TIGHTENING A WELFARE SYSTEM: THE EFFECTS OF BENEFIT DENIAL ON FUTURE WELFARE RECEIPT

by

David A. Green
Department of Economics
The University of British Columbia

and

William P. Warburton
Economic Analysis Branch
Ministry of Human Resources,
British Columbia

August 2001

Discussion Paper No.: 02-07
We derive Local Average Treatment Effect estimates of the impact of welfare benefit denial on future receipt using a unique experiment involving reassessment of some applicants who were originally slated to receive benefits. We find evidence of considerable heterogeneity among applicants. Our results support a model with a peripheral group who exhibit scarring effects from being granted benefits and a core group who do not. The core group moves quickly back onto welfare when they are denied benefits. Even for the peripheral group, benefit denial has intermediate term but not permanent impacts.

JEL codes: I38

Key words: welfare, structural dependence, LATE estimator

We would like to thank Paul Beaudry, Thomas Lemieux, and participants at labour economics seminars at the University of Toronto and the University of British Columbia and a session at the 1999 Canadian Economics Association Meetings for useful comments. Green gratefully acknowledges support from the Centre of Excellence on Equality, Security and Community. Any opinions expressed in this paper do not necessarily reflect those of the Ministry of Human Resources. Any remaining errors are our own.
Tightening a Welfare System:
The Effects of Benefit Denial on Future Welfare Receipt

David A. Green and William P. Warburton

In the past ten years, many jurisdictions in North America have reduced access to welfare benefits. At the heart of these reforms has been an argument that welfare receipt at one point in time induces more receipt in the future. This effect, often referred to as welfare scarring, is conjectured to happen either because welfare receipt causes poorer future labour market outcomes which in turn raises reliance on welfare, or because welfare receipt breaks down social barriers to future receipt. In either case, scarring effects imply that reforms that reduce the number of welfare recipients at a point in time will lead to lower expenditures on welfare in the future and, possibly, to better future outcomes for individuals whose access to welfare is restricted under the reforms.

Arguments that scarring effects exist are built upon the observation that individuals who receive welfare in a given period, t, are more likely to receive welfare in a future period than individuals who do not receive benefits in t. However, persistence of this type does not prove the existence of scarring effects. The persistence could occur because individuals who are denied benefits have different underlying characteristics compared to people who are granted benefits. In standard terminology, a positive correlation between current and future welfare use could arise from true structural dependence (scarring effects) or unobserved heterogeneity. However, if persistence in welfare receipt is generated from unobserved heterogeneity then welfare reforms do not help those eliminated from the welfare rolls by altering their future outcomes. Instead, the reforms may harm the individuals who most need the benefits without altering the conditions that led them to apply for welfare. Thus, establishing whether persistence in welfare receipt stems from scarring effects or unobserved heterogeneity is important for making appropriate welfare policy. In this paper, we investigate whether scarring effects exist for a group of welfare recipients who are affected by tightening in a welfare system. In particular, we study the impact of a policy under which some Income Assistance (IA) applicants in the Canadian province of British Columbia
We identify structural dependence using data from a quasi-experiment. Specifically, from October 1995 through June 1996, BC conducted an experiment in which 17 regional offices were given Verification Officers (VO's) who re-examined the files of IA applicants originally slated to be granted benefits. Data were kept on all applicants in these offices and in a matched set of "control" offices which were not given VO's. By comparing IA receipt in the months following the experiment for applicants at offices with and without VO's, we estimate the impact of benefit denial at a point in time on future IA use for a marginal set of applicants who would be granted benefits in a typical office but denied benefits when their files are examined by VO's. Thus, we use the introduction of VO's in some offices to create an instrument to separate structural dependence from unobserved heterogeneity. Since the VO's re-examined "marginal" cases, the estimates based on this instrument correspond to Local Average Treatment Effects (Imbens and Angrist (1994)). A key advantage of this approach is that we obtain impact estimates that are specific to the very group who would most likely be affected by any policy of tightening welfare eligibility.

Our approach differs from earlier attempts to separate scarring and unobserved heterogeneity effects in three aspects. First, most earlier papers rely on what amounts to functional form restrictions in duration model estimators to identify scarring effects, while we use an instrumental variables approach. Second, much of the existing literature focusses on duration dependence within welfare spells: whether being on welfare longer reduces the probability of leaving welfare. We study the impact of benefit denial on whether an individual receives IA in future periods, and thus, our estimates reflect both duration dependence and incidence dependence. The latter is defined as the impact of receipt of welfare at a point in time on the probability of starting a new welfare spell in the future. The third main difference in our work is that our data comes from a welfare system in which employable individuals without children are eligible for benefits and, in fact, constitute the majority of the caseload. This is in stark contrast to
the American system which focusses on lone parent families. This means our results are less relevant for discussions of the US welfare system than for discussions of systems that resemble Canada’s in covering more than lone parent families.

Our investigation generates five main results. First, applicants at offices with VO's experience lower future IA usage rates than applicants at non-VO offices. Thus, benefit denial reduces future IA use. Second, the measured benefit denial effect is essentially zero (and statistically insignificantly different from zero) by about 2 ½ years. Third, the persistence of this VO impact varies inversely with the proportion of cases reviewed by the VO. Fourth, first time IA users have larger estimated effects than repeat users. Fifth, the decline of estimated effects is dominated by declines in the usage rates of the marginal individuals who are granted benefits rather than by movements back onto IA by marginal individuals when they are initially denied benefits. We argue that these results fit with a model with a peripheral group which exhibits medium term scarring effects and a core group which exhibits few or no such effects. This implies that policies of tightening welfare systems through greater scrutiny, as in the introduction of VO’s, have medium but not long term impacts on the behaviour of some applicants. Thus, denying benefits in this way will save money, but by itself, benefit denial does not appear to set applicants on a whole new path with respect to welfare use.

The paper contains 5 sections. In section 2), we describe B.C.’s welfare system, the experiment we use to identify state dependence effects, and the data. In section 3), we present our instrumental variables estimates and evidence on the heterogeneity of benefit denial impacts. In section 4), we relate our results to the previous literature. Section 5) contains conclusions.

2) Background and Data

2.1) British Columbia's Income Assistance Programme

The social safety net in Canada primarily consists of separate unemployment insurance (UI) and income assistance (IA) components. The UI system in Canada delivers benefits based on the duration of employment in a qualifying period and is administered by the federal government.
The IA component is delivered by the provincial governments and has needs based eligibility requirements. In BC, IA applicants are first classified as either employable or unemployable, then their income levels are compared to a cut-off level specified for their employability status. Benefits are granted if the applicant's resources are below the cut-off. The benefit amounts are determined by employability status and family size, and are paid monthly. Perhaps of most interest for our purposes is the reasons benefits are denied. This information is not preserved in the system records, but a report from July, 1996 lists the top specified reasons as: income or assets in excess of the threshold (18% of rejections); family composition (probably another supporting adult present)(15%); not available/seeking employment (12%); and did not show for interview (10%).

