

Riddell, Chris; Riddell, W. Craig

Working Paper

When Can Experimental Evidence Mislead? A Re-Assessment of Canada's Self Sufficiency Project

IZA Discussion Papers, No. 9939

Provided in Cooperation with:

IZA – Institute of Labor Economics

Suggested Citation: Riddell, Chris; Riddell, W. Craig (2016) : When Can Experimental Evidence Mislead? A Re-Assessment of Canada's Self Sufficiency Project, IZA Discussion Papers, No. 9939, Institute for the Study of Labor (IZA), Bonn

This Version is available at:

<https://hdl.handle.net/10419/142378>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

IZA DP No. 9939

**When Can Experimental Evidence Mislead?
A Re-Assessment of Canada's Self Sufficiency Project**

Chris Riddell
W. Craig Riddell

May 2016

When Can Experimental Evidence Mislead? A Re-Assessment of Canada's Self Sufficiency Project

Chris Riddell

ILR School, Cornell University

W. Craig Riddell

*University of British Columbia
and IZA*

Discussion Paper No. 9939
May 2016

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

When Can Experimental Evidence Mislead? A Re-Assessment of Canada's Self Sufficiency Project*

The Self-Sufficiency Project was a well-known welfare-to-work experiment that provided a generous but time-limited financial incentive to leave welfare and enter the workforce. Experimental evidence showed large short-term impacts but no lasting effects. We argue that these conclusions need to be re-assessed. Policy changes implemented during the SSP implied the behavior of the control group did not provide an appropriate counterfactual. We estimate the impacts the financial incentive would have had in a stable policy environment. This re-assessment leads to significant changes in the lessons previously reached. Our study demonstrates that experimental findings need to be interpreted with care.

JEL Classification: C90, H53, I38

Keywords: welfare-to-work policies, social experiments, Self-Sufficiency Project

Corresponding author:

W. Craig Riddell
Department of Economics
University of British Columbia
#997 - 1873 East Mall
Vancouver, B.C.
Canada V6T 1Z1
E-mail: craig.riddell@ubc.ca

* We thank David Card and David Green for very helpful comments. We have also benefitted from comments of Tom Crossley, Josh Gottlieb, Hilary Hoynes, Jesse Rothstein, Chris Walters along with seminar participants at Berkeley, Santa Barbara, UBC, Brandeis, Essex, Minnesota, IZA/OECD/World Bank Social Safety Nets workshop in Paris, CLRSN workshop in Toronto, the IZA/IFAU Labour Market Policy Evaluation workshop in Uppsala, and the J-PAL Field Experiments and Social Policy conference in Paris.

1. Introduction

The increased emphasis on obtaining credible evidence has resulted in much greater use of randomized experiments in economics. Random assignment ensures that the treatment and control groups are statistically indistinguishable at the baseline. Thus the behavior of the control group provides an unbiased estimate of the counterfactual behavior of the treatment group and any difference in outcomes between treatments and controls can be attributed to the causal effects of the intervention. A further advantage is that experimental impact estimates are simple – often differences in mean outcomes between treatments and controls – and easily understood by experts and non-experts alike.

However, social experiments have limitations, some of which may affect the internal validity of the experimental evidence (Heckman and Smith, 1995). Non-random attrition can result in treatment and control groups that differ, even though the two groups had very similar characteristics at the baseline. Those assigned to the control group may obtain services similar to those provided to the treatment group, resulting in “substitution bias” that may result in under-estimating the impact of the intervention (Heckman, Hohmann, Khoo and Smith, 2000).

In this paper we illustrate another potential problem with social experiments – sometimes referred to as ‘contamination’ – that influences the interpretation of experimental results. Randomization ensures that the treatment and control groups are statistically equivalent at the baseline. However, once treatment begins the characteristics of the experimental groups will generally diverge. For example, in a randomized drug trial the health of the treatment group will improve relative to that of the control group if the treatment is effective. Subsequent events such as changes to the economic or policy environment may exert different impacts on the two groups. In such circumstances the behavior of the control group may no longer provide an appropriate counterfactual for the altered treatment group. Different experimental estimates could have been obtained if the social, economic or policy environment had evolved differently – raising questions about the external validity of the experiment. Because of the credibility associated with random assignment, there is a risk that experimental evidence may be interpreted too broadly or literally, rather than being viewed as being conditional on the evolution of events during and after the experiment.

This paper argues that the conclusions that have been reached on the basis of a well-known welfare-to-work experiment – the Self-Sufficiency Project – need to be re-assessed because of policy changes that took place during the SSP demonstration. The SSP was carried out in the 1990s in two Canadian provinces – British Columbia and New Brunswick. The experimental sample consisted of single parents who were long-term welfare recipients. The treatment was a generous but time-limited earnings supplement provided to treatment group members who left welfare and took up full-time work. One objective of the SSP was to rigorously test whether “making work pay” would lead to significant reductions in welfare use and increases in labor force participation among this population. A second objective was to investigate whether temporary financial incentives can lead to lasting reductions in welfare use. The results of the SSP Demonstration were striking. The financial incentive resulted in large impacts during the supplement period. However, treatment-control differences in employment and welfare receipt gradually faded and not long after the supplement period ended there were no significant treatment-control differences in employment and welfare receipt. The absence of lasting impacts of a generous but temporary earnings supplement reduced enthusiasm for this approach to welfare reform.

However, important developments took place in both provinces – events that call for a re-assessment of the SSP experimental findings. During the SSP Demonstration, BC introduced a major ‘work first’ welfare reform that made continuing receipt more difficult and created financial incentives to work. Subsequently New Brunswick’s social assistance program also underwent significant changes. In addition, in NB another welfare-to-work program – New Brunswick Works – operated at the same time as SSP, and members of the SSP experimental sample were eligible for and did participate in this alternative program. Participation in NB Works involved leaving welfare, and was more common among the SSP control group than among the treatment group. We show that the developments in both provinces raise questions about the interpretation of the experimental estimates. For BC and NB we estimate the impacts that the SSP treatment would have had in a stable policy environment. In the case of NB we also adjust the experimental estimates for participation in NB Works. In both provinces this re-

assessment leads to significant changes in the lessons previously reached on the basis of the SSP demonstration.

A key feature of the SSP was that intake and initial random assignment was staggered over time. The BC and NB welfare reforms therefore affected intake cohorts at different stages of their SSP treatment. We use this variation to estimate the impacts of the policy changes and to simulate what would have occurred in the absence of the reforms.

The paper makes three contributions. First, we show that time-limited earnings supplements that make work pay can have lasting effects on welfare use and labor force participation, a result of interest for welfare policy and more generally for earnings supplements for low income workers. Second, we provide a reminder that experimental findings need to be interpreted with care. Events that occur during or after the experiment may affect the validity of the experimental estimates. Finally, for the design of social experiments, our study illustrates the value of staggering entry over time. Doing so may allow identification of treatment effects in the event of unanticipated changes in the economic or policy environment during or after the experiment.

2. Contamination in Social Experiments

Consider a randomized drug trial to test a medication intended to lower blood pressure. If the drug is effective, once treatment is underway the control group will have a larger fraction of members with elevated blood pressure and the treatment group will contain a greater proportion of individuals with normal levels. After treatment is underway, nutritional scientists discover a new ‘super food’ that reduces blood pressure, and is particularly effective in lowering blood pressure from the ‘high’ range to the ‘normal’ range. Both experimental groups have access to this food. Even if both groups increase consumption of this super food to the same extent, the impact on blood pressure will be greater for the control group than the treatment group. In most circumstances the behavior of the control group will no longer provide an appropriate counterfactual for the impact of the medication being studied. The experimental estimates reflect a combination of two treatments: the blood pressure medication and the discovery of the super food and its properties. The experimental groups were randomly assigned for the blood pressure medication, but are not ‘as good as randomly assigned’ for the second treatment. The

experimental estimates are unlikely to provide an unbiased estimate of the impact of the blood pressure medication alone.¹

Figure 1 illustrates the potential for contamination in the context of welfare-to-work initiatives evaluated with an experimental design. The factors that influence the probability of leaving welfare are combined into an index referred to as “job readiness.” Those with higher values of job readiness are more likely to exit welfare. At the baseline t_0 , when random assignment takes place, all members of the experimental sample are receiving welfare. Half of the experimental sample is randomly assigned to the treatment group, which receives an intervention that provides an incentive to leave welfare and enter the workforce. The remainder is randomly assigned to the control group. Treatment status is independent of the observed and unobserved characteristics of both groups at time t_0 . Thus the distributions of job readiness in the treatment and control groups are identical (see the top panel in Figure 1).

For the purposes of exposition we assume that the treatment is effective, so that the welfare exit rate of the treatment group exceeds that of the control group in the post-baseline period. Subsequently, at time $t^* > t_0$, the existing welfare policy is changed in a way that encourages recipients to exit welfare and enter the workforce. This change can be thought of as a second treatment, the first being the incentive offered to the treatment group and the second being the new policy that applies to members of both the treatment and control groups who remain on welfare at time t^* (and to recipients not in the experimental sample). However, while the initial incentive treatment was independent of the characteristics of the treatment and control groups, the second policy change “treatment” is not independent of the characteristics of the two groups. In particular, at time t^* , compared to the control group a smaller proportion of the treatment group remains on welfare and those in the treatment group who still receive welfare have a lower average propensity to exit. This is illustrated in the bottom panel in Figure 1, where the area to the left of the vertical line indicates the fraction of each group that remains on

¹ As a second example, consider a randomized trial of a chemotherapy treatment carried out in a hospital or other institutional setting. We assume that the chemotherapy is effective, so in the absence of contamination the survival rate of the treatment group would exceed that of the controls. However, during the chemotherapy treatment an outbreak of C-difficile occurs in the institution. An unfortunate side effect of chemotherapy is a compromised immune system; as a consequence more treatment group members die from C-difficile than do control group members. Because of this contamination the experimental estimates will not yield an unbiased estimate of the impacts on survival of the chemotherapy treatment.

welfare at t^* . At time t^* the control group has more job ready welfare recipients than does the treatment group. Thus the policy change introduced at t^* is likely to have a larger impact on the exit rate of the control group than on that of the treatment group. In these circumstances the behavior of the control group after t^* may no longer provide an appropriate counterfactual.²

In the next two sections we present evidence that strongly suggests that this type of contamination occurred in the SSP demonstration.

3. Welfare Reform and the Self-Sufficiency Project

A frequent criticism of welfare programs is that they provide little incentive for recipients to seek employment. Under many such programs, recipients who enter the workforce are required to forego benefit payments by the amount of their labor market earnings -- implying that earnings are taxed at a rate of 100 percent. The implicit tax rate may even exceed 100% if, for example, those leaving welfare are no longer eligible for medical benefits or subsidized housing.

Several reforms have been proposed to deal with this incentive problem. One strategy is to raise the market wage of recipients through training and employment programs, thus making work more attractive relative to welfare. Another approach improves work incentives by reducing the implicit tax rate on market earnings. Examples of this approach include the negative income tax, earnings disregards, and income supplementation policies such as the Working Income Tax Benefit in Canada and the Earned Income Tax Credit in the U.S.³ A third strategy attempts to alter the preferences of recipients, either by raising the stigma associated with welfare receipt or enhancing the perceived value of work.

Some policies combine elements of two or more of these approaches. An interesting example is a temporary earnings supplement for welfare recipients who enter the workforce. During the period the supplement is in place, this policy has the work incentive features of many income supplementation schemes. Labor market earnings are implicitly taxed at a rate less than 100% and program participants receive income (market

² The policy change at time t^* may also influence the probability of re-entry on to welfare. This effect on the re-entry rate could also differ between the experimental treatment and control groups because of differences in the characteristics of the two groups at time t^* .

³ Earnings disregards refer to market earnings that welfare recipients are allowed to receive without a reduction in their benefits.

earnings plus the supplement) that exceeds welfare benefits. By encouraging recipients to leave welfare and enter the workforce for at least the period of the supplement, former welfare recipients may gain work experience and enhance their skills, thus raising their market earnings. The experience of working for an extended period of time may also alter individual's preferences between welfare and work. As a result of enhanced earnings capacity and/or altered preferences toward work, a temporary financial incentive may have lasting effects on welfare receipt and labor force participation.

