,267					
N (Treatment group)	3,728	2,262	541	1,071	77
Propensity	-8.4 (1.5)	5.0 (1.7)	-9.2 (2.4)	19.7 (1.9)	28.3 (6.0)
Taranaki					
N (Comparison) = 1,137					
N (Treatment group)	3,656	1,883	448	931	92
Propensity	-16.6 (1.6)	7.4 (2.0)	-17.4 (2.5)	8.6 (2.0)	18.3 (5.9)
Wellington					
N (Comparison) = 1,771					
N (Treatment group)	6,081	3,161	526	755	103
Propensity	-7.2 (1.2)	11.1 (1.6)	-13.3 (2.0)	17.7 (2.0)	12.6 (5.1)
Nelson					
N (Comparison) = 844					
N (Treatment group)	2,368	1,387	428	363	69
Propensity	-11.5 (1.8)	5.6 (2.2)	-12.5 (2.7)	26.3 (2.8)	25.6 (5.5)
Canterbury					
N (Comparison) = 1,904					
N (Treatment group)	6,150	3,208	896	831	200
Propensity	-10.6 (1.2)	7.6 (1.4)	-12.6 (1.8)	9.5 (2.0)	13.3 (3.5)
Southern					
N (Comparison) = 1,689					
N (Treatment group)	3,948	2,481	852	1,010	255
Propensity	-10.1 (1.4)	7.3 (1.6)	-11.8 (1.9)	22.0 (1.9)	16.8 (3.2)

Notes: Each cell in this table is from a separate regression. Cells in the same column relate to outcomes for the intervention specified at the head of the column. Cells in the same row are estimated using a subset of the treatment and comparison group, as defined in the row headings. 'Contact' time is time spent registered as unemployed *or* in receipt of some form of active labour market assistance.

Is the impact different for different outcome periods?

So far, we have used weeks of either unemployment or contact in the three years following the reference date as our measures of outcomes. In this section, we investigate whether measuring outcomes over a different time-span leads to different conclusions about the impact of interventions. Table 8 presents estimates of the impact of interventions on outcomes over six, 12, and 36 months. If interventions have a short-term impact on outcomes, we would expect that most of the difference between the treatment and comparison groups would occur within the first six months (or perhaps within the first year for interventions such as subsidies or training that may last for a few months). If the intervention has a permanent effect on a jobseeker's labour market experience, by enhancing skills, job search patterns, work experience, or attitudes, we might expect that the difference between outcomes for the treatment group and the comparison group would grow over time.

The pattern that we observe in Table 8 varies for the different types of interventions. For interviews and work experience, the impact over 12 months is approximately twice the six-month impact, suggesting a lasting impact. Over a three-year horizon, the impact is roughly twice (rather than three times) the one-year impact, suggesting a long-term but diminishing effect. For referrals to vacancies, the estimated impact is favourable, whereas it is unfavourable for interviews and work experience.

For training assistance, the unfavourable impact appears to grow over time. Six months after starting a spell of assistance, those receiving training spend about one week more in contact with NZES than does a comparison group. After one year, this has grown to around four weeks, and in three years the difference has extended to 20 weeks.

Finally, the case of subsidies is perhaps the most interesting. One year after being assigned to a subsidy, the treatment and comparison groups have had around the same amount of contact with NZES, despite those on subsidies having spent three weeks more time in contact in the first six months. After three years, the apparent benefit of having received a subsidy increases to around 10 weeks. This is the form of assistance that comes closest to our stylised 'effective intervention' pattern illustrated in Figure 1.

Is the impact different for different histories?