IA differs markedly from the welfare systems in US states in both size and scope. In the US systems, benefits are targeted mainly at lone parent families. In contrast, in the BC system benefits do not depend on the presence of children: single, employable adults and childless couples made up 68% of beneficiaries in 1992 (Barrett and Cragg(1998)). In fact, IA beneficiaries made up 9.7% of B.C.'s population in December 1995. On the other hand, IA resembles basic assistance systems in several European countries which also have universal, needs based social assistance systems backed by a separate UI system. Thus, while our results are difficult to apply directly to the US experience, they may be useful in considering outcomes in other developed economies.

The B.C. IA system underwent several changes in or near 1996 which were designed to encourage greater labour force attachment among recipients deemed to be employable. These included cutting benefits for and requiring job search from single employables beginning in January 1996. Beginning in August 1997, the government also adopted a stricter assessment of employability, moving borderline "unemployable" cases either into the employable or disabled categories. For the most part, these changes occurred just before the experiment we examine. This implies that we do not want to use before/after (the experiment) type variation in forming our estimates and makes the existence of a control sample more important.
2.2) The Early Detection and Prevention Project

The specific experiment we focus on is entitled the Early Detection and Prevention (EDP) project: a pilot project designed to assess the usefulness of extra scrutiny of IA applications. Under the EDP, the B.C. Ministry of Social Services selected 17 offices from around the province to act as test sites. Each office was then asked to name a control office which they viewed as a close match in terms of past levels and composition of IA caseloads. Each test site was given money to hire a Verification Officer (VO) to double check applicant files for the months of January through June of 1996. After the expiry of the experimental period, the VO funding was removed. The experiment was deemed a success and, as a result, VO’s were re-introduced throughout the province in late 1996, though at much lower levels than in the experiment.4

During the EDP, field officers were asked to pass files to the VO for applicants who were slated to get benefits but whom the field officer viewed as marginal cases. Marginal cases were ones where closer scrutiny might reveal reasons not to grant benefits, such as hidden assets or other supporting adults being present. The reassessments by the VO's were done before benefits were actually granted to the applicant. We call the offices given VO's, the VO offices and the matched comparison offices, the control offices. Identifiers were recorded for all applicants in both the VO and control offices for the months of January through June 1996. These identifiers are used to link to data on the individual's use of IA, UI and earnings from 1990 through to 2001.

We use the EDP to identify the causal effects of benefit denial at a point in time on future IA receipt. As is well known, comparing future benefit receipt of those who are granted and those who are denied benefits at a point in time does not necessarily identify that causal effect. Any observed correlations between current and future benefit status could reflect a combination of a true causal effect and persistence in other, non-IA generated, differences between the groups (e.g., persistent differences in support from extended family). In standard terminology, this is the problem of separating true state dependence from unobserved heterogeneity.

Our approach to estimating true state dependence effects is to compare IA recipiency rates
in future months for IA applicants at VO offices during the EDP project with those for applicants at control offices. As we demonstrate below, applicants at the two types of offices are well matched in terms of underlying, personal characteristics. Then, since (again, as we demonstrate below) VO’s induce higher rejection rates at VO offices, differences in future IA use between VO and control office applicants can be attributed to the effects of tightening the IA system.

One alternative approach to estimating causal benefit denial effects would be to perform a true experiment. In a true experiment, one would randomly deny or grant benefits to applicants in a given month and then compare future IA receipt between the recipient and denial samples.\textsuperscript{5} Our use of the EDP project differs from a true experiment in several ways. First, we do not have random assignment into our treatment and control samples. Rather, our approach is in the spirit of matching estimators where, in our case, matched offices are selected for the VO offices. Moreover, our treatment group (extra rejectees induced by VO’s) is highly selected. As we discuss in section 3, if reactions to benefit denial vary across individuals then the effects measured in a true random experiment and our experiment will be different. In that case, we contend, the measure obtained from our experiment is more relevant for considering welfare-tightening policies since it identifies the effect of denying benefits on the very group most likely to be affected by such tightening.

As a separate issue, the duration of the EDP implies that our experimental set-up is not perfect. In an optimal design, the treatment (the extra rejections) would occur only at an instant in time. This would ensure that the treatment and control group samples faced the same environment after the experiment. In contrast, the EDP lasted for six months. Further, unless they change IA districts, rejected applicants must reapply at the same office. Thus, VO office applicants faced an environment that included VO’s for a period after their initial application and review. The duration of this artificial environment varied with the timing of the initial review within the EDP window and with how quickly the VO's were reassigned after the EDP, but it does not appear to exceed six months in general. We follow individuals for almost 5 years after the EDP and thus can
analyse effects well beyond the initial period when there is some contamination of the experimental design. We examined the importance of the contamination by re-estimating our results using only applications from May and June of 1996 (the group who would face the least contamination). The estimated effects from this sample are somewhat smaller than those reported here for the first 6 months but fit very closely thereafter, suggesting that longer term contamination effects are small.

2.3) Data

The data we use is primarily from BC’s IA administrative files. For every applicant at the VO or control offices, we have information on IA receipt in each month from October 1990 to April 2001. We also know their age and gender from these files. The files are also linked to UI files, from which we construct UI benefit histories for each person, and Records of Employment (ROEs). ROEs are forms that all employers must file with the government for each employment separation. Based on the ROEs, we can construct earnings history variables. Finally, we also linked individual records to Census tract information based on the postal code of the office. The Census data pertains to June of 1996, the end of our experimental window. Using the Census data, we generate tract specific variables on: the proportion of families with income below Statistics Canada's Low Income Cut-Off, the proportion of the population who have not completed high school, and the unemployment rate. We do not know individual educational attainment, but Nakamura and Warburton (1999) find that education variables do not add much explanatory power in predictions of IA use when there are extensive individual histories of IA, UI, and earnings, and a set of matched controls. Our variable definitions are presented in appendix Table A1.

For each applicant in our sample, we know the percentage of files reviewed by a VO at the office where the individual applies in the month s/he applies and whether her/his specific file was reviewed. The percentage of files reviewed by VO's varies substantially across VO offices and over time within offices, from a low of 0% reviewed in some VO offices in some months to a high
of over 60% reviewed in others. The variation arises because VO's were sometimes temporarily assigned to other duties. In addition, because of some cases moving among offices, some control offices are recorded as having cases reviewed by VO's in some months. In the control sample, we keep only observations from control offices in months in which VO reviewed cases make up less than 2% of all applications. Similarly, we keep only observations from VO offices in months in which cases reviewed by VO's make up more than 2% of all applications. This ensures a cleaner separation between our treatment and control samples. However, including all VO and control office observations does not change basic data patterns or our conclusions. We restrict our attention to VO offices in which less than 40% of cases were reviewed in part because they match the control offices better and in part because it is not clear a meaningful extra review is happening in offices where VO’s examine such a high volume of files in each month.

