A Failure to Communicate:
The Labour Market Findings of the Negative Income Tax Experiments and Their Effects on Policy and Public Opinion

Karl Widerquist*

DRAFT, SEPTEMBER 2002. NOT TO BE QUOTED WITHOUT PERMISSION OF THE AUTHOR(S).

The views expressed in this paper are those of the author(s), and do not necessarily represent the views of BIEN or BIEN-Suisse.

* University of Oxford, United States.
Contents

Abstract ........................................................................................................................................ v

1. Introduction ................................................................................................................................ 1

2. The experiments ........................................................................................................................... 2

3. The work-disincentive results of the experiments ...................................................................... 9

4. What the experiments could not measure .................................................................................. 13

5. Political and media perceptions of the experiments ................................................................. 20

6. Conclusions ................................................................................................................................ 23

Bibliographical references A (a sampling of non-academic articles on the NIT experiments) ................................................................. 25

Bibliography B: Academic Articles on the NIT Experiments ........................................................ 27

List of tables

Table 1. Summary of the negative income tax experiments in the United States and Canada..................................................................................................................... 8

List of figures

Figure 1. Academic articles published each year on the NIT experiments (includes working papers, journal articles, and book chapters) .................................................................................. 9
Figure 2. The work disincentive effect ............................................................................................. 14
Figure 3. Workers receiving NIT ..................................................................................................... 14
Figure 4. Completely inelastic demand ............................................................................................ 17
Figure 5. Completely elastic demand ............................................................................................... 18
Figure 6. The range of possible market responses to a given horizontal shift in the supply of labour ........................................................................................................................................ 18
Abstract

This paper examines the labour market findings of the negative income tax experiments of the 1970s, and their impact on the public policy debate in the United States. Although both sides of the debate use statistics from the experiments to back up their arguments, the results are much more complex and less conclusive than one would hope. When the results were brought before Congress in the late 1970s, researchers failed to make the limits of their analysis clear giving the impression that the results were much more conclusive than they actually were.

Thanks to: Philippe Van Parijs, Michael Grossman, Robert Haveman, Robert Moffit, David Greenberg, and Robinson Hollister.
1. Introduction

Between 1968 and 1980, the United States Government conducted four negative income tax (NIT) experiments, and the Canadian government conducted one. They were designed to test the effects of a guaranteed income, which unconditionally assures all citizens some minimal level of income. The growing debate today about the basic income guarantee today is greatly affected by the labour market findings of those experiments. Although the modern basic income guarantee movement tends to focus on the basic income variant of the proposal rather than on the negative income tax as tested in the experiments the similarity between the two is so great that any conclusive findings from the experiments would be of great value for the current discussion. However, both basic income supporters and opponents quote the findings of these experiments with equal conviction.

At least 345 scholarly articles have been written on these experiments, but there is no clear consensus on what they implied for policy. The experimental results have been cited both by supporters and opponents of the redistribution of income as evidence for the workability or the unworkability of a negative income tax. For example, in 1993, long after the results were in and the initial flurry of articles was over, Hum and Simpson declared in the Journal of Labour Economics, “Few adverse effects have been found to date. Those adverse effects found, such as work response, are smaller than would have been expected without experimentation.” But in the same issue or the same journal, Anderson and Block speculated why social scientists continue to support the NIT “in the face of an avalanche of negative results” provided by the experiments.

Political perceptions of the experiments have been equally confused. The experiments received attention in the popular press in a few brief periods in the 1970s, most particularly in 1977 when Congressional hearings examined results from the experiments as part of their investigation of President Carter's ill-fated welfare reform proposal. The dozens of technical reports including large amounts
of data were simplified down to two statements: It decreased work effort and it increased divorce.

Dozens of editorials appeared in newspapers around the country criticizing the government for spending millions of dollars simply to show that people work less when you pay people not to work. The meaning of the results has been disputed by scholars, but the neither the results nor the disagreements about the results were understood by politicians or the media. Part of the reason for this misunderstanding is the natural difficulty of presenting complex technical results to a lay audience interested only in a bottom line. But part of the responsibility also rests with the scholars who presented bottom line results without clearly communicating just what these results did and did not show.

This paper examines the labour market results of the NIT experiments to determine what conclusions, if any, can be drawn from them conclusively, and how well these conclusions have been perceived by the media and the scholarly community. Part one summarizes the experiments. Part three discusses the ability of estimates of the work-disincentive effect to determine the market equilibrium outcome of a national policy.

2. The experiments

The United States Government sponsored 4 guaranteed income experiments between 1968 and 1980 (see Table 1.) The Canadian government got into the game with one experiment in the late 1970s. These experiments are known collectively as the income maintenance experiments, the guaranteed income experiments, or the NIT experiments. They began at a time when the elimination of poverty was the stated goal of the presidential administration, when there was a growing movement for economic rights, and when many social scientist and policymakers believed that social policy reform was heading in the direction of a guaranteed income. But by the time all of the results were available the movement for eliminating poverty had dwindled and the idea of “welfare reform” was beginning to be associated with dismantling rather than rationalizing the welfare system. To a large extent the NIT experiments simply outlived the movement that
spawned them, but to a small extent the experiments contributed to the demise of progressive social reform.