During the 1990s the Government of Canada funded an innovative demonstration project, the SSP, designed to provide evidence on the effects of a financial incentive on long-term welfare recipients.⁴ The SSP demonstration was carried out in British Columbia and New Brunswick, and focused on single parents with dependent children who had been on income assistance (IA)⁵ for at least 12 of the previous 13 months.⁶ Among those who agreed to participate, one-half were randomly assigned to the treatment or program group that was eligible for the earnings supplement; the rest were assigned to the control group. Random assignment took place between February 1992 and November 1995. Those in the treatment group were offered a financial incentive to leave welfare and take up full-time employment.⁷ The financial incentive was generous, approximately doubling income from work for the typical participant and providing total income substantially higher than welfare benefits.

The SSP demonstration incorporated two important time limits. Members of the program group were given up to 12 months following random assignment to obtain full-time employment. Once they had qualified, participants could continue to receive the supplement for three years providing they maintained full-time employment. Those in the control group could remain on welfare or enter the workforce. Card and Hyslop (2005) show that the two SSP time limits generated an “establishment” incentive to find a full-time job and exit welfare within 12 months after random assignment, and an “entitlement” incentive to choose work over welfare once eligibility was established.

⁴ For details of the SSP demonstration see the Final Report, Michalopoulos et. al (2002).

⁵ In Canada welfare programs are operated by the provinces, and are referred to as Income Assistance (IA) or Social Assistance (SA). IA/SA is not restricted to families with children.

⁶ About three-quarters of the SSP sample had been receiving IA for more than 2 years, and more than 40% had received IA for more than 3 years.

⁷ Full-time employment was defined as at least 30 hours per week and could be achieved by combining two or more part-time jobs.

A key objective of the SSP Demonstration was to determine whether financial incentives lead to reductions in welfare use among long-term IA recipients, and whether the magnitudes of program impacts on IA use and employment are sufficient to support this approach to welfare reform. Another key objective was to test whether a temporary financial incentive could have lasting effects on welfare receipt and work activity. The potential receipt of a substantial earnings supplement for up to three years was intended to provide such a test.

The experimental findings are summarized in the SSP Final Report (Michalopoulos et. al., 2002). Because of staggered entry, experimental impacts are typically reported in “SSP time” or time since baseline. More than one-third of the treatment group obtained full-time employment and qualified for the earnings supplement. During the eligibility period, the treatment group experienced substantial gains in employment and earnings and reduced welfare use relative to the control group. The largest impacts were observed during the first 12-15 months following random assignment. After this time the differences in outcomes between the treatment and control groups gradually narrowed. By the end of the 3-year period of supplement eligibility treatment-control differences in employment, earnings and welfare receipt were small. In particular, by the 54-month point – by which time the three-year supplement period had ended for all eligible participants -- there was no difference in full-time employment rates, part-time employment rates and average earnings between the two experimental groups (Michalopoulos et. al., 2002, chapter 3). Similarly, treatment-control differences in income assistance receipt had faded to zero by month 69 (Card and Hyslop, 2005).

There were, however, important differences between the two provinces in the impact of the SSP financial incentive. As illustrated in Figure 2a, income assistance rates of the program and control groups had fully converged in BC by month 54, the date of the final SSP survey, whereas a gap of 6 percentage points remained in NB (see Figure 2b).

Within the SSP treatment group, the characteristics of the single parents who obtained full-time employment and established eligibility for the earnings supplement differed from those of the recipients that did not establish eligibility. Specifically, the “eligible subset” had characteristics associated with higher wages and being more “job

ready.” Compared to the treatment group members who did not establish eligibility, the eligible subset were more highly educated, had more previous work experience, were more likely to be working at the baseline and were less likely to have been on income assistance continuously for the past three years (Card and Hyslop, 2005). Thus within the treatment group the most employable left welfare for full-time employment and established eligibility for the earnings supplement and the least job ready remained on income assistance.

The behavior of the SSP control group during the demonstration was also revealing. Despite the experiment’s focus on single parents with dependent children who had been on welfare for an extended period of time, these long-term IA recipients displayed a steady exit from welfare over time at a rate that, although modest, surprised many observers, including those familiar with previous research. For example, Barrett and Cragg (1998) used detailed administrative data from BC to analyze movements onto and off income assistance and found that single parents with dependent children had by far the longest IA spells and that there was evidence of negative duration dependence for this group. Since SSP eligibility was restricted to those who had been receiving IA for at least one year, slow exit from welfare was expected from the long-term recipients that were the focus of the SSP Demonstration.

The SSP Demonstration has received substantial attention in the research and policy literature – examples include Blank and Card (2000), Bitler, Gelbach and Hoynes (2008), Blundell (2002, 2006), Blundell and Hoynes (2004), Card and Hyslop (2005, 2009), Connolly and Gottschalk (2009), Lise, Seitz and Smith (2015) and Robins and Michalopoulos (2001). One reason was the generosity of the financial incentive, thus providing a striking test of the view – often expressed by economists – that “making work pay” would substantially alter behavior. For example, Blank, Card and Robins (2000) reviewed seven financial incentive welfare-to-work programs evaluated by random assignment in the US and Canada during the 1990s. The SSP earnings supplement was the most generous of this group. Another factor was the very large short-term response combined with the absence of a longer-term impact, a combination that Card and Hyslop (2005) shed valuable light on. A third feature that attracted both research and policy attention was the absence of wage progression among those in the program group who

qualified for the earnings supplement, despite their accumulation of work experience during the 3-year supplement period (Card and Hyslop, 2005; Blundell, 2002, 2006; Blundell and Hoynes, 2004). Within Canada there is evidence of SSP's influence on welfare policy. The province of Quebec introduced a program ("Action emploi") almost identical to the SSP for a subset of its welfare population (Lacroix, 2009). Many other provinces strengthened financial incentives to "make work pay" in the welfare reforms carried out in the 1990s and subsequently.

The purpose of this paper is to re-assess the key experimental findings of the SSP demonstration. We argue that the experiment was subject to potentially important contamination from policy changes that pose a threat to the validity of the experimental evidence. In particular, we provide strong evidence that the SSP financial incentive would have had lasting effects on welfare receipt and employment in a stable policy environment.

4. Contamination in British Columbia: Welfare Reform during the SSP

In the mid-1990s the B.C. government made sweeping changes to the provincial income assistance program (National Council of Welfare, 1997; Gornick, 1998).⁸ These changes strongly encouraged welfare recipients to leave income assistance and enter the workforce.⁹ In particular, "employable" recipients were required to actively seek work, and to remain eligible for IA could only refuse a job under a strict set of conditions. Some key changes affecting single parents with dependent children were introduced in December 1995 (when the definition of "employable" was broadened considerably) and January 1996 (when the exemption from the requirement to actively seek work was changed from exempting single parents with a dependent child under 12 years of age to those with a dependent child under 7 years of age). At the same time, new sanctions prohibited anyone quitting a job without just cause from receiving income assistance for 6 months. Thus former recipients who quit a job might not be able to return to income assistance, contrary to the original SSP design.

⁸ Although these changes culminated in the passage of the B.C. Benefits Act in October 1996, many important policy changes were introduced prior to the passage of the new Act through revisions to the regulations under the previous legislation governing the income assistance program.

⁹ Indeed, the new names for major components of the program – "Youth Works" for those age 19-24 and "Welfare to Work" for those age 25-59 made clear the shift in the intent of the program.

In April 1996 the earnings disregard was substantially reduced, making combining welfare with some work less attractive. The previous flat rate disregard of \$200 per month was eliminated and the “enhanced disregard” was limited to being received 12 months in a lifetime (Michalopoulos et al, 2002).¹⁰ In addition, the BC minimum wage increased for the first time since 1993 (i.e. at the beginning of the first SSP in-take) – by \$1 per hour (from \$6 to \$7).¹¹

Perhaps the most fundamental change was the introduction of the BC Family Bonus in July 1996. This new program paid up to \$103 per child each month to low-income families and reduced income assistance payments by the same amount, raising the incentives for recipients with children to work.¹² Other substantive changes came with the introduction of the Healthy Kids Act that raised the incentives to work (or to demonstrate that one is actively looking for work) by providing financial incentives that included sizeable day care subsidies and dental, vision and other extended health benefits to an individual’s dependents. Payments under the Healthy Kids Act began in January 1997.

As displayed in Figure 3, raw welfare caseloads declined substantially, beginning with the introduction of BC Benefits in January 1996. For single parents with dependent children particularly steep declines in IA caseloads are evident following the introduction of the Family Bonus in July 1996.¹³ Although the analysis is entirely descriptive, government reports and commissioned studies attribute the bulk of the decline to the Family Bonus (BC Ministry of Human Resources, 1997; Battle and Mendelson, 1997). While the various SRDC reports on the SSP do note many of the policy changes made in BC, to our knowledge no analysis of the impacts of these policy changes on the experimental findings has been carried out.¹⁴

¹⁰ Previously the enhanced disregard allowed a recipient to set aside up to 25% of earnings in excess of \$200 per month in 12 of every 36 months.

¹¹ The increase occurred in two phases: 50 cents on April 1996 and 50 cents on October 1996.

¹² The maximum monthly benefit of \$103 per child was paid to families with incomes of \$18,000 or less, decreasing to zero when income reaches \$34,000 for a family with 1 or 2 children. For larger families the payments decline less rapidly.

¹³ Figure 3 shows caseloads for all families, not our population of interest: single parents. Unfortunately, the data are not available for single parents. However, based on the evidence in Battle and Mendelson (1997), the trends are apparently very similar for single parents.

¹⁴ See, for example, the discussion in the Final Report (Michalopoulos et. al. 2002, pp. 6-9).

In Figures 4a and 4b we present evidence that these policy changes accelerated the exit from welfare of SSP participants, and that they appear to have had a different effect on the control group than on the treatment group. To do so, we show income assistance rates of SSP entering cohorts in *calendar time* as opposed to “SSP time” (i.e., time since baseline/onset). The analysis in this section uses the monthly welfare administrative data and the full SSP sample. The only information we require is IA status, treatment group status, province, and the individual’s baseline interview date.¹⁵ The distribution of observations by cohort is displayed in Appendix A. Purely for graphical purposes, we use the first six (January to June) cohorts of the 1993 cohorts (referred to as the ‘early cohort’), and the last six cohorts (July to December) of the 1994 cohorts (the ‘late cohort’). This allows for nearly a full calendar year difference between cohorts in terms of when the 1996 reforms affected SSP participants. While our empirical analysis uses a variety of robustness checks, Figures 4a and 4b specifically highlight (via the vertical line) July 1996, the date of the Family Bonus, which appears to be the key part of the reforms that accelerated exit from welfare among single parents.

Previous experimental analyses of impacts on outcomes such as income assistance receipt, employment and earnings examine behavior by month from random assignment rather than by calendar month – as displayed in the previous section in Figures 2a and 2b. Doing so is the appropriate approach for many purposes, but it obscures the possible effects of policy changes because it pools together cohorts that entered SSP at different points in time.

Figure 4a shows B.C. cohorts that were randomly assigned between January and June 1993. For this group, the one-year eligibility period had expired by June 1994. Note the large treatment-control gap in IA receipt that opens up during the period June 1993 to June 1994. After the initial eligibility period IA receipt continues to decline for both groups, and although the gap narrows somewhat it nonetheless remains substantial. IA receipt by the control group declines fairly steadily over the period from May/June 1993

¹⁵ The SSP surveys are very rich in detail but only go from the baseline to 4.5 years later (the final survey at 54 months). In contrast, the administrative data goes to 70 months, nearly 6 years post-baseline. Using the SSP surveys also limits the analysis to those who responded. In other analysis, we incorporated SSP survey data on household structure – in particular, changes over time in dependents and marital status – as the data also contains specific dates for changes in household composition (although only until the 54 month point). However, changes in household composition have no effect on our results, and thus it seems clear that using the full 70 months of administrative data is the appropriate choice.

to December 1995. However, in early 1996, and especially in July 1996, the control group's exit rate accelerates. By January 1997 the IA rates of the treatment and control groups are equal.