Our goal is to examine the effects of regular benefit receipt. For that reason, we cut from the sample, individuals who are under age 19 (because they are ineligible for regular benefits) and over age 60 (because they are paid under a special programme). We also cut individuals who get benefits under the "unemployable due to disability" category at any time in our sample period to make sure we do not face composition problems related to the reclassification of cases in 1997. The result of these sample cuts plus the cuts related to the percentage of files reviewed in VO and control offices is a sample containing a total of 21,859 applications, with 8,638 in VO offices and 13,221 in control offices. That the VO's had a substantial impact is seen by the fact that 42% of applications were rejected in VO offices versus 30% in control offices in the months when VO's were used. We do not know whether the extra rejections generated by the VO's were concentrated on any particular reason (e.g., assets and income in excess of the threshold).

Our empirical approach hinges on the month of application for IA. We compare IA use in months before the key application for applicants in the VO and control offices to investigate the usefulness of the controls. We examine IA rates in months after the key application to examine the impact of being denied IA on subsequent IA use. We define the key application as the individual's
first application in the experimental window (January, 1996 to June, 1996). Thus, if individuals are denied benefits in their first application in the window and are granted benefits later in the window, the latter is just counted as part of their subsequent IA history rather than initiating a separate observation in our dataset.

Finally, our VO and control samples consist of individuals whose key applications are in VO and control offices in the experimental window, regardless of what offices they might subsequently apply to. VO and control samples do not refer to IA usage in those specific offices by other individuals before or after the experimental window.

In Table 1, we present variable means for the control sample and two VO samples. Our main identifying assumption is that the IA usage observed in the control offices after the key application represents what would have happened in the VO offices if no VO had been introduced. This assumption seems more reasonable the more similar are the VO and control samples in observable characteristics related to IA use. We investigate this by comparing means of variables defined at the time of the key application for the overall pools of applicants at VO and control sample offices. These means are found in columns 1, 4 and 7 of Table 1.

Based on Table 1, the VO and Control samples are very similar. In each case, about a third of applicants are female, applicants are young, and neighbourhood characteristics are very similar. Income histories are also very similar across the samples, as reflected in the nearly identical values for months of UI receipt and proportion earning less than $1000 in the previous year. The variable, FIRST, which equals 1 if the individual has not received IA in the last 5 years, shows very similar proportions of applicants fitting our definition of first time users in our three types of offices. Further, the average number of months of IA received in each of the previous four years are nearly identical in the control and 20-39% review VO samples, and very similar in the 2-19% reviewed VO offices. The pairwise differences between variable means for each of the VO samples and the control sample are never statistically significant at a 5% significance level. Thus, the identifying assumption mentioned above appears plausible.
The remaining columns of Table 1 show variable mean values for subsets of individuals who are either granted or denied benefits at each type of office. In the control office sample, grantees are considerably older than denials, with 27% of grantees being age 19 to 24 compared to 35% of denials. Grantees are also much more likely to be repeat IA users, with only 23% being first time users compared to 44% of denials. Denials tend to have higher levels of earnings in the previous year, as one might expect given that evaluations are made on the basis of need. Finally, grantees and denials are nearly equal in terms of gender composition and the characteristics of the neighbourhoods in which they live. The sharp differences between grantees and denials in terms of their age, earnings and previous use of IA suggest that simply comparing future outcomes between these groups will not generate a clean estimate of the behavioural effects of benefit denial.

As discussed in section 2.2), using the EDP generates estimates of the effects of benefit denial for a marginal group who would receive benefits in a control office but be denied benefits in a VO office. Comparing characteristics of denials in control and VO offices provides some insight into who this marginal group is. Comparing columns 3, 6 and 9 of Table 1 suggests that the extra denials in VO offices tend to be younger males who had lower previous earnings and more previous contact with the IA system than denials in control offices.

More evidence on the marginal group affected by the experiment is provided in Table 2. Table 2 contains variable means for cases directly reviewed by VO's and is divided between those applicants for whom the initial determination (i.e., that they be granted benefits) is upheld by the VO and those who are denied benefits by the VO. Compared to all applicants in the VO offices, those reviewed in the 2-19% reviewed offices are more likely to be male, younger and to have lower earnings. Among the applicants reviewed by a VO, the ones denied benefits are even more likely to be male and younger and have less recent use of IA than those who are granted benefits. Differences between those reviewed versus not reviewed by VO's is smaller in the higher review rate offices. For example, there is a 10 point gap in the percentage of the sample who are age 19 to
24 in the 2-19% reviewed offices but only a 3 point gap in the 20-39% offices. We argue below that observed data patterns are consistent with VO's in the higher review rate offices scrutinizing each file less carefully. Thus, with more careful selection of files to review in lower review rate offices, there are greater differences between reviewed and non-reviewed applicants.

The variable means in Tables 1 and 2 provide some insight into the evaluation process. In both tables, differences in neighbourhood characteristics between grantees and denials are small, suggesting that officers focus on individual rather than community attributes in making their decisions. Differences in repeat use status, however, are very large, with applicants who have not used IA in the previous five years facing much higher probabilities of denial. This indicates there is persistence in access to benefits even after conditioning on application. Thus, the idea that the sole mechanism behind scarring effects is individuals continually reapplying for benefits may be inaccurate. Part of the persistence may be found in how the system itself deals with repeat users.

3.0) Estimating the Causal Effects of Benefits Denial

3.1) IA Usage Rates in VO and Control Samples

In this section, we present plots of IA usage rates for both VO and control office samples. This serves as a means of summarizing the dependent variable in our examination and allows us to check a key identifying assumption. We compare the IA rates of applicants at VO and control offices for the 4 years preceding and the 58 months after the key application.

Figure 1 contains plots of recipiency rates for applicants at control and VO. We record an individual as using IA in a month if they receive an IA cheque in that month. The control sample includes all applicants in control offices between January and June, 1996. Figure 1 shows similar patterns for both VO and control samples: a rising probability of receiving IA in each month up to about 1 year before the key application, followed by a sharp drop just before the key application, a sharp increase at the time of the application and then a gradual fall to levels below the pre-application levels.

Several features of the plots in figure 1 deserve comment. First, the usage rates are not zero
just before the key application because IA cheques are issued on a monthly basis. Thus, if a person finds a job and moves off IA in December then loses the job in January, she will need to re-apply for benefits (which will bring her into our dataset). If she is granted benefits in January then our measure of IA use (whether a cheque was issued) will equal one in both December and January even though there was a new application in January. Second, the rise in IA use at the time of the application occurs in part in the application month and in part in the month directly following because individuals who apply late in a month may not get their first cheque until the following month. Third, the plots show a wave pattern with peaks approximately every 12 months. This reflects a seasonal pattern, with peaks in the caseload in December and January of each year.

