Jain Family Institute

December 9, 2024

Guaranteed Income In The Wild: Summarizing Evidence From Pilot Studies and Implications for Policy

By Jack Landry

Contents

568 Broadway, Suite 601
New York, NY, 10012

Summary

Years of advocacy have led to the launch of over one hundred guaranteed income pilots across the US in the past six years. Broadly, advocates hoped these programs would demonstrate the effectiveness of cash in fighting poverty and catalyze momentum for larger scale policy. Now, as many wind down, some of the largest pilot research projects—including one evaluated by my colleagues at the Jain Family Institute—are reporting long-awaited results.

Analysts and policymakers with an interest in unconditional cash have struggled to reconcile competing, often contradictory claims about the results of these recent findings. One headline declares that "study finds no-strings cash leaves the poor worse off" while another asserts, "results further prove guaranteed income's critical role in supporting families." A third suggests that "the report card on guaranteed income is still incomplete."[1] Remarkably, all three headlines refer to the same underlying study. Supporters and detractors have been quick to interpret this round of results in their favor, whether as proof that guaranteed income is an urgently needed antipoverty tool at the national level or a waste of resources with perverse incentives.

How do we make sense of this confusing mass of competing claims? And what interpretations can we safely draw about what "works" when it comes to unconditional cash? This report aims to provide a comprehensive overview of the results from guaranteed income pilot programs and a new framework for understanding their implications for larger-scale policy. We analyze a subset of pilots that meet the highest standards for program evaluation—studies designed to best protect against erroneously finding that guaranteed income has a more positive or negative impact than reality. While we write from an institutional position supportive of guaranteed income based on pre-existing evidence, we try to give fair hearing to the claims of critics and scrutinize the claims of advocates.[2]

One of the most important takeaways from the pilot studies is a recognition of their limitations. Guaranteed income pilots are often framed as analogous to clinical drug trials: if a test on a small group of people shows positive effects, it should be adopted at scale. However, this analogy oversimplifies the complexities of adopting any policy at scale—be it a targeted guaranteed income or a true universal basic income (UBI). First, there are inherent difficulties in extrapolating from a small short-term pilot to a broad permanent program. Permanency and universality would have different

outcomes—both positive and negative—and require complex policy decisions, most notably around financing, which pilot studies need not address. Second, the desirability of such a policy is not solely determined by its impact—it is also shaped by values and ideology, unlike the more straightforward evaluation of a drug's effectiveness.

While pilots cannot offer a simple yes or no answer on the desirability of a basic income, they do offer valuable new evidence on the effects of temporary large-scale cash transfers. The results resist the claims of both guaranteed income's most ardent supporters and detractors.

The pilots do not show that a nine- to 36-month guaranteed income has wholesale transformational impacts on recipients. For instance, there is little evidence that cash causes recipients to create valuable new businesses or go back to school. Similarly, the poor physical and mental health outcomes associated with poverty show limited signs of improvement from a simple cash infusion. Overall, the results cast doubt on the idea that temporary guaranteed income can solve a poverty trap, by which a short-term cash transfer can alleviate problems of chronic scarcity and precarity such that it has large positive impacts across an array of different domains.

These results suggest a broadly applied guaranteed income is unlikely to generate enough economic gains to significantly offset its costs through reduced social spending or increased tax revenue. Large-scale policy will still have to answer difficult questions about financing. By the same token, the results show short-term cash is not effective for everything, and an untargeted guaranteed income cannot piggyback on initiatives that have aims distinct from income support, such as improving health or promoting entrepreneurship. Transfers directed at families with children have a more plausible case for long-run economic returns, though this case relies on pre-existing evidence that the pilots are too time-limited to substantiate.

Many observers had hoped for evidence that the economic returns might match or exceed the cost of such programs; from this vantage point, some have interpreted these findings as a disappointment. In our view, however, such expectations are misguided. The government does not function as a business, and the demand that guaranteed income "pay for itself" is an unrealistic criterion we rarely, if ever, apply to other social support programs.

In our view, the pilot results clarify that the case for guaranteed income should not be premised on claims about the extraordinary efficacy of cash.