So far, we have used information on each jobseeker's prior experience of unemployment and ALMP assistance over the three years prior to the reference date. One obvious question is whether the estimates of intervention effectiveness are different when we use a shorter record of prior experience – how much does the more distant history add to our identification of selection bias? Given the

TABLE 8: Estimated impacts of interventions – any contact (for different outcome windows)

Dependent variable: Weeks of contact in specified period after the reference date

<i>Estimate (standard error)</i>	<i>Vacancies</i>	<i>Interviews</i>	<i>Subsidies</i>	<i>Training</i>	<i>Work experience</i>
N (Comparison) = 20,664					
N (Treatment group)	59,767	32,285	7,988	12,576	1,786
Weeks of contact in the six months after reference date					
Naïve	-4.0 (0.1)	2.0 (0.1)	3.3 (0.1)	1.0 (0.1)	4.0 (0.2)
Regression	-1.4 (0.1)	1.8 (0.1)	3.4 (0.4)	1.3 (0.1)	4.0 (0.2)
Propensity	-2.8 (0.1)	2.3 (0.1)	2.8 (0.1)	1.2 (0.1)	3.9 (0.2)
Weeks of contact in the 12 months after reference date					
Naïve	-8.5 (0.1)	3.6 (0.1)	-0.4 (0.2)	3.8 (0.2)	8.1 (0.4)
Regression	-2.8 (0.2)	3.8 (0.2)	1.0 (0.3)	4.6 (0.2)	8.0 (0.4)
Propensity	-4.9 (0.2)	4.3 (0.1)	-1.5 (0.2)	4.5 (0.2)	7.8 (0.4)
Weeks of contact in the 36 months after reference date					
Naïve	-24.2 (0.4)	6.0 (0.4)	-10.9 (0.6)	19.0 (0.6)	16.1 (1.3)
Regression	-7.9 (0.4)	8.6 (0.5)	-6.2 (0.7)	20.1 (0.7)	15.1 (1.2)
Propensity	-11.7 (0.4)	6.3 (0.4)	-13.6 (0.6)	19.4 (0.5)	14.6 (1.2)

Notes: Each cell in this table is from a separate regression. Cells in the same column relate to outcomes for the intervention specified at the head of the column. Cells in the same row are derived using the same estimation technique.

'Contact' time is time spent registered as unemployed *or* in receipt of some form of active labour market assistance.

results in the previous section that interventions appear to have an impact on outcomes over three years, we might think that more distant history is important.

Table 9 compares estimates using histories from the previous 12 months with those using 36 months of history. The results barely change. The additional information about jobseeker heterogeneity that is contained in more distant histories does not change our estimates. Note that the regression and propensity

TABLE 9: Estimated impacts of interventions - any contact (with different 'history' windows)

Dependent variable: Weeks of contact in the 36 months after the reference date

Estimate (standard error) [Adj R ² for regression]	Vacancies	Interviews	Subsidies	Training	Work experience
N (Comparison) = 20,664					
N (Treatment group)	59,767	32,285	7,988	12,576	1,786
Using the previous 12 months of history					
Regression	-7.8 (0.4) [24.8]	9.1 (0.5) [15.8]	-6.1 (0.7) [20.7]	20.7 (0.7) [18.4]	15.7 (1.2) [22.0]
Propensity	-12.1 (0.4) [1.3]	7.3 (0.4) [0.5]	-14.8 (0.6) [1.9]	17.8 (0.5) [3.3]	15.1 (1.2) [0.7]
Using the previous 36 months of history					
Regression	-7.9 (0.4) [26.0]	8.6 (0.5) [16.5]	-6.2 (0.7) [21.6]	20.1 (0.7) [18.9]	15.1 (1.2) [22.8]
Propensity	-11.7 (0.4) [1.2]	6.3 (0.4) [0.4]	-13.6 (0.6) [1.7]	19.4 (0.5) [3.9]	14.6 (1.2) [0.6]

Notes: Each cell in this table is from a separate regression. Cells in the same column relate to outcomes for the intervention specified at the head of the column. Cells in the same row are derived using the same estimation technique.

'Contact' time is time spent registered as unemployed or in receipt of some form of active labour market assistance.

estimates reported in Table 9 contain a full set of covariates. What we are testing is the *additional* explanatory power of longer histories.