The VO and control office samples exhibit very similar patterns of IA receipt in the months preceding the key application. The largest pre-application difference between the control and VO lines occurs about 3 to 4 years before the key application. In those months, the IA rate is slightly under .01 higher in the control versus the VO sample, though the difference is never individually statistically significant from zero at conventional significance levels. In other months, differences in IA rates between either VO sample and the control sample are generally less than 0.005 and are never statistically significant. Thus, IA use before the experiment indicates that the VO and control samples are well matched. This reinforces our conclusions based on Table 1.

After the key application, the usage rates for the control and the VO samples differ substantially until between 2 and 2 ½ years after the key application. Usage rates are .14 higher in the control than the VO sample in the first post-application month, .04 higher at the 6 month post-application point, .02 at the 12 month point, .01 at the 24 month point, and .002 at the 30 month point.10 Thus, using VO's to reassess files creates substantial reductions in IA usage that dissipate gradually over time.

These plots can be used to assess whether introducing VO's reduces future caseloads and thus costs for the government.11 To do this, we again need the assumption that outcomes in the control offices after the key application represents what would have happened in the VO offices if
there were no VO's. Based on Figure 1, introducing VO's and having them re-assess between 2 and 39% of cases leads to a 7% reduction in the average number of months of IA used in the 58 months following the key application. The government's own assessment of the EDP project estimated that each VO generated savings of $21,000 per month in reduced benefits paid during the experimental period. These numbers indicate that VO's reduced recipiency rates. In the ensuing sections, we make use of that fact to identify the causal effects of benefit denial.

3.2) Estimators of Causal Effects

The simplest estimator of potential causal effects of benefit denial is calculated by taking the difference between the average values of the outcome measure for those granted and those denied benefits. The dashed line in Figure 2 presents this difference for each month after the key application. The plot is based solely on the control sample in order to portray the type of estimate one would obtain in the absence of an experiment. A symbol on a line at any point means the estimated effect at that point differs from zero by more than 2 standard errors. The dashed line indicates that those denied benefits are substantially less likely to use IA even 5 years after the key application. Denials have a probability of IA receipt that is lower than that for grantees by .31 at 6 months after the key application, .23 at 12 months, .12 at 36 months and .10 at 58 months.12

As we discussed in section 2), the results in Figure 2 do not necessarily reflect causal effects because of the standard problem of separating true state dependence from unobserved heterogeneity. We address this problem by using an instrument based on the EDP experiment.

In order to define our estimator, we first present some basic definitions. Let $Y_{0i}$ be an outcome of interest for person $i$ if s/he is granted benefits in the key application and let $Y_{1i}$ be the outcome for the same person if s/he is denied benefits. In our case, the $Y$'s are dummy variables taking a value of 1 if the individual receives IA benefits in some future month and 0 otherwise. To measure the impact of benefit denial, we would like to observe $\Delta_i = Y_{1i} - Y_{0i}$. A common goal in investigating scarring effects is to obtain a consistent estimate of the mean of $\Delta_i$. Unfortunately, we do not observe $Y_{1i}$ and $Y_{0i}$ for the same person: we observe $Y_{0i}$ only for grantees and $Y_{1i}$ only
for denials. In formal terms, we observe, $Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}$, where, $D_i$ is a dummy variable equalling 1 if the individual is denied benefits at the key application. Problems arise when $D_i$ is correlated with $Y_{1i}$ or $Y_{0i}$ and, potentially, $\Delta_i$ (Heckman(1997)).

Our approach is to find an instrument for $D_i$: a random variable, $Z_i$. Suppose such an instrument exists and can take two values, $z$ and $z'$. In our case, the value $z$ denotes cases that went through a control office and $z'$ denotes those going through a VO office. One can construct an average treatment effect estimator, $\hat{\delta}$, as,

$$1) \quad \hat{\delta} = \frac{E[Y_i | Z_i = z]}{Pr[D(z) = 1]} \& \frac{E[Y_i | Z_i = z']}{Pr[D(z') = 1]}$$

Intuitively, the estimator $\hat{\delta}$ compares IA usage rates in a future period for applicants at a regular office in our sample period with applicants at an office where the VO’s induced extra rejections. The estimator reveals an impact of current denials on future usage rates to the extent that future IA use differs and the difference is correlated with the extra rejections induced by the VO’s.

If $\Delta_i$ does not vary across individuals then $\hat{\delta}$ provides a consistent estimate of the effect of benefit denial today on the probability of IA receipt in a future period, providing $Z$ is a valid instrument. The variable $Z$ is a valid instrument if two main conditions are met: $Z$ affects $Y$ only through its affect on $D$; and $Z$ has a direct impact on $D$ (i.e., $Pr(D_i(z) = 1)$ must be a nontrivial function of $Z$). The second of these conditions is met in our case since, as discussed earlier, VO’s generated substantial increases in the rejection rate at VO offices relative to control offices. The first condition imposes two requirements in our case. First, there cannot be a systematic bias in the way VO’s were assigned across offices. If there were then the IV estimates would reflect the differences across the offices as well as any true causal effect. We argued in sections 2.3) and 3.1) that the control and VO samples are well matched both in observable characteristics and in past use of IA and, therefore, we claim that this requirement is met. The second requirement is that being reviewed by a VO must not directly affect future IA outcomes. As we discussed earlier, this
may not be true for the first 6 months because VO office applicants faced an environment which included VO’s during the remainder of the EDP while control office applicants did not. After that, however, this second requirement was likely met. In particular, whether an individual saw a VO in the past was not recorded on their file and thus could not directly affect their personal outcomes.

The solid line in Figure 2 corresponds to $\delta$ with $z'$ now corresponding to offices in which VO’s reviewed between 2 and 39 % of the cases. The line corresponds to $\delta$ times -1 and thus shows the reduction in the proportion using IA in each month. Again, a symbol at any point means that the measured effect in that month differs from zero by more than two standard errors. The line indicates that there are persistent but declining effects of benefit denial on future receipt. At month zero, we are comparing IA use for individuals denied benefits with those granted benefits so the measured difference is by definition, 1.0. By month 6 after the key application, rejectees have an IA usage rate that is still 0.27 lower than it would have been if they were initially granted benefits. The impact of the initial denial declines steadily to the 2 year point, when it takes a value of .06. Thereafter, the impact falls off rapidly and by 30 months rejectees have an IA rate that is only .01 less than it would have been had they been granted benefits initially, and the effect is not statistically significantly different from zero.

The main conclusion from the IV estimator is that denying benefits reduces future IA use but these effects decline over time and are effectively zero after 2 1/2 years. In comparison, the simple difference estimator line is somewhat higher for months 6 through 24. More importantly, though, the difference estimator shows persistent differences between grantees and rejectees even after nearly 5 years. The IV estimator results imply that short term (i.e, up to 6 month) effects in the difference estimator largely reflect true causal impacts but that the persistent differences between grantees and rejectees after the 3 year point are almost entirely due to unobserved differences between the two groups.