Figure 4b shows the IA rates of BC cohorts randomly assigned during the final half of 1994. This group had just completed their initial eligibility period when the new BC Benefits regime was introduced. Again, the exit rate among control group members accelerates with the introduction of the policy changes. The two figures are very similar in two respects: (i) a large treatment – control gap opens up during the 12-month eligibility period and (ii) the sharp decline in IA rates following January 1996, and in particular after the Family Bonus in July 1996. *We emphasize that only through the staggered nature of entering SSP cohorts can we illustrate this difference in the timing of individuals being 'affected' by the reforms.*

The visual evidence suggests that the 1996 policy changes had a larger impact on the control group than on the treatment group. Such a differential effect is consistent with the control group having a larger proportion of relatively employable members than the treatment group at the time the new policies were introduced. To check this hypothesis we constructed a 'job readiness index' as a weighted average of baseline characteristics that influence the likelihood of exiting IA, with the weights being estimated coefficients from a regression of the probability of receiving income assistance on these covariates.¹⁶ To avoid the index being influenced by the SSP financial incentive and the 1996 policy changes, the sample consisted of monthly observations on the BC control group (1993 entry cohort) up to December 1995. We scaled the predicted values for each individual in each month by the sample mean value of the probability in June 1993. Thus the index exceeds one for those with above average employability, and is less than one for those less likely than the average recipient to exit welfare. Figure 5 plots the time series behavior of this index for four subsets of the BC early cohort: Programs on and off IA, and Controls on and off IA.

The behavior of the index accords with the view that more employable recipients are more likely to leave welfare – the 'off IA' subgroups of programs and controls have

¹⁶ The variables included in the regression are: dummies for high school graduate, working full-time at baseline, 'likes to work', 'on welfare as a child', disability, foreign born, and continuous variables for years of work experience and its square. All variables are highly statistically significant.

noticeably higher values of the JR index than their 'On IA' counterparts. The 'Off IA' group is much smaller and the index thus noisier. During the 12-month eligibility period (that ended June 1994), gaps open up between programs and controls for both the 'Off IA' and 'On IA' groups. Within the 'Off IA' group, the more substantial decline in the index for the treatments reflects the fact that a greater fraction of the program group exited welfare during this period, driving down the index because at the margin less employable recipients entered the workforce. Similarly for the 'On IA' group, the decline in the index for treatments relative to controls reflects the more rapid exit from IA by program group members, leaving on IA a smaller pool of less employable recipients compared to the control group.

Note also that the measured employability of all four groups trends downward over time. This is what we would expect if at each point in time the most employable member of the 'On IA' group exits welfare. Doing so lowers the value of the index for both the On IA and Off IA groups because at the margin the most employable 'On IA' recipient is the least employable person in the 'Off IA' group.

Following the end of the eligibility period we see gradual convergence in job readiness between treatments and controls in the 'On IA' group, reflecting somewhat more rapid exit from IA in the control group, lowering the average job readiness of those still on IA. This convergence is dramatically hastened by the 1996 BC policy changes, especially the introduction of the BC Bonus in July 1996. The impacts of the policy changes are much more evident for the smaller 'Off IA' group, with the substantial treatment-control gap narrowing dramatically after the policy changes. The sharp decline in the index for the control group after the introduction of BC Bonus is consistent with the view that this legislation resulted in rapid exit from IA among the controls, reducing the average employability of this group. The impact of BC Bonus on the controls differs substantially from that on the treatments off IA, whose job readiness index is stable after mid-1996.

The evidence illustrated in the Figure 5 provides strong support for the hypothesis that, at the time of the policy changes, the control group had a greater proportion of relatively employable recipients, and was therefore more impacted by the new policies than the treatment group. Because of the differential effect of the policy changes on the

control group the experimental evidence is likely to under-estimate the true impact of the SSP financial incentive on income assistance receipt. Figures 4a and 4b also suggest that the behavior of the BC control group is likely to over-state the rate at which long-term welfare recipients leave IA and enter the workforce without financial or other incentives to do so. In our empirical work below we use New Brunswick as a falsification test since the time period for the reforms relevant for British Columbia should have had no effect on individuals in New Brunswick.

We now turn to a simple empirical framework for analyzing income assistance patterns in BC that will allow us to simulate what would have happened if the BC Benefits legislation had not been passed. Specifically, we estimate regressions of the following general form:

$$IA_{i,t} = \alpha + \delta * LAW_{i,t} + \lambda(SSP_TIME_{i,t}) + \psi(CALENDAR_MONTH_{i,t}) + \gamma(UNEMRATE_{i,t}) + \varepsilon_{i,t} \quad (1)$$

where $IA_{i,t}=1$ if individual i was receiving welfare in month t . Note that we estimate equation (1) separately by experimental group since the transitions over time are fundamentally different due to the incentives introduced by the earnings supplement. Our preferred specification includes a fixed effect for each individual. We use several different specifications for the policy change variable. The simplest specification defines $LAW=1$ for the January 1996 period onwards. Note that the point at which entering cohorts reach January 1996 will differ substantially (in terms of their lapsed time in the SSP program) depending on the cohort's baseline interview date. We then also test this specification using July 1996 onwards. Ultimately, it appears that the introduction of Family Bonus is the key date although the results are very similar whether we define the reforms as beginning in January or July 1996. It is likely given the discussion in the previous section that the impact of the new welfare regime will differ over time, and thus our preferred specification includes an interaction between LAW and a linear time trend. We also estimate a specification interacting LAW with a quadratic in time.

SSP_TIME refers to 69 monthly dummies that capture the welfare paths in SSP time; that is, time since baseline. $CALENDAR_MONTH$ are calendar monthly dummies,

and UNEMRATE is the prime-age unemployment rate measured in calendar monthly time. As noted, Appendix A shows baseline cohort dates in calendar time for the administrative data on welfare receipt.

Our identification strategy rests on the notion that baseline interview dates are exogenous to welfare receipt – that is, that individuals with early exposure to the new law can be used as a counterfactual for those with later exposure to the new law. This underlying stationarity assumption could be violated if there were seasonality effects or if labor market conditions changed. This could well be the case as the early 1990s recession was particularly severe in Canada and may have influenced labor market prospects for earlier cohorts (1993) to a greater extent than later cohorts (1994-95). The monthly unemployment rate – again, measured based on calendar time – is incorporated to capture seasonality and changes in labor market conditions. Finally, we estimate (1) for New Brunswick as a falsification test; LAW should not be statistically significant.

In addition to the ‘natural’ time trend in IA being different across experimental groups (due to the incentive effects of the SSP offer), the law is expected to affect the two groups differently as discussed above. The effect, if any, should be larger for the control group since relatively more of the program group left IA prior to the BC policy changes in response to the SSP offer. Note that there could be an effect for the program group since those who had not left IA became subject to the new welfare regime as was the case for individuals in the control group who had not yet left IA. If the BC Benefits Act (i.e., January 1996 on) had no differential effect across baseline cohorts then δ should be zero.

Before proceeding to the results, it is useful to emphasize what variation is used to identify the policy variable. The LAW variable only differs across baseline interview cohorts with respect to when the BC Benefits Act affects individuals in a given cohort. This is the same type of variation we see when examining Figures 4a and 4b; the IA path for the control group decreases much more quickly following January 1996 for all cohorts, but there is a substantial difference across cohorts in time since random assignment in terms of when the January 1996 changes affect them. Similarly, we could imagine conducting a series of ‘natural experiments’ where we compare, for example, the January 1993 cohort with the January 1994 cohort. In the latter case, the earlier cohort will have an extra year in their SSP participation before becoming covered by the new

legislation. Any difference in relative IA use in that year is what identifies δ . This is the only variation exploited in the regression equation above. We could also use New Brunswick as a comparison group, which would always equal zero for the interaction term. When we do this, our estimated policy effects are somewhat larger, but our preferred specification is (1) since this only relies on one source of variation: the timing of when a baseline cohort becomes covered under the legislation. We then use New Brunswick as a falsification test.

Table 1 presents the regression results and Figure 6 presents the predicted IA rates in the absence of the BC policy reforms. These counterfactual IA rates are shown in terms of time since onset, and thus can be compared to the SSP experimental data. The BC Benefits law reduced IA rates for both the Program Group and the Control Group, but by a larger amount for the Control Group. The top panel of Table 1 defines the reforms using January 1996; the effects are somewhat larger if we use July 1996 when Family Bonus was introduced as displayed in the bottom panel of Table 1. The data cannot distinguish between the two dates however.

The policy variable has essentially no effect in New Brunswick.¹⁷ We note that the New Brunswick regressions presented in Table 1 use only the later cohorts (1994-95) for the falsification test because the IA patterns for the early cohorts would mechanically bias the coefficients towards zero as we discuss further below. To briefly preview those results, we document remarkably flat IA patterns for most of the sample period for the early NB cohorts. This appears to be due to substitution bias towards NB Works as outlined in the next section. Because of this bias the coefficients for NB are even closer to zero if we include the 1992-93 cohort observations.¹⁸ In addition to substitution bias associated with NB Works, policy changes similar to those in BC were introduced in New Brunswick in mid-1997. We discuss these, and analyze their impacts, later in the paper. At this point, we note that the coefficient on the linear interaction term for NB controls is likely picking up the effects of mid-1997 policy changes. We believe this explains the marginally significant coefficient in the bottom panel of Table 1 (and the corresponding coefficient that is close to being significant in the top panel).

¹⁷ The only exception is the marginally significant coefficient for the controls in columns 11 and 12, discussed below.

¹⁸We also note that for New Brunswick most of the data was in the 1994-95 cohorts (see Appendix A).

As noted, Figure 6 asks the following counterfactual question: what would the pattern of IA rates have been if the BC Benefits Act had not been passed? We simulate this by predicting IA rates if the LAW variable was always zero. Figure 6 is based on the specification (3) from Table 1 and includes a fixed effect for each individual as well as the policy interaction term –this is our preferred specification. That said, we get the same basic story as shown in Figure 6 regardless of the specification; the main difference is in the precision of the estimates and thus the width of the confidence intervals.¹⁹

Not surprisingly, given the coefficients on the policy variables from Table 1, we see sizeable effects. We note that our model predicts IA rates to be identical to actual IA rates. By month 70 post random assignment, IA rates in BC are estimated to be 26 percentage points higher than the actual rates (71% vs. 45%) for the control group if BC Benefits had not been passed, and 19 percentage points higher for the program group (63% vs. 44%). In addition to suggesting that the BC policy reforms led to IA exit rates in the SSP that have been misinterpreted as being too high, the differential effects of the law across experimental groups are such that there may have been a lasting effect of the SSP in BC. We emphasize this is not 4.5 years post-baseline as in most of the previous literature but almost 6 years post-baseline. Overall, our simulations suggest that without BC Benefits, the experimental effect would have been quite constant around 6-9 percentage points, depending on the specification, over the final 3 years, suggestive of a lasting effect.

5. Substitution Bias in New Brunswick: NB Works

Together with the federal government, the province of New Brunswick launched New Brunswick Works in late 1992 (New Brunswick Department of Advanced Education and Labour, 1992; New Brunswick Department of Income Assistance, 1995; National Council of Welfare, 1997). This was a voluntary pilot project where individuals were removed from income assistance and inserted into programming where they received skills/job readiness/education upgrading, job placements, together with UI benefits and training grants for three years. Interestingly, NB Works and the SSP were jointly

¹⁹ The overall pattern in the experimental difference is somewhat larger using the Family Bonus/July 1996 date, and somewhat smaller (but also more precise) using just the dummy variable (as opposed to the specifications that also include the linear interaction) The overall findings are also similar (with even larger estimated experimental effects) if we use the quadratic interaction.

announced by the Minister of Labour at a press conference that received considerable media attention. Individuals were eligible for NB Works if they had been on welfare for at least 6 months, but preferably 12 months or more, and were deemed likely to become long-term welfare recipients in the absence of intervention. There were three intake periods of approximately 1000 individuals ranging from 1992 through to 1994.

Individuals in NB Works went through a three-year revolving set of employability programming. The timeline was as follows: an initial job readiness counseling and case development process for 2 months, a 24 month rotation of 3 months job placement (qualifying the individual for unemployment insurance), and 9 months of general skills upgrading and job preparation, a second set of job placements; concluding the program with a combination of assisted job search for unfunded employment and/or subsidized private sector placements. It is important to stress that individuals no longer received income assistance while they remained in the program, but rather were funded through the federal unemployment insurance system along with job training grants, and earnings from their job placements. Thus, from the perspective of this paper's objective of re-assessing the SSP, individuals in the SSP data who participated in NB Works were removed from income assistance; if the control group were more likely to enroll in NB Works we would have substitution bias in the SSP.

