I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14585

Welfare versus Work under a Negative Income Tax: Evidence from the Gary, Seattle, Denver and Manitoba Income Maintenance Experiments

Chris Riddell
W. Craig Riddell

JULY 2021

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14585

Welfare versus Work under a Negative Income Tax: Evidence from the Gary, Seattle, Denver and Manitoba Income Maintenance Experiments

Chris Riddell
University of Waterloo

W. Craig Riddell
Vancouver School of Economics and IZA

JULY 2021

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Welfare versus Work under a Negative Income Tax: Evidence from the Gary, Seattle, Denver and Manitoba Income Maintenance Experiments*

The Income Maintenance Experiments have received renewed attention due to growing international interest in a Basic Income. Proponents viewed a Negative Income Tax as a replacement for traditional welfare with stronger work incentives and reduced poverty. However, existing labor supply estimates for single parents are uniformly negative. We re-assess the experimental evidence and find randomization failure in two NITs (Gary and Seattle). In Denver and Manitoba, we find a positive labor supply response for those on welfare prior to random assignment. Our results provide strong evidence that a NIT can increase work activity among single parents on welfare.

JEL Classification: negative income tax, welfare, income maintenance experiments, labour supply

Keywords: C9, I38, J2

Corresponding author:
W. Craig Riddell
Professor Emeritus of Economics
Vancouver School of Economics
University of British Columbia
6000 Iona Drive
Vancouver, B.C. V6T 1L4
Canada
E-mail: craig.riddell@ubc.ca

* Our research on the Manitoba Income Maintenance Experiment (Mincome) was supported by the B.C. Government's Expert Panel on Basic Income. We thank David Card, David Green and Jesse Rothstein for useful comments, Greg Mason for assistance with the Mincome data, and Derek Hum and Wayne Simpson for helpful correspondence relating to our attempts to replicate their Mincome results.

1. Introduction

A universal basic income (UBI) or basic income (BI) program is receiving considerable attention in many countries. Numerous books and articles have recently been written by BI advocates in the U.S. (e.g. Lowrey, 2018; Murray, 2016; Yang, 2018), Canada (e.g. Forget, 2018; Segal, 2019) and Europe (e.g. Haagh, 2019; Van Parijs and Vanderborght, 2017). In addition, demonstration projects and experiments have been initiated in North America[1] as well as in European countries such as Germany, Finland, Spain and the Netherlands. Throughout the industrialized world, media mentions of the terms 'universal basic income' and 'basic income' have soared.[2]

A central issue that societies face when considering introducing a BI is its probable effect on work activity. The importance of understanding the likely labor supply impacts of various BI proposals has in turn resulted in greater attention to the lessons from past experience with financial transfers to the low-income population. This is especially the case for the guaranteed annual income or negative income tax experiments carried out as part of the 'War on Poverty' in the 1960s and 1970s.

During that period, the U.S. and Canada conducted a landmark series of negative income tax (NIT) experiments including four U.S. experiments -- (i) New Jersey; (ii) Rural Income Maintenance Experiment (RIME) carried out in rural counties in Iowa and North Carolina; (iii) Gary, Indiana; and (iv) Seattle-Denver (or SIME-DIME) -- and one Canadian experiment in the province of Manitoba (Mincome). These experiments involve a set of policy parameters close to those in many BI proposals, including those defined by Hoynes and Rothstein (2019) and the Stanford Basic Income Lab.[3] As such, they represent a key part of our current understanding of what the effects of a basic income approach would be on individual and household outcomes. Recent reviews of the NIT experiments have become commonplace in light of the basic income

[1]Stanford Basic Income Lab lists 11 BI experiments underway in North America, one (in Ontario) that was subsequently cancelled: https://basicincome.stanford.edu/research/basic-income-experiments/.

[2] For example, Hoynes and Rothschild (2019, Figure 1) provide evidence from the New York Times where mentions of these terms skyrocketed since 2015 relative to previous years.

[3] One potential difference in the negative income tax experiments relative to some basic income policies proposed today is the extent to which the cash benefit is clawed back with earned income. Not all BI/UBI proposals include a claw-back feature.

interest[4] and many policy pieces and media articles have noted the high-level results from these experiments.

The academic and policy literature on the NIT experiments is vast. Widerquist (2005) cites more than 200 scholarly studies published in books and academic journals and notes that "there are at least 200 more unpublished memorandums, reports, discussion papers and other unpublished works on the experiments as well" (Widerquist, 2005, p. 79). An unusual feature of this literature is its almost exclusive focus on the labor supply responses of the 'working poor' – effects that are predicted to be negative because for those who are employed both the income and substitution effects operate in the same direction. Very little attention has been given to possible positive impacts on labor supply, such as those that would be expected for welfare recipients facing high tax-back rates on earned income. This lack of attention is perhaps surprising given that early proponents of the NIT, such as Friedman (1962), advocated the NIT as a replacement for traditional welfare and other income support programs, emphasizing the NIT's stronger work incentives (Moffitt, 2003).[5]

Evidence on labor supply behavior is, of course, available from many other sources – see, e.g. the extensive survey by Blundell and Macurdy (1999) and the more recent review by Hoynes and Rothstein (2019). The strong renewal of interest in the results of the NIT experiments may reflect two key factors. One is the use of random assignment to treatment and control groups in a controlled setting – the 'gold standard' for credible evidence. In addition, the NIT experiments specifically focus on work behavior of the low-income population. Many other labor supply studies – e.g. on lottery winners (Cesarini et. al. 2017) or NY City taxi drivers (Ashenfelter, Doran and Schaller, 2010) examine demographic groups that are less likely to be affected by a BI policy.

This paper re-assesses the North American NITs with target populations that included sufficient single parents to permit empirical analysis: Gary, Seattle ('SIME'), Denver ('DIME') and

[4] Examples include Hoynes and Rothstein (2019), Widerquist (2005), Van Parijs and Vanderborght (2017), Marinescu (2018) and Stanford Basic Income Lab (2018).

[5] Most BI proposals are also intended to replace existing welfare and other income support programs.

Mincome.[6] We focus on the labor supply responses of single parents[7] as this group is the most likely to be receiving welfare during this time period.[8]

Our key contributions are twofold. First, we document carefully that random assignment appears to have failed in SIME and Gary. In SIME -- across multiple definitions of labor market attachment -- treatments and controls differed widely in their pre-random assignment characteristics (after controlling for the experimental stratifications, see below) with on average the treatment group working less, and more likely to be on welfare. In Gary, the treatment group was about 11 percentage points less likely to be on welfare prior to random assignment, and there is modest evidence of hours worked differences as well. In both Seattle and Gary, there was a steep downturn in the city's primary industry (aerospace in Seattle and steel in Gary) that coincided with the beginning of the experiment. In Seattle, the substantial deterioration in economic conditions was accompanied by differences between treatments and controls in enrollment dates (and thus calendar time differences in the start and end of the experiment), differences that provide a potential explanation for the failure of random assignment. While the U.S. NIT experiments did conduct considerable research on both misreporting of income and attrition bias, the possibility of a randomization failure is virtually non-existent in the extensive literature.[9] At the same time, we find no evidence of non-random

[6] There were no single parents in New Jersey, and sample sizes are too small in RIME.

[7] We adopt the term 'single parents' in the paper, but we follow the existing NIT literature by focusing on single female heads with a dependent child (as discussed in more detail in Section 3). There are a trivial number of single fathers in all four experiments.

[8] Historically, welfare in both countries was limited to single mothers and those unable to work. As discussed subsequently, this remained substantially the case when the NIT experiments were carried out. Welfare participation by two-parent families, such as under the AFDC-UP program in the U.S., was limited. In Manitoba, social assistance to two-parent families and single men and women was provided at the municipal rather than provincial level, and only for limited periods of time.

[9] In our detailed review of the Gary and SIME-DIME Final Reports, Brookings/Federal Reserve Bank of Boston 1986 Conference Proceedings (Mundell, 1987), the *Journal of Human Resources* special issue on SIME-DIME (Spiegelman and Yaeger (1980), and all subsequent published papers on SIME-DIME and Gary in economics journals, we found no discussion raising the possibility of a failure of random assignment. For example, Volume I, Part 1, chapter IV of the SIME-DIME Final Report (SRI International, 1983) contains an extensive discussion of threats to internal and external validity, but no mention of the possibility that random assignment failed to take place. A recent working paper by Price and Song (2018) finds that pre-random assignment hours worked (and thus earned income) in a subset of SIME – families with at least two children – differ between treatments and controls, while other variables examined do not. Although this suggests that random assignment may have failed it is not convincing evidence that this is the case, especially for the family types of interest in the NIT experiments. Along with Widerquist's summary of the literature (2005), we reviewed a large number of NIT survey papers released in more recent years, and none mention randomization failure (Hoynes and Rothstein 2019; Pasma 2010; Wiederspan et al 2015; Forget

attrition among single parents in SIME-DIME, in contrast to two-parent families where non-random attrition was evident, and appears to have affected the internal validity of the experimental estimates (Ashenfelter and Plant, 1990).

Moreover, we show that the oft-cited SIME-DIME pooled negative treatment effect on hours worked of single parents combined a sizeable, negative hours worked effect in SIME (an estimated impact that is based on treatment-control differences using non-randomly assigned data) with a small (and not statistically different from zero) effect in DIME. Using difference-in-difference estimators, we estimate that the effect of the SIME program is zero, a result we also obtain for Gary. Thus, overall, we find no credible evidence of a negative labor supply response for single mothers in any of these U.S. NIT experiments. This contrasts markedly with the consensus view that single female heads reduced labor supply. For instance, in his comprehensive review of the NIT literature Widerquist (2005) writes: "The response of wives and single mothers was somewhat larger in terms of hours, and substantially larger in percentage terms because they tended to work fewer hours to begin with. Wives reduced their work effort by 0–27% and *single mothers reduced their work effort by 15–30%*."

Our second key contribution is to test for heterogeneity in the NIT treatment effect on labor supply based on the individual's work and welfare patterns. Specifically, when sample sizes permit, we examine *separately* the labor supply response of those we label 'welfare recipients' (individuals receiving welfare over the pre-random assignment period) versus those we label 'workers' (individuals working over the pre-random assignment period). Because we are conditioning on a pre-random assignment characteristic, the 'working' and 'welfare' subsamples are both as good as randomly assigned to treatments and controls. For the Canadian experiment, we also provide the first credible evidence on labor supply impacts for the single parents group.[10] In both countries, our re-examination leads to markedly different conclusions about NIT impacts on the work activity of single parents than the current consensus view.

2018; Marinescu, 2018; Terwitte 2009; Abramowics 2019; Levine et al 2005; Paz-Banez et al 2020; Specianova 2018; Kearney and Mogstad 2019; Gibson et al 2018).

[10] We discuss the limitations of the sole prior Mincome labor supply study subsequently.

In Mincome, we find strong support for a positive labor supply response to the NIT. Based on the longitudinal survey data we estimate that treatments were about 14 percentage points more likely to be employed than controls over the roughly 3-year experimental period and annual hours worked were 210 hours greater for treatments. Relative to mean baseline hours worked, the treatment effect represents about a 34% increase. Results from hitherto untapped administrative data are virtually the same.

In DIME, we find a negative, but modest in size, labor supply response for 'workers' with an annualized *reduction* in hours of around 10%. Conversely, for the 'welfare participant' group, we find a large *increase* in hours of around 35%. The overall labor supply responses of the (smaller) 'welfare' and (larger) 'working' sub-groups in DIME approximately offset each other, producing the small, negative (but not statistically significant) estimates in the pooled experimental data.

Ultimately, our results suggest that a) the consensus view that the North American NITs can be considered internally valid experiments that yield unbiased ITTs and reduced labor supply of single parents appears to be misleading, and b) an NIT may increase labor supply for single parents on welfare – impacts that imply a guaranteed income can have offsetting positive and negative effects on work activity.

2. Theoretical Background and Consensus Impacts

Figure 1 illustrates the direction of predicted labor supply responses to introducing a NIT with guarantee level G = AC that exceeds the basic welfare benefit AB for a specific family type. In order to ensure sufficient take-up of the NIT offer, even the lowest guarantee level in each of the North American NITs exceeded the welfare benefit.[11] Participants in all of the North American NITs were forced to choose between welfare and the NIT.[12] In the absence of a NIT and welfare the budget constraint is the line AF, with slope equal to the hourly wage rate. Adding welfare that provides benefits AB yields the budget constraint ABDF in the case in which

[11] Special arrangements were also made to ensure that those leaving welfare for the NIT did not lose non-monetary benefits such as Medicaid and child care subsidies. For background on the U.S. and Manitoba welfare programs in effect at the time see Online Appendix: Data and Institutions sections A (U.S.) and C (Manitoba).
[12] NIT participants could return to welfare at any time.

the welfare program imposes a tax-back rate of 100% (i.e. reduces welfare benefits dollar-for-dollar on any market earnings) or ABD'F for a lower benefit reduction rate (BRR).[13] The 100% BRR was approximately the situation during Mincome, while during the U.S. NITs the federal benefit reduction rate was 67%.[14] Note, however, that there were numerous other differences between the U.S. and Manitoba that may have affected work incentives, as we document later. Finally, introducing a NIT with guarantee level AC > AB and a tax-back rate substantially below 100% yields the dashed budget constraint ACEF with the dashed NIT component CE. A 50% tax-back rate is used for illustration purposes.[15]

For low-income workers not receiving welfare – those in the segment DE on the budget constraint – static micro theory predicts a reduction in labor supply because both the income and substitution effects imply reduced hours of work. Individuals in this group – the "working poor" – experience an increase in income and work fewer hours. Some of those in segment EF – i.e. those working substantial hours such as multiple job holders – may also reduce work activity and accept less income, albeit a smaller income reduction than would have occurred without the NIT. However, for welfare recipients – those at point B – labor supply is predicted to remain unchanged (i.e., move from B to C as indicated by arrow 1) or increase (indicated by arrow 2) because the NIT has stronger work incentives than traditional welfare. Other things

[13] A similar figure also appears in Moffitt (2003), who presents simulations of the predicted effects of various NIT programs on hours of work of single mothers, the main group eligible for welfare in the U.S.

[14] The U.S. federal BRR was lowered from 100% to 67% in 1967 and returned to 100% in 1981 (Lurie, 1974). Although the Social Security Act governing AFDC is a federal law, U.S. states had substantial discretion in determining benefit levels. AFDC payments were based on 'financial requirements' (family 'need') minus 'countable income' and state regulations affected both. Hutchens (1978) estimated marginal effective tax rates (METR) on employment income in 20 U.S. states before and after the 1967 reduction in the federal rate and found substantial variation across states and over time, though the federal reduction did lower METRs on average and in most states examined.

[15] Welfare programs in both countries also included small earnings disregards as well as allowances for work-related expenses. In Manitoba, the disregard was $20/month, which –at the prevailing minimum wage at the beginning of the Mincome experiment of $2.30/hr – would allow for just under 9 hours of work per month. In the U.S. under the AFDC at this time, the disregard was $30/month, which—at the prevailing federal minimum wage of $1.60 at the beginning of SIME—would allow for just under 20 hours of work per month. The disregard introduces a non-linearity to the welfare budget constraints BD and BD' and a corner solution to the left of B at which the recipient combines welfare benefits and a small amount of work without any earnings being taxed back. The predicted effects of the NIT on labor supply are essentially unchanged.

equal, the incentive to enter the workforce will be stronger the higher the benefit reduction rate welfare recipients face on labor earnings and the lower the NIT tax-back rate.[16]

As noted previously, the central focus of the NIT literature has been on providing credible estimates of the magnitudes of potential adverse effects on work activity – i.e., the size of movements such as those depicted by arrows 3, 4 and 5 in Figure 1.[17] Our focus is on the labor supply responses of single parents, some of whom may be employed prior to random assignment (i.e., on the segment DF in Figure 1), but others who may be on welfare at point B.

Table 1 summarizes key characteristics of the income maintenance experiments we analyse. All were carried out in the 1970s. Target populations differed. Gary enrolled only Black two parent and single parent families, whereas SIME targeted Black and White two parent and single parent families and DIME also included Hispanics.[18] Mincome was unique in enrolling a representative sample of the low-income population, including single men and women. Sample sizes were much larger in Seattle and Denver than in the other studies.

A key feature of all the North American NIT experiments was the Conlisk-Watts assignment model for allocating families to treatment plans.[19] Prior to random assignment, families were stratified by family type (two-parent families, single parents with dependent children, and, in the Canadian case, single men and women); race (in Seattle and Denver), program length (SIME and DIME); location (in Gary and Mincome); and 'normal income' levels.[20] Each stratified sample was offered treatment plans that combined different guarantee levels *G* and implicit tax rates *t* in an attempt to facilitate estimates of the responsiveness of families to NIT plans with different incentives.