Does the impact change over time?

We might reasonably expect that the effectiveness of interventions would vary with the state of the labour market in which jobseekers are trying to find work. In the discussion above, we reported that variations in effectiveness across different NZES regions did not appear to be linked to local labour market characteristics. In this section, we examine the link between effectiveness and labour market conditions over time.

**TABLE 10: Estimated impacts of interventions – any contact
(for different years)**

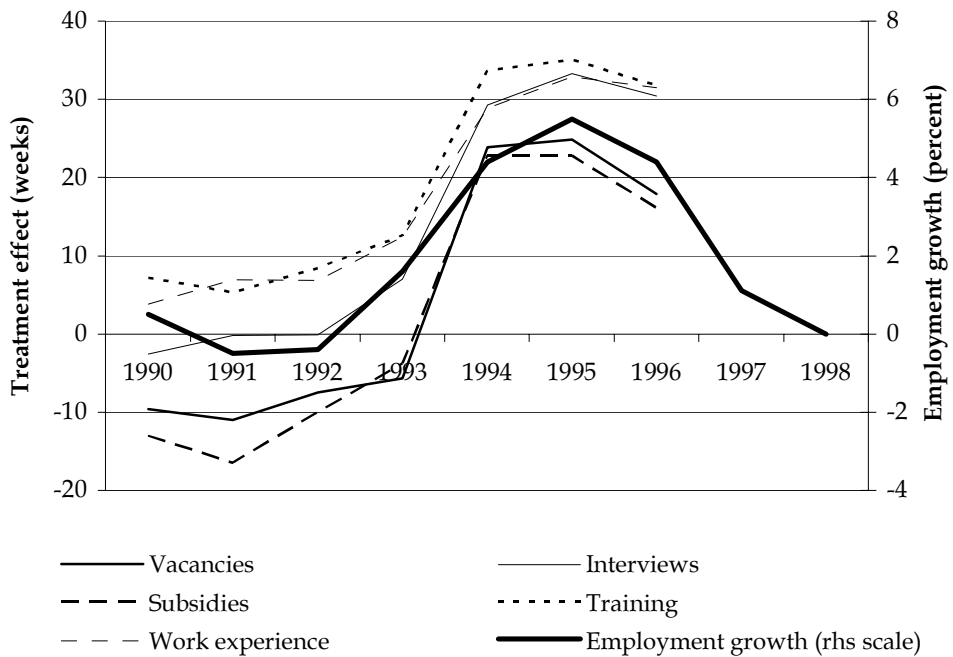
Dependent variable: Weeks of contact in specified period after the reference date

<i>Estimate (standard error)</i> [Adj R ² for regression]	<i>Vacancies</i>	<i>Interviews</i>	<i>Subsidies</i>	<i>Training</i>	<i>Work experience</i>
<i>Year of sample (Regression estimates)</i>					
1990	-9.6 (0.5) [23.0]	-2.6 (0.5) [23.2]	-13.0 (1.1) [23.8]	7.2 (1.1) [23.4]	3.8 (2.1) [24.3]
1991	-11.0 (0.4) [16.7]	-0.2 (0.4) [17.3]	-16.5 (0.8) [17.2]	5.3 (0.8) [17.4]	6.9 (0.0) [17.7]
1992	-7.5 (0.3) [22.9]	-0.1 (0.3) [22.9]	-10.0 (0.5) [24.4]	8.4 (0.5) [24.2]	6.8 (1.0) [25.1]
1993	-5.7 (0.3) [24.2]	7.0 (0.4) [14.5]	-3.8 (0.5) [19.2]	12.6 (0.5) [16.6]	12.4 (0.8) [20.8]
1994	23.9 (0.3) [24.1]	29.3 (0.3) [29.9]	22.8 (0.6) [22.0]	33.7 (0.4) [24.8]	28.9 (1.0) [21.6]
1995	24.9 (0.3) [22.7]	33.3 (0.3) [29.4]	22.8 (0.7) [19.2]	35.1 (0.5) [22.8]	32.9 (1.2) [18.9]
1996	17.9 (0.3) [20.7]	30.4 (0.4) [27.0]	16.1 (0.8) [18.8]	31.8 (0.5) [22.3]	31.5 (1.2) [19.0]

Notes: Each cell in this table is from a separate regression. Cells in the same column relate to outcomes for the intervention specified at the head of the column. Cells in the same row are derived using the same sample period.