3.3) Heterogeneity in Causal Effects

To this point, we have interpreted our IV estimates as capturing “the” effect of benefit
denial on future IA use. This is an accurate interpretation as long as \( \Delta_i \) does not vary across individuals, i.e., as long as benefit denial is expected to have the same average effect on everyone. However, it is more plausible to think that benefit denial will change future outcomes for some people but not for others (e.g., people with a persistent cause for their poverty). To the extent that such heterogeneity in impacts exists, we must change our interpretation of our IV estimates.

As noted in Imbens and Angrist (1994) and Heckman (1997), when \( \Delta_i \) varies with \( i \) in unobservable (to the econometrician) ways, the average treatment effect estimated by \( 1) \) can vary with the instruments. To see this, redefine the treatment indicator as conditional on specific realizations of \( Z: D_i(z) \). Thus, as the value of the instrument changes, the set of individuals who are given the treatment changes. Given heterogeneity in \( \Delta_i \), this means the estimated average treatment effect will change also, yielding different Local Average Treatment Effects (LATE).

If there is heterogeneity in \( \Delta \), \( 1) \) no longer necessarily provides a consistent estimate of any average treatment effect. Based on Imbens and Angrist (1994), we need to meet three conditions to achieve a consistent LATE estimator in our case: the two valid instrument conditions which we already argued are satisfied plus one other. The third condition is, either \( \text{Pr}[D_i(z) - D_i(z') = 1] = 0 \) or \( \text{Pr}[D_i(z) - D_i(z') = -1] = 0 \): the "monotonicity assumption". The necessity of the latter condition can be seen if one considers a VO who both denies benefits to applicants who would otherwise have received them and grants benefits to applicants who would otherwise be denied. Then, patterns in IA receipt in the VO sample in future months could not be solely attributed to benefit denials since the VO increased receipt for some applicants. In our case, because VO's were used to reassess files that were initially granted benefits, rather than all possible files, their introduction could only lead to an increase in denials. Thus, in the terminology set out above, \( \text{Pr}[D_i(z) - D_i(z') = 1] = 0 \) and estimation using equation \( 1) \) generates an estimate of \( \text{E}[Y_{1i} - Y_{0i} / (D_i(z) - D_i(z') = -1)] \): the effect of benefit denial on future IA use for those who would be granted benefits in a regular office but denied benefits in a VO office. The solid line in Figure 2 then reflects the impact of benefit denial on a marginal group who would be affected by the type of
system tightening that happens when VO’s are introduced rather than “the” effect of benefit denial for all applicants.

With heterogeneity in $\Delta$, the average estimated treatment impact will vary with how many cases are reviewed since the set of people denied benefits upon reassessment will change. To investigate this possibility, we define two further LATE estimators: one corresponding to offices where VO’s reviewed 2-19% of applications and one where they reviewed 20-39%. The results for both are plotted in Figure 3. As with the IV estimates in Figure 2, both lines show persistent but declining effects of benefit denial on the marginal group. In both cases, the estimated effect is not statistically significantly different from zero by the 3 year point. However, the 2-19% review sample estimates show much more persistence. For example, at the 12 month point, the marginal group who were rejected in the 2-19% review rate offices had IA usage rates that were .27 higher than they would have experienced if they had been granted benefits initially. In comparison, the denial impact was .14 at 12 months for the marginal group in the 20-39% review rate offices. The estimated impact for the 20-39% review rate offices is below .1 for all months after month 20 and is statistically significantly different from zero only in month 25. By month 31 the estimated impact in these offices is .0075. In contrast, the estimated impacts for the lower review rate offices are above .1 and statistically significantly different from zero until month 36.

The variation in impacts by review rate is important because it provides insight into the distribution of heterogeneity in $\Delta$ across the applicant population. Recall from section 2) that applicants from high and low review rate offices have very similar IA histories and observable characteristics. Thus, the differences portrayed in Figure 3 must arise because of differences in the review rates. Let us assume that VO’s worked from the cases that were most likely to be rejected upon closer examination to the ones least likely to be rejected, in order. In that case, the marginal cases whose average effects are being identified using the 2-19% review sample were the most likely to be rejected. The marginal cases in the 20-39% review sample will include those same cases plus an added group who are more needy. Since the 20-39% review sample effects are much
smaller than those from the 2-19% review sample, our reasoning implies that the cases making up the difference between the 2-19% and 20-39% samples must experience almost no effects of benefit denial after about 6 months. Thus, the results by office type indicate clear heterogeneity in responsiveness to denial and may support a model in which there are two groups of IA recipients: a peripheral group for whom benefit denial has structural dependence effects that persist for over 3 years; and a core group who show very limited scarring effects.

Benefit denial effects may also vary with observable characteristics of the applicants. Figure 4 contains plots of IV estimated effects for first time and repeat users who are age 19 to 24. We control for age in this way because age and repeat use status are highly correlated, making it hard to see the effects of one without controlling for the other. Recall that we define first time users as individuals who have no recorded IA receipt in the span of our data (5 years) before the key application. Repeat users do have prior recorded IA receipt. The estimates indicate that both groups have zero long term impacts of benefit denial but first time users experience much bigger effects between about 3 and 30 months than do repeat users. In an earlier version of the paper, we implemented a series of bivariate probits consisting of joint estimation of a process reflecting benefit receipt in a month and a process determining whether an individual was rejected at the initial application. This allowed us to examine the partial derivative effects of variables such as repeat use, age and gender. As in Figure 4, that exercise indicated that first time users have higher estimated effects in every month up to at least 2 ½ years. The bivariate probit estimates also indicated that denial effects rise with age and are somewhat larger for females.

3.4) Decomposing Movements in \( \Delta \)

The nature of the EDP experiment means that we can get estimates of \( E(Y_{1i}) \) and \( E(Y_{0i}) \) rather than just the expected value of their difference. To see this, consider the following identity defining the expected value of \( Y_{0i} \) for applicants in control offices who are granted benefits:
This equation says that control office applicants who are granted benefits can be broken into two sub-groups: the core grantees, who would be granted benefits whether they applied at a VO or control office; and the marginal grantees, who are granted benefits at the control office but would be denied benefits at the VO office. What we would like to observe is the first term after the plus sign on the right hand side of 5): the probability of future IA use for the marginal grantees. To get this, we make several simplifying assumptions. First, we assume that the proportion of control office applicants who are core grantees is the same as the proportion of VO office applicants who are granted benefits. This means that there is a set of cases where benefits are always granted, and we can see how large this set is by observing what proportion of applicants are granted benefits even after extra examination by VO's. Second, we assume that the expected value of $Y_{i0}$ for this core group can be observed as the average outcome for those who are granted benefits in VO offices. For these assumptions to be true we require that the VO's do not reassess initial benefit denials, moving some individuals into the grantee group who would not be there in a control office. That is, we require the “monotonicity” assumption for LATE estimators to hold.