---

[16] The cost and availability of child care is also likely to influence the choice between the movements depicted in arrows 1 and 2. We return to this issue later.

[17] Indeed, other than impacts on marital status, which also received considerable attention, labor supply has been one of the few outcomes extensively studied.

[18] In the official reports and the academic literature, these are generally referred to as 'single mothers', 'single female heads' or 'single female heads with a dependent child'. However, not all families classified in this group contain a dependent child. We discuss how these exceptions are dealt with in the data section.

[19] This model is designed to optimize the allocation of families with different pre-treatment income levels to the various treatment plans, taking account of the overall budget for the experiment. Pure random assignment of families to alternative treatment plans would result in some low-income families being offered very generous (high G, low t) treatment plans – resulting in very expensive observations. Essentially this assignment model reduces the likelihood that low-income families (and raises the likelihood that families with high pre-treatment income) are enrolled in generous treatment plans relative to what would occur under pure random assignment.

[20] Normal or permanent income was computed from pre-treatment surveys discussed subsequently.

An important consequence of the Conlisk-Watts assignment model is that for the sample as a whole there is non-random assignment to treatment and control groups. Rather, random assignment took place within combinations of the experimental stratifications noted above that were adopted for a particular experiment (see also Table 1). For single parents, this includes normal income in all experiments.[21] In order to obtain unbiased estimates of treatment effects it is therefore necessary to control for the appropriate stratification categories (Keeley and Robins, 1980; Keeley, 1981). Because the full sample for each income maintenance experiment is not randomly assigned, we use the term 'stratification group experiments' or, more simply, 'mini-experiments' to refer to the level at which random assignment takes place (e.g. Income Category 1, Blacks, 3-Year Program in the case of SIME or DIME).[22] For SIME and DIME, where the number of mini-experiments is particularly large, we show the counts for each single parent mini-experiment in Appendix Table A1. For single parents, Gary consists of ten mini-experiments, while Mincome consists of four. As the only data digitized for Mincome is the Winnipeg site, the mini-experiments for single parents in Mincome consist of only the 4 normal income categories (see Table 1). For single parents in Gary, there are 5 normal income categories, and 2 locations resulting in ten mini-experiments.

Table 2 summarizes the previous literature on labor supply responses of single parents[23] based on survey articles by Robins (1985), Burtless (1987) and Hum and Simpson (1993).[24] Several features are evident. All estimated impacts are negative in sign, although only those for Blacks in Seattle and Denver and Hispanics in Denver are statistically significant.[25] In addition,

[21] As discussed below, location was also a stratification in Mincome, but due to data availability we only use the Winnipeg (urban) data.

[22] Other consequences of the assignment model are that (i) sample sizes for individual mini-experiments are small in most cases, and (ii) there is unbalanced allocation to treatment and control groups – the sample size of the control group is typically much smaller than the treatment group.

[23] These estimates are averages of the experimental treatment-control differences and do not include the large number of structural model estimates. The main focus of the early NIT literature was on using structural labor supply models to estimate income and substitution effects, estimates that could then potentially be used to simulate economy-wide responses to the introduction of a NIT. Our focus in this paper is on the experimental estimates, an approach similar to Ashenfelter and Plant (1990).

[24] As discussed subsequently, the sole previous study of labor supply in Mincome (Hum and Simpson, 1991) pooled together single mothers and single adult women so there is no previous evidence for single parents.

[25] The previous literature typically reported results for Seattle and Denver pooled, as in Table 2. However, because separate estimates for Hispanics are not available for SIME, the SIME-DIME results for Hispanics come only from

estimated impacts on annual hours of work and the employment rate are generally small in size, again with the exception of those for Blacks in Seattle and Denver and Hispanics in Denver. More precisely estimated impacts on the intensive and extensive margins are consistent with the much larger sample sizes in SIME-DIME.[26] Overall, these findings from the Income Maintenance Experiments that included single parents reduced support for the view that the NIT constituted an alternative to welfare with stronger work incentives and potentially positive impacts on labor supply.

3. Data

(a) The Mincome Experiment[27]

Mincome was a joint federal-provincial initiative carried out in Manitoba in 1974-78. There were three sites: Winnipeg, the rural dispersed sites and the non-experimental 'saturation site' of the town of Dauphin in which all low-income families were eligible. We ignore the non-experimental Dauphin site as well as the rural site in this paper because the periodic surveys were never digitized for these sites. As the Canadian experiment has received far less attention from academics with only a single study of experimental labor supply impacts (Hum and Simpson, 1991), we spend more time on the background of Mincome with additional details provided in the Appendix (see Online Appendix: Data and Institutions section B).

A combination of declining interest in the concept of a guaranteed annual income and budgetary problems resulted in Mincome being shut down at the end of the operational phase in 1978 without any funding for research and analysis. No final report was produced and the survey and payment records remained mainly in hard copy form. In 1981 the federal government provided some funding to restore the Mincome data and promote its use. By 1983 the data that had been digitized, together with detailed codebooks, was available to

---

Denver. We argue later that there are good reasons to consider SIME and DIME as separate experiments, and therefore analyse them separately.

[26] Note, however, that standard errors were not clustered in these earlier studies and thus reported p-values are likely understated (and statistical significance overstated).

[27] This section provides a brief overview of Mincome. More detail is available in the various technical reports and studies referred to in Simpson, Mason and Godwin (2017).

researchers. Some research was subsequently carried out, but this was limited by the fact that interest in the guaranteed income policy had waned. As a consequence, Mincome remains substantially under-researched relative to the U.S. NIT experiments. Indeed, only one published study of labor supply effects using the Mincome survey data – that of Hum and Simpson (1991) – has been carried out.[28] We have been unable to replicate the results of this study and the published study does not provide sufficient detail about how the data were processed to be helpful. Neither the data used by the authors or their code are available. That said, Hum and Simpson pooled single women without kids with single mothers, and thus no estimates exist for single parents in Mincome.[29]

In addition to information from the two pre-random assignment surveys and 9 post-random assignment 'periodic' surveys, monthly administrative data from the payments system are also available (post random assignment only). A separate agency, Mincome Manitoba, was established to operate the payments system. Treatment group participants were required to submit monthly 'income reporting forms' (IRFs) as well as their employer's pay stubs and received monthly payments (depending on their earnings, guarantee level and tax rate) from Mincome Manitoba. Staff from the payments group were available in person to assist participants completing this form. Perhaps most importantly, Mincome Manitoba also filed annual income tax returns for participants and, after reconciliation, handled adjustments for under- or over-payments. One implication of this substantial monitoring of participant's

[28] Results from this study are also reported in Hum and Simpson (1993) that surveys evidence from the U.S. and Canadian income maintenance experiments. Calnitsky and Latner (2017) carry out a non-experimental analysis of the extensive labor supply margin using administrative data (see our discussion below) from the Dauphin site, which was not randomly assigned.

[29] We have tried to replicate the findings of Hum and Simpson across all family types, and have been unable to do so. For example, we cannot even replicate their sample. According to Mincome documentation (Mason 2016) 1074 intact households were enrolled and randomly assigned at the Winnipeg site, consisting of 704 treatments and 370 controls, sample sizes that match ours. However, Hum and Simpson (1991) report samples of 1187 intact families, 575 treatments and 612 controls. The reasons for the smaller number of treatments and much larger number of controls is unclear. As noted previously in footnote 21, one consequence of the Conlisk-Watts assignment model as implemented in all four U.S. NITs is a smaller number of controls than treatments, a feature that also holds in our sample and the Mincome documentation. The fact that the number of control families in their sample exceeds treatments together with the absence of reported balancing tests raise doubts about the validity of their sample.

employment income is that under-reporting of earnings, an important limitation of the U.S. NITs, is very unlikely in Mincome.[30]

We employ the various sources of information currently digitized for Mincome: the baseline and enrollment surveys conducted prior to random assignment, the post random assignment periodic surveys, and the monthly administrative data collected separately as part of the payments system. The baseline survey collected information on a limited set of individual and household characteristics including labor market information, income, and receipt of government transfers for the 1973 and 1974 years. Most of this information is annual such as weeks worked, income sources, and receipt of government transfers during those years. Only limited information from the enrolment surveys (that were conducted around the time of random assignment) has been digitized; we do not currently use these surveys other than for hours worked prior to random assignment. The post random assignment 'longitudinal labor surveys' `were collected approximately every four months for three years resulting in at most nine post random assignment observations for each participating family. To summarize: we have two 'point-in-time' pre-random assignment surveys, and then nine 'point-in-time' post-random assignment surveys which cover, on average, a three-year post-random assignment period.

We also note that in all four of the North American NITs there was variation in the calendar time date when both random assignment and these various surveys occurred. That is, like many large-scale social science experiments, there was staggered entry (i.e., the date of random assignment) into the experiment over calendar time. Unlike the U.S. NITs, there was a minimal amount of staggering in Mincome with about 85% of the sample beginning the experiment between March 1975 and May 1975. We also note, to preview the experience of the U.S. NITs, that there were no noticeable differences in the staggering between treatments and controls.

Our analysis is based on single parents that appear in both the baseline survey and the periodic surveys. This results in a sample size of 136 single parents.[31] Table 3 reports summary

[30] We discuss evidence on under-reporting in Gary and SIME-DIME subsequently. In Mincome, any mis-reporting of earnings for the purpose of obtaining larger NIT payments would also require cooperation of the employer and under-reporting income to the tax authorities, which risks serious penalties.

statistics across all four North American NITs. We note that the structure of the welfare and labor market variables is not the same between Mincome and the U.S. NITs as discussed in more detail below. We compute welfare and labor market variables to be as similar as possible. Briefly, the U.S. NITs contain monthly information pre-random assignment based on start and end dates of jobs and AFDC spells while Mincome is based on survey questions of whether the individual worked and received welfare payments during the calendar year prior to random assignment.

For Mincome, we see that 43% of the sample reported receiving welfare at some point during the pre-random assignment period. This variable is based on whether the individual received welfare at any point during 1973 or the survey window for 1974 (which ranged from May to December depending on the enrollment interview date). We suspect that the 'welfare receipt' variable understates total social assistance receipt in Manitoba, perhaps considerably. In particular, because Mother's Allowances remained a distinct program with different eligibility requirements and benefit levels, it is unclear if single mothers receiving Mothers' Allowance benefits would respond in the affirmative to a question regarding receiving 'welfare' benefits. The term 'Mothers Allowances' had a long history in the province, and in the 1970s continued not only to be the program's official name but also appears to be the term referred to by recipients, as evidenced by a recent interview of a single mother who attributes participating in Mincome's NIT with allowing her to enter the workforce and develop a successful career as a librarian (see Online Appendix: Data and Institutions section C).[32]

Moreover, our data also suggests that the welfare variable may not be capturing all forms of social assistance; in particular, we find a large number of 'Other' responses to the income module question on "other government assistance." Available responses to "other government assistance" were: Old Age Security ("OAS"), Canada Pension Plan ("CPP"), Guaranteed Income Supplement ("GIS"), Family Allowance and 'Other'. Roughly 90% of our sample reported receiving positive amounts of "other government assistance", and the mean

[31] The overall sample size of single parents appearing in the baseline and periodic surveys is 142; however, information on hours worked prior to random assignment is only available for 136 single parents and thus that is our sample at the time of random assignment.

[32] See https://www.huffingtonpost.ca/2014/12/23/mincome-in-dauphin-manitoba_n_6335682.html.

amount conditional on positive receipt is substantial. The first three sources (OAS, CPP, GIS) would only apply to individuals over 64 years of age, and so do not apply to our sample (maximum age is 57 for single parents in our data).[33] All of the single parents would receive the Family Allowance, but this alone cannot account for the dollar amounts we observe. To provide an alternative measure of welfare receipt, we compute a second welfare variable that includes individuals who report "other government assistance" in excess of their Family Allowance amount.[34] This yields welfare receipt of 68%, slightly higher than SIME and DIME.

About 66% of single parents worked at some point during the pre-random assignment period (lower than DIME by 6 percentage points, higher than SIME by 3 percentage points). The latter suggests strong labor market attachment among single parents in Manitoba; however, average monthly hours are substantially lower than what is observed in SIME (33% lower) and, especially, DIME (86% lower). Average monthly hours worked in Mincome implies only about 13 hours a week.[35] This is also consistent with the welfare variable reported in Table 3 being understated.

Finally, average age is similar across all the NITs. Single parents in Mincome are noticeably less educated than their counterparts in Seattle and Denver (but similar to Gary) and have more children under age 16 than in the U.S. NITs. We also note that the education variable has many missing values (especially in SIME and DIME), which appear to be non-randomly missing.[36]

(b) Seattle-Denver

We group Seattle-Denver (SIME-DIME) together for purposes of this section as the data structure is identical. Indeed, it has been common in the literature to refer to Seattle and Denver as essentially one experiment; previous analysis has pooled the two together, and much

[33] Unemployment insurance is a separate variable, and so likely does not explain the amounts we observe.
[34] The family allowance was $20/month per child. We create a second welfare variable that equals one if the individual reports receiving welfare based on the welfare variable, or reports receiving "other government assistance" in excess of $80 per child based on the second periodic survey (which was the enrolment survey, and had a 4-month survey window).
[35] Recall that a small earnings disregard existed at this time, equivalent to about 9 hours of work per week (see footnote 15).
[36] For these reasons we do not include years of education as a control variable in any of the regressions. Incidentally, the original NIT literature also did not include education as a control variable as far as we can tell.

of the literature refers to the U.S. NITs as consisting of four experiments (New Jersey, Rural, Gary and SIME-DIME). However, as we emphasized previously, there are fundamental differences between the Seattle and Denver experiments so we analyse them separately. Because SIME-DIME has received considerable attention in the academic and policy literature we restrict our discussion to key points about SIME-DIME's data structure that have received relatively little previous attention.

We use the SIME and DIME 16th Monthly Composite Principal Person Files. The original SIME-DIME data collection was similar to Mincome in that pre-random assignment surveys were collected followed by post-random assignment periodic surveys (also roughly 4 months apart). However, the data that was digitized for public use, the SIME and DIME 16th monthly composite files, have important differences from Mincome. First, the labor market information was primarily collected from job start and end dates; from those, a 72-month panel was constructed. Second, while both Mincome and SIME-DIME had staggered entry, SIME-DIME collected the 72 months of data over the same calendar time period. Thus, cohorts differed in the number of months that constitute pre-random assignment data. Moreover, SIME-DIME had both 3- and 5-year programs, and thus for the 3-year program, there is also post-experiment data. The number of months of post-experiment data also varies by entry cohort. This contrasts with Mincome where all single parents have two data points pre-random assignment, nine data points post-random assignment, and no information after the experiment ended. Table A2 in the Appendix shows the panel months (out of 72) that correspond to the experimental period (where single parents were eligible for the NIT) across entry cohorts.[37]

Note that DIME also had a 20-year program; however, there are no single parent members of the treatment group. All 33 observations on single parents in the 20-year DIME program are members of the control group. We thus exclude the 20-year program from our analysis.

[37] We note that-- as far as we can tell-- the fact that the length of the pre-experimental and experimental periods vary substantially across individuals (and thus should be coded at the individual level) was not incorporated into the original analyses of the 1970s-80s.

A key issue that arises in the SIME-DIME data is the extent to which a variety of family types are arguably mis-classified as 'single family heads'.[38] The literature has stated in many cases that the sample is 'single female heads with a dependent child'.[39] A close look at the data, however, reveals that there are a non-trivial number of observations that, at some point over the *pre-random assignment period*, do not satisfy these single parent definitions, and thus are appropriately deleted from the sample. In particular, we exclude the following cases: households where there was (at the time of random assignment) a second adult head[40], and households with no dependent child (i.e., children of the single female head are adults).[41] We also exclude observations from mini-experiments with no support (i.e., zero observations in either the control group or treatment group).[42] Ultimately, our sample consists of 869 single female heads for DIME and 713 for SIME (1582 total).[43]

Unfortunately, it is exceedingly difficult to reconcile the sample sizes of the early SIME-DIME studies with the public data. Based on our review including the SIME-DIME Final Report (SRI International, 1983), Brookings/Federal Reserve Bank of Boston 1986 Conference Proceedings (Mundell, 1987), and the special issue on SIME-DIME in the *Journal of Human*

[38] This discussion also applies to Mincome and Gary, but is trivial in terms of the number of observations affected. For instance, we find 6 cases in Mincome that do not satisfy the definition of a single female head with a dependent child (at the time of random assignment); these are all situations where the single female head has no dependent child (indeed, all these cases involve the 'child' being in their 30s or 40s). We exclude these observations. We define marriages and all dependent children leaving the household in the same manner as SIME-DIME.

