The Pros and Cons of a Negative Income Tax: New Evidence from Old Experiments

Chris Riddell
Department of Economics, University of Waterloo

W. Craig Riddell
Vancouver School of Economics, University of British Columbia

Preliminary and Incomplete—Please do not quote.

February 19, 2024

Abstract

Interest and research on a Basic Income continues to grow. A potentially important BI feature that has received relatively little attention in developed countries is the notion that recipient households can allocate time away from the labor market towards activities that yield family and social benefits rather than pure 'leisure'. We examine hitherto untapped data on two-parent families from the Manitoba Basic Annual Income Experiment including happiness, household time allocation, and social activities. We find that the treatment group reports higher levels of overall happiness, marital satisfaction, satisfaction with household duties, along with increased social activities. We find no effect on measures of financial well-being. Relatedly, we explore heterogeneity in the labor supply response in both the Denver and Manitoba NIT experiments. We find that the reduction in labor supply for both men and women is almost entirely restricted to families with children less than school age (<6). On the other hand, accounting for the low take-up rates in these experiments implies that the reductions in labor supply in both experiments are far greater than currently believed, with local average treatment effects implying a 40-50% reduction in both hours worked and the probability of working for women in both countries.

1. Introduction

Interest and research on a basic income (BI) or universal basic income (UBI) continues to grow. Numerous books, articles and policy studies by both proponents and skeptics have been devoted to the topic, and governments, research institutes and wealthy individuals have responded by initiating pilot projects to assess the feasibility and impacts of a BI/UBI.[1] In the U.S. alone there are about 30 BI pilot studies with a randomized control trial (RCT) design underway or recently completed – many sponsored by local governments and carried out with the assistance of organizations with expertise in social experiments (see, e.g., https://guaranteedincome.us/). Government initiatives in other developed countries include those in Canada, Finland, Italy, South Korea and Spain.[2] Review papers on this rapidly growing policy area include Widerquist (2005), Marinescu (2017) and Hoynes and Rothstein (2019) for developed countries and Banerjee et al (2019), Hanna and Olken (2018) and McGuire et al (2022) for developing countries.

Given the growing interest in a basic income, the early North American Negative Income Tax (NIT) experiments carried out in the 1970s have received renewed attention. In part this may reflect the fact that government-sponsored BI pilots such as those carried out in Barcelona (B-MINCOME) and Seoul (Seoul Safety Income Project) are testing an NIT design. The renewed interest may also arise because concrete BI proposals typically operate through the tax system and provide payments that are income-tested with NIT-features such as a minimum guarantee level and a tax-back rate on income above the basic income.[3] Accordingly, in recent BI review

---

[1] Recent and widely cited books include Forget (2018), Green at. al (2023), Haagh (2019), Lowrey (2018), Murray (2016), Van Parijs and Vanderborght (2017) and Yang (2018).

[2] The Finish BI experiment operated from 2017-18 and focused on unemployment insurance recipients, many long- term unemployed (Verho et. al. 2022). In Canada, the province of Ontario was the first to introduce a BI pilot project (https://www.ontario.ca/page/ontario-basic-income-pilot) but it was cancelled after a change in government. British Columbia appointed an Expert Panel on Basic Income that carried out extensive consultations and research and recommended substantial changes in existing income support programs rather than a BI program (Green, Kesselman and Tedds, 2020). Prince Edward Island's pilot recommended a Basic Income program funded principally by the federal government (https://www.gbireport.ca/). In 2019 the Italian government introduced a BI referred to as "Citizen's Income" (https://www.oecdbetterlifeindex.org/countries/italy/) that was replaced in 2023 by a less generous and more restrictive policy by the current government. In Spain the city of Barcelona implemented a BI experiment B-MINCOME between 2017 and 2019 (Riutort, Lain and Julia, 2023). The Seoul Safety Income Project, a three-year RCT with an NIT research design, began in 2021. See https://seoulsafetyincome.welfare.seoul.kr.

[3] In Canada, all fully developed BI proposals by academics operate by converting non-refundable tax credits (which low-income families individuals cannot take full advantage of, in contrast to high income earners who benefit fully) into refundable tax credits (Stevens and Simpson, 2017; Boadway, Cuff and Koebel, 2018a, 2018b); Koebel and Pohler, 2019; and Simpson and Stevens (2019). Refundable tax credits would be gradually reduced or clawed back for those with income above a certain threshold. Because individuals with no income would receive the full amount of the basic personal exemption, this creates a BI that operates through the personal income tax system.

papers, these experiments often make up the bulk of the evidence (e.g., Widerquist (2005), Van Parijs and Vanderborght (2017), Marinescu (2017), and Hoynes and Rothstein (2019)). These income maintenance experiments—the first large-scale social experiments in Economics – were carried out in the United States (New Jersey, Rural North Carolina/ Iowa, Gary, and Seattle/Denver ('SIME-DIME')), and Canada—the Manitoba Basic Annual Income Experiment (Mincome), which was conducted last and followed the SIME-DIME design.

The North American NIT literature has focused almost entirely on labor supply and marital status impacts – key components of the potential costs of a NIT. However, the recent BI literature also emphasizes broader outcomes that constitute potential benefits often highlighted by BI/UBI proponents such as subjective well-being, autonomy, financial security, educational attainment and health. Obtaining empirical evidence on such potential benefits is evident in studies carried out in developing countries (e.g. Haushofer and Shapiro 2016, Londono-Velez and Querubin 2022), in recent U.S. experiments (e.g., Jaroszewicz et al 2022; Pilkauskas et al 2022), and ongoing pilot studies such as those in South Korea (https://seoulsafetyincome.welfare.seoul.kr) and the U.S. (https://guaranteedincome.us/).

Our objective in this paper is to provide a better understanding of the 'pros and cons' of an NIT in a developed country context using data from the Income Maintenance Experiments carried out in North America in the 1970s. We focus on two-parent families, beginning our investigation with four NIT experiments carried out in metropolitan areas: Gary, Indiana; Seattle, Washington; Denver, Colorado and Winnipeg, Manitoba. We find lack of balance in the two-parent family samples in Gary and Seattle, as we found in an earlier paper on single parents (Riddell and Riddell, 2024).[4] Thus, we narrow our attention to analysing NIT impacts in Denver and Manitoba, where randomization was successful.

To assess potential benefits, we examine hitherto untapped data from the Manitoba Basic Annual Income Experiment (Mincome) on well-being and various dimensions of household time allocation including household production and social activities. We find that the treatment group reports higher levels of happiness, marital satisfaction, agreement and satisfaction with household duties, as well as increases in social activities (for women). If these results extrapolate to a modern labor supply setting, they suggest that an NIT may allow households to re-allocate

[4] The working paper by Price and Song (2018) also reports lack of balance in Seattle for their sample which consists of single and two-parent families with at least two children.

time in a manner that has important individual and family benefits, the types of consequences associated with greater autonomy often emphasized in the BI literature.

We also re-assess potential costs and explore—for the first time—heterogeneity in the labor supply responses in both the Denver and Manitoba experiments. Furthermore, we incorporate take-up of the experimental offer which has largely been ignored in the literature. Take-up was low in the NIT experiments (40-50%), but implementation of a basic income as a continuing government policy would likely involve much higher take-up rates. Thus, local average treatment effects from the NIT experiments are likely more relevant from a policy standpoint than the intent-to-treat (ITT) effects reported previously.

To summarize the labor supply results, our labor supply estimates are similar to those for Denver in the SIME-DIME Final Report (SRI International, 1983) – statistically significant reductions in hours worked and employment for both men and women. In Mincome we estimate that women in two-parent families reduced hours worked by 23% and employment by 14%. Both estimates are highly significant and larger than the small and insignificant results reported previously in the literature. Male labor supply estimates are also negative but not statistically significant. A key finding is that previously published pooled results mask important heterogeneity. In particular, our estimates show that for both men and women –and in both countries– the labor supply reductions are largely restricted to families with children less than six years old. In Denver, couples with no children experienced no reductions in work activity, as was the case for families with school-age children. These labor supply results appear consistent with our analysis on broader outcomes for Manitoba: households with young children on average reduced their labor supply, allocated that time towards other household activities, and were left approximately financially neutral.

However, once take-up of the NIT offer is accounted for—i.e., the labor supply response associated with actual *receipt* of the NIT supplement as opposed to the *offer* of the experimental earnings supplement—the reduction in labor supply in families with young children in both Denver and Manitoba is greater than one might expect given past reviews of the NIT literature—with estimates generally around a 50% reduction in both hours worked and the probability of working.

2. Basic Income Literature

The term Basic Income is used to describe a diverse set of policies intended to reduce poverty sand inequality. Typically, a BI refers to a cash benefit paid to recipients at regular intervals (e.g. monthly). These payments may replace parts of the existing social safety net, or supplement the income support system. Proposals differ, however, on several key dimensions. One is whether the BI is universal, i.e. received by all families, or income-tested, and limited to low-income families or paid to all but taxed-back according to family income. Another is whether the benefits are unconditional or conditional, e.g. requiring participation in the workforce or enrollment in education. While much discussion revolves around a basic income being 'universal', several authors question whether a UBI is fiscally realistic (Hoynes and Rothstein 2019; Green et al 2023).[5] Many recent studies, especially those in developing countries, but also those in developed countries during the Covid pandemic, examine unconditional cash payments targeted on low-income families (e.g. Haushofer and Shapiro 2016, Londono-Velez and Querubin 2022, Jaroszewicz et al 2022, Pilkauskas et al 2022). Being targeted on low-income families, eligibility is income-tested but benefit payments typically do not differ between those just below and far below the eligibility cut-off. Implementing such a policy on a wide-spread basis would face major challenges. For these reasons, interest in income-tested policies including, in particular, a NIT remains strong. Indeed, it is noteworthy that current large-scale, government-funded experimental pilots designed to reduce poverty and inequality — such as Barcelona's "B-MINCOME" and Korea's "Seoul Safety Income Program"—employ a NIT design. Clearly, a NIT-style basic income remains an important policy to understand.