'Contact' time is time spent registered as unemployed *or* in receipt of some form of active labour market assistance.

Using our dataset, we observe all interactions that jobseekers have with NZES between October 1988 and December 1997. However, it is not possible to produce estimates of the impact of interventions for all of these years. In order to produce estimates, we need to have measures of prior experience, and of subsequent outcomes. Therefore, we restrict both histories and outcome windows to 24 months rather than the 36 months that has been used as the benchmark so far. On this basis, we can produce estimates for March quarters from 1990 to 1996.

FIGURE 8: Employment growth and the effects of interventions over time

Regression estimates of effectiveness are presented in Table 10, and graphed in Figure 8.²⁷

Figure 8 shows a strong counter-cyclical pattern to programme effectiveness for all of the forms of assistance considered. Labour market interventions assist jobseekers less when employment is growing strongly. The changing relative effectiveness is due largely to poorer outcomes for the treatment groups rather than to significant improvements in outcomes for the comparison groups. For instance, between 1993 and 1994, when employment growth was strong, outcomes for the comparison group improved by 6.6 weeks of contact. Outcomes for the treatment groups deteriorated by between 13.8 weeks (training) and 30.9 weeks (referrals to vacancies).

We are unable to provide a good explanation of why the cyclical pattern is observed. A full examination of the cyclical behaviour of policy effectiveness remains as an important topic for future research.

The pattern is so significant and widespread that it suggests a change in the way that the data were recorded. No such change could be identified. There is

²⁷ Regression estimates are presented because of problems in achieving convergence in the logistic regression on which the propensity estimates are based. Given the similarity of regression and propensity estimates for the 1993 cohort, and for specifications in other years where convergence was achieved, we are confident that the regression results in Table 10 would be similar to propensity results, were they available.

some evidence that the sort of jobseekers who were receiving assistance altered, although this cannot explain the strong cyclical pattern in Figure 8. In 1990 and 1991, regression adjustment increased the estimated effectiveness of programmes, suggesting that assistance was being targeted at jobseekers with relatively poor prospects. In 1993, the opposite was the case, with regression estimates being less favourable than naïve differences, suggesting targeting toward jobseekers with relatively good prospects. Nineteen-ninety-four saw a return to modest targeting of more disadvantaged jobseekers but this did not drive the large deterioration in estimated effectiveness in that year as is evident in Figure 8. Similar deterioration is evident in the naïve as well as the regression estimates of effectiveness.

Delayed job-search has been suggested as a possible explanation of the observed cyclical patterns of effectiveness but does not provide a full explanation. Jobseekers who spend time receiving labour market assistance necessarily have less time to spend in job-search. When employment growth is strong, we expect job-search to be more productive, which would lead to a decline in the relative effectiveness of labour market assistance as jobseekers delay productive job-search. While plausible, this explanation cannot be a complete one because the same changes in relative effectiveness are evident for interviews and referrals to vacancies, which do not involve a significant time away from job-search. Moreover, referrals to vacancies are a form of (assisted) job-search, the effectiveness of which we would expect to increase.

Signalling has also been suggested as a possible explanation of the counter-cyclical pattern. Receipt of ALMP assistance may stigmatise jobseekers if employers perceive a jobseeker's receipt of ALMP assistance as a signal of undesirable characteristics. In times of high unemployment, the signal and, hence, the stigma, is less strong because assistance is extended to a broader group – a group that would include many more 'desirable' jobseekers.