[39] The general restriction to single female head with a dependent child is reasonable; for instance, among 'in-tact' families there are 10 single fathers vs. 563 single mothers in the case of SIME. Given the data construction, in-tact families are perhaps the cleanest sample; these are households of a single parent with a dependent child that have no change in any household composition from panel month 1 to 72. We replicate all of our results for this sub-sample, and reach the same conclusions.

[40] This primarily consists of cases where the spouse or partner had applied for eligibility but was still waiting for the decision, and cases where the spouse returned to the household over the pre-random assignment period.

[41] In the case of SIME this amounts to 70 observations that are actually households with two adult heads, and 50 observations where there is no dependent child (we define this as under age 16, but the majority of these are clearly adult children). Ultimately, we end up with a sample (at the time of random assignment) of 713 observations for SIME and so drop 137 observations (including those who do not appear in both the baseline and periodic surveys). The relative number of observations dropped for these reasons, and the category they fall in, are very similar for DIME.

[42] This is income category 6 for both SIME and DIME (there is also a single observation in DIME in income category 7). Income category 6 amounts to 13 observations for SIME. There is also no support in income category 1 for Whites in the 5-year program (we include this latter group in Appendix Table A1).

[43] We also exclude observations that never joined the experiment at the time of random assignment (5 observations). Finally, for DIME, we exclude the 32 observations in the 20-year program as there is no support (all 32 observations are Controls).

Resources (Spiegelman and Yaeger, 1980), we find sample sizes for "single female heads" ranging from 1459 to 1715. We can find no documentation that explicitly declares a starting sample nor what deletions were made.[44]

Finally, among the 'single female head with a dependent child' households at the time of random assignment, there are a few cases where the household type changes post random-assignment. The majority of these consist of the single mother marrying, with the others consisting of the dependent child leaving the household. We include the latter cases in the sample for the duration of single parent status, but treat the observations after the change in household type as attritors.[45]

Turning to the summary statistics, welfare use was high in SIME-DIME with around 60% of single parents in these experiments reporting welfare receipt in at least one month over the pre-random assignment period. Note, the pre-random assignment period varies from 10 to 22 months. In our balancing tests below, we focus on months 1-9 as these months are pre-random assignment observations for all individuals. Labor market attachment is higher in DIME with 71% of single parents working at least one month compared to 63% in SIME—consistent with very different economic conditions. Roughly 40% of single parents received welfare every month over the pre-random assignment period. We denote the latter group "welfare recipients". The "welfare recipient" group will, in our subsequent analysis, constitute those who are highly likely to be at the corner solution when the NIT is implemented.

[44] The SIME-DIME Final Report states that at the time of enrolment there were 2031 single heads eligible (SRI International, 1983, volume I, page 26); this may have included some single male heads. Not all of these individuals were actually enrolled (randomly assigned); the sample size with treatment assignment is listed as 1832 (Final Report, volume I, p. 25). Robins (1985) Table 1 lists this as 1872. Next, not all of individuals assigned to treatment status participated at all, which appears to reduce the sample in experimental year 2 to 1715 (see Final Report, volume I, Table 3.3). Finally, Table 3.6 of the Final Report reports estimation results for experimental years 1-4 with a sample size of 1459. The difference between experimental year 2 and the balanced sample for years 1-4 may reflect attrition. We note that sample sizes were not regularly reported in tables at that time. A 'number of enrolled families' was often reported early in studies (for all the NITs), but it is unknown what observations were excluded before analysis. Our calculations show that the 16th Monthly Composite file contains 1083 single female heads for DIME, and 850 for SIME for 1933 total female heads (this excludes single male heads). These are the original principal person records based on households with an experimental grouping of single female head. As would be expected given that some of the families were not in fact single female heads with a dependent child, welfare receipt is lower for the 'full' sample, and employment rates are higher. The results of the balancing tests (reported subsequently) remain essentially the same.

[45] For SIME as an example, this reduces the number of households from 713 to –by the end of the experiment— 689 single parents with the first 'attritors' beginning at panel month 26.

Figure 2 shows monthly trends in welfare receipt and take-up for the Seattle and Denver 3-year programs and Gary (i.e., all programs where post-experiment data is available). Vertical lines denote the beginning and end of treatment for the median entry cohort. The decline in welfare use in the treatment group largely mirrors take-up. On average take-up was 55% in DIME and 49% in SIME. The observed decline in welfare receipt after month 10 among treatments in SIME and DIME illustrates the staggered entry into the experiment coupled with the data structure noted above; the earliest cohorts begin treatment at month 10 while the last cohort begins treatment at month 22 (see Appendix Table A2). The other noteworthy feature of Figure 2 is the U-shaped behavior of welfare receipt for the treatment group.[46] Most treatment group members who took-up the NIT program and left AFDC appear to have reverted to their original income support use when the experiment ended. This pattern is particularly strong in Gary. The difference in take-up rates across experiments is also striking with take-up in Gary at virtually 100% while take-up in SIME and DIME is roughly half that. For comparison, take-up in Mincome was 55%, similar to SIME-DIME.

(c) Gary

We include Gary for completeness given that Gary contained a large sample of single parents (roughly 60% of Gary participants were single female heads). We also find several features of the Gary experiment –with respect to single parents—that have been downplayed in the previous literature and are worth emphasizing. That said, the Gary data has important caveats for our analysis; in particular, in addition to the failure of random assignment, there is very little pre-random assignment data.

The Gary data construction is broadly similar to the 3-year SIME-DIME case with the main data consisting of a 48-month panel, and where we observe a pre-random assignment

[46] Data at the end of the sample period should be viewed with some caution. As discussed later, attrition in the three-year SIME-DIME was relatively low until the end of the experimental period (about 15%), but then increased dramatically (to over 60% by month 70). While we find no evidence of attrition bias (see below), the sample sizes are very small at the end of the sample period.

period, the experimental period, and a post-experiment period.[47] As well, the relative lengths of each period depend on the entry cohort (date of random assignment or 'enrollment'). As seen in Table A2, the key difference in Gary is that due to the shorter panel (48 vs 72 months), the length of the pre-random assignment and post-experiment periods are considerably shorter.

The characteristics of the Gary sample are strikingly different; in particular, Gary appears to have a much higher proportion of welfare recipients. While the length of the pre-random assignment period is considerably shorter for Gary, welfare receipt is almost 80%. There is also low labor market attachment in Gary with only 20% of the sample working in at least one month.[48] Moreover, single female heads in Gary worked, on average, only about 7 hours per month –compared to 50 hours in Manitoba, roughly 70 hours in SIME, and nearly 90 hours in DIME.

Overall, with respect to single parents, the Gary NIT appears to consist of a dramatically different sample. While local labor market conditions at the time of random assignment in Gary were severe (see Online Appendix: Data and Institutions section D), so too were conditions in Seattle. However, Seattle did begin a recovery (despite a nation-wide recession) in early 1974 while Gary did not.

(d) Balancing Tests and Attrition

It is now common to present results from balancing tests for experiments; that is, examine the differences in the characteristics of treatments and controls prior to random assignment. As far as we can tell, no balancing tests were reported in the original NIT literature.[49]

---

[47] Unlike SIME-DIME, the public-use Gary file has a point-in-time baseline (the 'Economic Baseline') that was conducted prior to the periodic surveys (analogous to the baseline survey in MINCOME); however, there are widespread missing observations and thus we do not use it.

[48] The hours worked variable has quite a few missing observations in it for Gary unlike SIME-DIME. There is also a labor force status variable in Gary (as well as SIME and DIME), which is observed for the full 913 observations. If we define employment using the status variable, the mean employment rate is almost the same as presented in Table 2. For consistency across the experiments, we use positive hours worked to define the employment variable for Gary.

[49] Several authors (e.g. Keeley and Robins 1980; Robins and West 1980) point out that there were pre-experimental differences in labor supply of treatment and control families, as well as different trends in work activity during the experiment. However, as they also note, such differences are expected given the nature of the Conlisk-Watts assignment model. By "balancing tests" we refer to tests for treatment-control differences controlling for stratification categories, including normal income. Under such tests failure of random assignment could alter our interpretation of the treatment effects as causal impacts.

There is some reason for skepticism as to whether random assignment was in fact achieved in several NITs, especially in Seattle where economic conditions changed substantially during the enrollment and assignment to treatment period. Both Seattle and Denver were unable to enroll the number of families specified by the assignment model, and the extent of the gap between actual and theoretical assignment differed between treatments and controls.[50] In Seattle, changes in the assignment model during the enrollment process were an additional complication, one that contributed to the enrollment process taking longer in Seattle (SRI International 1983, volume II, Chapter 3). Inspection of the data shows that there were important differences in the dates on which treatment and control families were randomly assigned, perhaps as a result of these factors.[51]

In particular, we found that in SIME –but not in DIME or Gary—the entire control group was surveyed in the December 1970 cohort (see Appendix Table A1). That is, while the control group in SIME was matched with a particular treatment enrolment cohort for purposes of the assignment model, their actual labor supply data (i.e., when the panel started) was all based on December 1970 whereas the treatment group data was staggered over a one-year (October 1970 to October 1971) period with most enrollment surveys occurring in the Spring of 1971. Appendix Table A1 shows the differences in the median enrolment interview dates. Given rapidly changing economic conditions in Seattle, the differences in calendar time between treatments and controls may have been problematic. We note that there were non-trivial differences in enrolment dates for DIME as well; however, Denver was not experiencing similar labor market conditions.

We present results for balancing tests on labor market outcomes prior to random assignment in Tables 4-7.[52] In Table 4 we include estimated treatment – control differences

[50] For single parents the ratio of actual to theoretical assignment averaged 40.1% for controls and 59.9% for treatments, with a range from under 20% to over 80% across normal income categories (SRI International, 1983, Table 1.5).

[51] As stated in the Final Report, "It would, of course, had been better if the [time] span had been shorter at both sites so that all enrolled families could have faced the same labor market at the same point during their experimental treatment." SRI International, 1983, Volume II, p. 35)

[52] All of the North American NITs are light on non-labor market information, but we find differences in age between treatments and controls for SIME and Gary (see Table A3).

with and without controls for normal income categories to illustrate the importance of these controls under the Conlisk-Watts assignment model.[53]

Mincome sample sizes for single parents are particularly small, and there is some evidence in Table 4 that the treatment group had higher labor market attachment prior to random assignment. However, with controls for normal income no statistically significant differences are found and the magnitude of the differences are generally below 10% of the pre-random assignment mean. For SIME, we find substantial differences in labor market outcomes prior to random assignment.[54] We show monthly differences between treatments and controls for panel months 1-9 which corresponds to the pre-random assignment period for all entry (enrollment) cohorts. The mean differences for the 5-year program are generally small for employment and wages. A few differences are observed for welfare receipt although they are not statistically significant. Note sample sizes for the SIME 5-year program are small. For the 3-year program, however, there are quite systematic differences across all variables. Given the estimates in Table 5, we conclude there appears to be reasonably strong evidence that the treatment group in the SIME 3-year program had noticeably lower employment rates, and higher welfare receipt than the control group. Hours worked and total wages differences mirror the employment rate differences. We explore this in more detail later by examining the trends over time for each mini-experiment separately.

Conversely, in DIME, there are clearly no differences in the 3-year program for employment and wages. There is some evidence that the treatment group was more likely to be on welfare in the 3-year experiment, and less likely to be on welfare in the 5-year case, but no estimate is statistically significant. Finally, we also find substantial differences in welfare receipt prior to random assignment in Gary (see Table 7).

Lastly, we explore attrition and the possibility of attrition bias (results presented in Appendix Table A4). Attrition in the U.S. NITs has been discussed extensively previously, but prior work has a) pooled all family types and also pooled SIME-DIME; b) not necessarily distinguished between the experimental and post-experimental period; and c) not conducted tests on attrition bias that are more common today; in particular, examining whether attritors

[53] Similar comparisons with and without normal income controls for the U.S. NITs illustrate the same point.
[54] These differences are larger and more precisely estimated without normal income controls.

differ from non-attritors in pre-random assignment labor supply outcomes. Overall, attrition is not high for single parents with Mincome at 20%, and the U.S. NITs around 15% if we define drop-out at the end of the experiment (as opposed to the end of the data set). If we define drop-out as exiting the data at the end of the panel (i.e. month 72 for SIME-DIME, and month 48 for Gary depending on the enrolment date) attrition rates increase dramatically (see Table A4). We note that attrition rates are noticeably greater for two-head families. The latter two features of the data may explain the greater emphasis on attrition in the existing U.S. NIT literature. Given that the treatment effects should be estimated based on the experimental period, it is unclear whether drop-out from the data after the experiment ends is problematic.

There appears to be no consensus test for attrition bias. We perform the test outlined in Fitzgerald, Gottschalk and Moffitt (1998) by regressing pre-random assignment hours worked on an attritor dummy (and other individual characteristics). If attrition is independent of potential outcomes, it should not be correlated with pre-random assignment labor supply. Overall, we find no evidence of attrition bias in any of the North American NIT experiments for the single female head group.[55]

4. Results

This section reports experimental estimates for all four experiments and non-experimental estimates for SIME.[56] All regressions control for the random assignment stratification groupings which includes: normal income categories for all samples, race for SIME-DIME, location for Gary, and program length for SIME-DIME (along with the demographics from Table 3). Finally, all regressions include fixed effects for time; in Mincome we include dummies for the periodic survey (8 dummies), and for the U.S. NITs we include dummies for the panel month. Note that in the U.S. NITs we estimate the treatment effect only for the experimental period, which is

[55] In the original literature it was also common to look at the correlates of attrition. In results not shown we explored this as well since previous analyses pooled household types. The control group is more likely to drop out as noted previously in the literature, but only by about 2-3 percentage points (and the treatment – control difference is not statistically different from zero). For the U.S. NITs, the only group that differs in its attrition rate is the inside (urban) location group for Gary, with single parents in this group being 9 percentage points more likely to drop out.

[56] As noted previously, there is insufficient pre-random assignment data to credibly apply our empirical strategy (a difference-in-differences framework) for Gary.

defined at the individual level. Table A2 shows the relevant range of panel months. Because of sample size limitations we primarily use a single treatment dummy, as is common in earlier NIT studies. The estimates can thus be interpreted as the weighted average impact across the various treatment plans (combinations of G and t, see Table 1).

We begin with a discussion of Mincome. Treatment effects are reported in Table 8.[57] For single parents, the treatment effects estimated with the periodic surveys are positive along both the extensive and intensive margins. Specifically, we find that single parents randomly assigned to the treatment group were about 14 percentage points more likely to work relative to the control group, a large effect in magnitude, and hours worked were about 72 hours higher based on the four-month window measured in the surveys or roughly 210 hours annually. Relative to mean baseline hours worked, the treatment effect on hours translates into about a 34% increase. Recall that due to the small sample size (N=123) and the fact that we cannot reliably identify welfare recipients in the Mincome data (due to the Mothers Allowance and Winnipeg-specific programs), we cannot separate welfare recipients from workers (pre-random assignment) as we do below for DIME. From the framework outlined in Section 2 above, we believe the treatment effect here is understated for welfare recipients (those at the corner solution), and overstated for workers.

Turning to the results using the monthly administrative data on wages[58] from the payments file, we see that the findings outlined above appear robust.[59] The payments file not only consists of administrative data as opposed to survey-based data, but also differs from the survey-based sample in several respects (see Online Appendix: Institutions). First, non-in-tact households are included in the administrative records. Second, some households that never filed an IRF are included in the longitudinal labor survey results discussed above but would be

[57] The number of single parents in the sample declines from 136 (Table 3) to 126 (Table 8) due to 10 observations where the individual attrits before the first post-random assignment survey (periodic #3 for Mincome). This also occurs for SIME and DIME (compare Table 3 to Table 9) where a small number of single parents attrit before the experimental period begins (e.g., 713 observations down to 699 for SIME).

[58] There is no hours worked information in the payments file, only wage data.