Some caution is required in interpreting the vast NIT literature because the New Jersey experiment (that included only two-parent families in its target population) was adversely affected by a post-randomization change in its welfare system that "…created complex and

[5] For example, based on their simulations for the Canadian province British Columbia, Green et al (2023, p. 163) conclude that "UBIs are so much more costly than IBIs (more than twice as costly to achieve the same level of poverty reduction) that it is hard to conceive of them as a reasonable policy choice". In the US context Hoynes and Rothstein (2019) conclude that "A pure UBI (providing a set benefit to all regardless of income, age, etc.) funded to meet basic needs for a household without earnings would be extremely expensive, about twice the cost of all existing transfers in the United States. Funding it would require substantial new revenue." Hanna and Olken (2018) also find that income-tested BIs strongly dominate UBIs in the developing country context.

shifting incentives that are virtually impossible to quantify" (Pechman and Timpane, 1975, p. 5)[6] and the Gary and Seattle samples fail balancing tests, as demonstrated in the case of single parents samples in Riddell and Riddell (2024) and two-parent family samples in this paper.[7, 8] The fact that most studies pool the Seattle and Denver results as SIME-DIME further complicates interpreting previous evidence. Finally, for reasons discussed later, the Mincome labor supply evidence is based on a single study that cannot be replicated and yields results that are not credible.

We also comment on some of the recent experimental literature on unconditional cash transfers. While generally the basic income literature from developing countries is omitted from the wealthy country literature (e.g. Hoynes and Rothstein 2020, Marinescu 2018), we include some discussion of the former here given that the developing country literature has focused on a broader set of outcomes such as financial security, subjective well-being and health. We find it also informative to contrast the developing country results with recent U.S. randomized control trials testing unconditional cash payments.

The earlier NIT literature focused principally on labor supply and marital status.[9] The U.S. NITs did not contain questions on subjective well-being or questions comparable to Mincome on household production, which are an important focus in this paper. One broader outcome that is, indirectly, related to our study is children's educational outcomes. Overall, the evidence from multiple experiments suggests that the children of treatment households had greater school attendance than observed in the control group (see the review by Hanushek

[6] At that time the existing US welfare program (AFDC) was limited to single parents with dependent children. However, a federally sponsored program (AFDC-UP) that provided benefits to families with unemployed fathers also existed. New Jersey was chosen as the site for the first NIT experiment in part because it had not adopted AFDC-UP, so that the difference between treatment and control families would reflect the NIT offer alone. However, shortly after the experiment began, the state substantially altered its welfare system, including introducing a generous AFDC-UP program. Aaron (1975) provides a detailed analysis of the difficulties this created for interpreting the experimental estimates.

[7] Riddell and Riddell (2024) show that the single female head samples in Gary and SIME fail balancing for a variety of key measures (labor market and welfare participation), and provide evidence that the contamination is tied to a mis-balance of entry cohorts between treatments and controls, coupled with the large economic shocks that hit Seattle and Gary at the time the experiments began. Other suggestive evidence is provided by Price and Song (2018). In this paper we provide evidence that Gary and SIME also fail balancing within the two-heads experiment). The issues with SIME are important because the previous literature almost always combines SIME and DIME into one sample (the one exception is some of the labor supply estimates provided in the SIME-DIME Final Report, SRI International (1983)). SIME and DIME were actually two separate experiments conducted at different times and with different experimental stratifications.

[8] SIME-DIME included questions on child outcomes. Unfortunately, the child outcomes data was not digitized in the public use files and appears lost. Thus, it appears impossible to re-assess the evidence based on DIME data alone.

[9] Hanushek (1987) reviews the evidence on non-labor supply (and non-marital) outcomes which was restricted to studies on consumption patterns, housing, and human capital—the latter being primarily school enrolment and academic performance of the children.

1986).[10] There is also some evidence that the increase in probability of school attendance closely matches the decrease in probability of working (for women).[11]

The other literature on cash transfers that tends to be discussed in tandem with the NIT literature are the non-experimental studies on regular, localized cash transfers in wealthy countries including, in particular, the Alaska Permanent Fund Dividend literature, which has generated a large set of studies. Guettabi (2019) provides a review. It is unclear how relevant this literature is for our paper since the Alaska Fund is a universal program, whereas our focus is on programs targeted towards low-income households. As noted previously, a universal basic income that would bring about a meaningful reduction in poverty is unlikely to be fiscally realistic in high-income countries such as Canada and the United States. That said we note that, overall, the Alaska Dividend literature does suggest that the Fund's payments have had some positive impacts on children's health, crime and poverty. No study that we are aware of has examined subjective well-being, or time allocation and non-labor market activities.

In the literature on unconditional cash transfers, the experimental evidence could be described as mixed. Most experimental papers are in developing countries, examine unconditional payments, and find positive effects on a wide variety of outcomes. Over the last decade this literature has exploded —see Banerjee et al 2019 and McGuire et al 2022 for thorough reviews. Rather than providing an exhaustive review we focus on a) more recent studies contrasting selected experimental evidence in developing countries with recent U.S. evidence, and b) those that examine outcomes similar to ours.

Haushoffer and Shapiro (2016) find large positive effects on subjective well- being and financial well-being/security (food security in particular) in Kenya. However, in another Kenya experiment that compares an unconditional cash treatment with free health care of the same value, Haushofer et al (2019) find no effects of the cash payment on health outcomes and various measures of subjective well-being. Handa et al (2018) find positive effects on financial well-being (in particular measures of food security) in a government-backed experiment in Zambia. Banerjee et al (2020) examine an experiment of unconditional payments during the pandemic in Kenya, and find positive effects on measures of financial well-being (such as 'experiencing hunger') and health outcomes. An experimental study from Columbia examining cash payments

[10] The evidence is much more mixed on actual academic performance.
[11] Some caution is required with this evidence given that it relies in part on evidence from New Jersey, Seattle and in particular Gary (that fail balancing tests).

also made during the pandemic found positive (although small) effects on various measures of financial well-being/security such as making a loan payment (Londono-Velez and Querubin 2022). Overall, the experimental results from unconditional cash payments in developing countries is most consistent with positive effects on financial security with somewhat mixed but largely positive effects on subjective well-being and health outcomes. Perhaps not surprisingly, this literature is less informative about labor supply responses or alternative uses of time (i.e., non-labor market activities).

Conversely, two recent experimental studies carried out in the United States during the first year of the COVID pandemic find no effects on any of a large set of pre-registered outcomes. Specifically, Jaroszewicz et al (2022) find no effects of either a \$500 nor \$2000 one-time payment on financial well-being/security, health, or subjective well-being (all measured as indexes of multiple questions). Similarly, Pilkauskas et al (2022) find no effect of a \$1000 one-time payment on financial hardship, mental health or household outcomes comparable to ours including partner relationships (in one of the few studies to examine such an outcome). We note that, generally (with some notable exceptions), this experimental literature on unconditional cash payments (the U.S. studies in particular, and some of those from developing countries) examine outcomes over a much shorter time horizon than possible with the NIT experiments. Finally, we note that some studies (in particular, the two U.S. experiments) examine one-time payments [12] whereas other studies from developing countries along with both the original NIT experiments and current NIT pilots such as Barcelona and Seoul involve regular payments over multiple years (three to five years in the case of the North American NITs, 3 years for Seoul, 2 years for Barcelona).

3. Assignment Model

A key feature of all the North American NIT experiments was the Conlisk-Watts assignment model for allocating families to treatment plans.[13] Prior to random assignment,

[12] We note that this feature varies considerably across the developing country literature; indeed, Banerjee et al (2020) test three treatments, a one-time lump sum payment vs. two long-term payment streams.

[13] 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 guarantee G, low tax-back rate 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.

families were stratified by family type (two-parent families, single mothers 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.[14] 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.

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. For two-parent families, this includes normal income in all experiments. In order to obtain unbiased estimates of treatment effects it is therefore necessary to control for the appropriate stratification categories as well as interactions among these categories (see, e.g. Athey and Imbens, 2017).[15] Because the full sample for each income maintenance experiment is not randomly assigned, we use the term 'stratification group experiments' or 'mini-experiments' to refer to the level at which random assignment takes place. The number of mini-experiments varies substantially across the NITs. As the only data digitized for Mincome is the Winnipeg site, the mini-experiments for two parent families in Winnipeg consist of only the 4 normal income categories. DIME has the largest number of stratification groups; even within the two-parent family category there are 5 income categories, 3 races (Black, White and Hispanic), and 3 durations (3-years, 6-years and 20-years). One consequence of this model is that sample sizes are small for individual experiments. Anther consequence is that there is unbalanced allocation to treatment and control groups – the sample size of the control group is typically much smaller than the treatment group (approximately 60-40 in most cases). Perhaps the most important issue to note is that the early literature did not control for these stratifications properly.[16]

[14] Normal or permanent income was computed from pre-treatment surveys discussed subsequently.
[15] See, for example, Ashenfelter and Plant (1990). Their treatment effect estimates are a weighted average of mean treatment-control differences in each individual stratification group experiment which yields identical estimates to a regression-based approach with a full set of interactions among stratification groups.
[16] Specifically, the early literature simply included fixed effects for each separate stratification category. However, the researcher needs to include a dummy variable for each "mini-experiment" (or a full set of interactions between all stratification variables).

4. Data, Balance and Attrition

(a) The Mincome Experiment[17]

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' -- 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.[18]

A combination of declining interest in the concept of a guaranteed annual income, budgetary problems and changes in both the provincial and federal governments 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 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, until our recent paper on single parents (Riddell and Riddell 2024) only one published study of labor supply effects using the Mincome survey data (i.e., the randomized Winnipeg experiment) – that of Hum and Simpson (1991) – had been carried out.[19] 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. Of particular note, the sample sizes in Hum and Simpson do

[17] 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).

[18] One aspect of Mincome that has received considerable attention is the impact on health outcomes. An influential paper by Forget (2011) concluded that the NIT led to a 8.5% reduction in hospital use in the 'saturation site' Dauphin, Manitoba where all eligible residents could participate in the NIT supplement. Forget's conclusions have been questioned by Green (2021) for not taking into account pre-existing trends in hospital use, trends that continued after the NIT expired.

[19] 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.

not match the sample sizes in the Mincome official documentation, and are inconsistent with the assignment model. We outline our replication attempt in the Online Appendix (see section 3). Based on our doubts about the validity of their sample and the inability to replicate their findings, our view is that no credible evidence on labor supply impacts for two-parent families in Manitoba currently exist.

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). This is known as the 'Minc2' file. 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 employment income is that under-reporting of earnings, an important limitation of the U.S. NITs, is very unlikely in Mincome.[20] As discussed further below, the Minc2 file contains administrative earnings records at the *household* level as well as administrative data on NIT payments received by the household.

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 currently use only the data from the 'husband-wife' module discussed further below. The post random assignment 'longitudinal labor surveys' were collected

[20] We discuss evidence on under-reporting in the U.S. NITs 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.