The results demonstrate, instead, that cash transfers help mitigate the income instability that comes with the low-wage labor market, and affords participants greater capacity for spending. They reduce poverty and material hardship. And they provide some financial freedom to recipients who face severe financial strain. While these results are not particularly sensational, they are important justifications for any program that provides cash relief to low-income Americas.

Moreover, the pilots refute many of the negative predictions made by critics. Aggregating evidence across studies, the average impact on labor force participation and hours worked is consistent with guaranteed income having no appreciable work disincentive. And the studies provide the strongest evidence yet that recipients do not spend no-strings-attached cash on 'vice goods' like alcohol, drugs, or gambling.

While the pilots all test the impact of cash as an addition to status-quo programs, they have important implications for the design of incumbent safety net programs. Currently, the US conditions almost all cash aid on work, and provides most assistance in the form of in-kind vouchers. A large body of research has found that these restrictions impose high costs: they prevent eligible individuals from accessing benefits, inflate administrative expenses, and diminish the value of aid for recipients. These costs are premised on the assumption that unrestricted cash would be put to poor use or lead to decreased labor force participation—an idea that is soundly rejected by the pilot evidence. Policymakers can greatly improve the efficacy of the safety net by providing more benefits in cash.

Weighing the evidence: key methodological criteria for guaranteed income pilot studies

To inform this report, I focus on a small subset of the dozens of active and recently completed guaranteed income pilots whose research arms meet a common set of standards for rigor. All studies have strengths and weaknesses, and no single pilot result is definitive. Others will have different epistemologies for weighing different types of evidence. However, I rely on these particular studies because their practices best ensure that any effects are driven by the cash itself rather than confounding factors, which can make guaranteed income look both far better or far worse than reality.

Key methodological criteria for guaranteed income pilots

- ☐ Randomized control trial
- ☐ Minimal differential attrition
- ☐ Large sample
- ☐ Pre-analysis plan
- ☐ Clear evidence of successful randomization

Specific criteria are listed above and explained in more detail in the appendix. In addition to these methodological standards, I restrict attention to US based studies with sustained cash transfers (rather than one-offs) and broad-based samples. While there is much policy interest in one-off emergency cash assistance and programs for specific populations, this report analyzes studies testing a broadly applied basic income in the US. Studies meeting all these requirements are listed below.

Study Name	Details
Compton Pledge	**Amount**: $300/month no children, $540/month one child, $720/month two children, distributed either quarterly or bi-weekly **Duration**: Two years, data collected after 18 months **Location**: Compton, California **Sample selection**: At least one household member aged 23 to 57, and had a household income below 220% of the federal poverty threshold. **Average household size:** 4.4 **Sample size receiving cash and responding to surveys:** ~346
Chelsea Eats	**Amount**: $200/month one person household, $300/month two person household, $400/month three plus person household **Duration**: Nine months, data collected after six months **Location**: Chelsea, Massachusetts **Sample selection**: Lottery weighted towards those with more self-reported financial need **Average household size:** 3.2 **Sample size receiving cash and responding to surveys:** ~1,026
OpenResearch	**Amount**: $1,000/month **Duration**: Three years, most outcome data weighted 70% two years and five months after transfers 30% one year and five months after transfers **Location**: Select counties in Illinois in Texas **Sample selection**: Representative sample of low-income adults under age 40 **Average household size:** 2.9 **Sample size receiving cash and responding to surveys:** ~980
Baby's First Years	**Amount**: $333/month, provided in debit card labeled "4MyBaby" **Duration**: Four years and three months, data collected throughout **Location**: New York City, greater New Orleans, the Twin Cities, and the Omaha metropolitan area **Sample selection**: New mothers at or near the federal poverty line **Average household size:** Unreported **Sample size receiving cash and responding to surveys:** ~380

What can we learn from pilots?

Before delving into the details of the pilot results, it is important to be candid about their limitations.[3] Universal basic income envisions a permanent and continuous income transfer to all Americans. Pilots are short-term transfers to a handful of individuals, typically restricted to lower income households in a specific geography. This difference means that the pilot results cannot be straightforwardly extrapolated to predict the impacts of policy at scale.[4] Beyond empirical limitations, the desirability of guaranteed income depends on values and implementation specifics to which pilots cannot speak to.