[59] As noted previously, an additional potential benefit of the payments data is that under-reporting of income—an important caveat of the U.S. NITs—appears unlikely in the Mincome payments system since (i) pay stubs were required and carefully monitored and (ii) the payments staff completed tax returns for participants. That said, even the survey data in Mincome may be less likely to have under-reporting of income if participants believe Mincome staff will cross-reference survey responses with administrative information.

excluded from the administrative file. Third, the control group is different as additional control group households were included in the payments file ('IRF Controls') who were excluded from the surveys. Random assignment for the two household types was evidently conducted correctly for both, but we do have different control households. Finally, attrition rates were much lower in the payments data than the 20% in the surveys although unlike the surveys there were discontinuous spells, i.e., individuals not filing an IRF in a given month and thus no wage observations that month, but then subsequently filing. Overall, our evidence is that treatment effects for the probability of employment are almost identical to the survey-based evidence.

Table 9 reports estimates from the three U.S. NITs. As is common in the early literature, we also present results separately for the 3 and 5-year programs in SIME-DIME but we depart from the earlier literature in reporting separate results for Seattle and Denver. Treatment was associated with about an 11-percentage point decline in the probability of employment in SIME, and 3 percentage points in DIME (not statistically different from zero). In hours, the annualized coefficients imply a reduction in hours of around 200 hours in SIME, and 65 hours for DIME (although not statistically different from zero)—these are relative to pre- random assignment means of 830 annual hours worked in SIME, and 1060 hours worked in DIME. For SIME, the results are primarily driven by the 3-year program while for DIME there does not appear to be any difference between the 3 and 5-year programs. Estimated coefficients are small and just outside of conventional significance levels for Gary where single parents sample sizes are somewhat larger.[60]

Overall, we appear to largely replicate the results from earlier studies of the U.S. experiments, as summarized for example by Robins (1985), Burtless (1987) and Hum and Simpson (1993). Despite dramatically different economic conditions across Seattle and Denver and other differences in the experiments' operation, earlier studies pooled SIME and DIME. Based on Robins' (1985, Table 4) summary of the experimental findings, the estimated impact on the employment rate of single female heads is 8 percentage points for SIME-DIME pooled.[61]

[60] The sample size for Gary in Table 9 is larger than Table 3 because some individuals have no pre-random assignment data (other than the Economic Baseline, but as noted previously we do not use that data due to the large number of missing observations).

[61] Burtless (1987) and Hum and Simpson (1993) do not report estimates of employment rate impacts.

Standard errors were likely understated in the earlier studies due to not adjusting for clustering on the individual. Our weighted-average estimate for SIME-DIME on the probability of employment (7 percentage points) is very close. For annualized hours ITTs, our estimates (about 130 hours) appear to be somewhat lower than those of Robins (155 hours) and Burtless (144 hours) for SIME-DIME pooled. For Gary, the Robins (1985) average is an ITT on employment of -4 percentage points and -37 hours annualized compared to our ITTs of -3.4 percentage points and -15 hours worked. Overall, our tighter definition of the experimental period (defined at the individual-enrolment level), monthly fixed effects, and the possibility of a different single female head sample does not appear to result in much of a difference in estimated labor supply impacts relative to the early findings.[62] The key point we emphasize is that the treatment effect in DIME is negative, but small and not statistically different from zero. Recall from our previous discussion, we cannot find previous evidence on single female heads where SIME and DIME were examined *separately*. Thus, the consensus negative labor supply estimates for Seattle and Denver in the original literature were driven almost entirely by SIME.[63]

Given the results of the balancing tests, we explore each individual mini-experiment graphically in Figure 3 for SIME (see Appendix Figure 1 for Gary). The vertical bars in the figures show the experimental period using the median enrollment date for single parents in that particular mini-experiment. While sample sizes (shown in parentheses) in some cases are quite small (particularly for the SIME 5-year program), visual inspection of the figures reveals that many experiments do not appear to resemble random assignment either due to different trends in labor market outcomes pre-random assignment or, in particular, large differences in

[62] As discussed subsequently, we also estimated these regressions with and without cohort dummies (for DIME) to allow for the fact that some treatments and controls were assigned in different months, and thus subject to different economic conditions at random assignment, as well as calendar time controls (calendar month and calendar year). For SIME, cohort dummies are not possible as the entire control group was enrolled in a single cohort (with no matching treatment observations), and thus no within-cohort estimator is possible. Overall, cohort and calendar time controls make little difference to the treatment effects.

[63] As was also done in early studies for SIME-DIME specifically, we report in Appendix Table A5 the ITTs by program length and race. Thus, the only experimental stratification we are pooling (and controlling for in the regressions) is the normal income category. Table A5 shows that, in SIME, the negative labor supply response is similar across groups with the exception of the 5-year program for Whites. In DIME, we find a reduction in the probability of employment for Whites in the 3-year program, albeit one that is marginally significant at the 10% level; otherwise, ITTs are essentially zero for all other groups.

levels pre-random assignment. Indeed, the 5-year program for Whites appears to typify the SIME experiment: it is perhaps the only case of clear evidence of a negative labor supply response, but ironically, the original 'ITT' for this group was the one positive estimate.[64]

Given the patterns in hours worked prior to random assignment, a preferred analytical approach may be a difference-in-differences design. It is common in difference-in-differences to visually inspect pre-treatment outcomes to assess parallel trends. We focus on the SIME three-year program since sample sizes in the SIME 5-year program are likely too small to credibly assess parallel trends.[65] We comment on the visual patterns observed in the 5-year program below.

Nearly all experiments -- despite differences in levels -- have trends in hours worked pre-random assignment that are consistent with parallel trends with the exception of Income Cell 3 for Blacks (see Figure 3). Unfortunately, in Gary, there are insufficient pre-random assignment observations for a credible assessment of parallel trends.

For SIME we estimate canonical difference-in-differences models of the general form:

$$Hours_{it} = \alpha + \beta_1 (Treatment\ Group_i) + \beta_2 (Post\ NIT\ Offer_t) + \beta_3 (Treatment\ Group * Post\ NIT\ Offer_{it}) + \phi_i + \varepsilon_{it} \quad (1),$$

where ϕ_i are fixed effects for the mini-experiment, and where we also include interactions between the mini-experiment and *Post NIT Offer* to allow for different slopes. As well, we estimate equation (1) both separately by race, as well as pooling Whites and Blacks (and including an interaction term for race and *Post NIT Offer*). Other time and socio-economic covariates are identical to the specifications above.

The results are displayed in Table 10. The treatment effect for the 3-year program in SIME is small and statistically insignificant. For Whites in particular, the strong, negative coefficient reported in earlier studies (and in our replication above) is zero. For the 3-year program for Blacks, the negative labor supply response is driven entirely by Income Cell 3 where pre-random assignment trends are widely different (see Figure 3), and would almost

[64] While our focus will be on SIME, the same overall conclusion appears to hold for Gary (albeit with less pre-random assignment data to work with): a) hours worked (while much lower, on average, than in SIME) are constant or increasing for the treatment group in all cases, and b) negative ITTs (albeit not statistically different from zero) appear to be driven by pre-random assignment differences.

[65] For instance, Income Cells 1 and 5 for Blacks have only three treatment observations.

certainly be viewed as inconsistent with the parallel trends assumption in a standard difference-in-differences research design. When we exclude Income Cell 3, the difference-in-differences estimate is close to zero at -3.7 hours worked (and not statistically different from zero).

With respect to the 5-year program, our assessment of the visual trends is as follows. There does appear to be a negative labor response for Whites enrolled in the 5-year program (the opposite of the original estimates in the literature, which were positive). That said, the number of treatment observations for Whites in the 5-year program that is driving the result (income categories 4 and 5) is very small, and so it is unclear how compelling the evidence is. For the 5-year program for Blacks, there is perhaps some evidence of a small, negative labor supply response based on Income Cell 3, but given the offsetting patterns for Income Cells 2 and 4, any negative labor supply response for Blacks in the 5-year program appears considerably muted compared to the original results. Overall, combing through the difference-in-difference estimates from the 3-year program and a visual inspection of trends in the 5-year program, we conclude the consensus view of a negative labor supply response to the NIT in SIME for single female heads is misleading.

We now explore heterogeneity in the treatment effect based on welfare status pre-random assignment in DIME. We sub-sample based on our definition of a 'welfare recipient'; that is, we use those individuals who received welfare in every month over the pre-random assignment period as our test of the effect of the introduction of a NIT on labor supply for individuals at the corner solution.[66] The remainder are referred to as 'workers.' Because we are conditioning on a pre-random assignment characteristic, the 'working' and 'welfare' subsamples are both as good as randomly assigned to treatments and controls. We therefore focus on DIME given that random assignment appears to have been conducted correctly.

[66] Given sample size as well as data limitations we prefer a clear distinction between welfare recipients and workers in order to identify individuals who are highly likely to be at the corner solution when the NIT is introduced. Some single parents may appear in the data to combine both activities during a specific period of time. One reason is the earnings disregard. This type of 'corner solution' is consistent with the static labor supply model. Another potential reason would be obtaining or losing a job during the period, resulting in movements from welfare to work or vice versa. Distinguishing between these two types of combined activities as well as accounting for such dynamic behavior would require richer data than is available in these experiments.

Treatment effects are presented in Table 11. The results suggest that the response to the NIT program differed substantially across welfare participation. For 'welfare recipients' we find a positive labor supply response to the NIT of about 14 hours monthly (around 150 hours annualized). Our 'welfare recipient' group had average monthly hours worked over the pre-random assignment period of about 40 hours[67], and thus the estimate equates to about a 35% increase. For the 'worker' group we estimate a 10 hours reduction in hours worked (around 135 hours annualized).[68] This group had mean hours worked per month of 120 hours pre-random assignment; thus, the estimate implies a reduction of about 10%. The results in columns (3) and (4) suggest that the 5-year program is driving the estimates from the pooled sample for welfare recipients; both estimates in column (3) are just outside conventional significance levels.

To gauge whether the results in Table 11 appear consistent with the theory discussed in Section 2, we split the treatment group into a 'higher worker incentives' group –which equals one for individuals assigned to one of the lower NIT tax rates (50% and 70% variable with income[69])—and a 'lower work incentives group' (equals one for the higher NIT tax rates of 70% flat and 80% variable with income[70]). We take this parsimonious approach as the sample sizes in each treatment plan cell (a total of 11 plans) are too small to yield enough power to conduct hypothesis tests for equality of coefficients within groups. The number of observations is not evenly distributed across plans. We compute the weighted average NIT tax rate; the higher work incentive groups face a NIT tax rate of 54% while the NIT tax rate of the lower incentives group is 70%. Looking at columns 2, 5 and 6, the results are consistent with the theoretical predictions: individuals assigned to a plan with tax rates below the federal welfare clawback are driving the positive estimate on hours worked. For comparison purposes, the weighted average NIT tax rate in Mincome is 57%.

As a final comment, we caution readers to not place much emphasis on the similarity in the magnitudes of the percentage changes in labor supply in Mincome and DIME. The change in

[67] Recall from our previous discussion, combining work and welfare even in the same month was quite common in DIME

[68] Results for the probability of employment have the same pattern (see Appendix Table A6), but the change in behavior appears stronger along the intensive margin.

[69] At the median level of income, the 70% variable tax rate yields a 60% NIT tax rate.

[70] At the median level of income, the 80% variable tax rate yields a 70% NIT tax rate.

work incentives in Manitoba was clearly much greater since the treatment group moved from a 100% clawback to a 57% NIT tax rate (on average). The change in incentives was less extreme in Denver. However, there are numerous differences in the experiments that could plausibly work in opposite directions and may by coincidence have offsetting effects. First, our Mincome results –due to data constraints—come from welfare participants and workers combined, and thus constitute a lower bound of the treatment effect for welfare recipients that is not directly comparable to the labor supply response of welfare recipients in DIME. As well, Mincome single parents had substantially lower pre-random assignment work activity (88% lower hours worked), lower educational attainment (by two years), and more children below age 16. The U.S. earnings disregard for welfare recipients was much larger than that in Manitoba.

5. Conclusions

The consensus view of the NIT experiments for single parents has been that an NIT is associated with a reduction in hours worked. However, while theory predicts an unambiguous decline in hours conditional on working, this is not the case for those receiving welfare. Welfare participation was high for single parents in all of the relevant North American NITs. We re-assess the effects of the NIT by distinguishing between those working on a regular basis prior to random assignment from those principally collecting welfare. Because doing so conditions on a pre-random assignment characteristic, both groups are as good as randomly assigned. Our re-assessment reveals that in DIME—where there were no unusual economic shocks, and experimental evidence appears reliable—the small, negative (but not statistically different from zero) labor supply responses previously reported in the pooled data confounds two opposing effects: a reduction in hours worked of about 10% for 'workers', and a 35% increase in hours for 'welfare recipients'. Our results from the Mincome experiment—where welfare rules, work placement programs, and available child-care programs differed substantially—also yield positive labor supply responses to a NIT. Finally, we find that random assignment in Gary and SIME did not appear to hold. We also show (it appears for the first time) that the consensus negative labour supply response in SIME-DIME is based largely on the three-year SIME program where the treatment group worked less pre-random assignment. Difference-in-difference

estimates reveal no negative labor supply response to the NIT program in SIME while there was never compelling evidence of a negative labor supply response in DIME.

Overall, therefore, we find no compelling evidence from the North American NIT experiments that labor supply was reduced for the single parent group as a whole. On the contrary, our results suggest that a NIT may increase hours worked for single parents on welfare – impacts that imply a guaranteed income can have offsetting positive and negative effects on work activity and improved incentives relative to traditional welfare programs.

References

Abramowicz, M. 2019. "Tax Experimentation." *Florida Law Review*, Volume 71, pages 65-122.

Ashenfelter, O., K. Doran and B. Schaller (2010). "A Shred of Credible Evidence on the Long Run Elasticity of Labor Supply" *Economica*, Volume 77, pages 637-650.

Ashenfelter, O. and M. Plant. 1990. "Nonparametric Estimates of the Labor-Supply Effects of Negative Income Tax Programs" *Journal of Labor Economics*, Volume 8, pages S396-S415.

Barber, C. 1972. Welfare Policy in Manitoba: A Report to the Planning and Priorities Committee of Cabinet Secretariat, Province of Manitoba. Winnipeg: December 1972.

Bedard, J. 1994. Mothers' Experiences of the Social Assistance System. M.A. Thesis, Department of Sociology, University of Manitoba.

Blundell, R. and T. MaCurdy. 1999. "Labor Supply: A Review of Alternative Approaches" In *Handbook of Labor Economics*, Volume 3, pages 1559-1695. Edited by Orley Ashenfelter and David Card. Amsterdam: Elsevier.

Burtless, G. 1987. "The Work Response to a Guaranteed Income: A Survey of the Experimental Evidence" in Alicia H. Mundell (ed.) *Lessons from the Income Maintenance Experiments*. Boston: Federal Reserve Bank of Boston Conference Series No. 30, 22-52.

Calnitsky, D. and J.P. Latner. 2017. "Basic Income in a Small Town" *Social Problems*.

Cesarini, D., E. Lindqvist, M. Notowidigdo and R. Ostling. 2017. "The Effect of Wealth on Individual and Household Labor Supply: Evidence from Swedish lotteries" *American Economic Review*, Volume 107, pages 3917-46.

Fields, S. 2002. Mothers, Lone Parents and Welfare Reform: Winnipeg in the 1960s. M.A. Thesis, Department of History, University of Manitoba and University of Winnipeg.

Fitzgerald, J., Gottschalk, P., and R. Moffitt. 1998. "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics." *Journal of Human Resources*, Volume 33, pages 300-344.

Forget, E. 2018. *Basic Income for Canadians*. Toronto: James Lorimer.

Friedman, M. 1962. *Capitalism and Freedom*. Chicago: University of Chicago Press.

Gibson, M., Hearty, W. and P. Craig. 2018. "Universal Basic Income: A Scoping Review of Evidence on Impacts and Study Characteristics." What Works Scotland: Evidence Review.

Greenberg, D. and H. Halsey. 1983. “Systematic Misreporting and Effects of Income Maintenance Experiments on Work Effort: Evidence form the Seattle-Denver Experiment” *Journal of Labor Economics*, Volume 1, pages 380-407.

Greenberg, D., R. Moffitt and J. Friedmann. 1981. “Underreporting and Experimental Effects on Work Effort: Evidence from the Gary Income Maintenance Experiment” *Review of Economics and Statistics*, Volume 63, pages 581-89.

Haagh, L. 2019. *The Case for Universal Basic Income*. Wiley.

Hoynes, H. and J. Rothstein. 2019. “Universal Basic Income in the U.S. and advanced Countries” NBER Working Paper #25538.

Hum, D. and W. Simpson. 1991. *Income Maintenance, Work Effort and the Canadian Mincome Experiment*. Ottawa: Economic Council of Canada.