The primary aim of the NIT experiments was to test the effects of a guaranteed income on the work effort of recipients, and thereby to get some indication of the costs and feasibility of such a programme. Their secondary aim was to test the effects of a guaranteed income on any other effected variable the experimenters could measure. These variables included health statistics, educational attainment and performance, the divorce rate, and many others. But a discussion of these effects is beyond the scope of this paper.

The NIT experiments came about at a time when the negative income tax was being promoted by social scientists of various political backgrounds as a scientific solution to poverty. They were the first large-scale social experiment to use the scientific method of randomly assigning human subjects into treatment and control groups just as medical researchers do when testing drugs. Some social scientists have called the NIT experiments, “experiments in how to conduct experiments”. They have had much large influence in the conduct of social experiments than in the examination of the policy they were designed to test.

Table 1 summarizes the basic facts of the five NIT experiments. The first, the New Jersey Graduated Work Incentive Experiment (which is sometimes referred to as the New Jersey Negative Income Tax Experiment or simply the New Jersey Experiment), was conducted from 1968 to 1972. The researchers originally planned to conduct the entire experiment in New Jersey, but they were unable to find enough poor whites in New Jersey and had to open a second location in Wilkes-Barre, Pennsylvania in order to round out a racially representative sample. The treatment group originally consisted of 1,216 people and dwindled to 983 (due to drop outs) by the conclusion of the experiment. The sample size consisted of black, white and Latino, two-parent families with a male head, that were not approaching retirement, and with incomes below 150 per cent of the poverty line. Treatment group recipients received a guaranteed income for three years.

The Rural Income Maintenance Experiment (RIME) was conducted in rural parts of Iowa and North Carolina from 1970 to 1972. It functioned largely as a
supplement to the New Jersey experiment, which focused on an urban population. It began with 809 and finished with 729 experimental subjects. The treatment group received a guaranteed income for two years. Subjects met the same criteria as the New Jersey Experiment except that single parent, female-headed households were also included. Few, if any, Latinos were included in the sample. Both RIME and the New Jersey experiment began under the direction of Office of Economic Opportunity (OEO) and were completed by the Department of Heath, Education, and Welfare when OEO was disband.

The largest NIT experiment was the Seattle/Denver Income Maintenance Experiment (SIME/DIME), which had an experimental group of about 4,800 people in the Seattle and Denver metropolitan areas. The sampled included black, white, and Latino, families with at least one dependent and incomes below $11,000 for single-parent families and below $13,000 for two-parent families. The experiment began in 1970 and was originally planned to be completed within six years. However, researchers were interested in how the long-term effects of a permanent guaranteed income might be different from the short-term effects of a temporary guaranteed income experiment and so they obtained approval to extend the experiment for 20 years for a small group of subjects. This would have extended the project into the early 1990s, but it was eventually cancelled in 1980, so that a few subjects had guaranteed income for about nine years, during part of which time they were led to believe they would receive it for 20 years.

The Gary Income Maintenance Experiment (which is never abbreviated) was conducted between 1971 and 1974. Subjects were almost entirely black, single-parent families living in Gary, Indiana. The experimental group received a guaranteed income for three years. It began with a sample size of 1,799 families, which (due to a large drop-out rate) fell to 967 by the end of the experiment.

The Canadian government got into the business of conducting income maintenance experiments somewhat later. The Manitoba Basic Annual Income Experiment (Mincome) began in 1975 after most of the U.S. experiments were winding down. The sample included 1,300 urban and rural families in Winnipeg and Dolphin, Manitoba with incomes below $13,000 per year. By the time the
data collection concluded was completed in 1978, interest in the guaranteed income was seriously on the wane and the Canadian government cancelled the project before the data was analyzed. Fortunately, university-based researchers were eventually able to obtain and analyze the data, so that results are available today.

Two parameters are central to the design of any guaranteed income. The guarantee level or the minimum income level (G in Table 1) is the amount the recipient receives if she has no private income. The central goal of a guaranteed income programme is to ensure that no person’s (or no family’s) income falls below some given level for any reason. Theoretically, the guarantee level can be any number between zero and per capita GDP. A guarantee level that was too low would not significantly reduce poverty or increase income insecurity, but a guarantee level that was too high would have such strong work disincentive effects that the programme would not be affordable. The experiments intended to find out whether a guarantee level sufficient to seriously reduce or even eliminate poverty was feasible. For that reasons guarantee levels between 50 per cent and 150 per cent of the poverty line were tested.