Empirical limitations

Some have argued pilots are unfairly weighed toward showing only cash's benefits or the "best case" for the impact of guaranteed income.[5] There is an element of truth to this critique. Without any offsetting spending reductions or tax increases, a UBI would be inflationary. And while most plans for a UBI do include financing, any negative impact of raising taxes is not captured by these pilots, which are typically funded by philanthropy or short-term local government surpluses.[6]

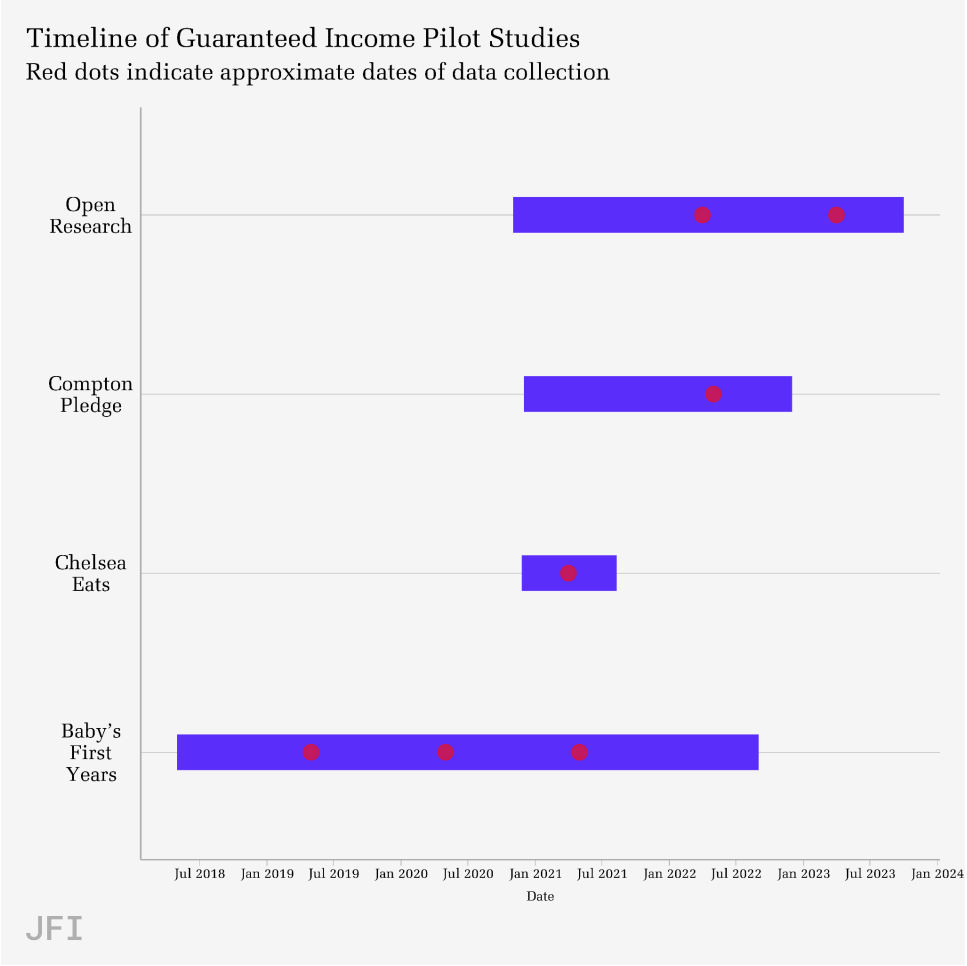
However, the elements not captured by short-term pilots also include many potential benefits to permanent policy. A nine month to three year-long guaranteed income (the range of lengths of pilots I review) might offer some reprieve from short-term financial pressures, but permanency would afford much stronger economic security. One recent review of various natural experiments that probed the effects of cash on children's outcomes noted that "all of the studies finding that family income is important are based on regularized income flows that parents could reasonably expect to persist for many years," and that "many of the studies that do not find evidence of substantial effects are based on research designs where the income event is either temporary or realized as an unanticipated lump-sum gain."[7] While guaranteed income pilots focus on adult outcomes, there is good reason to think the same dynamics should apply, and that cash transfers that last significantly longer than the duration of pilots are necessary to realize large long-run benefits. If it takes permanent policy to capture the positive impact of guaranteed income, pilot results could incorrectly be read as lackluster in demonstrating the efficacy of cash.