Hum, D. and W. Simpson. 1993. “Economic Response to a Guaranteed Annual Income: Experience from Canada and the United States” *Journal of Labor Economics*, Volume 11 , pages S263-S296.

Hutchens, R. 1978. ”Changes in AFDC tax rates 1967-1971” *Journal of Human Resources*, Volume 13, pages 60-74.

Kearney, M. and M. Mogstad. 2019. “Universal Basic Income as a Policy Response to Current Challenges.” The Aspen Group.

Keeley, M. and P. Robins. 1980. “Experimental Design, the Conlisk-Watts Assignment Model and the Proper Estimation of Behavioral Response” *Journal of Human Resources*, Volume 13, pages 3-36.

Keeley, M. 1981. *Labor Supply and Public Policy: A Critical View*. New York: Academic Press.

Levine, R., Watts, H., Hollister, R., Williams, A. and K. Widerquist. 2005. “A Restrospective on the Negative Income Tax Experiments: Looking Back at the Most Innovative Field Studies in Social Policy.” In: The Ethics and Economics of the Basic Income Guarantee. Routledge.

Lowrey, A. (2018). *Give People Money: How A Universal Basic Income Would End Poverty, Revolutionize Work, and Remake the World.* Crown Publishing.

Manitoba Department of Health and Social Development. *Annual Report* 1971 to 1978.

Marinescu, I. 2018. “No Strings Attached: The Behavioral Effects of U.S. Unconditional Cash Transfer Programs” NBER Working Paper #24337.

Mason, G. 2016. Mincome User Manual Dataverse. At: http:dataverse.lib.umanitoba.ca/dataset.xhtml?persistentId=doi:10.5203/FK2/XLOXQF.

McFarland, K. 2017. “Overview of Current Basic Income Related Experiments.” Basic Income Earth Network.

Moffitt, R. 1992. “Incentive Effects of the U.S. Welfare System: A Review” *Journal of Economic Literature*, Volume 30, pages 1-61.

Moffitt, R. 2003. “The Negative Income Tax and the Evolution of U.S. Welfare Policy” *Journal of Economic Perspectives*, Volume 17, pages 119-140.

Mundell, A. 1987. *Lessons from the Income Maintenance Experiments*. Boston: Federal Reserve Bank of Boston Conference Series No. 30

Murarka, B. and R. Spiegelman. 1978. “Sample Selection in the Seattle and Denver Income Maintenance Experiment” SRI Technical Memorandum No. 1. Menlo Park: SRI International.

Murray C. 2016. *In Our Hands: A Plan to Replace the Welfare State*. American Enterprise Institute.

New York Times. 1970. “Sharp Slump at Boeing Plant Puts Seattle Economy in Tailspin,” March 3rd.

New York Times. 1971a. “Layoffs Spread at Steel Mills,” August 7th.

New York Times. 1971b. “The Talk of Gary, Indiana.” December 14th.

Pasma C. 2010. “Working Through the Work Disincentive.” Citizens for Public Justice.

Paz-Banez, M, Asento-Coto, M., and C. Sanchez-Lopez. 2020. “Is there Empirical Evidence on How the Implementation of a Universal Basic Income Affects Labour Supply? A Systematic Review.” *Sustainability*, Volume 12, pages 1-36.

Price, D. and J. Song. 2018. “The Long-Term Effects of Cash Assistance.” Working Paper #621, Princeton University, Industrial Relations Section.

Robins, P. 1985. “A Comparison of the Labor Supply Findings of the Four Negative Income Tax Experiments” *Journal of Human Resources*, Volume 20, pages 567-82.

Seattle Times. 1996. “Lights out, Seattle,” November 3rd.

Segal, H. 2019. *Bootstraps Need Boots*. Vancouver: UBC Press.

Simpson, W., Mason, G., and R. Godwin. 2017. “The Manitoba Basic Annual Income Experiment: Lessons Learned 40 Years Later” *Canadian Public Policy*, Volume 43, pages 85-104.

Spiegelman, R. and K. Yaeger. 1980. “Overview.” *Journal of Human Resources*, Volume 15, pages 463-479.

SRI International. 1983. *Final Report of the Seattle-Denver Income Maintenance Experiment*. Menlo Park, Ca.: SRI International.

Stanford Basic Income Lab. 2018. https://basicincome.stanford.edu/

Sterling, R. 1992. *Legend and Legacy: The Story of Boeing and its People*. New York: St. Marten’s Press.

Terwitte, J. 2009. “Should Proponents of Basic Income Advocate Basic Income Social Experiments in Germany?” Working Paper, Hertie School of Governance.

Van Parijs, P. and Y. Vanderborght. 2017. *Basic Income: A Radical Proposal for a Free Society and a Sane Economy*. Cambridge, Mass.: Harvard University Press.

Widerquist, K. 2005. “A failure to communicate: what (if anything) can we learn from the negative income tax experiments?” *The Journal of Socio-Economics*, Volume 34, pages 49-81.

Wiederspan J., Rhodes, E. and H. Shaefer. 2015. “Expanding the Discourse on Antipoverty Policy: Reconsidering a Negative Income Tax. *Journal of Poverty*, Volume 19, pages 218-238.

Yang A. 2018. *The War on Normal People: The Truth About America’s Disappearing Jobs and Why Universal Basic Income Is Our Future*. Hachette Books.

Figure 1

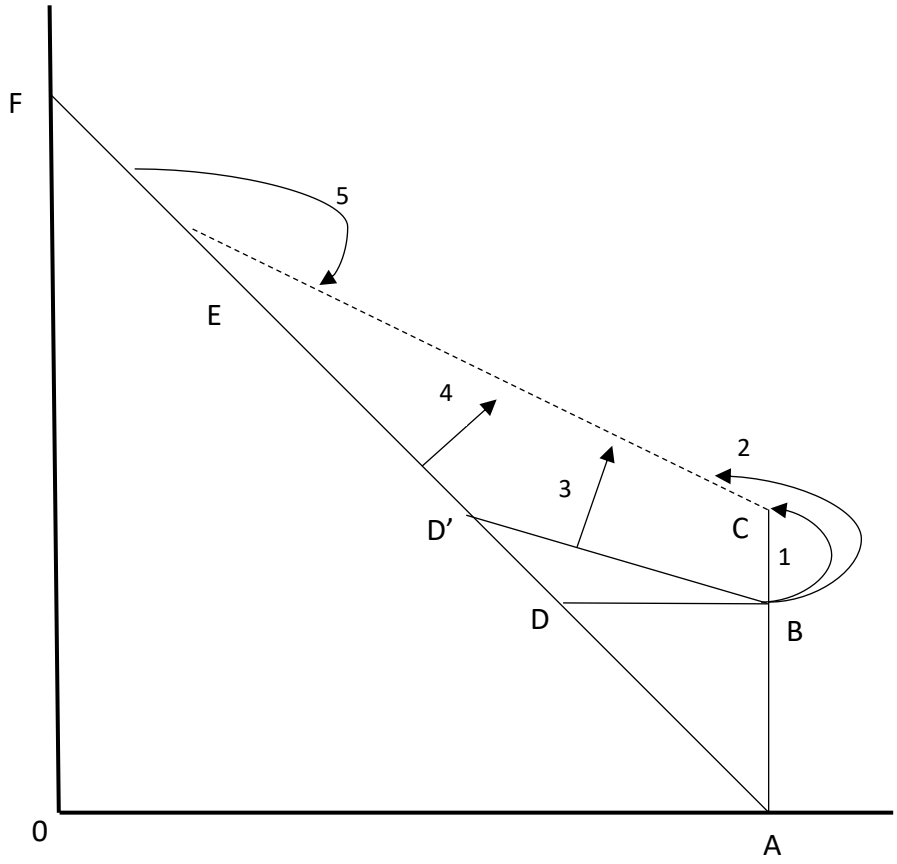
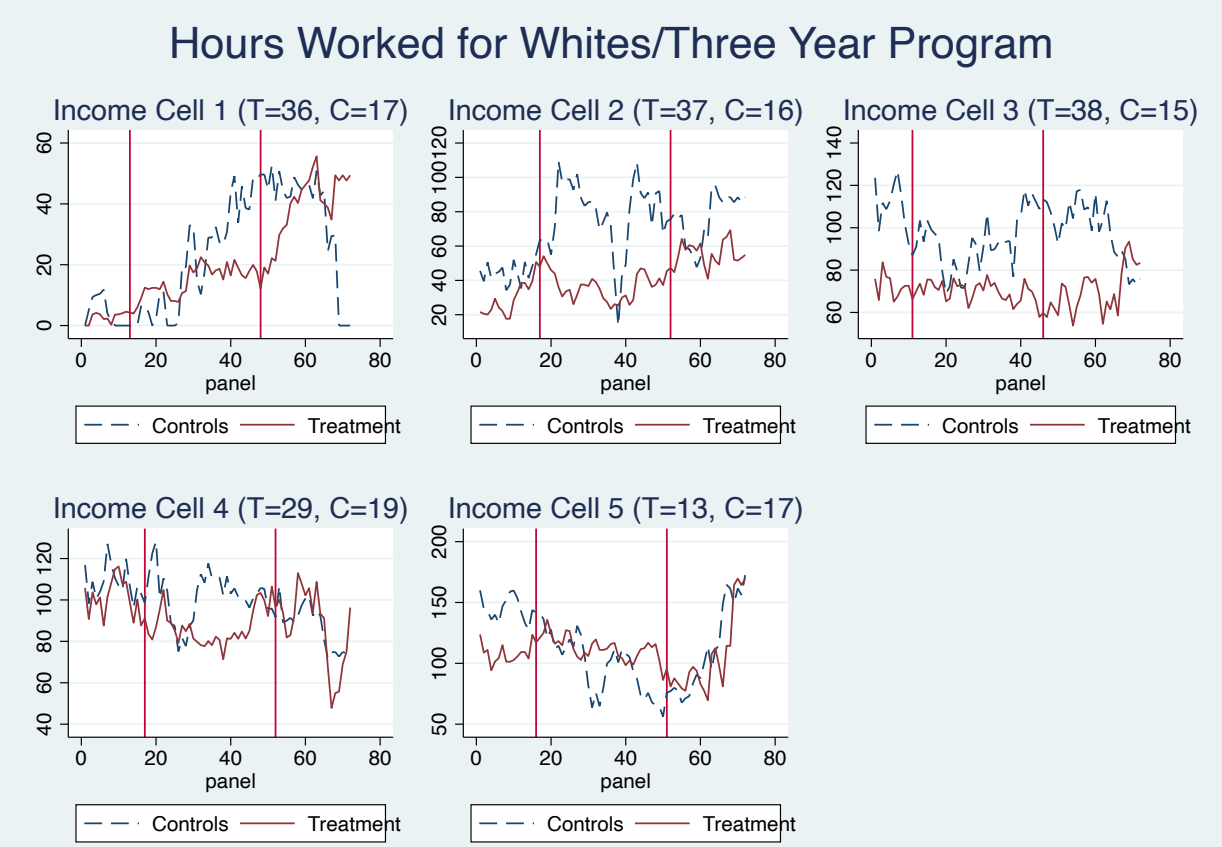
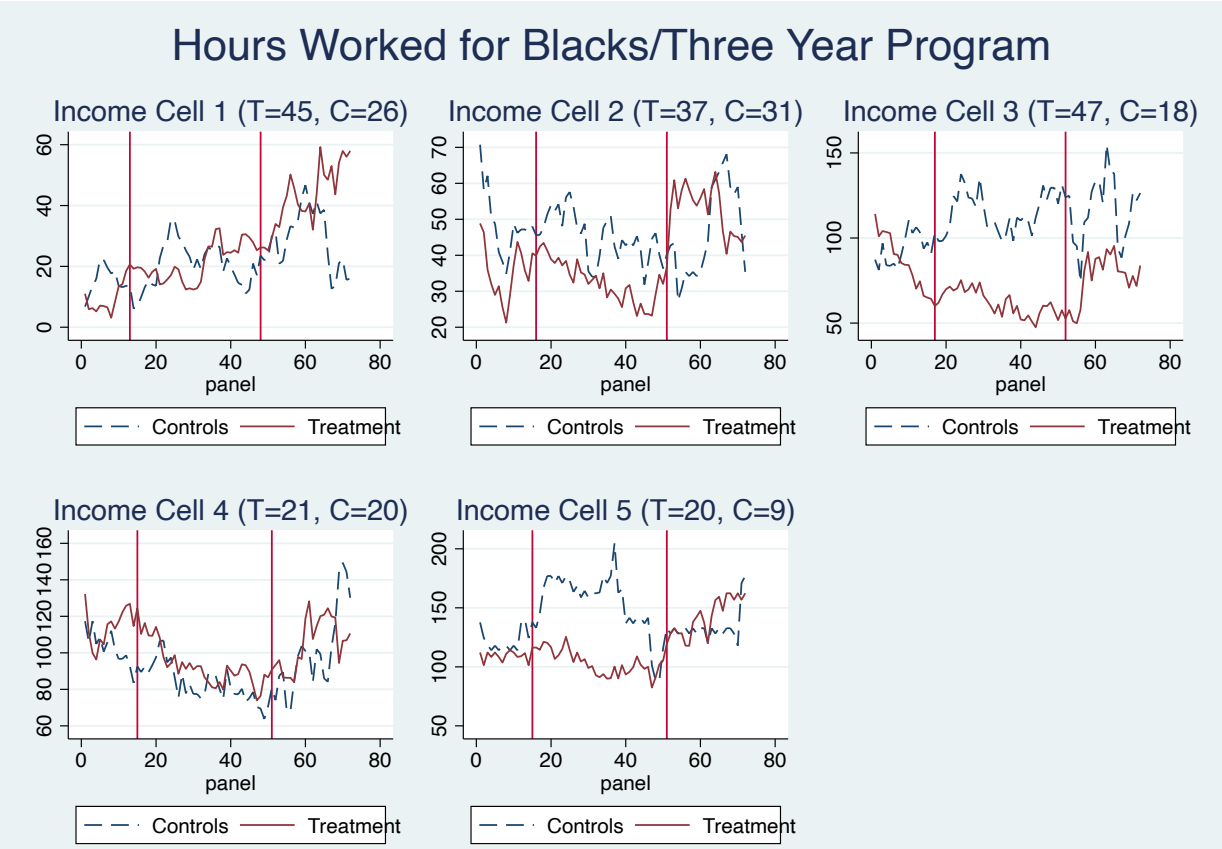


Figure 2

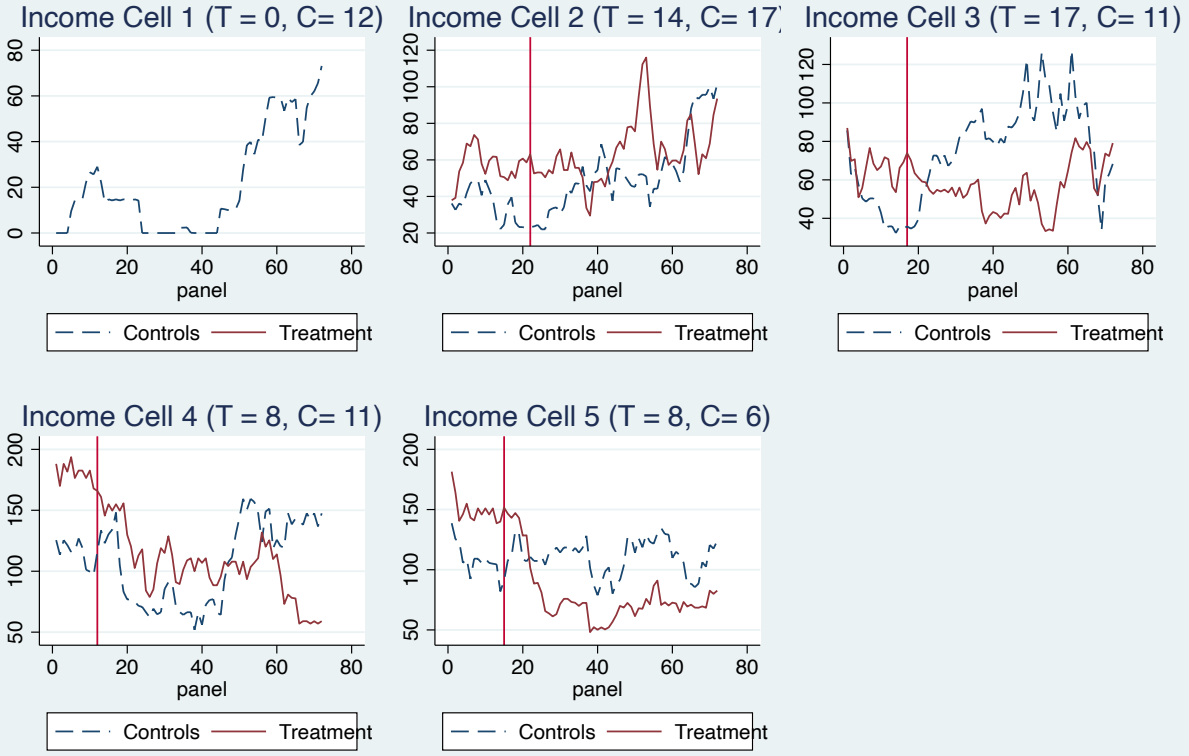

Welfare and Take-Up Rates in Gary
0
.2
.4
.6
.8
1
0
10
20
30
40
50
time
Welfare_Treatments
Takeup
Welfare and Take-Up Rates in SIME
0
.2
.4
.6
0
20
40
60
80
panel
Welfare_Treatments
Takeup
Welfare and Take-Up Rates in DIME
0
.2
.4
.6
0
20
40
60
80
panel
Welfare_Treatments
Takeup