The United States’ experiments all defined the guarantee level relative to the poverty line. A guarantee level of 1.0 or higher would eliminate poverty as defined by official statistics. The smaller the guarantee level is, the smaller the work disincentive and the smaller the cost of the programme will be, but the effect on the poverty rate will also be smaller. The larger the guarantee level is, the larger the effect on the poverty rate, but the higher the cost and the greater the work disincentive. The five experiments tested nine different guarantee levels. 0.5 (50 per cent of the poverty level) was tested in the New Jersey and Rural Income Maintenance Experiments. 0.75 was tested in all four of the United States experiments. 1.0 (just enough to eliminate official poverty) was tested in all of the United States experiments except SIME/DIME. 1.25 was tested in only in the New Jersey Experiment, and 1.26 and 1.48 were tested only in SIME/DIME. Mincome, which defined its guarantee level in Canadian dollars rather than relative to the poverty level, tested guarantee levels of $3,800, $4,800, and $5,800 per year.
The other central parameter of any guaranteed income system is the marginal tax rate (t in Table 1), also known as the “take-back rate.” The practical working of the marginal tax rate is slightly different if the guaranteed income is administered as a basic income rather than a negative income tax, but because all five of the experiments tested the negative income tax version, this small distinction is not important here. The take-back rate is the rate at which benefits are reduced as the recipient makes private income. That is, it is the effective income tax rate per dollar of private income for recipients of the negative income tax; hence the term marginal tax rate. A higher marginal tax rate is associated with a lower a overall tax-cost of programme\(^1\) but also with greater the work-disincentives, and a greater potential “poverty trap.” A lower marginal tax rates is associated with a higher overall cost of the programme, but also with greater work incentives. A lower marginal tax rate is also associated with a greater redistribution of income towards people with incomes above the poverty line. Redistribution to this group might be desirable in terms of equity (as a reward for low-wage workers), but to do so would greatly increase the cost of a programme primarily conceived as an anti-poverty policy\(^2\). For these reasons, it is important to know what kinds of take-back rates are feasible and the work-disincentive effects of each. The experimenters also tested nine different take-back rates. 0.3 (a 30 per cent marginal tax rate) was tested in the New Jersey and Rural Experiments. 0.35 was tested only in Mincome. 0.4 was tested only in Gary. 0.5 was tested in all of the experiments except Gary. 0.6 was tested only in Gary. 0.7 was tested in the New Jersey Experiment, RIME, and SIME/DIME. 0.75 was tested in Mincome. SIME/DIME tested two non-linear income functions with marginal tax rates of 0.7 minus 0.025 times income and 0.8 minus 0.025 times income. The effect of these two non-linear functions was to impose higher marginal tax rates on lower levels of income and lower marginal tax rates on higher levels of income.

\(^1\) Theoretically higher marginal tax rates could be associated with higher taxes costs if the supply of labour is highly elastic, but this was not expected and did not prove true in any of the experiments.

\(^2\) The basic income movement today puts less stress on the issue of poverty reduction and more stress on broader equity goals and to them this issue may be less important.
The use of so many different rates of G and t, reduced the numbers of subjects receiving each type of treatment, and therefore reduced the statistical reliability of the results for each. Some of this trade-off is worthwhile to allow for testing of a greater variety of potential parameters, but the experiments might have benefited from more coordinated effort to test a uniform group of parameters. A larger sample subject to three or four broadly spaced parameters might have been more beneficial than smaller groups subject to nine different and unevenly spaced parameters.

The primary goal of the experiments was to test the effects of G and t on work effort. The actual question was much more complex than what was usually reported. Most non-academic articles reported the simple summery statistics of how much less the treatment group worked than the control group but how the nine different levels of G and the nine different levels of t effected the work effort of men and women; primary, secondary, and tertiary household income earners; and whites, blacks, and Latinos in single-parent and two-parent families. From these results, researchers hoped to estimate the costs and effects of a national NIT programme and more generally to learn something about supply in the low-wage labour market. Table 1 summarizes the configuration of the experiments.
### Table 1. Summary of the negative income tax experiments in the United States and Canada

<table>
<thead>
<tr>
<th>Name</th>
<th>Site(s)</th>
<th>Data collection</th>
<th>Sample size</th>
<th>Sample characteristics</th>
<th>G*</th>
<th>t**</th>
</tr>
</thead>
<tbody>
<tr>
<td>The New Jersey Graduated Work Incentive Experiment</td>
<td>New Jersey and Pennsylvania</td>
<td>1968-1972</td>
<td>1,216 (983)</td>
<td>Black, white, and Latino, 2-parent families in urban areas with a male head aged 18-58 and income below 150% of the poverty line.</td>
<td>.5, .75, 1.0, 1.25</td>
<td>.3, .5, .7</td>
</tr>
<tr>
<td>The Rural Income-Maintenance Experiment (RIME)</td>
<td>Iowa and North Carolina</td>
<td>1970-1972</td>
<td>809 (729)</td>
<td>Both 2-parent families and female-headed households in rural areas with income below 150% of poverty line.</td>
<td>.5, .75, 1.00</td>
<td>.3, .5, .7</td>
</tr>
<tr>
<td>The Seattle/Denver Income-Maintenance Experiments (SIME/DIME)</td>
<td>Seattle and Denver</td>
<td>1970-1976, (some to 1980)</td>
<td>4,800</td>
<td>Black, white, and Latino families with at least one dependant and incomes below $11,000 for single parents, $13,000 for two parent families.</td>
<td>.75, 1.26, 1.48</td>
<td>.5, .7, .7-.025y</td>
</tr>
<tr>
<td>The Gary, Indiana Experiment</td>
<td>Gary, Indiana</td>
<td>1971-1974</td>
<td>1,799 (967)</td>
<td>Black households, primarily female-headed, head 18-58, income below 240% of poverty line.</td>
<td>.75, 1.0</td>
<td>.4, .6</td>
</tr>
<tr>
<td>The Manitoba Basic Annual Income Experiment (Mincome)</td>
<td>Winnipeg and Dauphin, Manitoba</td>
<td>1975-1978</td>
<td>1,300</td>
<td>Families with, head younger than 58 and income below $13,000 for a family of four.</td>
<td>$3,800, $4,800, $5,800 ***</td>
<td>.35, 5, .75</td>
</tr>
</tbody>
</table>