Using these assumptions, we can rewrite 2) as:

This equation says that the core grantees can be represented by those granted benefits in the
treatment offices and to the extent their average outcomes differ from those of grantees in the control offices, the differences must be driven by the average outcome for the marginal grantees. All elements on both sides of the equation are observable except for the expectation for the marginal grantees. Thus, an estimate of the probability of future IA use for the marginal grantees can be obtained by rearranging 3). Similarly, we can get an estimate of the probability of future IA use for the marginal group if they are denied benefits. To get this we assume that there is a core group of applicants who would be denied benefits even without extra screening and that this corresponds to those denied benefits in a control office. The IA behaviour of the marginal group when denied is then derived as a residual element in an equation similar to 3).

In Figure 5 we use the full 20-39% review sample to generate plots of the probability of IA use for the marginal group when granted and denied benefits. Interestingly, the figure indicates that, after the first six months, the fall in the estimated causal effect in Figure 2 is due mainly to the probability of using IA after being granted benefits falling rather than the probability of using IA after being denied benefits rising. When denied benefits, the marginal group initially moves back into the system to some extent but thereafter, their usage rates actually decline slightly with time from a high of about .20 at month 10 to a low of about .13 at the end of our sample period. In contrast, when granted benefits, their probability of being on IA falls rapidly in the first 6 months but slowly and steadily thereafter. This is new information relative to Figure 2, where we saw a declining effect but could not tell if it was due to rejectees moving back into the system or grantees moving out.16

4.0) Discussion and Previous Literature

In this section, we provide a brief interpretation of our results and relate them to work in earlier studies. In doing so, it is worth noting at the outset that the rapid decline in the measured effect of benefit denial is partly mechanical. In the months just after the key application, benefit recipients necessarily have higher recipiency rates because they are still receiving the IA benefits they were granted at the key application while the rejectees would have to re-apply to establish
eligibility. This mechanical movement in the measured impact meets some statistical definitions of state dependence (in that past rejections predict future receipt even in the absence of unobserved heterogeneity effects) but would not meet a behavioural definition. However, Barrett and Cragg (1998) report single IA spell survivor function values of approximately .12 at 12 months and .05 at 24 months for single males and females. These are well below the IA rates of .35 at 12 months and .26 at 24 months for marginal grantees shown in figure 5. Thus, our estimates do not appear to merely reflect grantees moving gradually out of the initial IA spell (the mechanical explanation). They suggest, at least, more persistence than we observe in typical IA spells.

To the extent there are true persistence effects, our results fit with a structural dependence model in which benefit denial effects dissipate over time. Certainly, the larger measured impacts for first time relative to repeat users fits with a structural dependence model. The conclusion that these effects dissipate must be treated with caution, however. In our data, a model with dissipating effects (e.g., one in which marginal grantees have less need of IA over time) cannot be identified from one with perfectly persistent effects. To see this, consider a world in which benefit denial effects are not cumulative past a certain point (e.g., having two denials has a greater impact on behaviour than having one denial but having three denials is no different from having two) and the marginal grantees have a non-zero probability of applying and being denied in the future. In that case, all the marginal grantees will eventually get the “treatment” of being denied and thus all will eventually adopt the patterns observed for the rejectees in our experiment. Convergence between the grantees and rejectees is inevitable, it is just a question of how long it takes. As a policy, that would mean that tightening the system just speeds up the inevitable. In behavioural terms, it would mean that benefit denial has long term behavioural impacts but our measurement makes it appear as though the impact does not last.

We also do not know the exact source of our measured structural dependence effects. In particular, a denial could alter an applicant’s behaviour, causing her to establish a labour market attachment that results in less future dependence on IA. Alternatively, the applicant may become
intimidated by a rejection and be afraid to reapply even if she becomes eligible. Still further, persistence could stem from memory inherent in the system itself. Even though data on denials is not kept, rejectees will likely reapply at the same office and possibly with the same officer. The rejectee could become labelled as a non-deserving person and have greater difficulty accessing benefits even if she becomes eligible.$^{17}$

The other key conclusion from our results is that there is substantial heterogeneity in structural dependence among applicants. The differences between the patterns derived from high and low review rate samples suggests there is a peripheral group for whom there are strong effects of benefit denial on future receipt and a core group for whom there may be virtually zero effects. Finally, the gradual decline in the IA rate for marginal group members when they are denied benefits and the continual decline when they are granted benefits fits with a scenario in which the marginal applicants need for assistance falls with time. Thus, the key application we study may have resulted from some important negative event. However, with time, that negative event may be superseded by other, more positive events, leading the marginal group members to make less use of IA over time whether or not they are initially granted benefits.

4.1) Previous Literature

A substantial portion of the existing literature studying welfare scarring effects focusses on duration dependence within single welfare spells. In this context, the issue of separating true state dependence from unobserved heterogeneity is often addressed by using estimators in which the heterogeneity distribution is estimated as a mixing distribution across underlying duration models. For example, Blank(1989) investigates U.S. welfare data allowing for differing forms for duration dependence and the heterogeneity distribution. In her preferred specification, she obtains estimates that point to two person types: one with a hazard rate out of welfare that is initially high but declines sharply with the length of the spell; and another with a low hazard rate that does not vary over the spell. The finding that there is considerable diversity in the way individuals interact with welfare echos results in other papers in the U.S. literature (see Moffitt(1992) for a thorough
review of U.S. studies). While many of these papers conclude that unobserved heterogeneity is more important than state dependence in explaining persistence patterns, Chay, Hoynes and Hislop(1999) find substantive evidence of state dependence using California administrative data and a “fixed effect” logit specification.

For Canada, Barrett(1998) examines B.C. IA spell duration data using econometric duration models that account for individual heterogeneity. His results for lone parents are very similar to those in Blank(1989): he finds evidence of two underlying groups, one exhibiting duration dependence and the other having a low, flat hazard rate. He concludes that "there is strong evidence that welfare participation does have a scarring effect on single men and women's subsequent labour market careers." (Barrett(1998), p. 16) Work on Quebec by Fortin and Lacroix(1997) supports this conclusion.

There is also some evidence on repeat use of welfare. Using B.C. data, both Bruce et. al. (1996) and Barrett and Cragg(1998) show that time between welfare spells tends to be quite short for single men and women. Thus, Barrett and Cragg characterize IA use by single individuals as consisting of repeated, though sporadic, use of short IA spells. They also show that the probability of returning to IA declines with the time since the last exit and that longer off-welfare spells correspond to higher hazard rates for exiting ensuing welfare spells. Thus, in this dimension of welfare use, too, there is potential evidence of scarring. Stark(1999) examines the probability of moving onto IA using survey data and shows that many of these conclusions persist even after accounting for individual heterogeneity.