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.[21]

A unique feature of Mincome was its inclusion of questions relating to happiness, marital satisfaction, satisfaction and agreement with household duties, and involvement in social activities such as church-going. The online appendix contains details on the questions asked and data availability.[22] The household/marital satisfaction questions were included as part of a "husband-wife" module that was first incorporated in the pre-random assignment enrollment survey—which we use in our balancing tests—and then administered again at the 9th periodic. (approximately 2.5 years later).

Our analysis is based on two-parent families that appear in both the pre-random assignment surveys and the periodic surveys. This results in a sample size of 489 couples. Some baseline summary statistics for Mincome are reported in Table 1 (see 'Mean Dependent Variable' row) while summary statistics for experimental period are presented in the treatment effect estimation tables.[23] Men are about 2 years older than their female partners and have slightly less education. Both have low educational attainment (less than 10 years). About 45% of women worked at some point during the pre-random assignment year versus 88% for their male counterparts. Weeks worked by males (38) is triple that of their female partners.

As noted above, we also utilize a separate administrative dataset ("Minc2") collected independently by Mincome Manitoba. Some households that completed the periodic surveys do not appear in the administrative data. This is more likely to be the case for the control group, however, some treatment group households are missing from Minc2. As well, similar to SIME-

[21] We also note that in all four 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, in contrast to the experience of some U.S. NITs, that there were no noticeable differences in the staggering between treatments and controls.

[22] Other than labor market information, virtually all questions in Mincome were not asked in every periodic. For example, the happiness question was asked three times. The specific questions and codebook for this set of questions was never digitized. We compiled the information in the Online Appendix from the Archives of Canada.

[23] The hours worked variables are not comparable across surveys, in particular the baseline survey (used in our balancing tests, see Table 1) definition differed from the both the enrollment survey (which we do not use for labor supply for this reason) and the post random assignment periodic surveys. The main differences were in the way casual work (often referred to as 'odd jobs') and paid hours not worked were incorporated into the hours worked variable. Thus, we use hours worked from the baseline in our balancing tests but then we use the control group mean for the experimental period for considering magnitude of the treatment effects (and present both summary statistics in the tables).

DIME there are additional control group households (known as "IRF controls") in Minc2 who did not participate in the surveys. Thus, Minc2 is a different sample of 343 households. We therefore conduct separate balancing tests for the Minc2 sample and also present summary statistics that can be compared to the main Mincome sample. These are contained in Appendix Table A1.

(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 of the literature refers to the U.S. NITs as consisting of four experiments (New Jersey, Rural, Gary and SIME-DIME). However, there are important 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.[24] The number of months of post-experiment data also varies by entry cohort. This contrasts with Mincome where all participants have two data points pre-random assignment, nine data points post-random assignment, and no information after the experiment ended.

[24] Note that DIME also had a small 20-year program (later abandoned) that we exclude from our analysis.

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.

(c) Balancing Tests and Attrition

Although it is now common to verify that the experimental sample is appropriately balanced, tests for balanced samples do not appear to have been reported in the original NIT literature.[25] Recent studies by Price and Song (2018) and Riddell and Riddell (2024) find that balance was not achieved for specific target populations in the Gary and Seattle experiments (single parents in Seattle and Gary in Riddell and Riddell, and families with at least two children in Seattle in Price and Song). This raises doubts about the likelihood that these experiments will yield unbiased estimates. Both cities experienced major downturns in the dominant industry (steel in Gary and aerospace in Seattle) around the time the experiments began, events that appear to have affected the treatment and control groups differently due to different timing of experimental in-take dates (see SRI International (1983, Vol II, Chap. 3) and Riddell and Riddell (2024) for details on problems encountered in enrollment and assignment to treatment in Seattle.)

Appendix Tables A2 and A3 present tests of balance in pre-treatment labor market outcomes for the two-parent family samples. These results confirm our doubts about the two parent family samples in these cities. In Seattle, treatment group males had lower earnings and higher welfare receipt while their female counterparts had lower earnings, hours worked and welfare receipt. These treatment – control differences are highly statistically significant. In Gary, treatment group families had more children, males had higher employment levels and hours worked, and females had lower levels of welfare receipt.

In contrast, our two parent family samples are balanced in Mincome (Table 1) and Denver (Table 2) based on demographic characteristics and labor market outcomes. In Mincome there are no statistically significant treatment – control differences prior to beginning of treatment and in Denver there is only one difference that is statistically significant (at the 10%

[25] 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 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.

level). Balancing tests pass for the Minc2 administrative records as well (see Appendix Table A1).

As noted above, the Mincome husband-wife module—which contains most of the key questions for non-economic outcomes (except 'happiness')—was administered pre-random assignment at the enrollment survey, and then administered again at the 9th periodic survey.[26] This allows us to conduct balancing tests on the various marital and household time allocation measures. These results are presented in Table 3, and we see that in the enrollment survey there are no differences between treatments and controls for any outcome.

We also check for evidence of non-random attrition, another potential source of bias. Attrition in the U.S. NITs has been discussed previously, and there is evidence of non-random attrition in SIME-DIME (Ashenfelter and Plant, 1990). We perform the test outlined in Fitzgerald, Gottschalk and Moffitt (1998) by regressing pre-random assignment hours worked and other measures of work activity 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.

Results are shown in Tables 4a (for Mincome) and 4b (for DIME). In the case of Mincome, for both males and females there are no statistically significant correlations between the attritor dummy and baseline measures of annual hours worked, weeks worked in the pre-random assignment year 1973, a dummy for worked in 1973, and (from the enrolment survey) the household satisfaction measures discussed above. Similarly, in Denver there are no significant correlations between the attritor dummy and hours worked and employment prior to random assignment. This suggests previous evidence of non-random attrition in SIME-DIME arises from the Seattle experiment – possibly for similar reasons underlying the unbalanced sample in SIME.

Due to the results of the balancing and attrition bias tests, the remainder of the paper will focus on DIME and Mincome where random assignment appears to hold and there is no evidence of non-random attrition.

[26] Unfortunately, in addition to the happiness question (only asked during the post random assignment period), we cannot do this analysis for church visits, as well as the financial security measures.

5. **Results**

(a) Subjective Well-Being and Time Use

Table 5 presents intent-to-treat ('ITT') estimates on our various measures of non-labor supply outcomes including happiness, satisfaction with contributions to household production and with the marriage, and our proxy for social activities (church attendance). For both men and women, we find an increase in happiness of 5 percentage points for women, 8 percentage points for men. These estimates amount to about 6-10% increases in happiness.[27]

On household production, both women and men in the treatment group report higher levels of *satisfaction* with their spouse's contribution to these activities. For women in particular, these are non-trivial estimates given that baseline satisfaction with their husband's help around the house are not as high as most of these indexes. The control group mean indicates that 62.5% of women are 'satisfied' with their husband's help; the ITT of 11 percentage points implies about an 18% increase. On *agreement* with household duties, the results generally point in the same direction with positive coefficients, but estimates are not statistically different from zero (for women the estimate of 9.2 percentage points is just outside of conventional levels with a t-stat of 1.55).

The estimated impacts on social activities (as proxied by church attendance)[28] suggest that households at least partly re-allocated labor market time towards additional personally rewarding activities beyond household production and pure 'leisure'. The results differ by gender with notable increases in church attendance only for women (of roughly a 20% increase in church attendance).

With respect to financial well-being/security, we examine two outcomes closely aligned with the previous literature: satisfaction with one's standard of living (asked in the husband-wife module), and the individual's views about the question: Do you feel your present financial situation is poor? Note that the latter is asked in the main survey at the same times as the

[27] Note that mean happiness levels at baseline are high – 87% for females and 80% for males. Thus a 6% to 10% increase is substantial. In part, the high levels may reflect the use of a 4-point scale that tends to yield higher scores than the 5-point scale.
[28] Attending church regularly was common in the early 1970s in central Canada (Ontario and the Prairie provinces) – the fraction of the population with no religious affiliation was less than 5% -- and was an important part of a community's social activity. There are also noteworthy cohort effects – those born prior to 1946 (i.e. most of the Mincome sample) maintained high levels of religious affiliation throughout life (Canada's Changing Religious Landscape, Pew Research Report, June 2013).

happiness variable (three data points), and thus the sample size is larger. The ITT estimates are presented in Table 6. Unlike most of the literature on unconditional payments in developing countries (but similar to the pandemic era U.S. experiments), we find no effect on financial security. There is perhaps some weak evidence for men of *increased financial insecurity*—similar to the recent U.S. experiment by Jaroszewica et al 2002. However, our results are not statistically significant. Certainly, for women it seems clear that the treatment and control groups report identical levels of financial well-being/security.

(b) Labor Supply

We now turn to the consequences of the NIT offer for work activity. As noted previously, there are currently no credible labor supply estimates for the Mincome experiment for two-parent families. This is important given that several reviews of the basic income and NIT literatures note that the estimated impacts on both hours worked and employment in Mincome are small in size and statistically insignificant (e.g., Widerquist 2005; Marinescu 2017)[29]. In addition, there is uncertainty about whether the labor supply estimates for Denver properly accounted for the experimental stratifications (Riddell and Riddell, 2024). Thus, we begin with results for the pooled sample before turning to our main interest: heterogeneity in the labor supply response based on the youngest child's age.

The Online Appendix (section 1) summarizes the previous literature on labor supply responses of males and females in two-parent families[30] based on survey articles by Robins (1985), Burtless (1987) and Hum and Simpson (1993).[31] Several features are evident. For men, all point estimates are negative on both the intensive and extensive margins, but only those for DIME, SIME and SIME_DIME pooled are statistically significant. The same is true for females with the exception of a small positive and statistically insignificant coefficient estimate for Gary. More precisely estimated impacts are consistent with the much larger sample sizes in SIME-

[29] For example, Widerquist (2005, p. 62) notes the "smaller response to the Canadian experiment" and attributes it to "the make-up of the sample and the treatments offered."

[30] 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.

[31] The sole previous study of labor supply in Mincome (Hum and Simpson, 1991) pooled together men in two -parent families with single men (21% of their sample) so no previous estimates exist for either as a separate group. Mincome was the only NIT to include single men and women in their target population.

DIME. Mincome stands out as reporting labor supply impacts that are very small and statistically insignificant for both men and women in two parent families.

Our estimated treatment effects for both Mincome and DIME pooled samples are reported in Table 7.[32] For females in two-person households in Mincome our ITT estimates imply a 23% decline in hours worked – an impact that is significant at 1% – and a smaller and less precisely estimated 14% decline in the employment rate. This estimated decline in hours of work is similar in percentage terms to those in the earlier literature for SIME-DIME, but much higher than the negligible and statistically insignificant estimates reported by Hum and Simpson (1991). Likewise, our estimated annual hours reduction (97) is similar to the 103 hours for DIME reported in the SIME-DIME Final Report (SRI International, 1983). The SIME-DIME Final Report doesn't provide separate employment rate estimates for DIME and SIME.