6.3 Staircasing

In this final section, we consider briefly the evidence for 'staircasing'. As noted earlier, interventions that prepare jobseekers for further assistance that will eventually lead them to employment will appear as ineffective on the basis of the outcome measures used in this study. This progression of jobseekers through a sequence of interventions is often referred to as 'staircasing'.

To provide an indication of the significance of staircasing, we present in Table 11 a summary of the degree to which prior receipt of each form of assistance increases a jobseeker's probability of receiving further assistance. Each entry is an estimate of the change in the odds of receiving a particular form of assistance that results from having previously received each of the forms of assistance that we consider. For instance, the first entry in the table indicates that the odds of being referred to a vacancy are a factor of 6.6 higher for jobseekers who have previously been referred to a vacancy. In terms of the likelihood of being referred, this

TABLE 11: Staircasing – effect of previous assistance on probability of receiving assistance

<i>Odds ratio</i>	<i>Odds of referral to vacancy</i>	<i>Odds of interview</i>	<i>Odds of subsidy</i>	<i>Odds of training</i>	<i>Odds of work experience</i>
<i>Assistance received in the previous three years</i>					
Referral to vacancy	6.5	1.4	2.1	1.8	1.7
Interview	1.8	0.9	2.5	6.9	2.1
Subsidy	1.2	1.0	2.5	0.8	0.3
Training	1.0	1.3	1.7	3.2	1.5
Work Experience	1.7	1.1	2.4	1.8	7.7

Note: Each entry in the table is the estimated ratio of the odds ($p/(1-p)$) of receiving the intervention listed in the column heading for a job-seeker who has received the form of assistance in the row heading, to the odds for a job-seeker who has not received such assistance. The entries are derived as $\exp(\beta_i)$ where β_i is the coefficient on a dummy variable indicating having received the (row) assistance from a logit regression that estimates the probability of receiving the (column) assistance. A number greater than one indicates that the odds are raised by having received the prior assistance.

implies that someone with a 10 percent chance of being referred to a vacancy for the first time would have a 40 percent chance of referral if they had previously been referred. The estimates are taken from a logistic regression that includes a full set of covariates, so the estimated impact controls for the fact that different jobseekers have different probabilities of referral.

There is clear evidence of staircasing. The interventions that are shown above to be associated with longer subsequent contact times do appear to lead to receipt of other interventions. Receiving assistance in the form of work experience increases greatly the odds of receiving further work experience assistance and, to a lesser extent, the odds of receiving subsidy or training assistance or being referred to a vacancy. Training and interviews both appear to lead to further training and, to a lesser extent, to subsidy and work experience assistance. A similar pattern of staircasing is evident for subsidy assistance, which is associated with less subsequent contact time. The strongest pattern of staircasing is that jobseekers receiving subsidy assistance are more likely to receive further subsidy assistance, and less likely to be referred to training.

7 Summary

Jobseekers who receive different forms of labour market assistance have markedly different subsequent experience. We have produced a range of estimates of the impact of five classes of intervention. The estimates differ in choice of outcome measure, the nature of the adjustments for possible selection bias, and the treatment of jobseeker heterogeneity.

The outcome measures that we use are not ideal because we do not have information on what jobseekers are doing when not in contact with NZES. The measures that we focus on are the time that jobseekers spend registered as unemployed or in receipt of some form of assistance in a period following receipt of assistance. Less subsequent contact is treated as a desirable outcome, although we are probably counting as good outcomes some instances where jobseekers leave the labour market. Our measure does not recognise the potentially useful contribution that interventions may have in moving jobseekers closer to independence, unless they actually achieve independence. Neither does the measure take any account of the relative costs of different interventions.