Figure 3: Hours Worked in SIME





Hours Worked for Whites/Five Year Program



Hours Worked for Blacks/Five Year Program

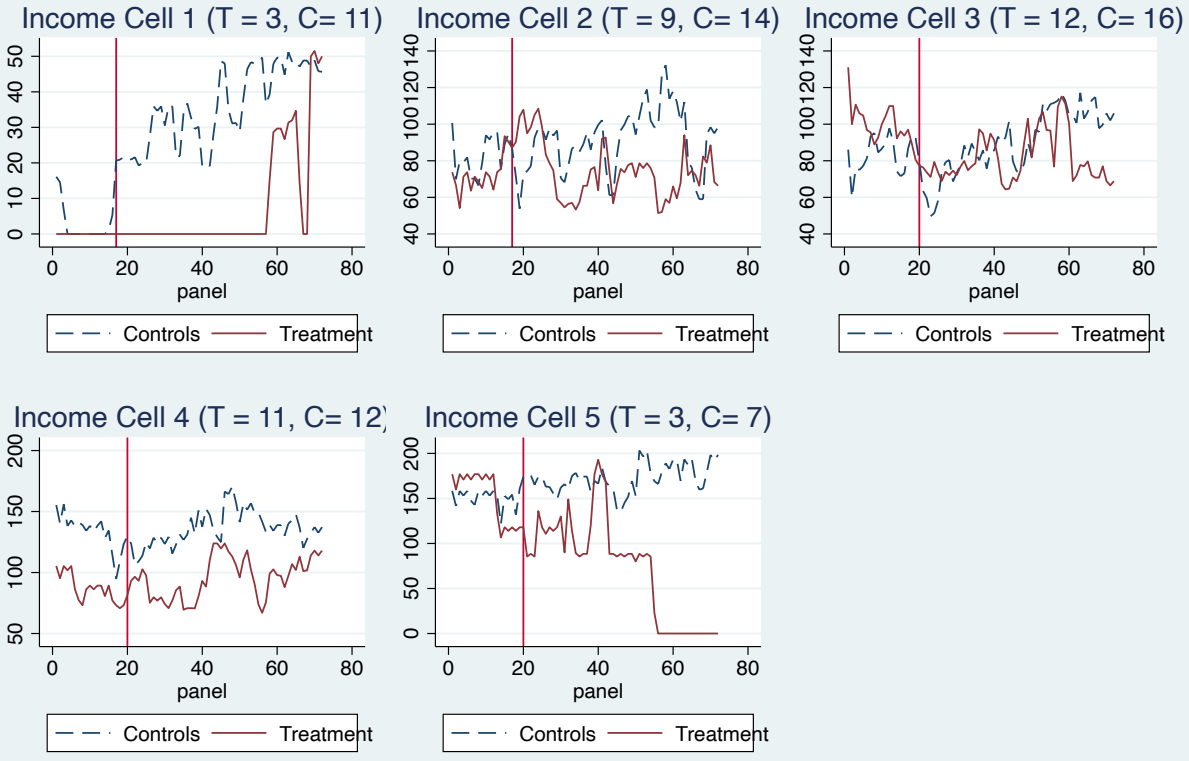


Table 1
Characteristics of Income Maintenance Experiments with Single Parents

Experiments	Dates of Operation	Population Enrolled	Experimental Stratifications	Number of Families Enrolled	NIT Treatment Plans	
					Guarantee	Tax Rate
Seattle	1971 – 1977	White and Black Two Parent and Single Parent Families	Race (2); Duration (2); Family Type (2); Income Level (5)	2042	0.90 1.20 1.30	0.50, 0.70*, 0.80* 0.50, 0.70*, 0.80* 0.50, 0.70*, 0.80*
Denver	1972 – 1978	White, Black, and Hispanic Two Parent and Single Parent Families	Race (3), Duration (3), Family Type (2); Income Level (5)	2758	0.90 1.20 1.30	0.50, 0.70*, 0.80* 0.50, 0.70*, 0.80* 0.50, 0.70*, 0.80*
Gary	1971 – 1974	Black Two Parent and Single Parent Families	Family Type (2); Income Level (5); Location (2)	1800	0.75 1.00	0.40, 0.60 0.40, 0.60
Mincome-Winnipeg	1974 – 1978	Two Parent and Single Parent Families and Single Men and Women	Income Level (4)	1274	0.50 0.60 0.70	0.35, 0.50 0.35, 0.50, 0.75 0.50, 0.75

NOTES: Durations were 3 and 5-year programs in Seattle, and 3, 5, and 20-year programs in Denver. Location was urban (inside) and rural (outside) in Gary. The guarantee is expressed as a fraction of the poverty line. *Seattle-Denver also tested tax rate that declined with annual income; this include all of the 80% plans and half of the 70% plans.

Table 2
Consensus Estimates of North American NIT Treatment Effects for Single Parents

	Gary	
	Annual Hours	**Employment Rate**
Robins (1985)	-37 [10%] (68)	-.04 (.03)
Burtless (1986)	-112 [30%]	
	Seattle-Denver	
	Annual Hours	**Employment Rate**
Robins (1985)		
All Programs, All Races	-155 [16%] (37)	-.08 (.02)
3-Year, All Races	-163 [16%] (39)	-.08 (.02)
Blacks, All Programs	-206 (52)	-.10 (.02)
Blacks, 3-Year	-232 (55)	-.11 (.03)
Whites, All Programs	-60 (61)	-.03 (.03)
Whites, 3-Year	-53 (65)	-.04 (.03)
Hispanics, All Programs	-207 (88)	-.10 (.04)
Hispanics, 3-Year	-189 (95)	-.09 (.05)
Burtless (1986)		
All Programs, All Races	-85 [9%]	
	Mincome*	
	Annual Hours	Employment rate
Hum and Simpson (1993)	-79 [7%]	N/A

NOTES: Standard errors are in the parentheses. Percentage changes are in [].

* Hum and Simpson pooled single mothers and single adult women (i.e., with no children). No evidence is available for single mothers as a separate group.

Table 3
Summary Statistics: Pre-Random Assignment

	SIME	DIME	Gary	MINCOME
Treatment Group	.572 (.019) [713]	.703 (.016) [837]	.577 (.016) [913]	.690 (.039) [136]
Employed at least One Month	.626 (.018) [713]	.708 (.016) [836]	.204 (.013) [884]	.661 (.041) [136]
Average Monthly Hours Worked	68.94 (2.72) [713]	87.66 (2.61) [836]	6.73 (.487) [884]	51.86 (3.93) [136]
Average Monthly Earnings	163.06 (7.24) [711]	200.86 (6.77) [835]	71.93 (4.86) [913]	-
Received Welfare at least One Month	.603 (.018) [713]	.587 (.017) [836]	.797 (.014) [853]	.429 (.043) [136]
Received Welfare Every Month	.377 (.018) [713]	.410 (.017) [836]	.773 (.014) [853]	-
Years of Schooling	11.16 (.079) [575]	11.20 (.071) [770]	9.92 (.089) [845]	9.72 (.207) [130]
Age	34.16 (.355) [713]	32.27 (.320) [836]	36.42 (.409) [913]	35.82 (.962) [136]
Number Children Under 16	2.25 (.049) [713]	2.16 (.044) [836]	2.33 (.062) [913]	2.85 (.155) [135]
Child Under 6	.560 (.019) [713]	.544 (.017) [836]	.315 (.015) [913]	.352 (.043) [122]
Normal Income Category 1	.210 (.015) [713]	.166 (.013) [837]	.390 (.016) [913]	.015 (.010) [136]
Normal Income Category 2	.245 (.016) [713]	.266 (.015) [837]	.258 (.014) [913]	.151 (.030) [136]
Normal Income Category 3	.244 (.016) [713]	.269 (.015) [837]	.181 (.013) [913]	.356 (.041) [136]
Normal Income Category 4	.184 (.015) [713]	.190 (.01) [837]	.155 (.012) [913]	.477 (.043) [136]
Normal Income Category 5	.116 (.012) [713]	.109 (.011) [837]	.016 (.004) [913]	-
Sample size	713	837	913	136

NOTES: Standard errors are in parentheses. Number of observations for item responses are in [].

Table 4
Balancing Tests: Labor Market Outcomes Pre-Random Assignment, Mincome

	Mean	**Estimated Coefficient on Treatment Group Dummy**	
		Including Controls for Income Cell	Excluding Controls for Income Cell
Received welfare during 1973 (MINC 1)	.390 (.042)	-.054 (.085)	-.105 (.091)
Received welfare during 1974 (MINC 1)	.375 (.042)	-.063 (.085)	-.092 (.091)
Worked positive weeks during 1973 (MINC 1)	.676 (.040)	.065 (.074)	.140* (.087)
Worked positive weeks during 1974 (MINC 1)	.661 (.040)	.060 (.073)	.180** (.087)
Hours worked, Periodic 1, Approximately 1-year period (MINC 4)	621.48 (48.19)	59.56 (87.73)	131.69 (105.63)
Hours worked, Periodic 2, Approximately 4 months period (MINC 4)	147.43 (16.38)	24.61 (34.52)	37.70 (35.69)
Employed, Positive Hours, Periodic 1 (MINC 4)	.654 (.041)	.058 (.071)	.179** (.088)
Employed, Positive Hours, Periodic 2 (MINC 4)	.632 (.041)	.067 (.078)	.145* (.089)

NOTES: Each cell in column three gives the estimate coefficient on a treatment group dummy from a separate regression. All regressions include dummies for the normal income category. Standard errors are in parentheses. The sample size is 136 observations in all cases.

Table 5
Balancing Tests for SIME on Labor Market Outcomes Pre-Random Assignment

	Three Year				Five Year			
	Employed	**Hours Worked**	**Total Wages**	**Received Welfare**	**Employed**	**Hours Worked**	**Total Wages**	**Received Welfare**
Month 1	-.065*	-9.04	-40.40**	.077**	.025	6.46	16.78	.046
	(.039)	(6.48)	(18.34)	(.038)	(.062)	(11.00)	(27.61)	(.063)
Month 2	-.050	-8.27	-34.24**	.046	.065	10.28	28.30	.005
	(.039)	(5.83)	(16.47)	(.039)	(.063)	(9.76)	(24.79)	(.064)
Month 3	-.065*	-13.25**	-46.12***	.065*	.031	4.43	14.10	.030
	(.040)	(6.53)	(18.72)	(.038)	(.063)	(10.85)	(26.96)	(.064)
Month 4	-.080**	-10.98*	-38.50**	.072*	.066	9.21	25.30	.059
	(.040)	(6.29)	(18.12)	(.039)	(.064)	(10.47)	(25.83)	(.065)
Month 5	-.054	-9.71	-34.61*	.065*	.066	10.78	35.32	.059
	(.040)	(6.54)	(18.81)	(.039)	(.065)	(11.10)	(26.82)	(.065)
Month 6	-.103***	-14.22**	-43.91**	.064*	.031	7.69	23.88	.042
	(.040)	(6.31)	(19.04)	(.038)	(.066)	(10.82)	(27.17)	(.066)
Month 7	-.082**	-14.76**	-51.29***	.050	.012	6.63	31.23	.037
	(.039)	(6.39)	(17.90)	(.038)	(.066)	(11.47)	(29.10)	(.067)
Month 8	-.094***	-13.41**	-45.77***	.042	-.010	4.01	24.88	.013
	(.038)	(6.17)	(17.40)	(.039)	(.066)	(11.45)	(29.01)	(.067)
Month 9	-.071*	-12.19**	-44.91***	.027	.016	5.90	26.26	.017
	(.038)	(6.03)	(17.36)	(.038)	(.067)	(11.03)	(27.10)	(.066)

NOTES: Each cell gives the estimated coefficient on a dummy variable =1 for the treatment group, =0 for the control group. All regressions include dummies for income category, and race. Standard errors are in parentheses. Sample sizes are 202 for the 5-year program, and 511 for the 3-year program.

Table 6
Balancing Tests for DIME on Labor Market Outcomes Pre-Random Assignment

	Three Year				Five Year			
	Employed	**Hours Worked**	**Total Wages**	**Received Welfare**	**Employed**	**Hours Worked**	**Total Wages**	**Received Welfare**
Month 1	-.007 (.041)	-2.44 (7.15)	-9.41 (16.77)	.065 (.044)	.002 (.058)	1.48 (9.88)	-2.40 (22.56)	-.070 (.061)
Month 2	.002 (.042)	-3.22 (6.32)	-6.92 (14.13)	.061 (.044)	.008 (.058)	0.99 (8.93)	-7.96 (20.09)	-.059 (.061)
Month 3	-.007 (.042)	-3.07 (7.05)	-4.55 (16.31)	.055 (.044)	.041 (.058)	2.72 (9.83)	-8.07 (22.45)	-.059 (.061)
Month 4	.004 (.042)	-1.63 (6.78)	-2.35 (15.22)	.055 (.044)	.034 (.058)	4.77 (9.44)	-3.28 (21.38)	-.072 (.061)
Month 5	.005 (.042)	-1.33 (6.98)	-1.99 (16.21)	.050 (.043)	.081 (.059)	11.03 (9.84)	11.61 (23.04)	-.084 (.061)
Month 6	.022 (.042)	3.37 (6.74)	8.91 (16.88)	.058 (.043)	.044 (.058)	6.95 (9.61)	0.26 (23.51)	-.053 (.061)
Month 7	.042 (.042)	1.94 (7.07)	5.43 (17.26)	.036 (.043)	.050 (.057)	6.29 (9.86)	-0.99 (23.46)	-.040 (.061)
Month 8	.032 (.042)	1.03 (7.03)	2.43 (17.76)	.041 (.043)	.042 (.057)	6.32 (9.75)	-0.39 (23.44)	-.017 (.060)
Month 9	.012 (.042)	-0.53 (7.03)	-1.81 (16.13)	.048 (.042)	.022 (.059)	6.08 (9.78)	0.64 (24.24)	-.012 (.060)

NOTES: Each cell gives the estimated coefficient on a dummy variable =1 for the treatment group, =0 for the control group. All regressions include dummies for income category, and race. Standard errors are in parentheses. Sample sizes for wage regressions are 553 for 3-year, and 283 for the 5-year program.

Table 7
Balancing Tests for Gary on Labor Market Outcomes Pre-Random Assignment

	Employed	Hours Worked	Monthly Wages	Received Welfare	Number Individuals
Month 1	-.021 (.026)	-1.19 (.981)	-9.37 (11.51)	-.120*** (.028)	880
Month 2	-.035 (.031)	1.98* (1.15)	-12.84 (14.43)	-.114*** (.031)	641
Month 3	-.036 (.032)	-2.08* (1.19)	-11.19 (14.31)	-.100*** (.032)	582
Month 4	-.009 (.046)	-1.18 (1.73)	-3.16 (21.55)	-.130*** (.045)	324
Mean Months 1 – 4	.201	6.83	80.09	.78	

NOTES: Each cell gives the estimated coefficient on a dummy variable =1 for the treatment group, =0 for the control group. All regressions include dummies for income category, and location. Standard errors are in parentheses. Sample sizes are based on the employment column; sample sizes are smaller for welfare receipt in Gary due to missing data. Note, 65 single parent households in Gary have zero pre-random assignment observations.