* G = the Guarantee level(s). Guarantee levels for the U.S. Experiments are reported as a percent of the poverty line, in Canada they are presented in Canadian Dollars.
** t = the marginal tax rate or the “take back rate.”
*** Canadian dollars

3. **The work-disincentive results of the experiments**

   Since 1966, the NIT experiments have been the subject of at least 340 scholarly articles, including working papers, journal articles, and book chapters (see bibliography B for a list). There is some overlap in this number because in some cases the same or a very similar article was published in all three forms. Most of these articles were published in the late 1970s as the experiments reached completion, but a trickle of articles reassessing the experiments continues today. Figure 1 shows the number of articles published each year on the experiments. About half of these articles deal with theoretical, methodological, and interpretational issues. Of those that report empirical findings, nearly half deal with the work disincentive effect.

**Figure 1. Academic articles published each year on the NIT experiments (includes working papers, journal articles, and book chapters)**

Source: Bibliography B, based on an extensive, but by no means exhaustive survey of the available literature.
Many of the researchers who conducted the experiments and others who examined the data, were strong backers of the programme and viewed the results as proving the feasibility of the NIT, but other researchers, as well as some politicians and members of the media, saw the results as proving the very opposite, that a national guaranteed income could not or should not be adopted. The experimental results seem to be a scientific Rorschach test in which an observer can see whatever she wants to see. The most important reason for this disagreement is that the most general result of the experiment was what everyone expected before the experiment was conducted: the treatment group worked less than the control group, all else equal. This agreed; the central question was how much less for each demographic group and for each level of G and t, and whether this work-disincentive effect fell into the acceptable range for a viable national NIT policy, but the experiments alone could not answer this question.

Two limitations of the NIT experiments kept them from giving a definitive answer to this question. First, there is no clearly agreed, objective criteria for how much effect on work effort is acceptable, allowing researchers with differing political views to draw opposite conclusions from the same results. Second, the experiments did not replicate the conditions of a national policy on a smaller scale, allowing researchers to claim more meaning to the effects of a national policy than the experiments warranted.

The work disincentive results met two criteria that were important to guaranteed income supporters.

First, the fear that a negative income tax would cause some segment of the population to completely withdraw from the labour force was not confirmed by the experiments. None of the experiments found evidence of such behaviour; the lower work effort of the treatment group relative to the control group took the form of increased weeks of unemployment between jobs, or fewer hours worked per week, but not the wholesale labour market withdrawal critics feared. The distinction between these two different types of work disincentive effects has not been well understood and opponents of the basic income guarantee continue to voice this fear despite the lack of experimental evidence for it.
Second, the cost of the programme was not so great as to make necessarily make the programme technically untenable. Critics of the guaranteed income have feared that work effort reductions would greatly increase the cost of the programme, requiring a large increase in taxes, which would further discourage work, and ultimately lead to the collapse of the programme. Certainly, whether this would happen depends on the level of the guarantee. It would inevitably happen if the minimum income was set near per capita national income, leaving a negligible work incentive, and certainly it would not happen with a guarantee level of $1 a year which would have a negligible work disincentive, but the results implied that a guarantee level between 50 per cent and 150 per cent of the official poverty level would be financially tenable. Although the work disincentives of the programme would increase the cost of an NIT over what it would have been if it had no effect on hours worked, it was not so much that it could make the programme unaffordable.

But, inevitably the experiments gave two important labour market results that NIT opponents could point to: there was a statistically significant work disincentive effect and that work disincentive increased the cost of the programme over what it would have been if work hours were unaffected by the NIT. Although these results were completely expected, they were reported in the press (see part 5) as if they were the critical findings of the experiments, and they largely shaped political and media perceptions that the experiments proved the failure of the guaranteed income.

Because the work disincentive effects of the NIT were greater than negligible but not so large as to make the programme unaffordable, the meaning of the figures depends on how large is large and how small is small. The work-disincentive effect seems to have been just enough that supporters can claim it to be small and opponents can claim it to be large.