Even the macroeconomic effects of policy at scale are not all negative. In a small-scale pilot, if some people work less, they can be readily replaced by

others. However, if transfers were applied at scale, a drop in labor force participation would likely cause businesses to respond by increasing wages, as the labor market's new equilibrium shifts along the labor demand curve.[8]

Beyond the limitations of scale and duration, all the pilots I discuss were conducted, at least in part, during the significant disruptions of the Covid-19 pandemic. Specific study timelines are shown in the figure below. Given that participants were all from low-income households, they were especially likely to have been severely impacted; for example, the Compton Pledge reported that 7% of its participants had a household member die from Covid-19.

It is plausible that the pandemic context may have either amplified or dampened the effects of cash support. During a time of rising economic hardship, cash assistance could be especially valuable. Conversely, the pandemic also saw an unprecedented fiscal response that temporarily reduced poverty rates, potentially lessening the added value of the cash support provided by these pilots.[9] The disruptions to everyday life during the pandemic may have also reduced the relative impact of what would ordinarily be a significant change in a household's financial situation. Overall, all the short-term pilot studies fail to capture or include many elements that could both strengthen and weaken the argument for UBI.



Other limitations

Beyond the limits of the pilots as empirical evidence, the desirability (or lack thereof) of a universal basic income is in part a question of values—debates that cannot be settled by new research. The empirical evidence pilots deliver must be interpreted through a normative lens. Imagine the pilots show people work slightly less after receiving cash. Some people who view UBI as a potential answer to a post-work future might not view this as relevant at all. Others might be ideologically opposed to a transfer payment regardless of its effects on work. Of course, many people occupy a more middle ground, where the specifics of the evidence on the costs and benefits (like the magnitude of the effect on a reduction in work) will inform their policy preferences. However, even in this middle ground, different people can have the same

view of the evidence and come to different conclusions about the desirability of guaranteed income based on their values.[10]

Finally, the desirability of UBI can depend heavily on policy implementation specifics. Some debates about implementation are mostly ideological—some on the right are drawn to proposals that would replace the existing safety net with a much smaller and less progressive UBI, while some on the left sees UBI as a way to make society more redistributive. However, other specifics in large part hinge on empirical evidence not on offer here. For instance, some people are drawn toward universal basic income as a fix to alleged poverty traps—where safety net programs quickly stop providing benefits if people earn more money from work. To them, UBI offers a way to create a generous safety net without these perverse incentives. Evidence that the incumbent safety net creates poverty traps that can be solved by a UBI is mostly theoretical, and the pilots do not add to it. Pilots offer evidence about the impact of cash, but offer little evidence about how exactly a cash program would fit within the existing safety net.[11]

In sum, pilots can at best partly answer some empirical questions about the effects of a UBI, and cannot settle more normative debates over whether such a policy is desirable or what it should specifically look like. These are critical limitations to keep in mind when discussing all pilot results. No experiment can give an up or down answer on whether a UBI is a good idea.

Strengths of pilots

So what are pilots good for? In short, they give some of the clearest evidence of the impacts of large, short-term cash infusions. What happens to people when they get a cash transfer worth 20–40% of their household income for a year or more?

While there is extensive prior evidence on the effects of cash, these often capitalize on “natural experiments” where happenstances of history gives one group cash and another nothing.[12] Since researchers do not control these interventions, outcome variables can be limited to what happened to have been collected by other data sources. There is also more uncertainty about whether any difference between the group that gets the cash and the group that does not is driven by the effects of the cash itself rather than some pre-existing difference between the groups. Finally, the means by which cash is delivered in these “natural experiments” is not necessarily equivalent to a guaranteed income. For instance, lottery winnings or tax refunds might have

different effects than a continuous transfer. Pilot randomized control trials give researchers greater control over the nature of the intervention, rich outcome variables of their choosing, and greater confidence that any effects are driven by the cash itself.