While to the best of our knowledge no published academic work has evaluated the NB Works program -- and unfortunately program and administrative data are thin -- there was substantial evaluation work commissioned by the NB government from 1993 to 1996 (e.g., PGF Consultants 1993; Johnson 1993). Much of this work was qualitative including interviews of case managers and program participants. It seems clear that the NB Works pilot was met with enormous enthusiasm in its early days. Moreover, many participants specifically indicated that they saw the program as a way out of welfare. For individuals in the SSP experimental sample the NB Works program was thus an obvious alternative to the SSP as a means of leaving welfare. The program was voluntary, but individuals in the SSP treatment group would be less inclined than those in the control group to enroll in NB Works because (i) unless the timing worked out (unlikely for most) it would typically be difficult to qualify for the SSP earnings supplement (i.e., achieve at least one month of full-time work within the one year qualification period), and (ii) given the

amount of time spent out of work in NB Works, an individual who qualified for the supplement would have dramatically less (supplement) earnings potential over their three years of entitlement. This is not to suggest some individuals in the SSP treatment group would not enroll in NB Works since many were unable to find full-time work and thereby qualify for the supplement. Indeed, NB Works was specifically designed for the least employable recipients. But NB Works offered an alternative route out of social assistance, and the design of the SSP was such that, a priori, control group members were more likely to enroll.

The timing of NB Works was such that the early SSP cohorts (November 1992 to June 1993)²⁰ were more likely to be influenced by the NB Works program. First, in-take into NB Works ended in early 1994, and thus is expected to have at most a modest effect on the 1994-95 SSP cohorts. Second, enthusiasm was high early on with NB Works, and there is evidence that the program even permeated the standard media. Certainly, it would be reasonable to think that NB Works was in the mindset of case officers, who in turn would transfer this knowledge to welfare users. Figure 7 provides evidence consistent NB Works having different impacts on the early and late SSP cohorts. Figure 7a shows that the treatment-control gap had disappeared for the early cohorts by month 54, whereas a difference of 8 percentage points existed for the later cohorts at that point (Figure 7b).

Table 2 presents the experimental effect of being randomly assigned to the SSP treatment group on participation in the NB Program. The raw experimental effect is startling. Overall, participation in NB Works by the control group was remarkably high: nearly 1 in 5 individuals randomly assigned to the SSP control group participated in NB Works. Participation in NB Works by the treatment group was much lower with an experimental impact of around 12 percentage points for the early cohorts. There was a small and statistically insignificant treatment –control difference for the later cohorts. Interestingly, participation among the program group was roughly the same over time; the reason for the overall experimental effect is driven by a very high participation rate among the control group early on. This suggests that any substitution bias would principally affect experimental estimates for the 1992-93 cohorts.

²⁰ There was no intake in NB in the latter half of 1993 (see Appendix A).

Table 3 presents regression results that estimate the effects of participating in NB Works on IA rates. Based on earlier work with the SSP data (e.g., Card and Hyslop 2005; Riddell and Riddell 2014), there is reason to be concerned about measurement error with the NB Works questions. The questions about participation in NB Works in the SSP surveys began with the general question: “Have you ever participated in NB Works?” We use the response to this question for our primary NB Works variable, as in Table 2 above.²¹ To correct for measurement error, we also estimate IV regressions that use the response to the question on whether the individual took a NB Works course as an instrument for the response to the participation question. Because we must use the survey data to examine NB Works, our sample period for this analysis is the 54 months covered by the four SSP surveys as opposed to the 70 months of administrative data used for our analysis of BC. The Table 3 results show a negative correlation between participation in NB Works and IA receipt. We remind readers that there is a mechanical negative relationship between NB Works and IA: the NB Works program automatically removed the individual from IA and funneled them into training, paid internships and the unemployment insurance system for the remainder of the year – for a total of three years for those that remained in the program. Aside from individuals dropping out of NB Works, there should be a coefficient of unity. Thus, our analysis here is simply estimating the dropout rate, and in particular allowing the dropout propensity to vary with individual characteristics and SSP program characteristics. The estimated coefficients –indeed, even the IV coefficients that deal with measurement error – thus may appear somewhat small (in absolute value). Taken literally, our IV estimates suggest a dropout rate around 50%, while the government evaluations of NB Works indicated a dropout rate in the 25-35% range. Thus our estimate may be understated.

²¹ We use a single cross-sectional indicator, which equals one if the individual responded in any survey that they had ever been enrolled in NB Works. There is not complete longitudinal information in the NB Works questions: interestingly, nothing about NB Works was asked in the SSP baseline survey despite the two programs being announced at the same time. The first question about NB Works appeared in the 18-month survey. Further, very few individuals responded to the 54-month survey question on participation in the program (even though the question was ‘Have you ever been...’ and not ‘Since the last survey...’). We suspect this is because, in calendar time, for those would have been able to enroll in NB Works 54 months post baseline was well after the NB Works program had ended. In any event, the results are robust to various definitions of NB Works participation (i.e., using only the 18 month survey question vs. combining two or all three post baseline surveys —as we do for the results presented in the paper.

Figure 8 uses the coefficients from the early cohorts IV regression in Table 3 to predict what IA rates for these cohorts would have been if NB Works had not existed. Rather than the treatment-control gap having converging to zero at the 54-month point (as shown in Figure 7a), we find that there would have been a 6 percentage point difference. This estimated differential is similar in size to our simulated impact for BC as well as the actual experimental difference for NB's later cohorts at 54 months (see Figure 7b).

In addition to an alternative welfare-to-work program operating in NB during part of the time period covered by SSP, important policy changes that raised the incentives to work for NB families with children were also introduced during the SSP time period. At the federal level a variety of changes occurred with the Canada Child Tax Benefit (CCTB). In particular, in July 1997, the CCTB working income supplement was increased substantially from \$500 per family (irrespective of number of children) to \$605 for the first child, \$405 for the second child, and \$330 for each additional child. One month prior to this federal change, New Brunswick introduced its own working income supplement of \$250. Only families with children who were working could receive either of these supplements. Families in New Brunswick could receive both working supplements (federal and provincial) with no claw-back.

To assess the consequences of these changes, we conduct an analysis similar to that for BC for New Brunswick's later (1994-95) cohorts. Details are reported in Appendix B and briefly summarized here. The changes at the federal level potentially affect behavior of SSP participants in both provinces, so we replicate our estimation of specifications (2) and (3) in Table 1, but include an additional time-based parameter for July 1997, the date of the working income supplement increases. The results are displayed in Appendix Table A1. The BC policy variables are similar to previous estimates for BC, and completely zero for NB. We also note that the negative coefficient on the BC linear interaction term for NB seen earlier in Table 1 is now gone. The working income supplement variables are strongly negative for NB and greater (in absolute value) for the control group than the program group. We find no effect of the increase in the federal working income supplement for BC. It is likely that those families with children able to respond to the work incentives of BC Benefits (which carried through to January 1997 with the Healthy Kids Act) and, in particular, the BC Family

Bonus had already done so. We also show in calendar time the convergence across different cohorts to the working income supplement changes in Appendix B Figures A1-A3. Note that the convergence in Figure A2 is similar to the other groups, but the original experimental effect prior to the reforms was much larger. Finally, we predict the pattern in IA rates in the absence of the working income supplement increases, as displayed in Figure A4. Our findings suggest a similar, although somewhat larger, longer-run effect of the SSP financial incentive on the NB later cohorts than we found for BC.

Recall that our estimates for the NB early cohorts in the absence of NB Works indicated an experimental impact of 6 percentage points. Combining the two sets of NB results, our predictions for NB are qualitatively similar but quantitatively somewhat larger than those for BC.

6. Conclusions

Three major conclusions have been reached on the basis of the experimental evidence from the SSP demonstration: (i) long-term IA recipients do respond to financial incentives to leave welfare and enter the workforce, (ii) over time a substantial fraction of IA recipients who are single parents with dependent children leave welfare on their own, without financial or other incentives to do so, and (iii) temporary financial incentives to leave IA have temporary but no lasting impacts on welfare receipt and labor force participation.

This study makes two principal contributions:

- (i) Our investigation contributes to the ongoing debate over the advantages and disadvantages of experimental versus non-experimental methods for analyzing social policies. By analyzing the consequences of a substantial policy change that was introduced during the period that a major demonstration project was underway, we provide new evidence on the extent to which the validity of the findings of social experiments may be threatened by such policy changes.
- (ii) Our study sheds new light on the conclusions that have previously been reached on the basis of the SSP evidence. On the basis of our investigations, we believe that the SSP experimental estimates are unduly pessimistic about the long-term consequences of temporary financial incentives on welfare dependence. It also appears that the observed

behavior of the experimental control group over-estimates the extent to which long-term welfare recipients with dependent children leave IA in the absence of incentives to do so.

In addition, our study points out the potential value of staggered entry in the design of social experiments. It is difficult to forecast whether policy or other changes will occur during or after an experiment, and even more difficult to assess the likelihood that changes that may contaminate the experimental findings will take place. As in the analysis in this paper, staggered entry may allow researchers to estimate what the experimental findings would have been in a stable policy environment.

References

Barrett, Garry and Michael Cragg. "An Untold Story: The Characteristics of Welfare Use in British Columbia" *Canadian Journal of Economics* 40 (February 1998) 165-88.

Battle, Ken and Michael Mendelson. *Child Benefit Reform in Canada: an evaluative framework and future directions*. Ottawa: Caledon Institute of Social Policy, 1997, pp. 59-68.

BC Ministry of Human Resources. *Monthly Statistical Report*, various issues. 1997.

Bitler, Marianne P., Jonah B. Gelbach and Hilary W. Hoynes. "Distributional impacts of the Self-Sufficiency Project" *Journal of Public Economics* 92 (2008) 742-765.

Blank, Rebecca M. and David Card (editors). *Finding Jobs: Work and Welfare Reform*. New York: Russell Sage Foundation, 2000.

Blank, Rebecca M., David Card and Philip K. Robins. "Financial Incentives for Increasing Work and Income among Low-Income Families" in *Finding Jobs: Work and Welfare Reform*. New York: Russell Sage Foundation, 2000, pp. 373-419.

Blundell, Richard. "Welfare-to-Work: Which Policies Work and Why?" Keynes Lecture in Economics, *Proceedings of The British Academy*, Vol.117, 2002, 477-524.

Blundell, Richard. "Earned Income Tax Credit Policies: Impact and Optimality" The Adam Smith Lecture, *Labour Economics* 13 (2006) 423-44.

Blundell, Richard and Hilary Hoynes. "Has 'In-Work' Benefit Reform Helped the Labour Market?" in R. Blundell, D. Card and R. Freeman (eds), *Seeking a Premier League Economy*. Chicago: University of Chicago Press, 2004, pp 411-460.

Card, David and Dean Hyslop. "Estimating the effects of a time-limited earnings subsidy for welfare-leavers." *Econometrica* 73 (2005) 1723-1770.

Card, David and Dean Hyslop. "The Dynamic Effects of an Earnings Subsidy for Long-term Welfare Recipients: Evidence from the SSP Applicant Experiment." *Journal of Econometrics* 153, November 2009.

Connolly, Helen and Peter Gottschalk. "Do earnings subsidies affect job choice? The impact of SSP subsidies on job turnover and wage growth" *Canadian Journal of Economics* 42 (November 2009) 1276-1304.

Dickenson, Paul. "First Steps in Developing an Evaluation Framework for NB Works," NB Department of Income Assistance, 1992.

Gornick, Carolyne. *Welfare-To-Work Programs: A National Inventory*. Ottawa: Canadian Council on Social Development, 1998.

Heckman, J.J., N. Hohmann, M. Khoo and J. Smith, "Substitution and Drop Out Bias in Social Experiments: A Study of an Influential Social Experiment" *Quarterly Journal of Economics* 115, No. 2 (May, 2000) 651-694.

Heckman, James J. and Jeffrey Smith "Assessing the Case for Social Experiments." *Journal of Economic Perspectives* 9 (1995) 85-110.

Johnson, Marc. "New Brunswick Works: A Report on Evaluative Research Surveying Dropouts, Case-study Managers and Teachers, New Brunswick Department of Income Assistance, 1993

Lacroix, Guy. "Assessing the Impact of a Wage Subsidy for Single Parents on Social Assistance" CLSRN Working Paper No. 26, May 2009.

Lise, Jeremy, Shannon Seitz and Jeffrey Smith. "Evaluating search and matching models using experimental data" *IZA Journal of Labor Economics* 2015, 4:16.

Michalopoulos, Charles, et al. *Making Work Pay: Final Report on the Self-Sufficiency Project for Long-Term Welfare Recipients*. Ottawa: Social Research and Demonstration Corporation, 2002.

National Council of Welfare. *Another Look at Welfare Reform*. Ottawa: National Council of Welfare, Autumn 1997.