Table 8
Treatment Effects in Mincome

	Periodic Surveys (MINC4)		Payments Administrative File (MINC2)
	Employed	**Hours**	**Employed**
Treatment Group	.139** (.068)	72.43* (39.53)	.137* (.072)
8 Survey (time) fixed effects	Yes	Yes	No
34 Month fixed effects	No	No	Yes
Sample size	1018	1018	3484
Number of individuals	126	126	127

NOTES: Standard errors are in parentheses and are clustered on the individual. All regressions include dummies for the normal income category and household demographics. For the periodic surveys, employed=1 if positive hours, =0 if no hours worked. For the payments file, employed=1 if wages are positive, =0 for no wages in a given month.

Table 9
Treatment Effects in the U.S. NITs

	SIME (pooled)	**SIME 3 year**	**SIME 5 year**	**DIME (Pooled)**	**DIME (3 Year)**	**DIME (5 Year)**	**Gary**
Employed	-.108*** (.030)	-.122*** (.034)	-.089 (.058)	-.032 (.027)	-.029 (.033)	-.030 (.047)	-.034 (.022)
Hours Worked	-16.77*** (4.90)	-19.68*** (5.66)	-.12.10 (9.78)	-5.47 (4.60)	-4.66 (5.62)	-5.44 (8.01)	-1.29 (.834)
Sample Size	24 271	16 207	8 064	31 189	17 173	14 016	31962
Number Individuals	699	500	199	821	538	283	945

NOTES: All regressions include dummies for race (where appropriate), income category, program length (where appropriate), panel month, and household demographics. Gary includes a dummy for inside location. Standard errors are clustered on the individual.

Table 10
Treatment Effects in U.S. NITs by Experimental Cell

	SIME				**DIME**					
	Three Year		**Five Year**		**Three Year**			**Five Year**		
	Whites	**Blacks**	**Whites**	**Blacks**	**Whites**	**Blacks**	**Hispanics**	**Whites**	**Blacks**	**Hispanics**
Employed	-.104** (.051)	-.144*** (.045)	.017 (.089)	-.187** (.080)	-.096* (.052)	.010 (.051)	-.020 (.066)	-.031 (.081)	.033 (.073)	-.076 (.097)
Hours Worked	-16.78** (8.59)	-23.53*** (7.54)	5.61 (14.93)	-27.36** (13.33)	-14.80 (9.55)	2.73 (8.61)	-5.87 (10.97)	-13.07 (14.42)	2.94 (12.42)	-0.74 (14.92)
Individuals	229	271	104	95	139	242	157	86	120	77
Sample Size	7292	8915	4304	3760	4509	7630	5034	4159	6013	3844

NOTES: All regressions include controls for income category, panel month, and household demographics. Standard errors are in parentheses and clustered on the individual.

Table 11
Difference-in-Differences Estimates for SIME

	Whites, 3 Year, Full Sample		Blacks, 3 Year, Full Sample		Blacks, 3 Year, Excluding Income Cell 3		Blacks + Whites, 3 Year, Full Sample		Blacks + Whites, 3 Year, Excluding Income Cell 3 (Blacks)	
	Employed	**Hours Worked**	**Employed**	**Hours Worked**	**Employed**	**Hours Worked**	**Employed**	**Hours Worked**	**Employed**	**Hours Worked**
Treatment Group	-.011 (.036)	-2.82 (6.29)	-.052** (.024)	-8.73 (5.17)	-.069** (.026)	-11.09* (5.77)	-.040 (.026)	-7.16 (4.76)	-.039 (.027)	-6.84 (5.12)
Post Treat. Period	.090 (.065)	-14.20 (11.15)	.177** (.058)	-29.03** (11.54)	.132* (.061)	23.37 (13.57)	.120*** (.044)	19.44** (7.91)	.091* (.045)	15.84* (8.42)
D-in-D Estimate	-.009 (.051)	-0.861 (9.52)	-.091 (.061)	-15.27 (11.58)	-.022 (.041)	-3.67 (7.73)	-.044 (.041)	-7.66 (7.53)	-.012 (.035)	-1.98 (6.79)
Sample Size	10 781		12 874		9 708		23 655		20489	
Number Individuals	237		274		209		511		436	

NOTES: All regressions include controls for income cell (i.e., mini-experiment), interactions between income cell and post-treatment dummy, panel month, calendar month and calendar year dummies, and demographics. Regressions that pool races also include a dummy for race and interaction term between race and the post-treatment dummy. Standard errors are in parentheses and clustered on the income cell * post treatment period (and income cell * post treatment period * Race for the pooled Blacks + Whites sample).

Table 12
Treatment Effects for Hours Worked in DIME: Workers vs. Welfare Recipients

	(1) Full Sample		(2) Full Sample		(3) Three Year		(4) Five Year		(5) Three Year		(6) Five Year	
	Workers	**Welfare Recipients**	**Workers**	**Welfare Recipients**	**Workers**	**Welfare Recipients**	**Workers**	**Welfare Recipients**	**Workers**	**Welfare Recipients**	**Workers**	**Welfare Recipients**
Treatment Group	-10.19* (6.00)	13.67** (6.41)	-	-	-10.77 (7.16)	12.74 (8.62)	-8.53 (11.84)	23.93** (11.86)	-	-	-	-
Treatment: Higher Incentives	-	-	-13.23* (7.39)	20.45*** (7.88)	-	-	-	-	-19.69** (8.42)	19.23** (9.60)	-3.02 (14.96)	32.22** (14.23)
Treatment: Lower Incentives	-	-	-8.89 (6.63)	4.96 (7.37)	-	-	-	-	-4.43 (8.19)	6.93 (9.34)	-9.12 (12.34)	9.69 (12.45)
Sample Size	18 220	12 910	18 220	12 910	9 661	7 453	8, 559	5 457	9 661	7 453	8, 559	5 457
Number Individuals	482	337	482	337	307	229	175	108	307	229	175	108

NOTES: Standard errors are in parentheses and are clustered on the individual. All regressions include dummies for the income category, panel month, race, program length (when appropriate), enrolment cohort, and demographics.

Appendix Figure 1: Hours Worked in Gary

Table A1
Mini-Experiment Characteristics: SIME and DIME

DIME				SIME			
Mini-Experiment	**# Individuals @ month 1**	**# Individuals @ month 50**	**Median Enrollment Interview**	Mini-Experiment	**# Individuals @ month 1**	**# Individuals @ month 50**	**Median Enrollment Interview**
Income1/Whites/3year				*Income1/Whites/3year*			
Controls	10	8	May 1972	Controls	17	12	December 1970
Treatments	16	12	July 1972	Treatments	36	33	January 1971
Income2/Whites/3year				Income2/Whites/3year			
Controls	4	1	January 1972	Controls	16	10	December 1970
Treatments	30	26	April 1972	Treatments	37	32	May 1971
Income3/Whites/3year				Income3/Whites/3year			
Controls	8	5	May 1972	Controls	15	13	December 1970
Treatments	28	22	April 1972	Treatments	38	31	March 1971
Income4/Whites/3year				Income4/Whites/3year			
Controls	8	6	January 1972	Controls	19	14	December 1970
Treatments	21	19	March 1972	Treatments	29	26	April 1971
Income5/Whites/3year				Income5/Whites/3year			
Controls	12	10	January 1972	Controls	17	11	December 1970
Treatments	6	6	December 1971	Treatments	13	12	April 1971
Income1/Hispanics/3year				*Income1/Blacks/3year*			
Controls	17	15	March 1972	Controls	26	25	December 1970
Treatments	18	14	January 1972	Treatments	45	50	January 1971
Income2/*Hispanics*/3year				Income2/ *Blacks* /3year			
Controls	4	3	March 1972	Controls	31	22	December 1970
Treatments	40	32	March 1972	Treatments	37	32	April 1971
Income3/*Hispanics*/3year				Income3/ *Blacks* /3year			
Controls	9	6	March 1972	Controls	18	17	December 1970
Treatments	34	28	December 1971	Treatments	47	42	March 1971
Income4/*Hispanics*/3year				Income4/ *Blacks* /3year			
Controls	14	11	February 1972	Controls	20	15	December 1970
Treatments	18	15	January 1972	Treatments	21	18	March 1971
Income5/*Hispanics*/3year				Income5/ *Blacks* /3year			
Controls	6	3	October 1971	Controls	9	5	December 1970
Treatments	4	2	January 1972	Treatments	20	14	March 1971
Income1/Blacks/3year				*Income1/Whites/5year*			
Controls	13	9	February 1972	Controls	12	10	December 1970
Treatments	23	20	July 1972	Treatments	0	0	-
Income2/ *Blacks* /3year				Income2/Whites/5year			
Controls	15	10	February 1972	Controls	17	17	December 1970
Treatments	53	42	July 1972	Treatments	14	13	October 1971
Income3/ *Blacks* /3year				Income3/Whites/5year			
Controls	16	8	December 1971	Controls	11	11	December 1970
Treatments	58	50	February 1972	Treatments	17	17	May 1971
Income4/ *Blacks* /3year				Income4/Whites/5year			
Controls	14	10	January 1972	Controls	11	8	December 1970
Treatments	27	23	November 1971	Treatments	8	8	December 1970
Income5/ *Blacks* /3year				Income5/Whites/5year			
Controls	12	8	February 1972	Controls	6	6	December 1970
Treatments	15	12	January 1972	Treatments	8	7	March 1971
Income1/Whites/5year				*Income1/Blacks/5year*			
Controls	9	8	March 1972	Controls	11	11	December 1970
Treatments	9	9	March 1972	Treatments	3	3	May 1971
Income2/Whites/5year				Income2/ *Blacks* /5year			
Controls	3	2	January 1972	Controls	14	10	December 1970
Treatments	17	13	July 1972	Treatments	9	9	September 1971

Income3/Whites/5year				Income3/ *Blacks* /5year			
Controls	7	6	March 1972	Controls	16	15	December 1970
Treatments	14	10	March 1972	Treatments	12	11	August 1971
Income4/Whites/5year				Income4/ *Blacks* /5year			
Controls	9	9	December 1971	Controls	12	12	December 1970
Treatments	13	12	May 1972	Treatments	11	10	August 1971
Income5/Whites/5year				Income5/ *Blacks* /5year			
Controls	11	9	December 1971	Controls	7	6	December 1970
Treatments	4	4	March 1972	Treatments	3	2	August 1971
Income1/Hispanics/5year							
Controls	2	2	April 1972				
Treatments	12	11	July 1972				
Income2/*Hispanics*/5year							
Controls	9	6	August 1972				
Treatments	19	18	January 1972				
Income3/*Hispanics*/5year							
Controls	5	5	February 1972				
Treatments	14	11	November 1971				
Income4/*Hispanics*/5year							
Controls	6	5	January 1972				
Treatments	9	7	January 1972				
Income5/*Hispanics*/5year							
Controls	5	4	December 1971				
Treatments	4	3	July 1972				
Income1/Blacks/5year							
Controls	7	6	August 1972				
Treatments	8	6	February 1972				
Income2/ *Blacks* /5year							
Controls	11	11	February 1972				
Treatments	22	18	February 1972				
Income3/ *Blacks* /5year							
Controls	8	7	May 1972				
Treatments	29	26	January 1972				
Income4/ *Blacks* /5year							
Controls	12	11	December 1971				
Treatments	16	13	July 1972				
Income5/ *Blacks* /5year							
Controls	15	13	January 1972				
Treatments	6	6	November 1971				

Table A2
Experimental Periods in US NITs by Cohort

SIME			DIME			Gary	
Cohort	**3 Year**	**5 Year**	**Cohort**	**3 Year**	**5 Year**	**Cohort**	**3 Year**
October 1970	Months 10 – 45	-	November 1971	Months 11 – 46	Months 11 – 72	January 1971	Months 1-36
November 1970	Months 11 – 46	Months 11 – 72	December 1971	Months 12 – 47	Months 12 – 72	February 1971	Months 2-37
December 1970	Months 12 – 47	Months 12 – 72	January 1972	Months 13 – 48	Months 13 – 72	March 1971	Months 3-38
January 1971	Months 13 – 48	Months 13 – 72	February 1972	Months 14 – 49	Months 14 – 72	April 1971	Months 4-39
February 1971	Months 14 – 49	-	March 1972	Months 15 – 50	Months 15 – 72	May 1971	Months 5-40
March 1971	Months 15 – 50	Months 15 – 72	April 1972	Months 16 – 51	Months 16 – 72	June 1971	Months 6-41
April 1971	Months 16 – 51	Months 16 – 72	May 1972	Months 17 – 52	Months 17 – 72	July 1971	Months 7-42
May 1971	Months 17 – 52	Months 17 – 72	June 1972	Months 18 – 53	Months 18 – 72	August 1971	Months 8-43
June 1971	Months 18 – 53	Months 18 – 72	July 1972	Months 19 – 54	Months 19 – 72	September 1971	Months 9-44
July 1971	Months 19 – 54	-	August 1972	Months 20 – 55	Months 20 – 72	October 1971	Months 10-45
August 1971	Months 20 – 55	Months 20 – 72	September 1972	Months 21 – 56	-	November 1971	Months 11-46
September 1971	Months 21 – 56	Months 21 – 72	-	-	-	December 1971	Months 12-47
October 1971	Months 22 – 57	Months 22 – 72	-	-	-	January 1972	Months 13-48

Table A3
Balancing Tests: Non-Labor Market Outcomes

	Mincome	**SIME**	**DIME**	**Gary**
Age	-2.05 (1.85)	-1.82** (.807)	-.134 (.771)	1.95** (.868)
Years Schooling	-.389 (.457)	-.097 (.167)	.009 (.172)	-.153 (.188)
Number Children Under 16	-.127 (.310)	-.047 (.109)	.025 (.100)	-.034 (.131)
Dummy for Child Under 6	.104 (.096)	.049 (.041)	.008 (.041)	-.048 (.032)

NOTES: Each cell gives the estimate coefficient on a treatment group dummy from a separate regression. Standard errors are in parentheses.

Table A4
Tests for Attrition Bias: Hours Worked Pre-Random Assignment

	Mincome	DIME	SIME	Gary
Attritor	11.40 (33.99)	-1.25 (5.21)	-3.26 (5.67)	.721 (.961)
N	136	837	713	913
Attrition Rate	20%	3 Year@ mo. 50: 15% 3 Year @ mo. 72: 63% 5 Year: 18%	3 Year@ mo. 50: 14% 3 Year @ mo. 72: 64% 5 Year: 14%	@ mo. 40: 14% @ mo. 48: 28%

NOTES: Standard errors are in parentheses. Results are based on regressions of a dummy=1 for participants who would drop-out on hours worked pre-random assignment. All regressions include controls for random assignment, experimental stratification group (income category, program length where appropriate, location where appropriate, race where appropriate), and household demographics.

Table A5
Treatment Effects in U.S. NITs by Experimental Cell

	SIME				DIME					
	Three Year		Five Year		Three Year			Five Year		
	Whites	Blacks	Whites	Blacks	Whites	Blacks	Hispanics	Whites	Blacks	Hispanics
Employed	-.104** (.051)	-.144*** (.045)	.017 (.089)	-.187** (.080)	-.096* (.052)	.010 (.051)	-.020 (.066)	-.031 (.081)	.033 (.073)	-.076 (.097)
Hours Worked	-16.78** (8.59)	-23.53*** (7.54)	5.61 (14.93)	-27.36** (13.33)	-14.80 (9.55)	2.73 (8.61)	-5.87 (10.97)	-13.07 (14.42)	2.94 (12.42)	-0.74 (14.92)
Individuals	229	271	104	95	139	242	157	86	120	77
Sample Size	7292	8915	4304	3760	4509	7630	5034	4159	6013	3844

NOTES: All regressions include controls for income category, panel month, and household demographics. Standard errors are in parentheses and clustered on the individual.