While a small-scale experiment cannot capture the macroeconomic effects of policy, it can inform predictions about its impacts. For instance, if labor supply effects are small in the experiments, these second order macro effects are of minimal importance. Similarly, while the aggregate effects of a UBI on consumption and investment won't be realized in a small-scale experiment, the microeconomic changes identified by pilots can inform the evidence on how much consumption and investment would change.

Importantly, many plausible near-term implementations of guaranteed income are not large enough to cause significant macroeconomic impacts. For instance, the expanded Child Tax Credit costs about $100 billion dollars per year—not an insignificant sum, but small relative to the size of the American economy or proposals for a true UBI. The nonpartisan Joint Committee on Taxation estimated that the extra cost of the program from any macroeconomic feedback effects were small (less than 10%) in relation to the direct cost of sending the money.[13]

It's more difficult for short-term experiments to speak to the potential differential effects of more permanent policy. Short-term pilots simply cannot answer questions about the long-term impacts of continual cash transfers. A detailed examination of different outcomes can, however, point to some categories of results that are more amenable to immediate improvement from a cash infusion. Studies that track outcomes at multiple points over time also offer clues about how effects may change over time.

Pilot results

A brief note on technicalities

When summarizing results, I try to attend to the magnitude and precision of estimated effects. Just because a study estimates a statistically significant change does not mean the difference is substantively important—it depends on how big the change is. And likewise, just because a study estimates cash had no (statistically significant) effect on something, it matters how precise the answer is. For some outcomes, cash may have had an important positive impact, but the study sample was not large enough to detect it.

Without getting into statistical technicalities, if I say something is statistically significant, it means an effect is unlikely to be a result of a happenstance difference between control and treatment group. When I say a study can rule out effects beyond a certain size, it means effects larger than that are unlikely to have occurred by chance, but effects smaller than that are within the range of uncertainty for the study. For those interested in the technicalities, I use a significance threshold of $p < .05$ and say a study can rule out effects above or below the 95% confidence interval.[14]

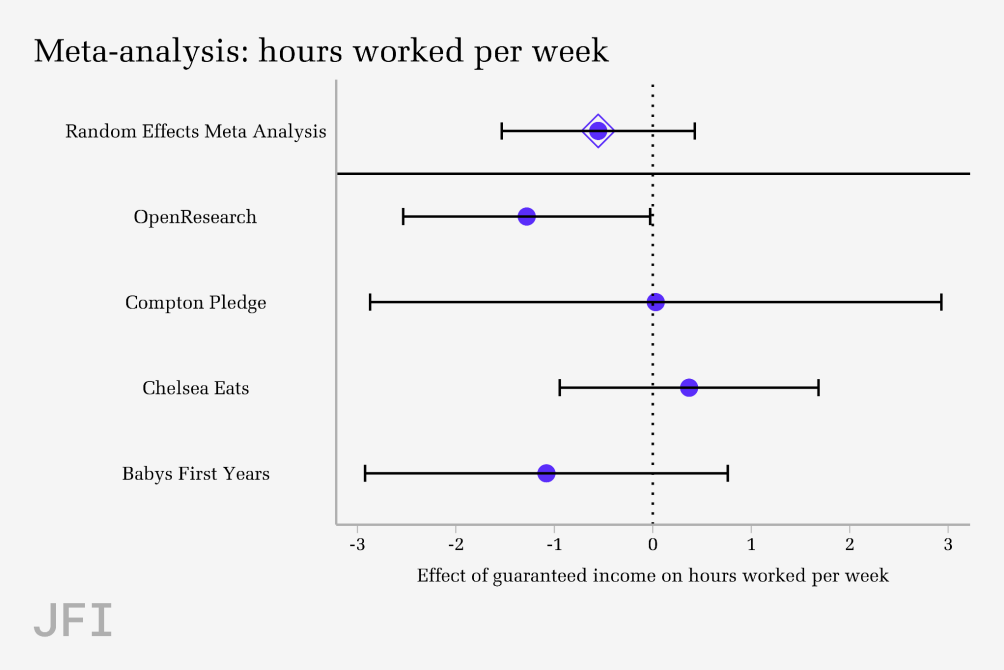
For several reasons, I generally do not summarize effects on different subgroups. First, the average effect of cash on low-income households is of first order importance. Both advocates and opponents of cash assistance mostly make claims about the effect of cash in general rather than on specific groups. Second, such analysis is likely to result in conclusions that reflect noise or chance rather than genuine subgroup effects. There are many possible subgroups to analyze, and sample sizes (when drilling down into subgroups) are small. Moreover, studies do not perform the subgroup analysis in the same way, the subgroup analysis is rarely pre-specified, and several studies don't do subgroup analysis at all. This dynamic means that highlighting subgroup results where cash seems particularly effective in a single study is likely to be the product of random chance rather than a real difference in impact across groups.[15]