New Brunswick Department of Advanced Education and Labour. *NB Works: A Joint Pilot/Demonstration Project New Brunswick/Canada*, 1992.

New Brunswick Department of Income Assistance. *Annual Reports (various years)*. Fredericton: Queen's Printer.

New Brunswick Department of Income Assistance, *Outline of an Evaluation Framework for NB Works*, SP Research Associates, 1995.

PGF Consultants. *New Brunswick Works: A Report on Participants' Evaluation*. New Brunswick Department of Income Assistance, 1993.

Riddell, Chris and W. Craig Riddell. "The Pitfalls of Work Requirements in Welfare-to-Work Policies: Experimental Evidence on Human Capital Accumulation in the Self-Sufficiency Project" *Journal of Public Economics* 117 (2014) 39-49.

Robins, Philip K. and Charles Michalopoulos. "Using Financial Incentives to Encourage Welfare Recipients to Become Economically Self-Sufficient" *Economic Policy Review*, Federal Reserve Bank of New York, September 2001, pp. 105-126.

Table 1 – Estimated Coefficients for the Probability of Being on Welfare

	British Columbia						New Brunswick Falsification Test					
	(1)		(2)		(3)		(4)		(5)		(6)	
	Programs	Controls	Programs	Controls	Programs	Controls	Programs	Controls	Programs	Controls	Programs	Controls
BC Benefits	-.036 (.013)	-.066 (.020)	-.005 (.010)	-.014 (.014)	-.005 (.009)	-.010 (.012)	.025 (.017)	.003 (.016)	.024 (.016)	.019 (.013)	.013 (.016)	.014 (.012)
Linear interaction	-	-	-.0043 (.0011)	-.0068 (.0013)	-.0044 (.0011)	-.0062 (.0012)	-	-	.0005 (.0015)	-.0037 (.0022)	-.0028 (.0017)	-.0038 (.0024)
Fixed effects	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
R-squared	.09	.12	.10	.13	.10	.13	.06	.09	.06	.09	.06	.09
N	106960	104580	106960	104580	106960	104580	68462	68594	68462	68594	68462	68594
	British Columbia						New Brunswick Falsification Test					
	(7)		(8)		(9)		(10)		(11)		(12)	
	Programs	Controls	Programs	Controls	Programs	Controls	Programs	Controls	Programs	Controls	Programs	Controls
BC Bonus	-.064 (.015)	-.105 (.021)	-.034 (.010)	-.055 (.015)	-.030 (.010)	-.050 (.014)	.023 (.017)	-.004 (.019)	.020 (.013)	.020 (.012)	.008 (.012)	.019 (.011)
Linear interaction	-	-	-.0039 (.0011)	-.0062 (.0015)	-.0034 (.0011)	-.0050 (.001)	-	-	.0000 (.0018)	-.0042 (.0023)	-.0032 (.0018)	-.0041 (.0022)
Fixed effects	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
R-squared	.09	.12	.10	.13	.10	.13	.06	.09	.06	.09	.06	.09
N	106960	104580	106960	104580	106960	104580	68462	68594	68462	68594	68462	68594

NOTES: All specifications include 69 SSP time dummies (or dummies for time since onset) as well as calendar month dummies and the monthly (calendar time) unemployment rate. Standard errors are in parentheses and are adjusted for clustering on the baseline interview date. BC Benefits uses January 1996 as the beginning of the reforms; BC Family Bonus uses July 1996. New Brunswick regressions are for the 1994-95 cohorts only.

Table 2
Experimental Effect of SSP Supplement on NB Works Participation

Control Group	Program Group	Difference	Difference (with controls)
(a) Early Cohorts			
.188 (.023)	.071 (.016)	.117 (.027)	.117 (.027)
(b) Later Cohorts			
.098 (.010)	.080 (.009)	.018 (.014)	.018 (.014)
(c) All Cohorts			
.122 (.010)	.078 (.009)	.044 (.013)	.044 (.013)

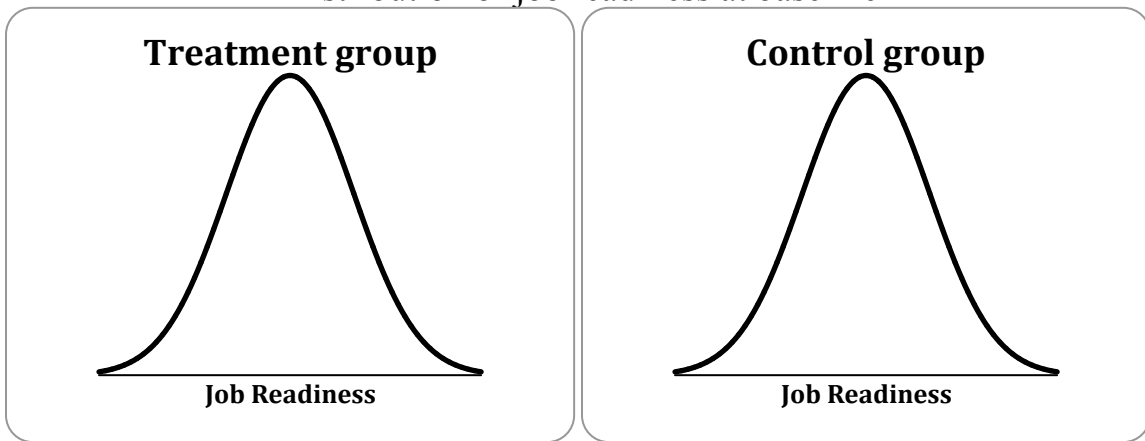
Table 3
Estimated Coefficients for Effect of NB Works on Probability of Being on Welfare

	Early Cohorts		Later Cohorts		Pooled	
	OLS	IV	OLS	IV	OLS	IV
Enrolled in NB Works	-.168 (.028)	-.509 (.141)	-.047 (.018)	-.299 (.090)	-.089 (.015)	-.353 (.075)
Time dummies	Yes					
Control group* time dummies	Yes					
Baseline controls	Yes					
R-squared	.20	.16	.21	.19	.21	.18
N	2136		6464		8600	
First stage F-stat	-	22.9	-	40.3	-	57.2

NOTES: Standard errors are in parentheses, and are clustered on the individual.