For males our Mincome ITT estimates are negative but not statistically significant on both the hours and employment rate dimensions. Hum and Simpson (1991) report annual hours estimates that are smaller in size but also statistically insignificant. The estimates for DIME are much more precise, likely due to the larger sample sizes. For women in DIME, the ITT estimates imply a decline in hours worked and the probability of working of 20-21% for women—very close to the estimates for Mincome. For men, the respective estimates are a 7% decline in hours and 3.5% decline in the probability of working. These estimates are very close to the DIME-only sample in the SIME-DIME Final Report.

Our labor supply results for Mincome and Denver suggest that the NIT may have not only relaxed the budget constraint for these low-income families but also the time constraint, allowing more time for activities such as household production, social engagement and educational improvement. To investigate whether this re-allocation of time was more prevalent in families with younger children we examine heterogeneity in the labor supply response by age of the youngest child. Table 8 presents the estimated ITTs for Mincome, and Table 9 presents the same for DIME. Given the sample sizes, we sub-sample into two categories: (i) households with at least one child less than school age (<6), vs (ii) households with no children less than school age.

[32] As noted previously, there are three surveys annually in Mincome and monthly surveys in DIME. To ease comparison, we include a row that reports the size of the hours of work coefficient in annual hours equivalents.

Overall, the results are broadly similar between Mincome and DIME. Turning to the specifics, for Mincome, we find decreases in labor supply—along both margins—for women *and men* in households with at least one child less than school age. Moreover, the magnitudes of the effects are almost identical across the two countries. In Manitoba, women in the treatment group report a 25% decline in hours worked while women in DIME report at 26.5% decline in hours. On the probability of working, women in Mincome were about 9 percentage points less likely to work—a 21% decline—while women in DIME were about 6 percentage points less likely to work—a 27% decline. We note that while the employment rates make it appear that women in Mincome had greater labor force attachment, they worked substantially fewer hours (conditional on working).

For households with older children (6+ years), the estimates for women are very different. We find no difference between treatments and controls on the probability of working in either country. For hours worked, there is modest evidence of a 23% decline in hours in Mincome (statistically significant at the 10% level, of course sample sizes are small in Mincome), while there is no decline in hours for women in DIME.

Turning to men, differences according to age of youngest child are broadly similar to those for women. In Mincome, the current consensus result for men is that the NIT offer has no effect on labor supply (see Online Appendix Table 1)—as we also find in Table 7—a result that differs from DIME and has attracted attention in the basic income literature. The pooled sample does appear to mask important heterogeneity however. Men in Manitoba households with a young child reduced labor supply by 9% and the probability of working by 6.5% (relative to the control group mean). Both estimates are very similar to DIME (at 11% and 6% respectively). There is no labor supply response (in fact, positive coefficients) for men in Mincome with older children. In the case of DIME, we find comparable—although slightly lower—labor supply response for men with older children.

Finally, it is interesting to note that the DIME sample is unusual in that we observe two-head households with no children. The NIT experiments are typically viewed as only including households with dependents.[33] It therefore remains unknown how single adults or couples without children would respond to a NIT offer, which is notable given that a basic income

[33] As far as we can tell from official documentation, only Mincome had singles with no dependents as an experimental stratification. There are no two headed families without dependents in Mincome. It is unclear if families with two heads without dependent children met the eligibility criteria in SIME-DIME.

program would likely include such households in the eligible population. While the sample size is small, we observe— in DIME—133 households with no dependent children[34]. We find no reduction in labor supply for both men and women in this sub-sample (in fact, positive coefficients).

(c)Take-up/LATEs:

When considering the external validity of ITT estimates we need to consider take-up of treatment, and compare that with the presumed take-up if treatment was implemented on a widespread and continuing basis. Take-up in the North American NIT experiments has received virtually no attention in the recent literature, but authors of both the SIME-DIME Final Report and internal government documents for Mincome comment on the low take-up—in particular how low take-up was relative to expectations.[35] Figure 1 shows monthly take-up rates for DIME and Mincome. For DIME, the data contains a (self-reported) monthly variable for NIT payment received. Take-up in DIME is fairly constant at around 50%. Note that the initial rise in take-up for DIME over months 1-7 is due primarily to the variation in experiment start dates noted previously.

For Mincome, calculating take-up is more complicated. The administrative data (Minc2) contains monthly NIT payments; however, as noted in the previous section, the Minc2 file does not contain the full sample. As we documented previously, the characteristics of the Minc2 sample are highly comparable to the full sample at the household level, and are also balanced. We also remind readers that the Minc2 data only contains household characteristics. For Figure 1 we use the monthly administrative information; note this data is not self-reported by participants, but is from the independent agency Mincome Manitoba and thus month=1 if a payment was issued to the household in that month. Take-up is more erratic than in DIME although sample sizes are smaller. It also seems plausible that the self-reported nature of DIME 'smooths' the data. The overall mean take-up in Mincome is 47%, while in DIME take-up is 52% for women

[34] This is 133 households who had no dependent children for the duration of the experiment, not just over the pre random assignment (enrollment) period.

[35] We have reviewed the internal governments documentation—in particular correspondence between the federal government, government of Manitoba and research team of Mincome—at the Archives of Canada as part of this research and the low take-up was clearly the number one concern with Mincome. Indeed, the governments agreed to add the second Winnipeg experiment (known as the "Supplemental Experiment") due to the low take-up. That said, take-up in Mincome was broadly similar to both SIME and DIME.

and 45% for men. The lower take-up rate reported by men in DIME likely reflects the fact that NIT payments were typically made to the woman.

As a check we calculated a second take-up variable for Mincome. A Mincome program module was administered as part of periodic 8, which contains a question on NIT payment receipt. Periodic 8 was administered about 2.5 years into the experiment, and thus some attrition had occurred by this point in the experiment. Using the self-reported take-up variable from periodic 8 we compute a take-up rate of 43%— 4 percentage points lower than the administrative data.

It is unclear why take-up was low in Mincome and SIME-DIME. Both Mincome and SIME-DIME asked a large set of questions testing respondent's knowledge of the NIT parameters. Evidence from these questions for Mincome suggests that while a strong majority of the treatment group understood the basic elements of a NIT, most members of the treatment group substantially underestimated the benefit level for a given earnings amount, which could explain low take-up (Bennett 1986).[36] There was also a Mincome program survey administered to the treatment group that was never digitized. We gathered new (previously not digitized) data from the Mincome program survey and it is clear that a non-trivial number of participants either a) did not trust the experiment because it was run by the government and/or b) did not want to go through the details of the experiment because it was not permanent. If a NIT-type basic income was adopted today, for instance, as a national program it is plausible that it would not suffer from these various concerns that limited take-up.

If take-up were to be greater for a NIT implemented as an ongoing national program, ITTs may not be the most appropriate treatment effect. Rather, estimating the effect of actual receipt of the program—as opposed to the randomly assigned offer of the program —using random assignment as an instrument would be a more appropriate indication of the effects of a national program. The Wald Estimator is the simplest instrumental variable estimator; the denominator is the take-up rate (and the numerator is the ITT), and thus we can see that local average treatment effects for the impact of actual program receipt will be substantially higher than the ITTs discussed above. To have point estimates and standard errors, we present estimates of LATEs for labor supply in Table 10 and by age of youngest child in Tables 11 and 12 for

[36] We are unaware of any evidence from SIME-DIME on these questions, and unfortunately responses to these questions were not digitized in the public use file (the only data that now exists from SIME-DIME).

Mincome and DIME respectively. These results estimate the effects of actual receipt of the NIT using random assignment to treatment as an instrumental variable.

Beginning with the results for the full sample, the LATE estimate along the extensive margin is a 12 percentage points decline in the probability of working for DIME, and a 19 percentage points decline in the probability of working for Mincome. Recall that employment rates are lower in DIME. In percentage terms (relative to the control group), the reductions in the likelihood of working are about 45% (Mincome) and 41% (DIME)—far greater than the 14% (Manitoba) and 21% (DIME) effects based on the ITTs. The percentage declines for hours worked are comparable. The LATEs for Mincome in the estimates utilize the survey-based measure (see below for the analysis with the administrative data), and thus are likely somewhat overstated in magnitude given that take-up is 4 percentage points lower in the self-reported periodic than in the administrative records.

The results above mask the heterogeneity by child age. In the case of DIME—with larger sample sizes—the LATEs for women with children <6 suggest a 50% reduction in labor supply along both margins. Again, the estimates for women with children older than school age are negative in sign, but t-statistics are very small. For Mincome, the LATE of 26 percentage points for women with young children implies a 58% reduction in the likelihood of working (likely overstated due to an understated take-up rate). Similar to DIME, estimates for women with older children are not statistically significant. Overall, the results are very similar across the two countries.

The estimate for men with young children—while not surprisingly far lower in magnitude than for women—also indicate a much greater labor supply response than the ITT estimates and the consensus view in the literature. The results here are also similar across the two countries. The LATEs imply reductions in labor supply of about 15% for the probability of working in both countries.

Finally, we exploit the administrative nature of the Minc2 payments file to also examine the effects of actual receipt of the NIT on household labor force participation. The Minc2 file contains the total earnings amount paid to the household (i.e., total household wages earned) based on pay slips. We can therefore analyze the extensive margin. The literature has always separated men from women—here we ask what effects NIT receipt has, if any, on withdrawal from the labor market by the entire household. The estimates are presented in the first column of table 10. The mean household employment rate is 88% in Mincome. OLS estimates suggest that the treatment group household was 5.6 percentage points less likely to be working. Accounting for take-up, we find that the effect of receiving the NIT reduced the probability of labor force participation by the entire household by 12.5 percentage points, a 14% reduction. While there is no estimate in the literature to compare this to, the magnitude of the estimate—at the household level—strikes us as very large.

6. Conclusions

The North American NIT experiments have received renewed attention given the interest in a basic income. The Canadian 'experiment' has received particular attention—this appears to be due primarily to the non-experimental Dauphin saturation site. However, the reality is very little is known about the Canadian randomized experiment. Moreover, the North American NITs still have value given that it is unclear if a universal/unconditional basic income is financially realistic—at least in a North American setting.