Labor force participation

Across studies, effects on labor force participation are small. The most negative impact on labor force participation (having any kind of job) across studies is five percentage points—however, that effect itself is not statistically significant, and the effect on average hours worked in the same study is a positive .03 hours. Most other studies I review have similar stories—where any impact on working is small—even positive—and largely not distinguishable from random noise. If impacts on labor force participation were large, the evidence would not be nearly as equivocal.

The best evidence for the impact of unconditional cash on labor force participation combines evidence from multiple studies. Unlike many other outcome measures, questions about labor force participation are standard enough that it is straightforward to aggregate the evidence with meta-analysis, which combines the estimated impact on labor force participation from each study to come up with the best estimate of the average impact of guaranteed income on labor force participation across

studies.[16] It is not a simple average—larger studies with more statistical precision get more weight.



The results (displayed above) show that aggregating across studies, the expected change in hours worked in response to a typical cash transfer is a decrease of about half an hour per week.[17] The impact on employment is -1.26%—in other words, for every hundred people given an unconditional cash transfer, just one person would stop working.[18]

While guaranteed income programs do not appear to meaningfully reduce labor supply, there is also little evidence that they improve recipients' employment opportunities. Some researchers have hypothesized that a lack of money can be a barrier to finding consistent employment due to a lack of transportation or childcare.[19] Alternatively, cash transfers might indirectly support increases in employment via more indirect mechanisms like improved health, mental outlook, or moving to an area with more job opportunities. While cash may cause some individuals to work more, the pilots I review do not show that a short-term guaranteed income increases employment on average. The one study that investigates barriers to employment in detail (rather than just testing this idea via impact on average employment), OpenResearch, finds no evidence that people receiving

$1,000/month are less likely to say they missed work due to lack of childcare, illness or transportation, and actually finds a four percentage point increase in self-reported disability. Across studies, there is no evidence that wages increase. And the OpenResearch study finds no impact on a large index of measures (over 35 questions) around the recipients' quality of employment.[20]

Takeaway: Evidence shows that guaranteed income has minimal impact on labor force participation or hours worked.