Figure 1

Distribution of job readiness at baseline



Distribution of job readiness at t^*

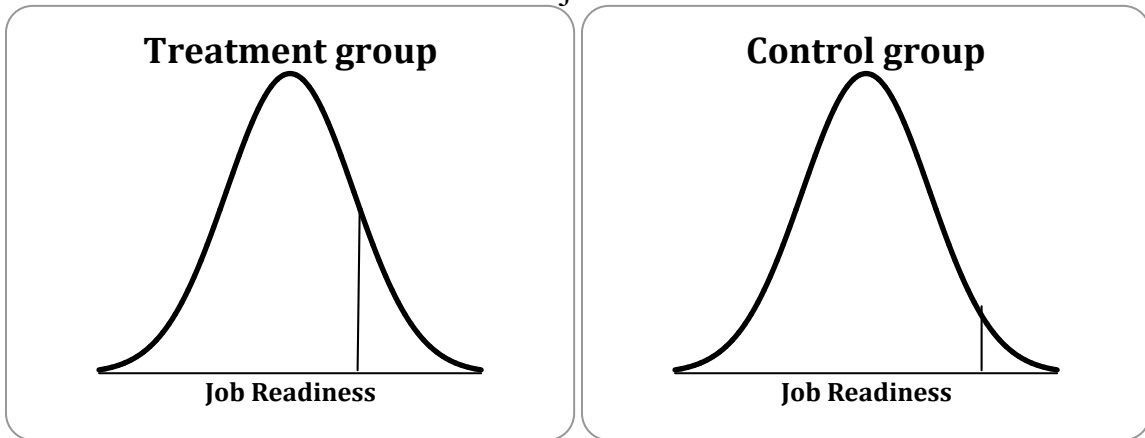


Figure 2a – Income Assistance Rates in British Columbia

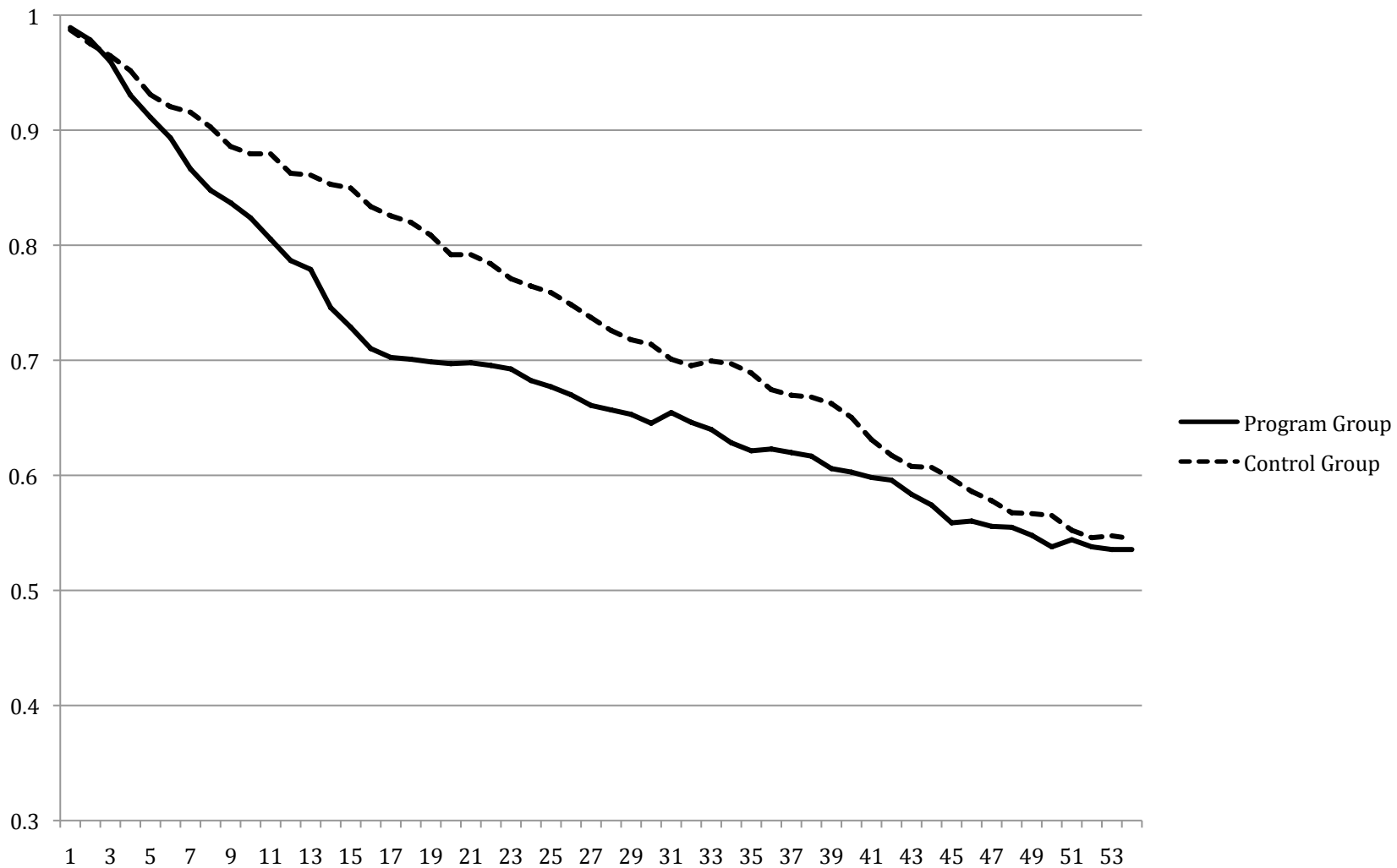


Figure 2b – Income Assistance Rates in New Brunswick

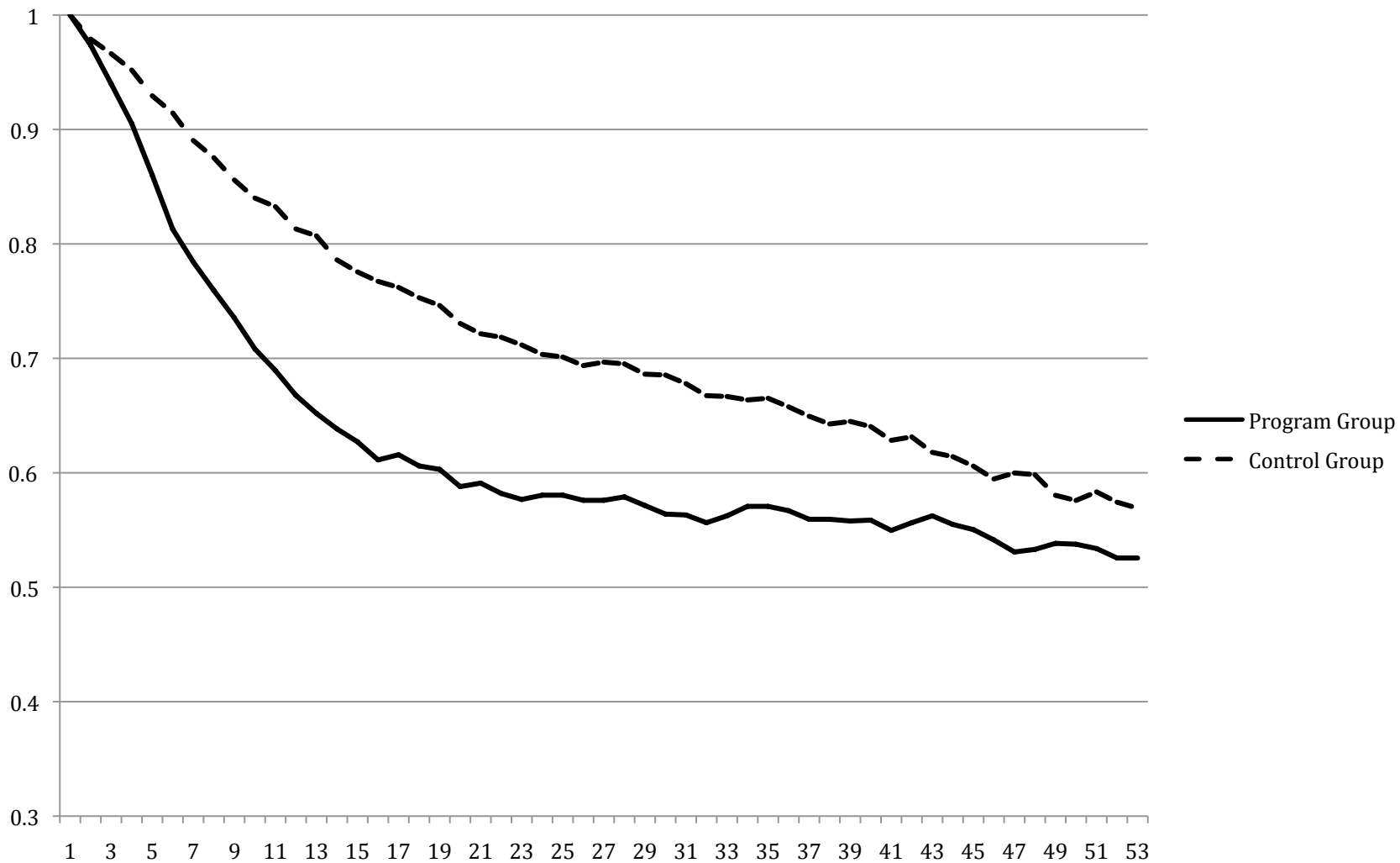


Figure 3 – Income Assistance Caseloads for British Columbia, All Families with Children