Researches who look at the percentage decline in work effort can claim minimal effects. The basic findings were that male heads of households’ (men with wives and children) hours worked declined slightly if at all (relative to the control group). From 0 per cent to 9 per cent depending on the study and the data.
Work effort of female heads of households (single mothers) declined slightly more. Hours worked of married woman and hours worked of teenagers living with parent(s) declined more substantially - in the neighbourhood of 20 to 30 per cent. The decline in work effort among teenagers was not associated with increased hours of schooling, but it was associated with increased school performance in some studies. These figures are not terribly disturbing to guaranteed income supporters. Heads of households taking more time to look for work in between jobs, but not dropping out of the labour force is just the kind of result most supporters had hoped for. Parents spending more time with their children, and teenagers spending more time on their studies, are the kinds of benefits a guaranteed income would hope to give.

But as Keeley, Robins, Spiegelman, and West (1978) and Burtless (1986) independently concluded, even these modest reductions in work effort can cause significant increases in the cost of the programme (relative to what the cost would be if there was no work-effort response), because those whose earnings decline receive a larger share of the payments. Keeley, Robins, Spiegelman, and West found that work disincentive effects increased the cost of the programme by more than 50 per cent, and Burtless found that in some cases a tax expenditure of $3 would be required to raise the incomes of recipients by $1 (the rest going to work time reduction). Opponents of the guaranteed income have used these figures to claim that the project is a political non-starter.

But the meaning of these figures can be easily overblown. Imagine a minimum wage experiment in which the government picked out 1,000 low-wage workers and applied the minimum wage only to them. Such an experiment would no doubt find an enormous increase in unemployment in this group because they would be unable to compete with the lower cost of the millions of workers not receiving a minimum wage. But a minimum wage applied to all workers shows a small or even negligible relationship to the average unemployment rate, because the demand for labour responds differently to a change in supply of the market as a whole than it does to a change in supply small group in the market. To a lesser extent this problem affects the NIT experiments in the same way. There is much more that the experiments cannot show than they can. Part 3 discusses the limits
to what the NIT experiments could show about the market effects of a guaranteed income.

4. **What the experiments could not measure**

All of these precise and technical estimates are small in comparison to what we simply do not know about the effects of a national guaranteed income programme even if the responses of the experiment group are completely representative of the national response. At best the experiments measured the short run horizontal shift in labour supply caused by the experiments (the shift from A to B in Figure 2). But what we really want to know is the long run market response to a permanent national negative income tax (the shift from point A to point C in Figure 3). Without knowing the market response it is impossible to accurately estimate the cost of a national programme or its effects on work hours and poverty. To determine the shift from A to C, researchers would have to know several important pieces of information that the NIT experiments could not measure: First, how does the long-run shift in supply caused by a permanent national programme differ from the short run shift caused by a temporary experiment? Second, even if the experiments perfectly measured the long run supply shift caused by a national programme, to determine the market outcome we would also need to know the long run elasticities of both the supply of and the demand for labour (i.e. the shape of the supply and demand curves).

The vertical axis shows the wage (W), the horizontal axis shows the hours worked (H). The work disincentive effect causes the supply among the experimental group to shift from S to S₁. Because the experimental group is small in comparison to the size of the market the wage (W) is fixed and the shift from point A to point B shows only a decline in hours and no increase in the wage. The experiments give no information about the demand for labour or about the shape of the supply curve, only the size of the shift in supply (Figure 2).
If all workers received the NIT, the market would respond. The market outcome would go from a to C instead of a to B. This would reduce the drop in work hours, increase the income of recipients, and decrease the cost of the programme in terms of tax dollars and efficiency (Figure 3).

SIME/DIME made some attempt to measure the long-run shift in supply, but as all the researchers involved knew, these experiments were unable to make any estimates of the elasticities of supply or demand. The rest of the section examines these issues in turn.

The difference between the long-run supply shift of a national policy and the shift measured in temporary experiments has been well discussed. This is a
problem that is unique to social science experiments because they deal with human behaviour rather than human biology. If a thousand people in an experiment respond similarly to a vaccine, medical researchers can be reasonably sure that a million people will respond similarly in a national programme. But this may not be true in a social experiment. People behave differently in different circumstances. As Harold Watts described it, an experimental plan that recipients know will be in place for only a few years, is the equivalent of putting leisure time on sale. When laundry soap is on sale, people buy more of it, and we can expect a similar response from a temporary guaranteed income. People, who might want to take a few weeks or months off work sometime in the next ten years, might as well take it while the experiment is going on. Therefore one would expect that the experiments overestimated the decline in work effort.