Upward mobility

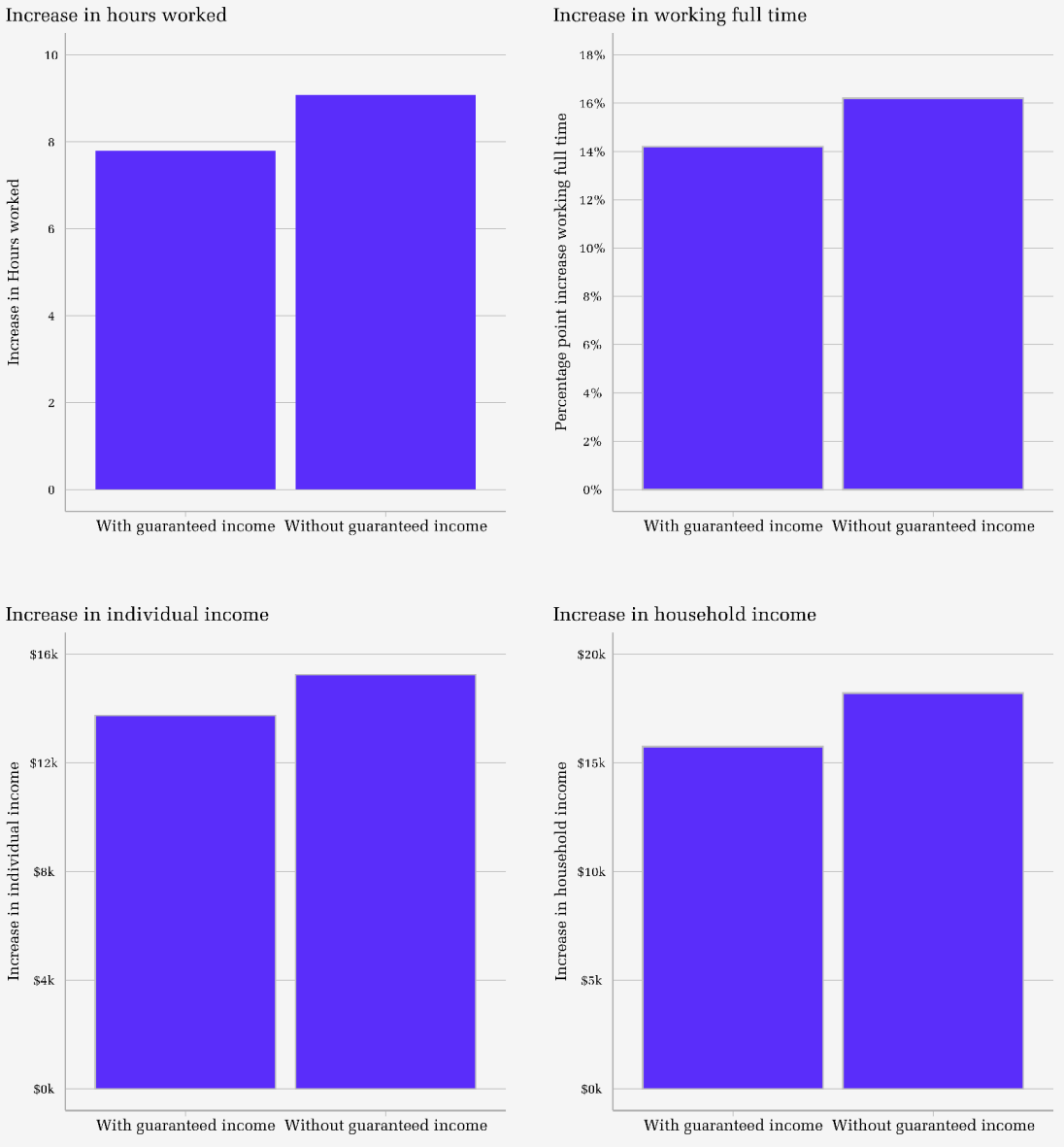
The aggregate small effects on work across pilots implies that large-scale cash transfers do not impact upward mobility via disconnection from the labor force. But perhaps even more instructive is to examine changes in income over time for pilot participants.

The long timeline of the OpenResearch study tests this dynamic best. The three-year $1,000/month pilot is the longest-running and transfers the highest amount to recipients among the studies I review, so it represents the best test of unconditional cash inhibiting upward mobility. Shown below, all measures of labor market activity and income saw marked increases throughout the study period. Full-time work grew over 14 percentage points, and household income grew almost $15,000.

While those receiving cash had slightly smaller improvements, these differences were small relative to large improvements from the start to the end of the study. This demonstrates that even if a larger long-term guaranteed income reduces recipients' labor force participation, this effect does not meaningfully hinder upward mobility.

Takeaway: Guaranteed income does not significantly inhibit participants' upward mobility—both cash recipients and non-recipients sharply increase income and employment rates over time.

Change in income and laborforce participation metrics with vs. without guaranteed income



JFI

Source: OpenResearch study, Vivalt et. al. 2024

Consumption

Most studies find large increases in consumption. OpenResearch found increases of $300 per month, Baby's First Years $150 per month, and Chelsea Eats $150 per month.[21] The OpenResearch team convincingly cites evidence that surveys of consumption substantially understate total spending, so the true increase may be quite a bit larger—as much as $800 per month.. As this finding should apply across studies, one should not read too much into the topline numbers. The lone exception to the consumption pattern is the Compton Pledge, which puzzlingly finds that cash recipients reported spending over $300 less each month. This cannot be explained by a decline in labor supply, and is unlikely to be explained by random chance (it's actually their most statistically significant result across all their outcome measures).