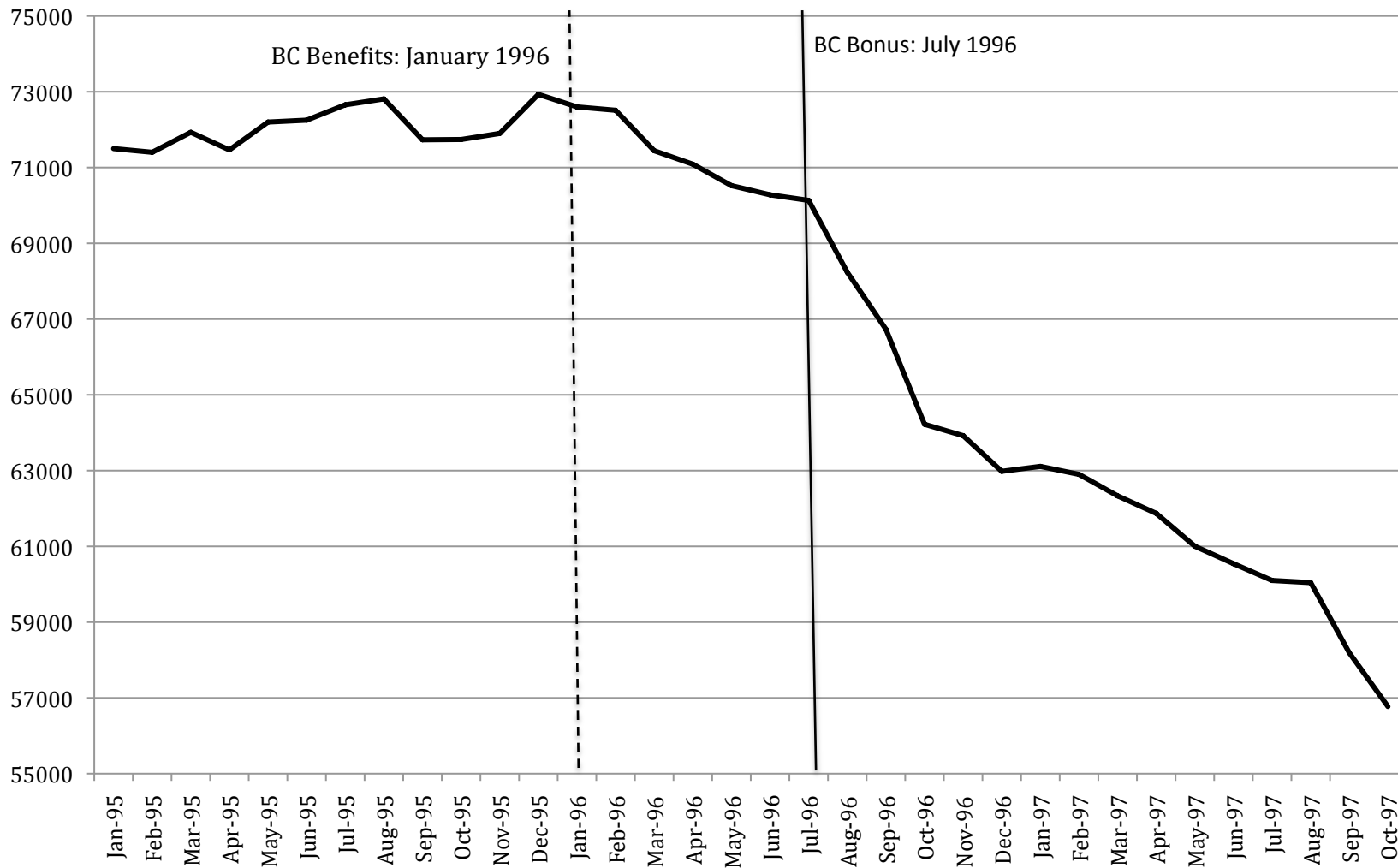


Figure 4a – Income Assistance Rates of Early BC Cohorts in Calendar Time

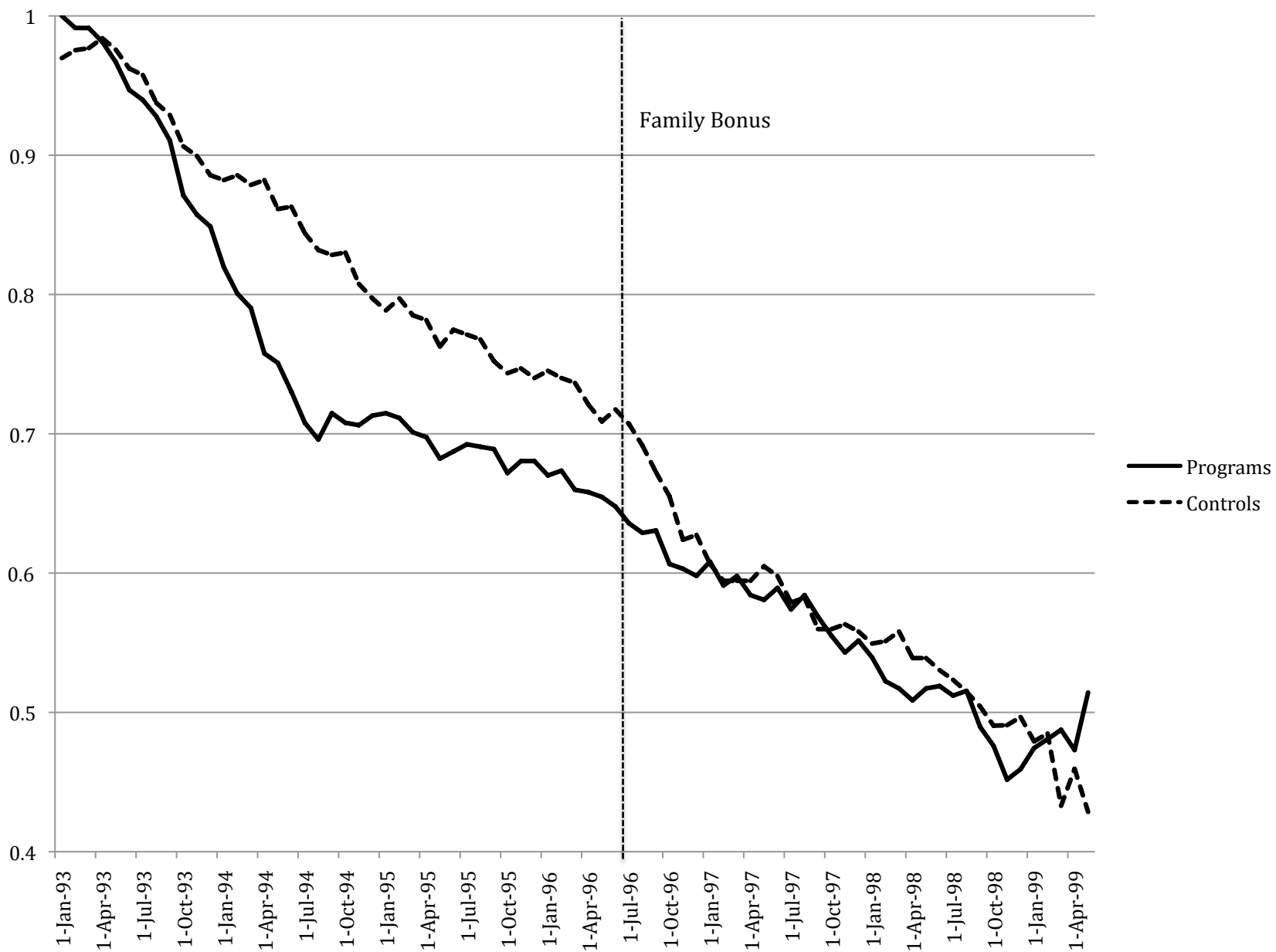


Figure 4b – Income Assistance Rates of Later BC Cohorts in Calendar Time

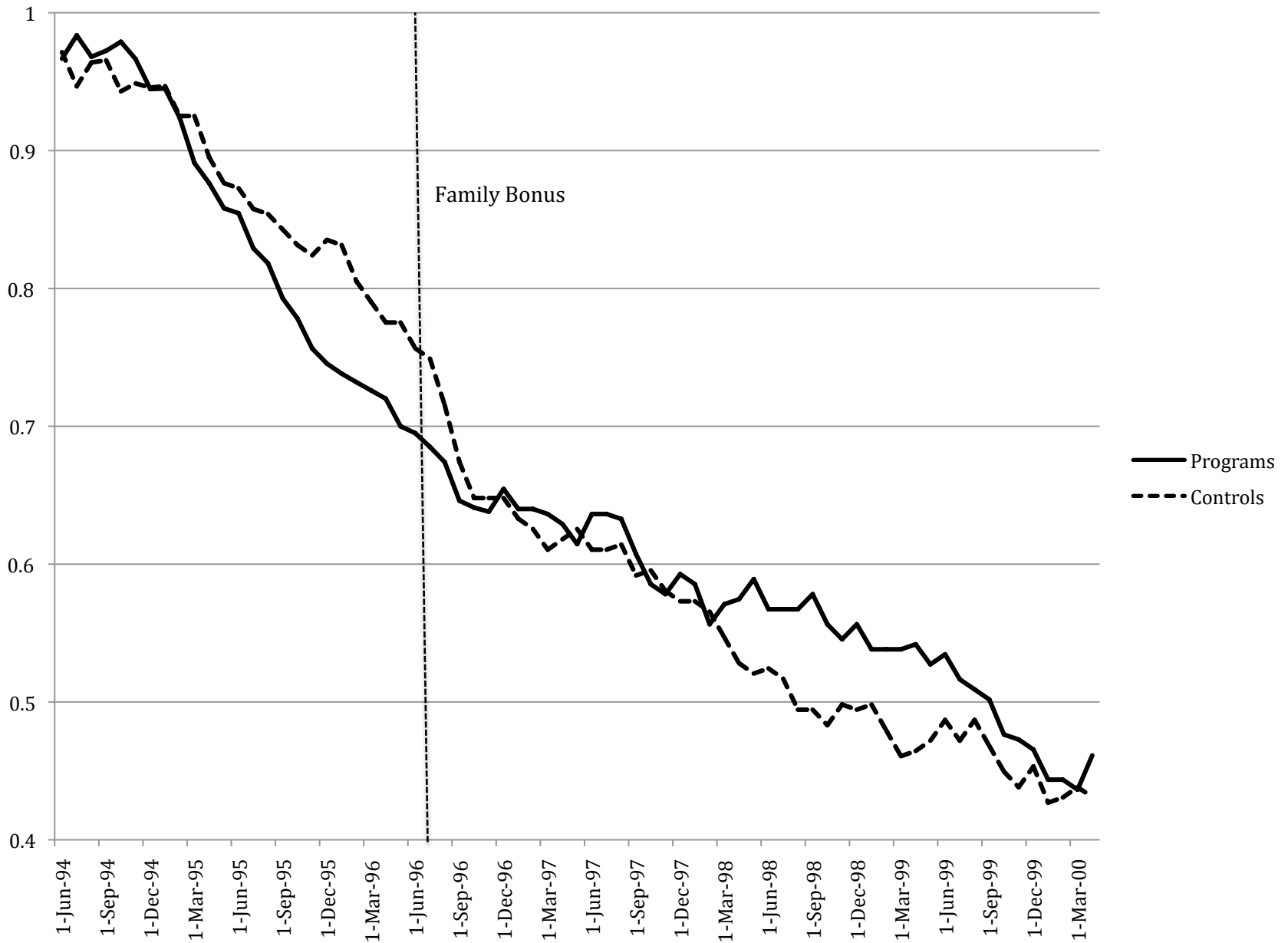


Figure 5 – Job Readiness in British Columbia

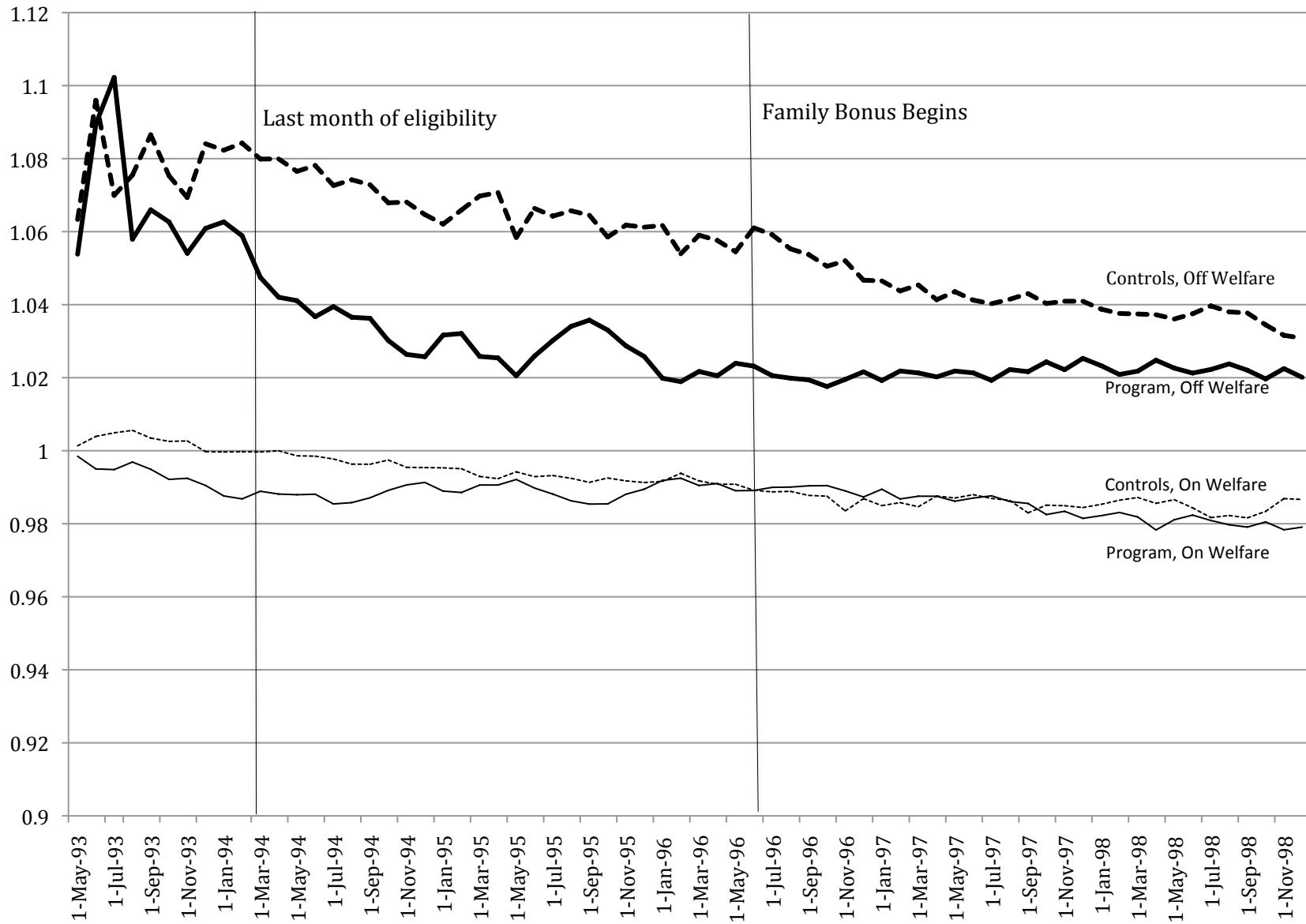


Figure 6 – Predicted Income Assistance Rates for British Columbia in the Absence of BC Benefits

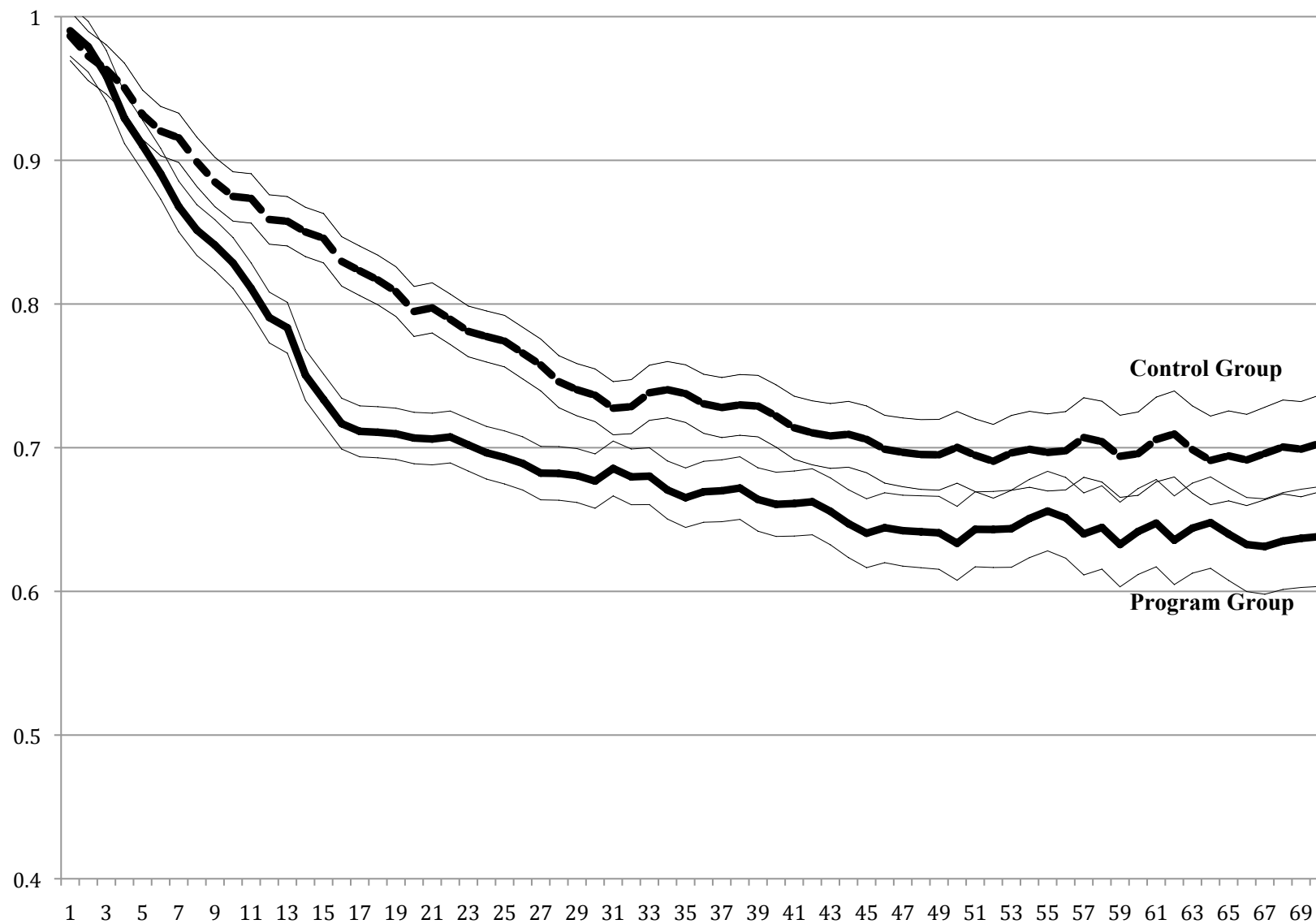


Figure 7a– Income Assistance Rates in New Brunswick, ‘Early Cohorts’

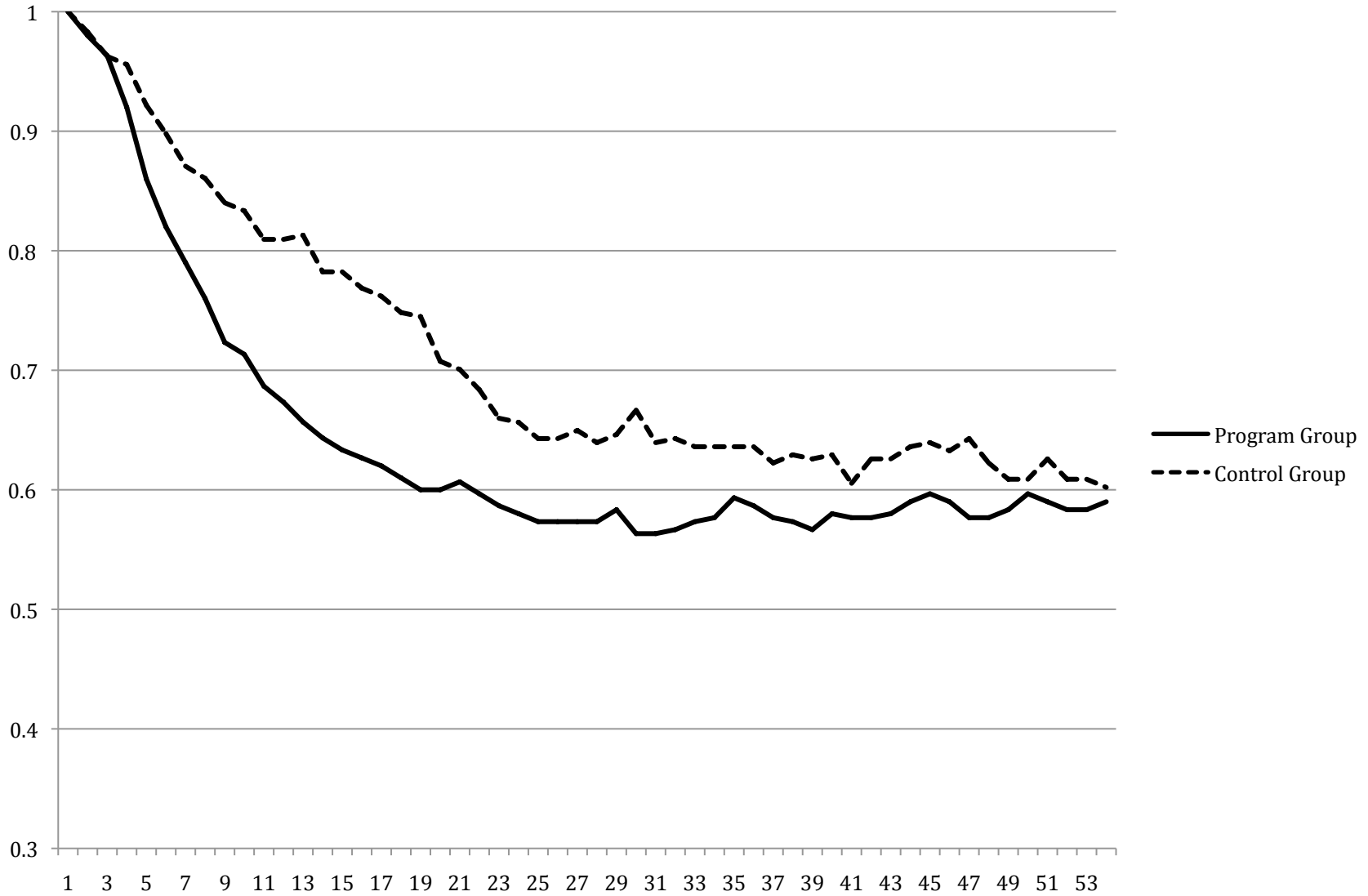


Figure 7b– Income Assistance Rates in New Brunswick, ‘Later Cohorts’