Some recent large-scale, government-backed income-tested BI pilots—Barcelona's B-Mincome, and South Korea' Seoul Safety Income Program—chose to test a BI with an NIT design. Recent research has begun to move past labor supply, and instead emphasize the possible effects of a basic income on various dimensions of well-being. Evidence from B-Mincome along with some studies of unconditional income payments—in countries such as Columbia and Kenya, largely as part of pandemic-relief programs—have found positive effects. One important feature of a basic income that has received relatively little attention is the notion that households could allocate time away from the labor market and towards other activities that may yield important benefits.

Our focus is on the behavior of male and female heads in two-parent families in the NIT experiments caried out Gary, Indiana, Seattle, Washington, Denver, Colorado and Winnipeg, Manitoba (Mincome) in the 1970s. We first establish that the experimental samples for two-headed families in Gary and Seattle fail balance tests, raising doubts about the likelihood that these samples will yield unbiased estimates. Together with similar findings of lack of balance in these two NITs for single parents and for families with two or more children, our results indicate that unbalanced samples are pervasive in these experiments. In contrast, we conclude that the Denver and Manitoba are balanced and also appear to be unaffected by non-random attrition.

`We examine hitherto untapped data from the Manitoba Basic Annual Income Experiment on well-being and measures of household time allocation including household production and social activities. We find that the treatment group reports higher levels of happiness, marital satisfaction, agreement and satisfaction with household duties, in addition to increases in social activities. These results extrapolated to a modern labor supply setting suggest that an NIT income payment may allow households to re-allocate time in a manner that increases family well-being.

Next, we examine labor supply. For Mincome we provide the first credible estimates of labor supply impacts for two- parent families in the Canadian NIT experiment. For females our Mincome estimates imply reductions in annual hours worked of 23% and in the employment rate of 14%, estimates that are statistically significant and similar in magnitude to those for Denver. Our estimates contrast with those of the sole prior labor supply study for Mincome that estimated impacts that were very small and statistically insignificant.

Given these results, we explore heterogeneity in the labor supply responses in both the Manitoba and Denver experiments. We find that (a) the broad reduction in labor supply (i.e., along both the intensive and extensive margins) for both men and women is almost entirely restricted to families with children less than six years old; (b) couples with no children in Denver had no reduction in labor supply; and (c) once take-up is accounted for, the reductions in labor supply—i.e., those associated with actual receipt of the NIT program as opposed to the offer of the experimental treatment—are far greater (generally around a 50% reduction in both hours worked and the probability of working) than previously realized. All of the labor supply results are remarkably similar across the two countries.

References

Aaron, Henry J. 1975. Cautionary notes on the experiment. In Work incentives and income guarantees: The New Jersey negative income tax experiment, ed. Joseph A. Pechman and P. Michael Timpane. Washington, D.C.: The Brookings Institution.

Ashenfelter, Orley. 1978. The Labor Supply Response of Wage Earners. In Welfare in Rural Areas: The North Carolina – Iowa Income Maintenance Experiment, ed. John L. Palmer and Joseph A. Pechman. Washington, D.C.: The Brookings Institution.

Ashenfelter, Orley, Kirk Doran and Bruce Schaller. 2010. A Shred of Credible Evidence on the Long Run Elasticity of Labor Supply. Economica, 77: 637-650.

Ashenfelter, Orley and Mark W. Plant. 1990. Nonparametric Estimates of the Labor-Supply Effects of Negative Income Tax Programs. Journal of Labor Economics, 8: S396-S415.

Athey, Susan and Guido W. Imbens. 2017. The Econometrics of Randomized Experiments. In Handbook of Economic Field Experiments, Volume 1, ed. Esther Duflo and Abhijit Banergee . Amsterdam: Elsevier.

Banerjee, A., P. Neihaus and T. Suri. 2019. Universal Basic Income in the Developing World. Annual Review of Economics 11 961-985.

Banerjee, Abhijit, Michael Faye, Alan Krueger, Paul Niehaus and Tavneet Suri. 2020. Effects of a Universal Basic Income during the Pandemic. Working paper, December 8.

Blundell, Richard and Thomas MaCurdy. 1999. Labor Supply: A Review of Alternative Approaches. In Handbook of Labor Economics, Volume 3, ed. Orley Ashenfelter and David Card. Amsterdam: Elsevier.

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

Calnitsky, David and Jonathan P. Latner. 2017. Basic Income in a Small Town. Social Problems, 64: 373-397.

Cesarini, David, Erik Lindqvist, Matthew Notowidigdo and Robert Ostling. 2017. The Effect of Wealth on Individual and Household Labor Supply: Evidence from Swedish lotteries. American Economic Review, 107: 3917-46.

Eissa, Nada and Jeffrey B. Liebman. 1996. Labor Supply Response to the Earned Income Tax Credit. The Quarterly Journal of Economics, 111: 605-637.

Fitzgerald, John, Peter Gottschalk and Robert Moffitt. 1998. An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics. Journal of Human Resources 33: 300-344.

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

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

Green, David A., Jonathan Rhys Kesselman and Lindsay M. Tedds. 2020. Covering All the Basics: Reforms for a More Just Society. Final Report of the British Columbia Expert Panel on Basic Income. Victoria: Government of British Columbia.

Green, David A. et. al. 2023. Basic Income and a Just Society: Policy Choices for Canada's Safety Net. Montreal: Institute for Research on Public Policy.

Green, David A., J. Rhys Kesselman, Daniel Perrin, Gillian Petit and Lindsay M. Tedds. 2023. "Effects of a Basic Income on Paid and Unpaid Work" chapter 14 in Green et.al. 2023, pp. 273-304.

Greenberg, David and Harlan Halsey. 1983. Systematic Misreporting and Effects of Income Maintenance Experiments on Work Effort: Evidence from the Seattle-Denver Experiment. Journal of Labor Economics, 1: 380-407.

Greenberg, David, Robert Moffitt and John Friedmann. 1981. Underreporting and Experimental Effects on Work Effort: Evidence from the Gary Income Maintenance Experiment. Review of Economics and Statistics, 63: 581-89.

Guettabi, Mouchine. 2019. What do we know about the effects of the Alaska Permanent Fund Dividend? Institute of Social and Economic Research, University of Alaska Anchorage.

Haagh, Louise. 2019. The Case for Universal Basic Income. London: Polity Press.

Handa, Sudhanshu, Kuisa Natali, David Seidenfeld, Gelson Tembo and Benjamin Davis. 2018. Can unconditional cash transfers raise long-term living standards? Evidence from Zambia. Journal of Development Economics 133, 42-65.

Hanna, Rema and Benjamin A. Olken. 2018. Universal Basic Incomes versus Targeted Transfers: Anti-Poverty Programs in Developing Countries. Journal of Economic Perspectives 32 (4) 201-226.

Hanushek, Eric A. 1987. Non-Labor-Supply Responses to the Income Maintenance Experiments. In Munnell, Alicia H., ed. Lessons from the Income Maintenance Experiments. Boston: Federal Reserve Bank of Boston Conference Series No. 30. pp 106-121.

Haushofer, Johannes and Jeremy Shapiro. 2016. "The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya" Quarterly Journal of Economics 1973-2042.

Haushofer, Johannes, Matthieu Chemin, Chaning Jang and Justin Abraham. 2019. Economic and Psychological Effects of Health Insurance and Cash Transfers: Evidence from a Randomized Experiment in Kenya. Working paper, November 11.

Heim, Bradley T. 2007. "The Incredible Shrinking Elasticities: Married Female Labor Supply, 1978-2002" Journal of Human Resources 42 (4) 881-918.

Hoynes, Hilary and Jesse Rothstein. 2019. Universal Basic Income in the U.S. and Advanced Countries. Annual Review of Economics, 11:929-958.

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

Hum, Derek and Wayne Simpson. 1993. Economic Response to a Guaranteed Annual Income: Experience from Canada and the United States. Journal of Labor Economics, 11: S263-S296.

Hutchens, Robert. 1978. Changes in AFDC tax rates 1967-1971. Journal of Human Resources, 13: 60-74.

Jaroszewicz, Ania, Jon M. Jachimowicz, Oliver P. Hauser and Julian Jamison. 2022. How Effective is (More) Money? Randomizing Unconditional Cash Transfer Amounts in the US. Working Paper, July 2022.

Keeley, Michael C. and Philip K. Robins. 1980. Experimental Design, the Conlisk-Watts Assignment Model and the Proper Estimation of Behavioral Response. Journal of Human Resources, 13: 3-36.

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

Londono-Velez, Juliana and Pablo Querubin. 2022. The Impact of Emergency Cash Assistance in a Pandemic: Experimental Evidence from Columbia. Review of Economics and Statistics 104 (1) January 157-165.

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

Marinescu, Ioana. 2017. No Strings Attached: The Behavioral Effects of U.S. Unconditional Cash Transfer Programs. Roosevelt Institute, May.

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

McGuire, Joel, Caspar Kaiser and Anders M. Bach-Mortensen. 2022. A systematic review and meta-analysis of the impact of cash transfers on subjective well-being and mental health in low- and middle-income countries. Nature Human Behaviour 6 (March) 359-370.

Meyer, Bruce D. and Dan T. Rosenbaum. Welfare, the Earned Income Tax Credit and the Labor Supply of Single Mothers. The Quarterly Journal of Economics, 116: 1063-1114.

Moffitt, Robert A. 1992. Incentive Effects of the U.S. Welfare System: A Review. Journal of Economic Literature, 30: 1-61.

Moffitt, Robert A. 2003. The Negative Income Tax and the Evolution of U.S. Welfare Policy. Journal of Economic Perspectives, 17: 119-140.

Moffitt, Robert A. and Kenneth C. Kehrer. 1981. The Effect of Tax and Transfer Programs on Labor Supply: The Evidence from the Income Maintenance Experiments. Research in Labor Economics, 4: 103-150.

Munnell, Alicia H., ed. 1987. Lessons from the Income Maintenance Experiments. Boston: Federal Reserve Bank of Boston Conference Series No. 30.

Murarka, Bina A. and Robert G. Spiegelman. 1978. Sample Selection in the Seattle and Denver Income Maintenance Experiment. SRI Technical Memorandum No. 1. Menlo Park: SRI International.

Murray Charles. 2016. In Our Hands: A Plan to Replace the Welfare State. Washington, D.C: AEI Press.

Pechman, Joseph A. and P. Michael Timpane. 1975. Introduction and summary. In Work incentives and income guarantees: The New Jersey negative income tax experiment, ed. Joseph A. Pechman, and P. Michael Timpane. Washington, D.C.: The Brookings Institution.