In sum, the existing literature indicates that state dependence does exist but that it may be concentrated in a subset of the population. Heterogeneity in responses to welfare also appears to be a significant component of welfare patterns. We also find response heterogeneity but find that structural dependence effects from the benefit denials induced by greater scrutiny of applications are not permanent. The advantage of the approach in this paper is that our instrument allows us to focus on the specific set of applicants who are most likely to be affected by tightening the system.
5.0) Conclusions

In this paper, we use data from a unique experiment to analyse the effects of denying access to welfare benefits at a point in time on future welfare use. In the experiment, Verification Officers (VO's) are assigned to one welfare office in a pair of matched offices. The VO's reassess files of welfare applicants who have been initially slated to receive benefits but are viewed as potentially marginal by the field officer. Using comparisons between the VO offices, which are assigned VO's, and the control offices, which are not, we derive Local Average Treatment Effect estimates of the impact of denying benefits at a point in time on subsequent IA use for the marginal group who are denied benefits because of the VO's. A key advantage of this approach is that we obtain impact estimates that are specific to the very group who would most likely be affected by a policy of tightening welfare eligibility.

We find substantial medium term effects of denial on future welfare receipt, with the measured effect being approximately zero by the 2 ½ year point. The measured impacts are greater for applicants who have never received welfare before and for applicants in offices with lower assessment rates. Finally, a decomposition of the measured differentials indicates that the observed convergence occurs mainly because usage rates of the initial grantees decline to the level of the initial rejectees. IA use for both grantees and rejectees declines over time.

We argue that these results indicate that benefit denial has medium term structural dependence effects. They also point to a scenario in which the applicant population contains a peripheral group which exhibits some behavioural response to being denied benefits and a core group which does not. Thus, the type of tightening undertaken in this experiment can reduce future welfare use for some recipients but it would have to be carefully targeted in order not to simply cause hardship for cases who cannot avoid their need for welfare. Even for the applicants in the peripheral group, the effects do not appear to be permanent. Thus, tightening welfare systems through greater scrutiny, as was done in this case, does not appear to set denied applicants on a permanently different path with respect to future welfare use.
Bibliography


Table 1
Sample Characteristics: Mean Values and Sample Sizes

<table>
<thead>
<tr>
<th>Variable</th>
<th>Control Offices</th>
<th>VO Offices with 2-19% Reviewed</th>
<th>VO Offices with 20-39% Reviewed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Granted</td>
<td>Denied</td>
</tr>
<tr>
<td></td>
<td>0.30</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>DENIAL</td>
<td>0.35</td>
<td>0.35</td>
<td>0.36</td>
</tr>
<tr>
<td>FEM</td>
<td>0.30</td>
<td>0.27</td>
<td>0.35</td>
</tr>
<tr>
<td>AG1924</td>
<td>0.36</td>
<td>0.37</td>
<td>0.32</td>
</tr>
<tr>
<td>AGG35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.33</td>
</tr>
<tr>
<td>FIRST</td>
<td>0.29</td>
<td>0.23</td>
<td>0.44</td>
</tr>
<tr>
<td>IAL1</td>
<td>3.1</td>
<td>3.5</td>
<td>2.0</td>
</tr>
<tr>
<td>IAL2</td>
<td>2.9</td>
<td>3.3</td>
<td>1.9</td>
</tr>
<tr>
<td>IAL3</td>
<td>2.6</td>
<td>3.0</td>
<td>1.8</td>
</tr>
<tr>
<td>IAL4</td>
<td>2.1</td>
<td>2.5</td>
<td>1.4</td>
</tr>
<tr>
<td>UIL1</td>
<td>1.6</td>
<td>1.6</td>
<td>1.5</td>
</tr>
<tr>
<td>ERN15</td>
<td>0.48</td>
<td>0.50</td>
<td>0.43</td>
</tr>
<tr>
<td>ERN15</td>
<td>0.16</td>
<td>0.16</td>
<td>0.15</td>
</tr>
<tr>
<td>ERN15</td>
<td>0.23</td>
<td>0.23</td>
<td>0.25</td>
</tr>
<tr>
<td>ERN15</td>
<td>0.23</td>
<td>0.23</td>
<td>0.25</td>
</tr>
<tr>
<td>PRLWY</td>
<td>0.13</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td>DRPOUT</td>
<td>0.20</td>
<td>0.20</td>
<td>0.19</td>
</tr>
<tr>
<td>URATE</td>
<td>11.5</td>
<td>11.4</td>
<td>11.4</td>
</tr>
<tr>
<td>Number of obs.</td>
<td>13221</td>
<td>9298</td>
<td>3923</td>
</tr>
</tbody>
</table>
Table 2
Sample Characteristics for Cases Reviewed by VO’s

<table>
<thead>
<tr>
<th>Variable</th>
<th>2-19% Granted</th>
<th>2-19% Denied</th>
<th>20-39% Granted</th>
<th>20-39% Denied</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEM</td>
<td>.28</td>
<td>.26</td>
<td>.29</td>
<td>.27</td>
</tr>
<tr>
<td>AG1924</td>
<td>.24</td>
<td>.34</td>
<td>.24</td>
<td>.27</td>
</tr>
<tr>
<td>AG2534</td>
<td>.40</td>
<td>.30</td>
<td>.39</td>
<td>.40</td>
</tr>
<tr>
<td>AGG35</td>
<td>.36</td>
<td>.36</td>
<td>.37</td>
<td>.33</td>
</tr>
<tr>
<td>FIRST</td>
<td>.16</td>
<td>.27</td>
<td>.20</td>
<td>.30</td>
</tr>
<tr>
<td>IAL1</td>
<td>4.46</td>
<td>3.43</td>
<td>4.01</td>
<td>3.35</td>
</tr>
<tr>
<td>IAL2</td>
<td>3.91</td>
<td>3.06</td>
<td>3.88</td>
<td>3.10</td>
</tr>
<tr>
<td>IAL3</td>
<td>3.58</td>
<td>2.73</td>
<td>3.32</td>
<td>2.72</td>
</tr>
<tr>
<td>IAL4</td>
<td>3.18</td>
<td>2.23</td>
<td>2.82</td>
<td>2.08</td>
</tr>
<tr>
<td>UIL1</td>
<td>1.21</td>
<td>1.13</td>
<td>1.39</td>
<td>1.12</td>
</tr>
<tr>
<td>ERNLT1</td>
<td>.66</td>
<td>.54</td>
<td>.53</td>
<td>.54</td>
</tr>
<tr>
<td>ERN15</td>
<td>.13</td>
<td>.18</td>
<td>.18</td>
<td>.16</td>
</tr>
<tr>
<td>ERN515</td>
<td>.17</td>
<td>.20</td>
<td>.20</td>
<td>.19</td>
</tr>
<tr>
<td>ERNG15</td>
<td>.04</td>
<td>.09</td>
<td>.09</td>
<td>.12</td>
</tr>
<tr>
<td>PRLWY</td>
<td>.18</td>
<td>.17</td>
<td>.22</td>
<td>.22</td>
</tr>
<tr>
<td>DRPOUT</td>
<td>.36</td>
<td>.34</td>
<td>.36</td>
<td>.36</td>
</tr>
<tr>
<td>URATE</td>
<td>11.4</td>
<td>10.7</td>
<td>11.7</td>
<td>11.5</td>
</tr>
<tr>
<td># of obs</td>
<td>276</td>
<td>261</td>
<td>956</td>
<td>639</td>
</tr>
</tbody>
</table>
Appendix A
Data Definitions