However, NIT opponents can make the opposite claim just as logically. Because the experiments are only temporary, and recipients know that they must return to the workforce eventually, and will be less likely to drop out for fear of losing work experience or losing their place in line for promotion. Further, some NIT opponents have argued that a national guaranteed income would create a “culture of poverty” in which a subculture develops in which no one is ever expected to get a job. These claims are only speculation, but the NIT experiments were unable to shed any light on whether they are true or not. The only evidence provided by the experiments comes from the SIME/DIME recipients who received an NIT for nine years. These recipients did not behave terribly different from other experimental recipients, but they were only led to believe that their income would be permanent for part of that time, and it is uncertain whether they believed it. And clearly they would have been wise not to believe it. But even if the experiment had gone on for the full 20 years it could not have estimated whether a subculture of dropouts would develop if a national programme were put in place.

Therefore both those who want to believe that the long-run supply shift will be larger than the experiments showed and those who want to believe it is smaller have some theoretical justification for their claims, but that does not mean that the two cancel out; one effect could be much larger than the other. It simply means
that what we actually know about the labour market from these experiments is smaller relative to what we would like to know. The experiments simply do not give any information to answer these questions. It does mean, however, that those who make claims that the long-run effect is certainly larger than the experimental effect (Burless, 1986; Anderson and Block, 1993) are making unwarranted claims that are not supported by evidence.

The ability of the experiments to correctly measure the shift in supply is further complicated by the representativeness of the sample. Only families with low incomes were sampled. Most sampled only families with incomes below 150 per cent of the poverty line, and only SIME/DIME sampled families with incomes as high as 240 per cent of the poverty line. The higher one’s private income the less likely she would be to reduce their work effort in response to a programme giving them the possibility of a sub-poverty income without working. Because most of the population makes more than 150 per cent of the poverty line, the effects on the entire labour market would probably be much smaller than the effects measured in the experiment. However, arguably even a small decline in work effort among this group might be considered a serious problem. The experimenters were well aware that 7 per cent decline in work effort among recipients of an NIT would not mean that labour would decline by 7 per cent if an NIT programme was introduced, but if politicians and pundits understood this distinction they did not make it clear.

Further, very few if any single, childless individuals were sampled. Aside from single parents, this is the group that would be most likely to be expected to drop out of the labour force in response to a guaranteed income. Single parents at the time were eligible for a relatively generous (by current standards) AFDC programmes, and so the relative effect of the NIT would be reduced. Single, childless individuals are not eligible for any non-work-based benefits, and therefore, had they been sampled, one would have expected a larger response from this group, and researches may have found some of the wholesale withdrawals from the labour market that NIT opponents most fear.
These problems effect the accuracy with which the experiment measure the shift in the supply curve (the shift from A to B in Figure 1), but the shift in the supply curve does not give the market outcome, which is seriously effected by the elasticity of both supply and demand. The intuitive reason for this is that when supply decreases, demand responds by offering a higher price to elicit a return of the lost quantity supplied. How large this response is depends critically on the elasticities of both functions.

Examining the extreme cases can show the range of possible outcomes. A shown in Figure 4, if the demand for labour is perfectly elastic (if firms will hire any amount of labour at the going wage, but will not pay even a cent more for any amount of labour), the market equilibrium will be entirely determined by the horizontal shift in the supply of labour regardless of its elasticity. Figure 5 shows the effects of a perfectly inelastic demand for labour. If this is the case, firms need a fixed amount of workers and will pay anything to get it. If so, no amount of labour-disincentive effect will cause any long-run decrease in work effort, firms will pay workers whatever it taxes to keep doing the same amount of work. If so, the entire result of the work-disincentive effect would be to raise wages, and there would be no equilibrium decline in hours worked. If demand is completely inelastic, there is no equilibrium reduction in work hours (Figure 4). If demand is completely elastic (Figure 5), there is no change in the wage, and the full reduction in work hours in the experiments would occur in the market.

**Figure 4. Completely inelastic demand**
The more general results are that the equilibrium level of work effort will be somewhere between the initial equilibrium (point A) and the horizontal shift in supply (point B), and that the equilibrium wage will be as high or higher than the initial wage. In other words, the market equilibrium will be somewhere in the shaded area in Figure 6. Without information on elasticities, it is impossible to say where in this region the equilibrium would be. Thus, instead of identifying the equilibrium outcome of a negative income tax, the experiments identified only the lower left hand border of a region of possible outcomes.

It should be noted that it is theoretically possible for the equilibrium point to be in the region to the upper left of point B if the labour supply is backward bending. However, a backward bending labour supply can be safely ignored in the case of a substantial guaranteed income because backward bending requires that workers’ demand for goods is so inelastic that a decrease in wages will cause them to work more hours to maintain their level of consumption. That is quite
reasonable for someone who’s labour is the primary or their only source of income, but if a guaranteed income was in place, the lower the wage, the smaller portion of the worker’s income comes from that wage and it becomes unreasonable to believe that workers will work more and more to maintain the level of such a small part of their income.

If a backward-bending labour supply curve is ruled out, the experiments found upper-bound estimates for the decline in hours worked, upper-bound estimates for the cost of the programme, and lower-bound estimates for the effect of the programme on the income of recipients.