Jobseekers who are referred to vacancies or who receive wage subsidies have less subsequent contact with the public employment service. The improvement in outcomes for those receiving wage subsidies becomes evident only a year or more after commencing a subsidised placement. Initially, the time spent in the subsidised job leads to their having more contact time than a comparable group of jobseekers.

Jobseekers receiving training assistance, a work experience placement, or attending an interview with the employment service have more subsequent contact time. Some of this effect can be attributed to the fact that these forms of treatment increase the likelihood that jobseekers are then given other forms of assistance. This may reflect 'staircasing', as jobseekers move through a sequence of interventions that move them closer to independence and employment.

There is surprisingly little variation in estimated impacts for different sub-groups of jobseekers. Broadly speaking, interventions that are relatively effective for one group of jobseekers are also relatively effective for other jobseekers. There are some exceptions, which are noted in the text.

The impact of interventions does, however, appear to vary over time, with all interventions being less effective when employment growth is small. A more detailed examination of this pattern is beyond the scope of this paper and is left as a challenge for future research.

References

- Angrist, J and Krueger, A (1999) 'Empirical Strategies in Labour Economics, Chapter 23 of Ashenfelter, O and Card, D (eds) *Handbook of Labor Economics: Volume 3A*, Amsterdam, Elsevier, pp 1277-1366.
- Dehejia, R and Wahba, S (1999) 'Causal Effects in Non-Experimental Studies: Re-evaluating the Evaluation of Training Programs', *Journal of the American Statistical Association*, Vol 94, No 448, (December 1999), pp 1053-1062.
- Department of Work and Income New Zealand (1994-2000) *Research and Evaluation at Work and Income New Zealand* http://www.winz.govt.nz/text_only/publications_and_reports/research/.
- Gobbi, M and Rea, D (2000) 'Unemployment Dynamics in New Zealand'. Paper presented to the 9th Labour Employment and Work Conference, Victoria University of Wellington, November 2000.

- Greene, W (1997) *Econometric Analysis*, New Jersey, Prentice Hall.
- Heckman, J (2001) 'Micro-data, Heterogeneity and the Evaluation of Public Policy: Nobel Lecture', *Journal of Political Economy*, 109(4), pp 673-748.
- Heckman JJ; Lalonde RJ and Smith JA (1999) 'The Economics and Econometrics of Active Labour Market Programs', in AC Ashenfelter and D Card (eds) *Handbook of Labor Economics: Volume 3A*, Amsterdam, Elsevier, pp 1865-2097.
- Heckman, J; Tobias J and Vytlačil, E (2001) 'Four Parameters of Interest in the Evaluation of Social Programs', *Southern Economic Journal*, 68(2), pp 211-223.
- Heckman, J and Vytlačil, E (2001) 'Policy-relevant Treatment Effects', *American Economic Review: Papers and Proceedings*, 91(2), pp 107-11.
- Martin, J (2000) 'What Works Among Active Labour Market Policies: Evidence from OECD countries' experiences', *OECD Economic Studies*, 30: 2000(I), Organisation for Economic Cooperation and Development, Paris.
- New Zealand Department of Labour (1985) *Studies of Employment and Training Programmes*, Wellington.
- Robinson P (2000) 'Active Labour-Market Policies: A Case of Evidence-Based Policy Making?', *Oxford Review of Economic Policy*, 16(1), pp 13-26.
- Rosenbaum, P and Rubin, D (1983) 'The Central Role of the Propensity Score in Observational Studies for Causal Effects', *Biometrika*, 70(1), pp 41-55.
- Rosenbaum, P and Rubin, D (1985) 'Constructing a Control Group Using Multi-variate Matching Methods that Include the Propensity Score', *American Statistician*, 39, pp 33-38.
- Rubin D (1974) 'Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies', *Journal of Educational Psychology*, 66, pp 688-701.
- Smith, J and Todd, P (2001) 'Reconciling Conflicting Evidence on the Performance of Propensity-score Matching Methods', *American Economic Review: Papers and Proceedings*, 91(2), pp 112-124.