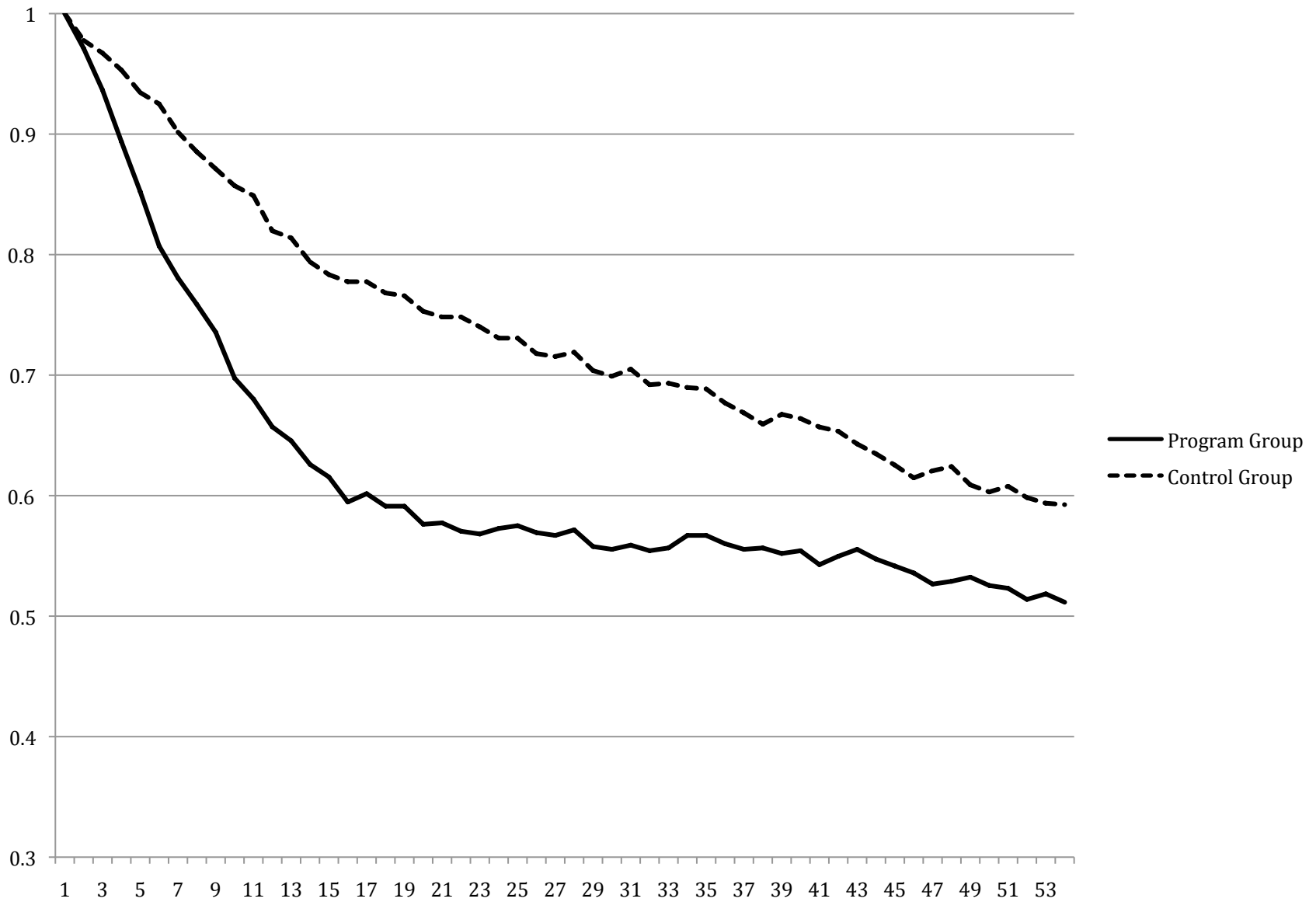
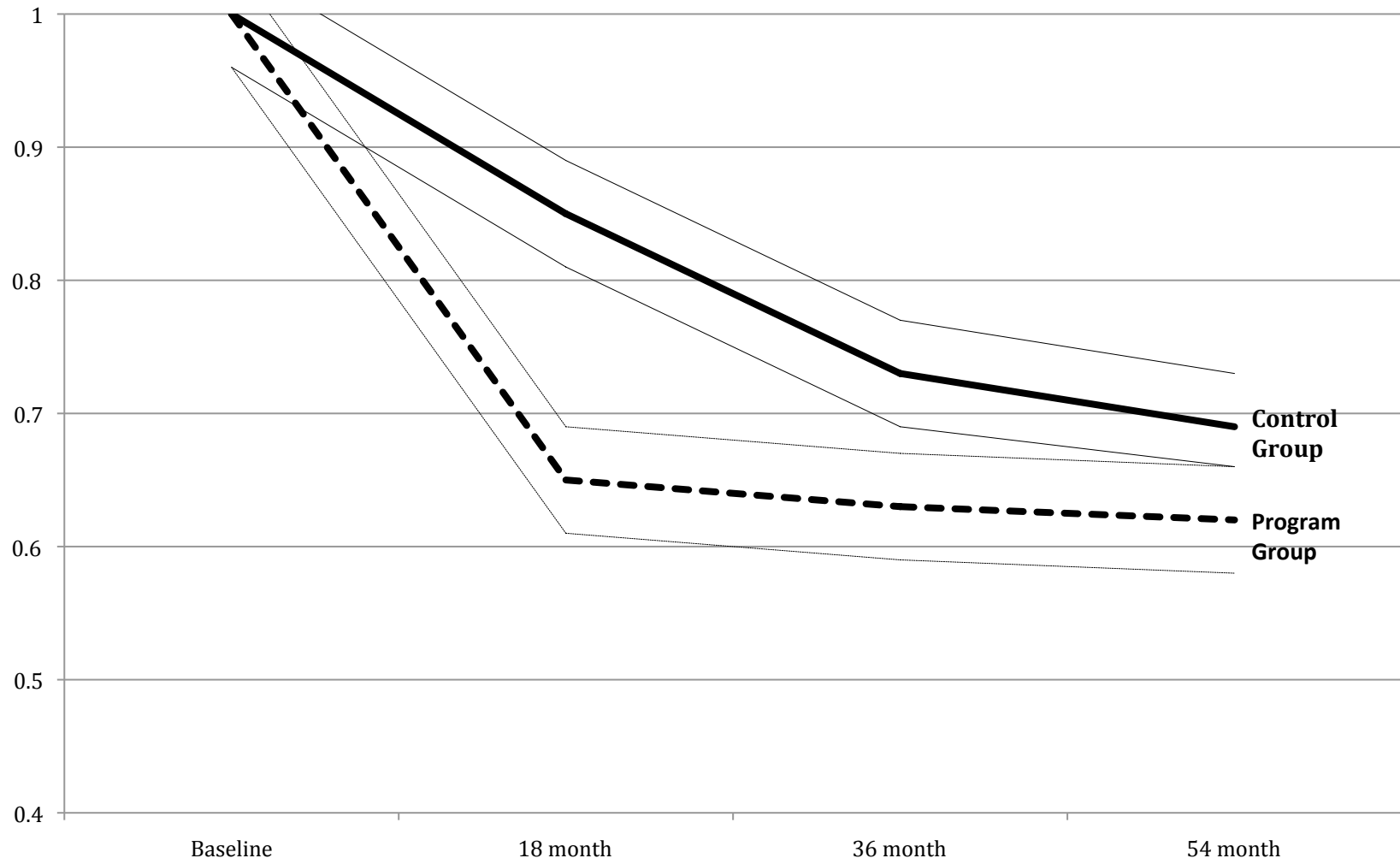


Figure 8 – Predicted IA Rates for New Brunswick’s Early Cohorts in the Absence of NB Works



Appendix A: Distribution of Observations by Baseline Date and Province

Date of random assignment	BC	NB
November 1992	0	89
December 1992	0	94
January 1993	84	99
February 1993	221	111
March 1993	246	105
April 1993	240	111
May 1993	237	41
June 1993	201	30
July 1993	91	0
August 1993	86	0
September 1993	11	0
October 1993	1	0
January 1994	125	128
February 1994	118	133
March 1994	127	154
April 1994	128	153
May 1994	143	126
June 1994	124	129
July 1994	131	127
August 1994	134	133
September 1994	142	151
October 1994	144	149
November 1994	152	184
December 1994	112	183
January 1995	15	172
February 1995	10	180
March 1995	0	173
Total SSP observations	3023	2955

Appendix B – Extensions to New Brunswick Analysis

Table A1 – Estimated Coefficients for the Probability of Being on Welfare, Extensions for the Working Income Supplement

	British Columbia				New Brunswick			
	(1)		(2)		(3)		(4)	
	Programs	Controls	Programs	Controls	Programs	Controls	Programs	Controls
BC Benefits	.006 (.011)	.015 (.006)	.006 (.009)	.015 (.006)	.014 (.016)	-.001 (.010)	.004 (.016)	.000 (.009)
BC Linear interaction	-.0058*** (.0012)	-.0083*** (.0013)	-.0059*** (.0012)	-.0083*** (.0013)	.0028 (.0017)	.0010 (.0019)	-.0006 (.0022)	.0015 (.0017)
Working Supplement	.000 (.010)	.011 (.012)	.000 (.009)	.014 (.010)	.020 (.014)	.021 (.014)	.020 (.014)	.021 (.013)
Supplement Linear Interaction	.0020* (.0010)	.0021 (.0016)	.0020* (.0010)	.0021 (.0015)	-.0043* (.0021)	-.0073*** (.0020)	-.0038* (.0022)	-.0073*** (.0020)
Fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
R-squared	.10	.13	.10	.13	.06	.09	.06	.09
N	106960	104580	106960	104580	68462	68594	68462	68594

NOTES: All specifications include 69 SSP time dummies (or dummies for time since onset) as well as calendar month dummies and the monthly (calendar time) unemployment rate. Standard errors are in parentheses and are adjusted for clustering on the baseline interview date. BC Benefits uses January 1996 as the beginning of the BC reforms; the Working Income Supplement uses July 1997. New Brunswick analysis is for the 1994-95 cohorts only.

Figure A1 – Income Assistance Rates in New Brunswick, Calendar Time, Cohorts January 1994 to May 1994



Figure A2 – Income Assistance Rates in New Brunswick, Calendar Time, Cohorts June 1994 to October 1994



Figure A3 – Income Assistance Rates in New Brunswick, Calendar Time, Cohorts November 1994 to March 1995



Figure A4 – Predicted Income Assistance Rates in New Brunswick in the Absence of the Working Income Supplement