All of the researchers who worked with the data from these experiments were aware of this shortcoming of the available results. It is extremely basic economics, and it was pointed out as early as 1971 (Browning). Yet, few of the researchers who wrote on the NIT experiments treated this issue with more than a passing mention. It seems to me that there are two reasonable ways to present results under these circumstances. One option is to obtain the best available estimates for the elasticities and simulate the outcome. The other is to present range of possibilities. This would mean showing ranges like the one in Figure 5, and pointing out that to the extent the experiments were capable of correctly estimating the long-run shift in supply, they obtained only an upper-bound estimate of the effect on hours worked, a lower-bound estimate of the effect on the income of recipients, and an upper-bound estimate of the cost of the programme in terms of tax dollars. A few researchers took the first option; I know of none who took the other option.

The most common way of handling the results was to ignore or to effectively ignore the need to know the demand response. Some did not mention the need to know the demand response at all; others mentioned it only in passing as if it would have only a minor effect on the estimates, but nearly all presented their figures as if they were estimates of the equilibrium outcome rather than of the lower bound of possible outcomes. Whether a caveat was made or not, presenting estimates of boundaries of a range of possible outcomes as if they are point estimates of the actual values is clearly misleading. Certainly, economists
understood this, but it is clear from looking at the Congressional testimony and from examining articles in the popular media that the policymakers and pundits did not understand these issues. The technical experts failed in their responsibility to make their lay audience understand the meaning of the results they were presented with.

5. Political and media perceptions of the experiments

Hopefully, parts 2 and 3 have demonstrated that the findings of the NIT experiments are far more complex, subtle, and ambiguous than one might be led to believe by figures such as an X per cent decline in hours worked. But as this section shows, the complexity of the results was largely lost on the politicians and media to whom the findings were reported. Bibliography A contains a survey of about 50 articles from the popular media on the experiments.

The experiments gained significant attention in the press only twice: in 1970 and 1972, when Nixon’s Family Assistance Plan (FAP) was under debate in Congress, and in 1977 and 1978 when Carter’s Programme for Better Jobs and Income (PBJI) was under consideration. Both plans had elements of a negative income tax, but neither were a pure guaranteed income, although FAP was considerably closer to it than PBJI. In 1970, the first experiment had only been under way for two years and researchers believed that they were at least three years away from being able to produce meaningful results. But at the insistence of the administration and some members of Congress, the researchers released preliminary reports showing no evidence of any work disincentive effect. Some members of Congress (rightly) could not believe the result, and commissioned a review of the results from an independent auditor that concluded the results were “premature,” which was just what the researchers had initially warned.

The results of the fourth and largest experiment, SIME/DIME, were released while Congress was debating PBJI, and the existence of work disincentive effects caused quite a stir. Never mind that everyone going into the experiments agreed that there would be some work disincentive effect; when the results were
publicized, members of Congress were appalled and columnists across the country responded with a chorus of negative editorials decrying the guaranteed income, and ridiculing the government for spending millions of dollars to find out whether people work less if you pay them not to work.

The United Press International (1977) simply got the facts wrong saying that the SIME/DIME study showed that “adults might abandon efforts to find work,” when in fact no such evidence was found. The UPI apparently did not understand the difference between a decline in work hours while continuing to work, and abandoning the labour market. The Rocky Mountain News claimed that the NIT “saps the recipients desire to work.” Jones (1977) writing for the Seattle Times presented a relatively well-rounded understanding of the results, but despite this, simply concluded that the existence of a decline in work effort was enough to “cast doubt” on the plan. Similarly Rich (November 18, 1978) implied that the evidence presented that the NIT “might cause recipients to work less,” is enough to disqualify the programme from consideration. Raspberry (1978) declared the experiments a failure simply because people worked less.

Senator Daniel Patrick Moynihan who had written a book in support of the guaranteed income a few years early and who had been one of the architects of FAP, recanted his support for the guaranteed income as a result of the SIME/DIME findings. He is a sociologist and would be expected to have a sophisticated understanding of statistical data, but he implied in a letter to William F. Buckley later published by the National Review that the mere existence of a work disincentive effect was an important factor in his recantation. He stated, “But were we wrong about a guaranteed Income! Seemingly it is calamitous. It increases family dissolution by some 70 percent, decreases work, etc. Such is now the state of the science, and it seems to me we are honour bound to abide by it for the moment.” He held Congressional hearings on the results in November of 1977 to discuss the evidence, but media reports and politicians comments on the results did not betray a real understanding of the results.

Headlines such as “Income Plan Linked to Less Work,” and “Guaranteed Income Against Work Ethic” appeared in newspapers. The Knight News Service
Jodie Allen of the Labour Department commenting on Spiegelman’s cost estimates saying, “It could easily turn out that the government might spend billions of dollars on benefit payments and have little effect on the families’ incomes. Instead, most of the (government) expenditures would offset reductions in earnings” apparently in complete ignorance of the possibility of a demand response and the affect it could have to increase family earnings and reduce government expenditures. Only a few exceptions such as Carl Rowan for the Washington Star (1978) considered that it might be acceptable for people working in bad jobs to work less, but he could not figure out why the government would spend so much money to find out whether people work less when you pay them to stay home.