Pilkauskas, Natasha V., Brian A. Jacob, Elizabeth Rhodes, Katherine Richard and H. Luke Shaefer. 2022. The COVID Cash Transfer Study: The Impacts of an Unconditional Cash Transfer on the Wellbeing of Low-Income Families. May

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

Riddell, Chris and W. Craig Riddell. 2024. Welfare versus Work under a Negative Income tax: Evidence from the Gary, Seattle, Denver and Manitoba Income Maintenance Experiments. Journal of Labor Economics 42 (2) April.

Riutort, Sebastia, Bru Lain and Albert Julia. 2023. Basic Income at Municipal Level: Insights from the Barcelona B-MINCOME Pilot. Basic Income Studies 18 (1) 1-30.

Robins, Philip. 1985. A Comparison of the Labor Supply Findings of the Four Negative Income Tax Experiments. Journal of Human Resources, 20: 567-82.

Robins, Philip K. and Richard W. West. 1980. Program Participation and Labor Supply Response. Journal of Human Resources, 15: 499-523.

Simpson, Wayne, Greg Mason and Ryan Godwin. 2017. The Manitoba Basic Annual Income Experiment: Lessons Learned 40 Years Later. Canadian Public Policy, 43: 85-104.

Spiegelman, Robert G. and K. E. Yaeger. 1980. Overview. Journal of Human Resources, 15: 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/

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

Verho, Juoko, Kari Hamalainen and Ohto Kanninen. 2022. “Removing Welfare Traps: Employment Responses in the Finnish Basic Income Experiment. American Economic Journal: Economic Policy. 14(1) 501-522.

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

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

Table 1
Balancing Tests for Mincome

	Women						
	Age	Years of Education	Working in 1973	Weeks Worked 1973	Annual Hours Worked	Number of Children	Age of Youngest Child
Treatment Group	-1.45 (.922)	.022 (.281)	-.014 (.045)	-.934 (1.65)	-37.09 (38.46)	.153 (.158)	-.571 (.470)
Mean Dependent Variable	31.73 (.437)	9.77 (.133)	.450 (.022)	12.53 (.811)	275.39 (19.39)	2.61 (.074)	4.75 (.223)
Number of Individuals	485	484	482	485	489	489	445
	Men						
Treatment Group	-.992 (.956)	.359 (.357)	.030 (.025)	.539 (1.53)	54.80 (51.22)	.153 (.158)	-.571 (.470)
Mean Dependent Variable	34.61 (.454)	9.59 (.180)	.884 (.014)	37.94 (.851)	951.50 (27.96)	2.61 (.074)	4.75 (.223)
Number of Individuals	489	488	483	489	489	489	445

NOTES—Standard errors are in parentheses. All regressions include controls for experimental stratification (family size-adjusted pre-random assignment income). Estimation is by OLS. Statistical significance indicates as follows: * for 10% level, ** for 5% level, and *** for 1% level.

Table 2
Balancing Tests for DIME

	Women						Men					
	Received Welfare	Employed	Hours Worked	Earnings	Age	Number Children	Received Welfare	Employed	Hours Worked	Earnings	Age	Number Children
Treatment Group	.025* (.015)	.008 (.019)	2.33 (2.94)	2.72 (6.60)	-.019 (.488)	-.096 (.077)	.022 (.015)	.009 (.014)	-3.92 (3.19)	-5.67 (12.05)	.123 (.501)	-.106 (.075)
Mean Dependent Variable	.139	.289	40.86	83.05	29.35	2.14	.137	.842	150.88	470.55	32.14	2.14
Number of Individuals	1459	1459	1459	1458	1459	1459	1503	1503	1503	1499	1503	1503
Sample Size	21829	21814	21814	21684	1459	1459	22518	22518	22518	22077	1503	1503

NOTES—Standard errors are in parentheses, and are clustered on the individual for regressions based on multiple months pre-random assignment. All regressions include controls for experimental stratification (interactions of family size-adjusted pre-random assignment income and race). Estimation is by OLS. Statistical significance indicates as follows: * for 10% level, ** for 5% level, and *** for 1% level.

Table 3
Balancing Tests for Mincome: Marital Satisfaction and Household Time Allocation

	Women			Men		
	Agree on who does housework	Satisfied with husband's help around house	Happy with marriage	Agree on who does housework	Satisfied with wife as homemaker	Happy with marriage
Treatment Group	-.011 (.050)	-.010 (.045)	.009 (.032)	-.023 (.049)	-.018 (.031)	-.037 (.027)
Mean Dep. Variable (Controls)	.515	.725	.875	.596	.906	.943
Number of Individuals	434	434	434	430	430	430
Sample Size	434	434	434	430	430	430

NOTES—Standard errors are in parentheses. All regressions include controls for experimental stratification (family size-adjusted pre-random assignment income). Estimation is by OLS. Statistical significance indicates as follows: * for 10% level, ** for 5% level, and *** for 1% level.

Table 4a
Tests for Attrition Bias: Mincome

	Women					
	Working in 1973	Weeks Worked in 1973	Annual Hours Worked	Agree on who does housework	Satisfied with husband's help around house	Happy with marriage
Attritor	-.049 (.046)	2.41 (1.70)	-24.59 (40.71)	.012 (.052)	-.062 (.048)	-.025 (.034)
Number of observations	498	501	498	436	436	437
	Men					
	Working in 1973	Weeks Worked in 1973	Annual Hours Worked	Agree on who does housework	Satisfied with wife as homemaker	Happy with marriage
Attritor	-.012 (.028)	-.931 (1.81)	56.24 (59.99)	.021 (.052)	-.035 (.033)	-.021 (.030)
Number of observations	499	505	502	434	435	434

NOTES—Standard errors are in parentheses. Estimation is by OLS. All regressions include demographic controls (number of children, age of youngest child, education, age).

Table 4b
Tests for Attrition Bias: DIME

	Women		Men	
	Employed	Hours Worked	Employed	Hours Worked
Attritor	-.007 (.020)	-2.03 (3.18)	-.014 (.015)	-3.52 (3.41)
Number of Individuals	1459	1459	1459	1459
Sample Size	21814	21814	21814	21814

NOTES—Standard errors are in parentheses. Estimation is by OLS. All regressions include demographic controls (number of children, age of youngest child, education, age). Labor market outcomes are measured over the pre-experimental period.

Table 5
Estimated Treatment Effects: Happiness and Household Time Allocation

	Women					Men				
	Happy	Agree on who does housework	Satisfied with husband's help around house	Happy with marriage	Church visits	Happy	Agree on who does housework	Satisfied with wife as homemaker	Happy with marriage	Church visits
Treatment Group	.050** (.024)	.092 (.060)	.110** (.056)	.095** (.041)	1.01*** (.263)	.080*** (.026)	.030 (.061)	.072* (.040)	.150*** (.059)	-.078 (.286)
Mean Dep. Variable (Controls)	.871	.467	.625	.809	4.80	.796	.533	.836	.819	4.60
Number of Individuals	480	295	295	295	374	488	294	294	296	345
Sample Size	1133	295	295	295	766	1156	294	294	296	704

NOTES—Standard errors are in parentheses, and are clustered on the individual where appropriate. Statistical significance denoted by *** for the 1% level, ** for the 5% level and * for the 10% level. The online appendix (section 3) provides definitions of all variables. All regressions include fixed effects for the experimental stratification (income cell) and survey/time, in addition to demographic controls (number of children, age of youngest child, education, age, work pre random assignment). Estimation is by OLS.

Table 6
Estimated Treatment Effects: Financial Well-being

	Women		Men	
	Unhappy with standard of living	Present financial situation is poor	Unhappy with standard of living	Present financial situation is poor
Treatment Group	.001 (.040)	.009 (.031)	.036 (.044)	.050 (.032)
Mean Dependent Variable (Controls)	.113	.356	.123	.357
Number of Individuals	296	489	295	489
Sample Size	296	1142	295	1146

NOTES—Standard errors are in parentheses, and are clustered on the individual where appropriate. Statistical significance denoted by *** for the 1% level, ** for the 5% level and * for the 10% level. The online appendix (section 3) provides definition of all variables. All regressions include fixed effects for the experimental stratification (income cell) and survey/time, in addition to demographic controls (number of children, age of youngest child, education, age, work pre random assignment). Estimation is by OLS.

Table 7
Estimated Treatment Effects: Labor Supply

	Mincome, Women		Mincome, Men		DIME, Women		DIME, Men	
	Hours Worked	Employed	Hours Worked	Employed	Hours Worked	Employed	Hours Worked	Employed
Treatment Group	-48.35*** (19.67)	-.065* (.039)	-27.47 (19.00)	-.036 (.024)	-8.11*** (2.89)	-.061*** (.017)	-11.10*** (3.06)	-.030** (.013)
Mean Dep. Variable (Controls)	212.2	.449	505.12	.846	40.86	.288	150.88	.842
Coefficient in annual hours	144	-	90	-	96	-	132	-
Number of Individuals	418	418	420	420	1321	1321	1332	1332
Sample Size	3159	3159	3162	3162	52856	52856	51720	51720

NOTES—Standard errors are in parentheses, and are clustered on the individual. Statistical significance denoted by *** for the 1% level, ** for the 5% level and * for the 10% level. All regressions include fixed effects for the experimental stratification (income cell for Mincome, income cell*race*program length for DIME) and survey/time, in addition to demographic controls (number of children, age of youngest child, education, age, work pre random assignment). Estimation is by OLS.

Table 8
Estimated Treatment Effects: Labor Supply, Heterogeneity by Household Composition, Mincome

	Women				Men			
	(1) Youngest child under 6		(2) Youngest child 6+		(3) Youngest child under 6		(4) Youngest child 6+	
	Hours Worked	Employed	Hours Worked	Employed	Hours Worked	Employed	Hours Worked	Employed
Treatment Group	-47.31** (22.32)	-.090* (.048)	-61.62* (34.46)	-.043 (.061)	-47.66** (21.07)	-.062** (.026)	36.88 (40.02)	.046 (.052)
Mean Dep. Variable (Controls)	181.43	.422	262.83	.492	513.78	.897	490.86	.762
Number of Individuals	270		148		272		148	
Sample Size	2119		1040		2130		1032	

NOTES—Standard errors are in parentheses, and are clustered on the individual. Statistical significance denoted by *** for the 1% level, ** for the 5% level and * for the 10% level. All regressions include fixed effects for the experimental stratification (income cell) and survey/time, in addition to demographic controls (education, age, work pre random assignment).