Table A1
Variable Definitions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AG1924, AG2534, AGG35</td>
<td>= dummy variables which equal 1 if the individual's age in the key application month is 19 to 24, 25 to 34 and over 35, respectively</td>
</tr>
<tr>
<td>FEM</td>
<td>= 1 if the individual is female, 0 otherwise</td>
</tr>
<tr>
<td>FIRST</td>
<td>= 1 if the individual has not received IA before the key application</td>
</tr>
<tr>
<td>IAL1, IAL2, IAL3, IAL4</td>
<td>= number of months of IA receipt in the periods 1 to 12, 13 to 24, 25 to and 37 to 48 months preceding the key application, respectively</td>
</tr>
<tr>
<td>UIL1</td>
<td>= number of months of UI receipt in the 12 months preceding the key application</td>
</tr>
<tr>
<td>ERNLT1</td>
<td>= dummy variables which equal 1 if ROE based earnings in the 12 months preceding the key application (ERN) are such that ERN#$1000, $1000&lt;ERN#$5000, $5000&lt;ERN#$15000, and $15000&lt;ERN</td>
</tr>
<tr>
<td>ERN15, ERN515</td>
<td></td>
</tr>
<tr>
<td>ERNG15</td>
<td></td>
</tr>
<tr>
<td>PRLWY</td>
<td>= proportion of families in Census tract with income below Statistics Canada's Low Income Cut-Off</td>
</tr>
<tr>
<td>DRPOUT</td>
<td>= proportion of population in the Census tract who have not finished high school</td>
</tr>
<tr>
<td>URATE</td>
<td>= unemployment rate in the Census tract in June 1996</td>
</tr>
</tbody>
</table>
Endnotes

1. We use the term welfare when referring to income support schemes in general and IA for BC's system in particular.

2. Unemployables include individuals who cannot work due to medical conditions, single parents with one dependent child under the age of six or two or more dependent children under the age of twelve, individuals over age 65, or a single parent with a disabled child.

3. Benefit amounts remained constant in nominal dollars from January 1996 until the end of our sample period (March, 1999).

4. During the EDP, the test offices averaged 59 cases reviewed by a VO per office per month. In the first 6 months of 1997, all offices in BC together averaged approximately 11 cases reviewed by a VO per office per month.

5. If anticipation of IA receipt also affects application propensities then a more general experiment might go farther and randomize benefits across individuals who have not even so far applied for IA.

6. Individuals who are denied benefits can re-apply at other offices. One potential concern is that those rejected by VO's could reapply, effectively undoing the experiment. However, the percentage of rejectees from VO and control offices who got benefits in the 8 months following the key application is very similar (23% for rejectees from VO offices versus 22% for rejectees from control offices). Thus, there is no evidence that rejectees from treatment offices undid the experiment by disproportionately getting back onto IA.

7. Note, however, that if an individual holds a job in a calendar year but does not separate from that job, an ROE based measure will record zero earnings. Since all individuals will have to separate from any significant earnings job before applying for IA, our earnings measure will be accurate for the twelve months just before the key application.

8. No office assesses 2-39% of applications in every month of the experiment. In constructing the 2-39% review sample, we use the cases for which the key application occurred in a VO office in a month in which the VO assessed 2-39% of all applications.

9. Unfortunately, the exact list for matching offices does not exist and attempts to create specific control samples for each set of VO offices was not successful. Throughout the paper we use the sample from all control offices for comparisons.

10. The standard errors associated with the differences in IA rates at those points are (respectively): .0067, .0066, .0065, .0061, and .0058.


12. All the differences are statistically significantly different from zero at conventional levels.
13. Imbens and Angrist (1994) mention an example in which applicants for a social programme are screened by two different officials. The instrument then takes different values depending on which official screens the case. They point out that even if one official is on average less generous, it is unlikely that all cases accepted by the less generous official would also be accepted by the more generous official. Thus, the third LATE estimator assumption estimator would be violated. This is different from our case, where the VO reassesses cases that have already been granted benefits.

14. We do not include the bivariate probit estimates in this version of the paper because they add little in the way of insight over what can be obtained from the simpler plots presented here. The bivariate probit results are available from the authors upon request.

15. Differences between effects for 19 to 24 year old applicants and applicants over age 35 generally were between .05 and .10. Differences between males and females were generally approximately .04.

16. We also tried a matching approach to identifying the outcomes for the marginal group. In particular, we used the set of applicants actually denied by VO’s to represent the marginal group when rejected and a matched sample of grantees drawn from control offices to represent the marginal group when granted benefits. To select the latter, comparison group, we first ran a probit on the probability of being rejected by a VO using the sample of applicants at VO offices who were originally slated to be granted benefits. We then formed fitted probabilities for the sample of grantees from control offices and used nearest neighbour propensity score matching to select the comparison sample. Unfortunately, even with an extensive set of IA history controls, we were unable to get a pseudo-$R^2$ in the initial probit that was higher than .03. This means the “marginal granted” sample selected in this way was little more than a random sample of grantees from the control offices. Thus, there seems to us to be little to be learned from studying them and we do not present the results here, though they are available upon request.

17. Note that this is an argument about what happens to all rejectees and is distinct from the assumption that being reviewed by a VO does not directly change future outcomes that is necessary for the validity of the IV estimator.
Figure 1
Proportion Using IA by Month: Treatment and Control Offices
Figure 2
Estimates of the Reduction in the Probability of Using IA Due to Denial

- Difference Estimator Based on Control Sample
- IV Estimator Using 2–39% Reviewed Offices
Figure 3
IV Estimates of the Reduction in the Probability of Using IA Due to Denial, by Office Type

[Graph showing trends over time with two lines indicating different percentages reviewed]
Figure 4
IV Estimates, By Repeat Use Status, Age 19–24,
Reduction in the Probability of Using IA Due to Denial

Month Relative to Application

Proportion

First Time Users
Repeat Users
Figure 5
Proportion on IA: Marginal Group
From VO Offices with 2 – 39% Reviewed

Proportion

Month Relative to Application

- Granted
- Denied