Spiegelman, one of the directors of SIME/DIME, defended the experiments in the Washington Star (1978), saying that the experiments provided much needed costs estimates that demonstrated the feasibility of the NIT. He said that the decline in work effort was not dramatic, but he did not offer an explanation for why so many commentators believed the results were dramatic and why they drew such different conclusions than he did. Demokovich (1978) is one of the few popular writers who considered the reduced work effort small, but the more common reaction was shown in the Denver Post (Brimberg 1980). Citing only that a work disincentive effect existed, not its size, Brimberg quoted Senator Bill Armstrong of Colorado as saying the experiment was, “An acknowledge failure. Let’s admit it, learn from it, and move on.”

It may be an impossible task to communicate such complexities to an audience interested only in sound bytes or in the bottom line, but I cannot help thinking that social scientists have a responsibility to do a better job than we did in this instance. None of the articles in the popular media that I was able to find betrayed any understanding that the experiments measured only the horizontal shift in the labour supply function. None seemed to understand the elementary economic principle that a change in supply necessitates a demand response that will greatly affect the equilibrium outcome. The understandings of the NIT experiments displayed in the popular press were so superficial, that it is
reasonable to accuse social scientists of failing to communicate the meaning of their results.

6. Conclusions

Even if the public had been made to understand more of the complexities of results, as long as there is a significant political block that believes any work disincentive is unacceptable, the NIT experiments were bound to give ammunition to NIT opponents. To that extent it was a mistake for any guarantee income supporters to agree to the experiments in the first place. One writer asked what would have happened if the introduction of Social Security had been preceded by a retirement insurance experiment. It would certainly have shown that it caused people to save less for their retirement and to retire sooner than they otherwise would have, giving considerable ammunition to Social Security opponents. But to those who believe that low-wage workers need more power in the labour market, and that a basic income guarantee can give them that power, if it is affordable, the NIT experiments demonstrated the feasibility of a desirable programme. And, therefore, the NIT experiments, as long as they are discussed, will always mean different things to different people. But these differences are more philosophical than scientific. None of the facts of the findings are persuasive enough that they should cause either supporters or opponents to change their minds.

Why was the supply response to the NIT experiments so widely ignored? One reason is that scientists like to focus on the results, not the limits of their research. Another reason is that they probably assumed this fact was too obvious to be bothered with among social scientists and too difficult to be dealt with by a lay audience. Perhaps, opponents didn’t want to bring it up because it waters down their argument that the work disincentive is “large” and the costs are “high.” Perhaps, supporters didn’t want to bring it up because it is much more difficult to make a case for NIT based on the desirable effect on wages that a work disincentive might have, than to make the case that the work-disincentive is “small.” Using the small argument requires only an objective look at empirical evidence. But using the desirability argument requires not only empirical data that the experiments could not produce, but also a much more complex ethical
argument. It affronts those who want to keep wages low to keep profits high and those who espouse the extreme version of the work ethic stating that everyone must at all times work. Although the basic income guarantee is not work-ethic friendly in that sense it is worker friendly because it allows those who do work to command higher wages, and it gives those who might consider not working a positive incentive (rather than a punitive incentive) to work.

Is it worth it to allow some to drop out to increase the wages of those who don’t? That answer depends on how big the increase in wages is and how big the decline in work effort is, whether it means few hours worked by many or dropping out by some, and on other issues. If dropping out means leaving the labour force to become an unhappy, idle soul who drains the resources of others, perhaps not. If dropping out means pursing artistic, educational, spiritual, entrepreneurial, or care-giving activities that will ultimately benefit others, perhaps so. These issues, both positive and normative, are the issues that separate supporters from opponents of the idea. The NIT experiments were able to shed only a small amount of light on a few of these issues. The most important questions went unanswered. The NIT experiments we able to indicate only very tentatively that a basic income guarantee is financially feasible at a cost of certain side effects that people with differing political beliefs may take to be desirable or disastrous. To claim more would be to overstate the evidence.
<table>
<thead>
<tr>
<th>Year</th>
<th>Article</th>
</tr>
</thead>
</table>


Bibliography B: Academic Articles on the NIT Experiments

(Alphabetical by author)

<table>
<thead>
<tr>
<th>Year</th>
<th>Citation</th>
</tr>
</thead>
</table>


Also, a negative income tax does not entail a cash transfer to wealthier individuals, so it costs the government less, according to this 2006 article in Basic Income Studies. Writing in 2011 for the journal Canadian Public Policy, Evelyn Forget, an economist at the University of Manitoba, found that the basic income moderately reduced poverty and improved residents’ health. Though Forget found some reduction in the number of hours people worked, much of that time was dedicated to education. These normative issues separate supporters from opponents of the basic income guarantee, and therefore, the NIT experiments, as long as they are discussed, will always mean different things to different people.