Table 9
Estimated Treatment Effects: Labor Supply, Heterogeneity by Household Composition, DIME

	Women						Men					
	Youngest child < 6		Youngest child 6+		No Children		Youngest child < 6		Youngest child 6+		No Children	
	Hours Worked	Employed	Hours Worked	Employed	Hours Worked	Employed	Hours Worked	Employed	Hours Worked	Employed	Hours Worked	Employed
Treatment Group	-8.95*** (3.20)	-.062*** (.019)	-4.03 (6.65)	-.028 (.041)	2.66 (11.44)	-.015 (.066)	-15.73*** (3.52)	-.048*** (.016)	-14.30** (6.59)	-.047* (.027)	10.01	.056 (.070)
Mean Dependent Variable (Controls)	33.9	.223	53.4	.379	58.1	.408	143.5	.821	147.5	.849	122.5	.736
Number of Individuals	998	998	300	300	126	126	1029	1029	311	311	133	133
Sample Size	39439	39439	11500	11500	4897	4897	39114	39114	11594	11594	4958	4958

NOTES—Standard errors are in parentheses, and are clustered on the individual. Statistical significance denoted by *** for the 1% level, ** for the 5% level and * for the 10% level. All regressions include fixed effects for the experimental stratification (income cell*race*program length) and survey/time, in addition to demographic controls (education, age, work pre random assignment). Estimation is by OLS.

Table 10
Estimated Treatment Effects: Labor Supply, Local Average Treatment Effects, Alternative Definitions/Samples

	Mincome, Household Employed, Minc2 Admin Data, Monthly panel		Mincome, Men (Periodic 8 Self-report)		Mincome, Women (Periodic 8 Self-report)		DIME, Women		DIME, Men	
	Employed (OLS)	Employed (LATE)	Hours Worked	Employed	Hours Worked	Employed	Hours Worked	Employed	Hours Worked	Employed
Received NIT payment	-.056** (.027)	-.125** (.058)	-55.84 (48.47)	-.076 (.061)	-132.8*** (50.73)	-.190* (.100)	-15.4*** (5.46)	-.117*** (.032)	-23.9*** (6.49)	-.066** (.028)
First stage F statistic	-	54.2	60.9		57.3		85.0		84.8	
Mean Take- up	.48		.43		.43		.52		.45	
Number Individuals	328		332		331		1321		1332	
Sample size	9409		2770		2768		52856		51720	

NOTES—Mean household employment rates from Minc2 is 0.88. Standard errors are in parentheses, and are clustered on the individual. Statistical significance denoted by *** for the 1% level, ** for the 5% level and * for the 10% level. Estimation is by 2SLS, where receipt of the negative income tax is instrumented with random assignment. All regressions include fixed effects for the experimental stratification (income cell for Mincome, income cell*race*duration for DIME) and survey/time, in addition to demographic controls (number of children, age of youngest child, education, age, work pre random assignment).

Table 11

Estimated Treatment Effects: Labor Supply, Heterogeneity by Household Composition, Mincome, Local Average Treatment Effects

	Women				Men			
	(1) Youngest child under 6		(2) Youngest child 6+		(3) Youngest child under 6		(4) Youngest child 6+	
	Hours Worked	Employed	Hours Worked	Employed	Hours Worked	Employed	Hours Worked	Employed
Received NIT payment	-138.73*** (50.90)	-.258** (.121)	-124.11 (106.29)	-.103 (.189)	-96.94** (46.23)	-.124** (.057)	119.24 (126.40)	.136 (.165)
First-stage F-statistic	46.5		16.9		46.3		19.8	
Number of Individuals	222		109		223		109	
Sample Size	1879		889		1881		889	

NOTES—Standard errors are in parentheses, and are clustered on the individual. Statistical significance denoted by *** for the 1% level, ** for the 5% level and * for the 10% level. All regressions include fixed effects for the experimental stratification (income cell) and survey/time, in addition to demographic controls (education, age, work pre random assignment).

Table 12
Estimated Treatment Effects: Heterogeneity by Household Composition, DIME, Local Average Treatment Effects

	Women						Men					
	Youngest child < 6		Youngest child 6+		No Children		Youngest child < 6		Youngest child 6+		No Children	
	Hours Worked	Employed	Hours Worked	Employed	Hours Worked	Employed	Hours Worked	Employed	Hours Worked	Employed	Hours Worked	Employed
Received NIT payment	-15.68*** (5.56)	-.110*** (.034)	-9.49 (15.60)	-.066 (.096)	7.89 (33.34)	-.046 (.190)	-32.11*** (7.09)	-.098*** (.031)	-38.72** (16.89)	-.128* (.069)	28.57 (42.9)	.159 (.208)
First Stage F-Statistic	424.6	424.6	343.5	343.5	11.8	11.8	86.6	86.6	685.4	685.4	41.3	41.3
Number of Individuals	998	998	300	300	126	126	1029	1029	311	311	133	133
Sample Size	39439	39439	11500	11500	4897	4897	39114	39114	11594	11594	4958	4958

NOTES—Standard errors are in parentheses, and are clustered on the individual. Statistical significance denoted by *** for the 1% level, ** for the 5% level and * for the 10% level. All regressions include fixed effects for the experimental stratification (income cell) and survey/time, in addition to demographic controls (education, age, work pre random assignment). Estimation is by 2SLS where receipt of the NIT is instrumented with random assignment.

Figure 1
Take-up Rates in Mincome and DIME

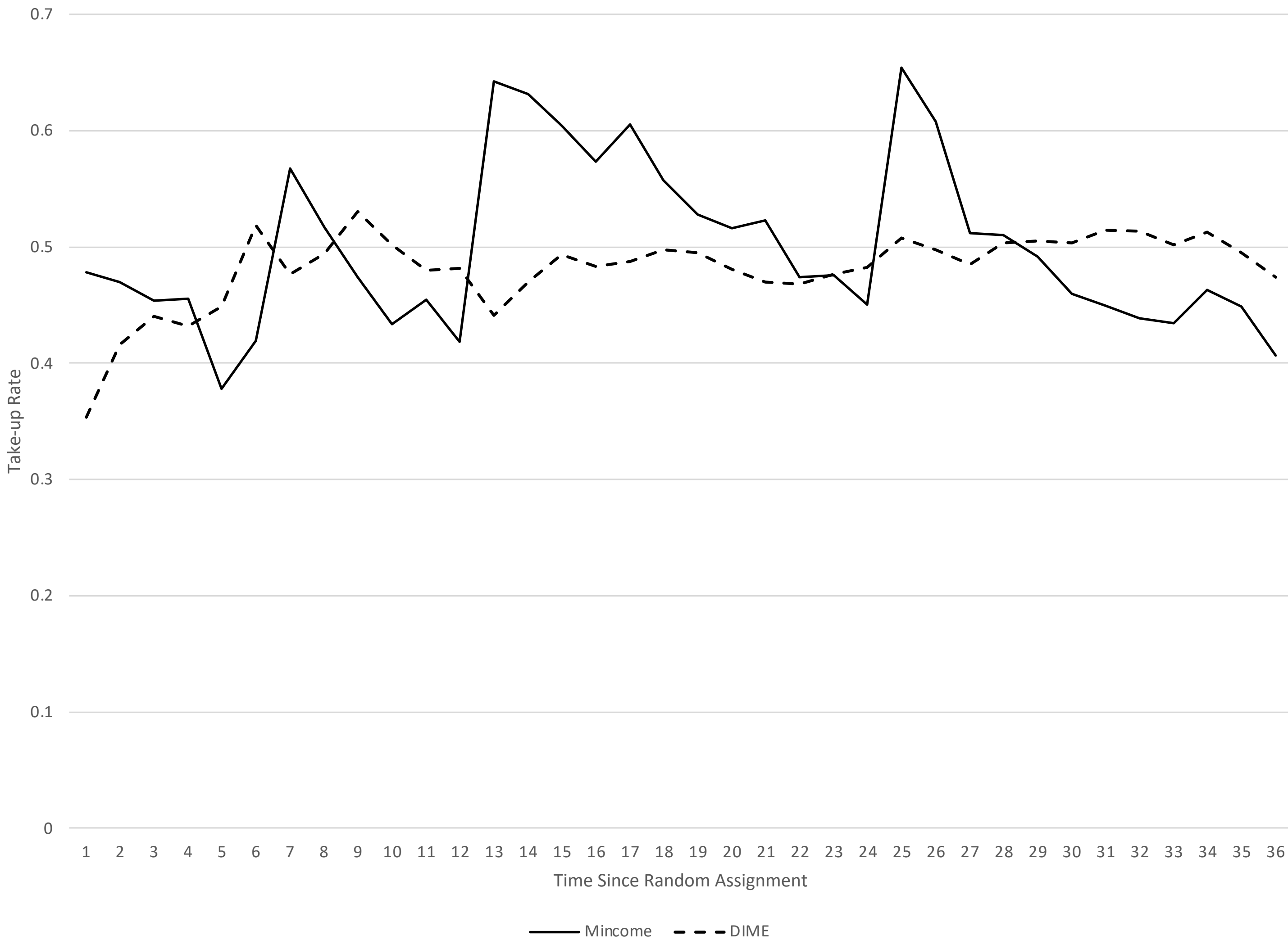

Table A1
Balancing Tests for Mincome: Minc2 Administrative Data Sample

	Women						
	Age	Years of Education	Working in 1973	Weeks Worked 1973	Annual Hours Worked	Number of Children	Age of Youngest Child
Treatment Group	-1.25 (.1.25)	-.269 (.411)	-.038 (.069)	-2.52 (2.43)	-77.65 (57.88)	.178 (.230)	-.415 (.622)
Mean Dependent Variable	30.83 (.482)	9.94 (.158)	.458 (.027)	12.52 (.988)	292.82 (24.42)	2.66 (.090)	4.39 (.247)
Number of Individuals	333	332	331	333	336	336	307
	Men						
Treatment Group	-.873 (1.34)	.532 (.532)	.023 (.037)	-.024 (2.25)	25.15 (78.63)	.179 (.230)	-.415 (.622)
Mean Dependent Variable	33.92 (.523)	9.88 (.205)	.902 (.015)	38.08 (.977)	980.50 (34.05)	2.66 (.089)	4.29 (.248)
Number of Individuals	336	335	333	335	336	336	307

NOTES—Standard errors are in parentheses. All regressions include controls for experimental stratification (family size-adjusted pre-random assignment income). Estimation is by OLS. Statistical significance indicates as follows: * for 10% level, ** for 5% level, and *** for 1% level.