Table A6

Treatment Effects for the Probability of Employment in DIME: Workers vs. Welfare Recipients

	(1) Full Sample		(2) Full Sample		(3) Three Year		(4) Five Year		(5) Three Year		(6) Five Year	
	Workers	**Welfare Recipients**	**Workers**	**Welfare Recipients**	**Workers**	**Welfare Recipients**	**Workers**	**Welfare Recipients**	**Workers**	**Welfare Recipients**	**Workers**	**Welfare Recipients**
Treatment Group	-.050 (.034)	.055 (.043)	-	-	-.053 (.041)	.060 (.054)	-.040 (.066)	.100 (.079)	-	-	-	-
Treatment: Higher Incentives	-	-	-.078* (.043)	.104** (.050)	-	-	-	-	-.106** (.050)	.092 (.062)	-.024 (.084)	.185** (.090)
Treatment: Lower Incentives	-	-	-.033 (.037)	.013 (.047)	-	-	-	-	-.014 (.046)	.033 (.059)	-.048 (.068)	.019 (.083)
Sample Size	18 220	12 910	18 220	12 910	9 661	7 453	8, 559	5 457	9 661	7 453	8, 559	5 457
Number Individuals	482	337	482	337	307	229	175	108	307	229	175	108

Appendix: Data and Institutions

A. Welfare in the United States

In the United States, the program in effect at the time of the NITs was the Aid to Families with Dependent Children ("AFDC"). Lurie (1974) provides a detailed review of AFDC from 1935 until 1973. The stated purpose was to provide financial assistance to needy dependent children.[1] The federal program made no provision for assisting a parent or other relative in the household although it did specify that the child must live with a parent or other close relatives to be eligible for federal aid. It was not until 1950 that the federal government began to share in the maintenance costs of a caretaker relative. Congress later allowed states to claim federal reimbursement for assisting other persons under the AFDC program; for instance, the child of an unemployed parent and that parent (AFDC-Unemployed Parent), effective in 1961.[2]

With respect to work incentives, as part of the inclusion of unemployed parents in the program, states were required to deny assistance to families if the unemployed parent refused to accept work without "good cause" beginning in 1961. In 1968 Congress required states to establish a work and training program called Work Incentive (WIN) for "appropriate" AFDC recipients; however, the focus was on unemployed fathers.[3] When setting up WIN to take effect in 1968, Congress offered a financial incentive for AFDC adults to work in the form of a permanent disregard of a portion of earnings. Previously only work expenses were deducted

[1] States were required to establish a standard of need, limitations on the possession of personal or real property, rules for the treatment of any earned or unearned income, and a payment standard. The standard of need was the maximum amount of income allowed for a family to be considered "needy." In recent years the standard of need was almost always considerably higher than the amount actually provided in assistance for any given family size. For example, in July 1994 the average of the states' stated need standards, weighted according the share of the total caseload, was $688 per month. However, the average payment standard was $420 per month.

[2] There were optional provisions that the states could choose to adopt depending on their policy priorities. A state might choose to participate in the Unemployed Parent program for several years and then reverse that decision. For example, in 1978 twenty-eight states participated in the AFDC-UP program and in 1982 that number had dropped to twenty-three states.

[3] The Family Support Act of 1988 replaced WIN with the Job Opportunities and Basic Skills Training program (JOBS) in a new part IV-F of the Social Security Act. It required states, to the extent resources allowed, to engage most mothers with no child below age 3 in education, work, or training under JOBS.

from adult earnings and the remainder was counted against the AFDC check (payment standard) in most states. The new law required states to disregard the first $30 earned and one-third of the remaining monthly earnings; thus, at the time the U.S. NIT experiments began there already existed a NIT for welfare recipients with a 67% federal tax-back rate. However, as noted in the text (section 2) and in Hutchens (1978), states had considerable discretion in determining AFDC benefit levels and there was considerable variation across states and over time in marginal effective tax rates on earned income.[4]

B. Mincome Experiment

Mincome had a budget that was fixed in nominal dollars and was not adjusted during a period of high inflation. In order to maintain high standards those responsible for the experiment chose not to 'cut corners' on operations and data collection. As a consequence, when the budget was exhausted the project was shut down without any funding for research and analysis. Thus, unlike the U.S. NITs and most major social science experiments well known to economists, no Mincome final report was produced. After the 1974 to 1978 operational phase ended, Mincome staff completed technical documentation and coded some survey information electronically. The hard copy records (completed surveys and payments and taxation records) and computer tapes were turned over to the federal government. In 1981 the federal government provided limited funding to create the Institute for Social and Economic Research (ISER) at the University of Manitoba with the purpose of restoring the Mincome data and promoting its use to the research and policy community. Important parts of the data, however, remain available only in hard copy form at Library and Archives Canada.

All of the North American NITs followed a broadly similar data collection process. Several pre-random assignment surveys were carried out, followed after random assignment by nine periodic surveys. The initial 'screener' survey (not publicly available) in late 1973 collected a limited amount of income and family composition information designed to establish an eligible

[4] In 1981 Congress repealed the permanent work incentive (disregard of one-third of every extra dollar), confining it to the first 4 months of a job.

population. Those deemed potentially eligible were approached for the baseline survey (known as MINC1), which was administered between May and December of 1974. The baseline survey was followed by an enrolment survey which was generally carried out at the time (or just prior) to random assignment. Following random assignment, 9 periodic surveys were carried out, on average, every 4 months over the period 1975-77. We note that the specific chronology (i.e., calendar date) of these surveys varied across individuals, similar to the U.S. NITs.

The post-random assignment periodic surveys are publicly available only for Winnipeg, and restrict the sample to 'in-tact' households – that is, households with no change in marital status during the experiment. All individuals who originally indicated they would participate responded to the Baseline Survey, and thus we can do a simple test of sample selection bias by examining characteristics of the individuals from Winnipeg who responded to the baseline survey but never appear in the longitudinal labour surveys (this is most likely the non-in-tact households). There were few such individuals from the single parent group and their characteristics appear quite similar to those included in the surveys (in contrast, about one-quarter of two-headed households appear to be excluded by this restriction).

Because the digitized information from the periodic surveys is limited to 'in-tact' households, a selected (and potentially non-random) subset of observations, the administrative payments data – which in principle constitutes the complete experimental sample – is a potentially important source of information to complement that from the surveys.[5, 6]

The baseline survey is known as MINC1, the payments files MINC2, and the longitudinal labour file MINC4. There actually appears to be three pre-random assignment surveys included in the digitized files. Mincome documentation is somewhat light on the details. There are two pre-random assignment surveys available in the MINC4 file (i.e., Periodic #1 and #2 out of 11). One would think based on Mincome documentation that these consist of the baseline and the

[5] To our knowledge, other than the recent work by Calnitsky and Latner (2017) – which focuses on Dauphin, a small sub-set of the overall data, and uses a difference-in-differences design to deal with non-random assignment in Dauphin – the data from the payments system has not been analyzed. A disadvantage of the payments data is that we only observe wages, and not hours worked; hence our analysis with the payments file will focus on whether the individual is employed or not.

[6] The administrative and survey-based samples differ in two additional respects. As noted, some treatment group families responded to the periodic surveys but do not appear in the administrative data because they never filed an IRF. In addition, 'IRF controls' were excluded from the surveys.

enrollment surveys. Note that only labor supply information is included in the MINC4 file (i.e., not socio-economics background for instance). MINC1 is the baseline and contains a limited set of socio-economic information. However, a careful comparison of the interview dates for the Periodic #1 from MINC4 and MINC1 reveals that for a strong majority of the sample, these are different surveys. After reviewing background information on the early administration of Mincome at the National Archives, there are multiple references to several pre-random assignment surveys. All individuals in MINC4 also responded to MINC1 (where we access household characteristics).

We note, however, that some treatment group individuals or households who responded to the longitudinal labor surveys never filed an IRF. This is peculiar as (i) you were not supposed to be eligible to even participate in the experiment (aside from take-up) without filing the IRF, and (ii) you could not receive (even potentially) payment without filing an IRF. There does appear to be a legitimate question as to whether individuals assigned to the treatment group but who never filed an IRF realized they were in the treatment group.

There were two types of control group participants: those who also agreed to fill out these IRF forms (known as IRF controls) as well as the surveys discussed above, and a second sub-group (known as PC controls) who only were asked to complete the surveys. IRF forms were also used in the U.S. experiments; however, none of the U.S. experiments utilized a separate agency to compile the administrative data, nor complete participants' tax forms. As well, in DIME (but it appears not SIME or Gary) there were also two types of controls as in Mincome. Overall, only Gary had administrative data available; however, generally half the observations are missing.

C. Social Assistance and the Mothers Allowance in Manitoba Circa 1970s

Here we outline salient features of social assistance in Manitoba during the 1970s. Historically, social assistance in Canada was provided at the local level, relying mainly on private and church charity. Provinces first entered this field toward the end of World War I with Mother's Allowances programs for needy mothers with dependent children whose fathers were killed during the war.

These Mothers Allowances programs in Canada and the U.S. were the forerunners of government welfare/social assistance programs in North America.

During the 1970s social assistance in Manitoba remained an area of shared responsibility with municipalities providing support to "employables" whose welfare spells were expected to be brief and the provincial government supporting longer term cases -- those unable to or not expected to work. "Employables" included the employable unemployed, short-term handicapped, transients, and mothers with dependent children recently deserted by the husband or in prison with a short-term sentence. Longer term cases that came under the provincial legislation included single mothers with dependent children (after a waiting period), the elderly, disabled and blind.[7] Under the provincial legislation benefits for lone parents remained a distinct program and was still referred to as 'Mother's Allowances.'

When Mother's Allowances was first introduced single mothers were not expected to work – indeed, were discouraged from working outside the home, although they could engage in home-based work such as taking in boarders, knitting and sewing. The philosophy appeared to be that the state provided some income support and in exchange the mother's job was to raise responsible future citizens (Fields, 2002). However, by the 1960s and 1970s attitudes had changed, in part reflecting rising female participation in the labor force. The composition of sole support mothers had also shifted – with separated/divorced and never married representing the largest groups and widowed and abandoned being a much smaller fraction of the total. In Manitoba the rising number of Mother's Allowances cases was also a growing policy concern (Manitoba Department of Health and Social Development, 1992).

In his report to the provincial government on welfare policy, the economist Clarence Barber (1972) noted that single mothers constituted a significant part of overall social assistance costs and one that was growing as a proportion of total expenditures. He regarded sole support mothers as the most employable group receiving provincial income support, followed by those with a disability. However, he noted that the existing system provided almost no incentive to earn additional income because – apart from a small earnings exemption –

[7] Eligibility under the provincial Social Allowances Act was more restrictive prior to the 1970s. For example, an unmarried mother with one child was not eligible for assistance until 1970 (Bedard, 1994).

recipients faced a 100% tax back rate on market earnings.[8] He also characterized existing day care services as 'completely inadequate' – strictly commercially operated and having too few spaces available to allow a significant number of single mothers to enter the workforce. Provincial social assistance also provided some health services – dental, drugs and optical --for recipients and their dependents. These non-cash benefits would cease upon leaving social assistance to enter the workforce, potentially resulting in a tax rate exceeding 100%.[9]

As part of his report, Barber carried out a case review and small survey of single parents on social assistance. Most had very low education and few specialized skills but over 80% had worked in the past, albeit not recently. The survey also provided limited evidence on willingness to work, evidence that Barber interpreted as indicating that a significant number (but less than a majority) would work part-time or full-time if suitable incentives were provided.

Shortly after Barber's report the provincial government began introducing incentives for social assistance recipients to work. Their ability to do so was greatly enhanced by the federal government's recently introduced Canada Assistance Plan that provided for 50% cost-sharing with provinces of income support benefits and related programs. In September 1973 a new Work Incentive program that "ensures some financial gain to recipients from seeking and obtaining work" was introduced and expanded in subsequent years (Manitoba Department of Health and Social Development, Annual Reports, 1973- 1978).[10] Mothers Allowances and Long-Term Disability cases represented the two largest groups in the Work Incentive program caseload. Additional programs designed to help those having difficulty finding and retaining work were introduced in 1975. These programs were available to those receiving and not receiving Social Assistance. The earnings disregard was also increased from $20 per month at

[8] Barber (1972) reports the earnings disregard was $20 per month. In addition, there was an allowance of $7.50 per month for work-related expenses. At the minimum wage of $2.30 (as of January 1975) this allowed up to 12 hours per week free of claw-back.

[9] According to statistics in the annual reports of the Department of Health and Social Development these health benefits cost 7% to 8% of social assistance benefits over the 1972 to 1978 period.

[10] Data provided in the Department's Annual Report from 1975 to 1978 indicate that the WI program was short-term in nature, and thus not a continuing incentive such as an earnings disregard. Each year the new case intake and case outflow exceed, often substantially, the stock of cases at the beginning and end of the year. This suggests interventions such as job search assistance or mentoring of job search and job retention.

the beginning of Mincome to $50 per month at the end (Bedard, 1994), equivalent to about 17 hours of minimum wage work per week in the latter years of the experiment.[11]

Perhaps the most significant policy development affecting single mothers during the Mincome experiment was the introduction of the Child Care program in September 1974, a program that was enhanced substantially in subsequent years. Under this cost-shared program the provincial government provided advisory services to both group day care and family day care providers. Maximum fee levels for full day and half daycare were established for day care providers to be eligible under the program. Subsidies to parents were based on an income test and 'social need' with single parents rated the highest priority and two-parent families with both parents working the next highest priority.

Child care subsidies under this program were highly progressive. At family income levels approximating those paid to social assistance recipients child care was fully subsidized. As family income rose subsidies were reduced in a manner similar to a NIT with a 50% tax-back rate. The break-even point was reached when family income was approximately equal to the average income in Manitoba for that family type. In November 1975 subsidies to families using day care were made more progressive by changing tax-back rates to 0%, 0.25% and 0.75%.

Ministry reports document the dramatic growth in availability and use of day care under the Child Care program.[12] In interpreting this strong growth, it is important to keep in mind that subsidized child care was available to all low-income families, not just those in the Mincome experiment, and within Mincome was available to families in both the treatment and control groups. Thus, our estimated treatment effects should be valid experimental estimates of the impact of the NIT on labor supply. Although the internal validity of our estimates is not compromised by the introduction and expansion of the child care program, external validity may be affected. Our experimental estimates may not generalize to an environment in which subsidized child care is not widely available. Similarly, because the Work Incentive program was

[11] The Manitoba minimum wage was raided to $2.95 in September 1976 and remained at that level until the end of the experiment.

[12] Licensed space in group care centres increased from 374 in 1974 to 4814 in 1978, while those in family day care rose from 14 to 575 over the same period (Manitoba Department of Health and Social Development, Annual Reports, 1974 to 1978).

available to both the treatment and control groups, its introduction during Mincome should not affect the validity of our experimental estimates.

D. Economic Conditions and Seattle WA and Gary IN

Economic conditions at the time of random assignment, and throughout the experiments, varied widely in the U.S. NITs. In Seattle, the Boeing Company played a major role in the city; indeed, Seattle was referred to as the "world's largest company town" (New York Times, 1970). Beginning in the late 1960s, all three of Boeing's aerospace industry components went into downturn simultaneously.[13] The nation as a whole entered a recession in 1970, which caused new orders for airliners to dry up as well. This brought extensive layoffs at Boeing, along with severe distress within its home city of Seattle (Seattle Times, 1996).

Then, in 1971, the U.S. government cut funding to the company's supersonic transport (SST) program, and the project was cancelled soon after. By the time that layoffs ceased, Boeing had lost more than 60 percent of its workforce (Serling 1992). Because the aerospace company was the region's largest employer, unemployment rates increased dramatically. Overall, due to changing external demand and the cancellation of the SST program, Boeing's workforce was cut from 80,400 to 37,200 between early 1970 and October 1971 (Seattle Times, 1996; Serling 1992).[14] In 1970, Seattle's unemployment rate was 10 percent compared to a national average of 4.5 percent (New York Times, 1970). After the cancellation of the SST program, unemployment in Seattle reached 13.8 percent; at the peak of the recession, general Puget Sound unemployment stood at 17 percent (Seattle Times, 1996).

In Gary, the United States Steel Corp. played a similarly important role as the largest employer in that city. Gary is also largely a one-company town. The New York Times reported in December 1971 that around 25,000 persons were likely to be out of work in the city of 175,000 population, with unemployment ultimately reaching a record level of 41 per cent (New York

[13] The waning of the Vietnam War brought a sharp falloff in military procurement, which dropped from $23.3 billion in 1968 to $18.4 billion only three years later. The waning of Apollo led to a similar falloff in NASA employment, which plunged from 394,000 in 1966 to 144,000 in 1971.

[14] Layoffs hit every department. The number of hourly workers declined from 40,000 to 15,000. The number of engineers and scientists, which had been near 15,000, dropped by more than half. Office staff was cut from 24,000 to 9,000. Managerial positions were slashed all the way up the line, and even the top executives took pay cuts of up to 25 percent (Serling 1992).

Times 1971a; 1971b). As noted the Gary sample consists only of Blacks. The layoffs in the steel industry were based on seniority. Evidence suggest that the downturn in steel was particularly hard on Gary's black community, as many black workers had been employed for a shorter time than their white colleagues.[15] The layoffs may also explain Gary's higher welfare receipt as it was also reported that welfare caseloads had doubled between May and December of 1971.

[15] The NY Times reported that of total unemployed throughout the city, more than 75 per cent were black.