Table A2
Balancing Tests: Gary

	Women					Men				
	Received Welfare	Employed	Hours Worked	Age	Number Children	Received Welfare	Employed	Hours Worked	Age	Number Children
Treatment Group	-.055** (.024)	-.002 (.046)	-.698 (1.81)	-.018 (.880)	.304** (.157)	.009 (.009)	.062*** (.024)	3.78*** (1.07)	-.840 (1.22)	.312** (.161)
Mean Dependent Variable	.069	.219	7.75	39.13	2.17	.014	.890	35.50	42.29	2.17
Number of Individuals	540	564	564	582	582	550	541	541	557	557
Sample Size	2195	2275	2275	582	582	2263	2193	2193	557	557

NOTES—Standard errors are in parentheses, and are clustered on the individual for regressions based on multiple months pre-random assignment. All regressions include controls for experimental stratification (interactions of family size-adjusted pre-random assignment income and locations). Estimation is by OLS. Statistical significance indicates as follows: * for 10% level, ** for 5% level, and *** for 1% level.

Table A3
Balancing Tests: SIME

	Women						Men					
	Received Welfare	Employed	Hours Worked	Earnings	Age	Number Children	Received Welfare	Employed	Hours Worked	Earnings	Age	Number Children
Treatment Group	-.045*** (.012)	-.025 (.023)	-6.06* (3.58)	-47.11*** (8.21)	-.605 (.590)	.086 (.085)	.044*** (.012)	-.013 (.019)	-2.60 (.2.98)	-159.05*** (14.23)	.158 (.627)	.083 (.086)
Mean Dependent Variable	.091	.300	42.84	85.81	32.77	1.99	.090	.731	123.92	392.43	35.56	1.99
Number of Individuals	1157	1157	1157	1156	1157	1157	1156	1156	1156	1156	1156	1156
Sample Size	19073	19073	19073	18944	1157	1157	19062	19062	19062	19062	1156	1156

NOTES—Standard errors are in parentheses, and are clustered on the individual for regressions based on multiple months pre-random assignment. All regressions include controls for experimental stratification (interactions of family size-adjusted pre-random assignment income and race). Estimation is by OLS. Statistical significance indicates as follows: * for 10% level, ** for 5% level, and *** for 1% level.

ONLINE APPENDIX

1. Experimental Estimates of Treatment Effects for Two-Parent Families[1]

	Men		Women	
	Annual hours	Employment rate	Annual hours	Employment rate
		GARY		
Robins (1985)	-35.4	-0.01	-57.6	-0.03
	(65.1)	(.03)	(62.7)	(.04)
Burtless (1986)[2]	-114 (6.5%)	N/R	+14 (5.0%)	N/R
		SIME – DIME		
Robins (1985)	-112.8***	-0.04***	-141.2***	-0.08***
	(30.1)	(.01)	(34.5)	(.02)
Burtless (1986)	-133 (7.1%)	N/R	-101 (14.2%)	N/R
SRI International (1983)	-133.1***	-0.05**	-101.4***	-0.11**
	(37.4)	(.01)	(35.8)	(.02)
		MINCOME		
Hum and Simpson (1993)[3]	-17(1%)		-15 (3%)	
		DIME		
SRI International (1983)	-149.6***	N/R	-103.2*	N/R
	(54.6)		(51.6)	
SRI International (1983)		SIME		
	-123.0**	N/R	-100.2*	N/R
	(50.8)		(49.3)	

[1] Standard errors are in parentheses below the estimated coefficients. Percentage changes relative to baseline are in parentheses beside estimated coefficients. ***, ** and * indicate statistical significance at 1%, 5% and 10% levels.
[2] Burtless and Hum and Simpson do not report employment rate estimates (denoted N/R) or standard errors.
[3] Estimates for men include single men (21% of all males in sample).

2. Well-being, Time Use and Related Survey Questions

Marital satisfaction and Time allocation:

Helping with work around the house.

Always disagree	1
Almost always disagree	2
Occasionally disagree	3
Almost always agree	4
Always agree	5

Men: As someone who is a good homemaker.
Women:

Very dissatisfied	1
Dissatisfied	2
Neither satisfied nor dissatisfied	3
Satisfied	4
Very satisfied	5

In general, how happy would you say you are with your marriage?

Very unhappy	1
Unhappy	2
Neither happy nor unhappy	3
Fairly happy	4
Very happy	5

Happiness:

Taken altogether, how would you say things are these days…would you say you were very happy, fairly happy, not too happy, or not happy at all?

Very happy	1
Fairly happy	2
Not too happy	3
Not happy at all	4

Standard of Living:

How happy are you with your present standard of living?

Very unhappy; wish I was much better off	1
Unhappy	2
Neither happy not unhappy	3
Happy	4
Very happy; very satisfied with what I've got	5

Financially poor:

I feel that my present financial situation is poor.

Strong agree	1
Agree somewhat	2
Neither agree nor disagree	3
Disagree somewhat	4
Strongly disagree	5

Church attendance:

How frequently do you go to worship services?

Never	1
Several times a year	2
Once a month	3
Two-three times a month	4
Every week	5
More than once a week	6

3. Replication of Hum and Simpson

a) Sample sizes:

We have tried to replicate the findings of Hum and Simpson across all family types, and have been unable to do so. Indeed, 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. These sample sizes match ours (which are based on the public use file, see the References). 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 are unclear. As noted in the main texrt, 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 raises doubts about the validity of their sample.

One potential issue that could explain the discrepancy are households that were deemed eligible for Mincome–this means that they appear in the 'baseline survey' (known as Minc1 in the public use files)—but were never enrolled in the experiment.[4]

[4] Mincome was unusual by modern social experiment standards in that there were *two* major pre-random assignment surveys: the baseline survey which was administered to all households deemed eligible for the NIT following the screening survey, and then the enrollment survey which was only administered to those enrolled in the experiment.

Such households were coded as a -1 in the treatment status variables (two such variables are available, ‘treat’ and ‘plan’). A possibility is that these households were included in the Hum and Simpson analysis which would both result in a larger sample size and also relatively more controls. Normally, a researcher would be unable to accidentally include such observations since they would have no post random assignment data, but due to a separate error with the hours worked variables (see below) these observations could have been included in estimation.

The table below (“Mincome Cell Counts across Samples”) shows cell counts –for the two head households and, for information purposes, all households—for different sample construction. Minc1 is the baseline which includes many households that were ultimately not enrolled in the Winnipeg experiment. The Minc1-Minc4 merge is the data that merges the post random assignment (labor market data only) periodics (often referred to as the longitudinal labor file) with the baseline data. In the table we show cell counts with and without dropping the non-enrolled where the non-enrolled are counted as part of the control group when included in the sample. We compare these different samples to the Hum and Simpson cell counts. While we still cannot replicate their sample, including the non-enrolled as control group households does substantially increase the number of control group observations, and move the total sample size much closer to Hum and Simpson. Below, we provide further discussion but we note here that these non-enrolled households appear in the data as 0s for hours worked, and not missing data. This therefore appears to be one plausible source of discrepancy in samples.

b) Hours worked errors:

Related to above, in attempting to replicate Hum and Simpson we also discovered an error in the hours worked variable (in the public use file). Specifically, hours worked that are missing were not coded as -9 (as the official Mincome documentation indicates), but rather are coded as a 0, i.e., the same number as individuals who did not work (for both men and women). This was never noted in the Hum and Simpson work (nor elsewhere in the Mincome literature such as Simpson 2017). Given that a classic econometric model for estimating treatment effects is a post-random assignment outcomes (such as hours worked) as a function of a treatment dummy (in this case experimental stratifications) and —to reduce residual variance—pre-random assignment characteristics, the result of this error is that the non-enrolled (i.e., who never participated in the experiment) could be included in the analysis and count as zero hours worked if coded as part of the control group. It is unclear ex ante how this would bias treatment effects, but clear it substantially reduces the mean hours worked and mean employment rate. To explore the implications of this error for treatment effects we estimate labor supply treatment effects for a sample that includes the non-enrolled and compare that to our estimates (see table below "Estimated Treatment Effects in Mincome: Hours Worked Measurement Checks".) Estimation reveals that this error does bias the treatment effects towards zero for both men and women, but we still obtain statistically significant effects for women on hours worked (although not for the probability of employment). Taken together, the non-enrolled/hours worked issues may explain part of the discrepancy between our estimates and Hum and Simpson, but not all of the differences.

Mincome Cell Counts across Samples

	(1) Minc1		(2) Minc1-Minc4 Merge		(3) Hum and Simpson (1991 page 58 Table 7-1)
	Drop -1	Inc -1	Drop -1	Inc -1	
G=$3800; t=0.35	39	39	32	32	29
G=$4800; t=0.35	45	45	40	40	39
G=$3800; t=0.50	55	55	47	47	41
G=$4800; t=0.50	72	72	61	61	59
G=$5800; t=0.50	43	43	38	38	35
G=$3800; t=0.75	50	50	44	44	40
G=$4800; t=0.75	30	30	26	26	24
G=$5800; t=0.75	46	46	37	37	35
Total Treatments	378	378	325	325	302
Control Group	218	414	181	376	348
Sample Size	598	794	506	701	650

NOTES: Minc1 counts are based on the stand-alone Baseline Survey (known as "Minc1" in the public use files). Minc1-Minc4 Merge is based on the households who appear in both Minc1 and the post random assignment longitudinal labour file (known as "Minc4 in the public use files). For the columns 'Drop -1' we exclude those households who were deemed eligible for Mincome following the screening survey and subsequently were administered the Baseline Survey, but ultimately not enrolled in the experiment. For the columns 'Inc -1' we include the non-enrolled as members of the control group.

Estimated Treatment Effects in Mincome: Hours Worked Measurement Checks

	Women				Men			
	(1) With Zero Hours/Missing Fix (Table 7)		(2) Without Zero Hours/Missing Fix		(3) With Zero Hours/Missing Fix (Table 7)		(4) Without Zero Hours/Missing Fix	
	Hours Worked	Employed	Hours Worked	Employed	Hours Worked	Employed	Hours Worked	Employed
Treatment Group	-48.35*** (19.67)	-.065* (.039)	-28.43* (16.29)	-.052 (.033)	-29.47 (19.01)	-.036 (.024)	-19.02 (23.80)	-.023 (.038)
Mean Dependent Variable	61.1	.387	61.2	.397	951.5	.869	945.9	.864
Number of Individuals	418		481		420		482	
Sample Size	3159		4329		3162		4338	

NOTES—Standard errors are in parentheses, and are clustered on the individual. Statistical significance denoted by *** for the 1% level, ** for the 5% level and * for the 10% level. All regressions include fixed effects for the experimental stratification (income cell) and survey/time, in addition to demographic controls (number of children, age of youngest child, education, age, work pre random assignment). Estimation is by OLS.