While aggregate consumption is very important, a perhaps more interesting question is what people spend money on? Although spending categories are measured inconsistently across studies, cash transfers appear to increase spending across a broad range of categories. OpenResearch finds the largest increases (in absolute dollars) are in food, rent, and car related expenses. Health, child-related expenses, other transportation expenses, and gifts or loans to family and charity also see large increases. Baby's First Years has a smaller sample and smaller transfer, so when broken down into distinct spending categories, most increases are insignificant. However, they do have a battery of specific questions about child-related spending, and this category accounts for about a third of the overall increase in spending. Chelsea Eats only measures food spending—the increase in spending they find ($150/month) is large in comparison to the transfer, especially if consumption is underreported.[22] The Compton Pledge finds reduced spending is split evenly between housing (rent and utilities payments fall) and non-housing categories.

One thing that studies consistently do not find are increases in spending on "temptation goods"—things like drugs and alcohol, tobacco, and gambling. Baby's First Years found no statistically significant changes in alcohol, tobacco, or opioid use and spending; the Compton Pledge found a small statistically significant decrease in tobacco spending and no chance in alcohol spending; OpenResearch found a marginally statistically significant ($p < .1$) increase in spending of just $13 per month on their combined measure of alcohol, tobacco, marijuana, and gambling, less than 2% of the $1,000 per month transfer. Perhaps more importantly, the OpenResearch study found

large significant decreases in problematic drinking ("Drinking/hangovers interfered with responsibilities") and use of "painkillers not prescribed to you." Furthermore, non-trivial fractions of people in the control group in each study did admit to some of these potentially stigmatized activities, so the results are not attributable to reluctance to admit to spending on these categories.

Another important component of consumption is not just how much you spend, but how much your spending varies. OpenResearch's frequent surveys uniquely (among the studies I review) enable them to empirically assess the impact of transfers on consumption volatility, and they find a large (20%) statically significant decrease.

Takeaway: Most studies report large increases in consumption, particularly on essential items like food, rent, and transportation. There is virtually no evidence of increases in spending on drugs, alcohol, tobacco, and gambling.

Savings, assets, debt, and net worth

Across studies, there is little evidence for lasting improvements in recipient's net financial position via accumulating significant assets or savings or via reducing debt. However, one important caveat to this conclusion is that it is difficult to study outcomes like debt, savings, and assets because there is so much natural variation between households. For instance, the Compton Pledge authors speculate that the reduced spending they identify is plowed into debt reduction, and they do show cash recipients have about $2,000 less in total debt. However, the debt reduction finding is not close to statistically significant, and the study cannot definitively preclude potential debt reductions as large as $5,000—a large sum relative to the average transfer size of less than $500 per month. The story for Baby's First Years is similar—there is no evidence of debt reduction or asset accumulation, but the findings are imprecise.[23]

OpenResearch's study is the exception to this pattern. The combination of large sample size, large transfer, repeated surveys, and linking to administrative credit report data offer more definitive answers about recipients' financial potion, which find no evidence of substantively important changes for most outcomes. For instance, they find credit scores increase six points—a finding that is statically significant but not

substantively important. Similarly, they can track measures like delinquencies, total balances past due, and other negative credit marks and find no significant changes, ruling out small changes. Car ownership may increase slightly—they estimate a two percentage points, up from the control group car ownership rate of 78%. While the estimates of overall net worth (assets plus savings minus debt) are fairly imprecise (they are unable to rule out improvements of less than almost $6,000), they are more meaningful given the total net transfer ($34,200 over three years after accounting for the $50/month given to the control group) and the lack of significant change is consistent with other outcome measures. The one item that does see some significant improvement is savings—recipients have saved between $800 to $2,000 more than the control group on average. Overall, there is no evidence that guaranteed income recipients significantly improve their financial position, but this conclusion is mostly supported by just one study, since these outcomes are difficult to get precise answers on.

Takeaway: There is little evidence guaranteed income significantly impacts recipients' long-term financial positions or net worth.

Other measures of financial well-being

Studies show small improvements in financial well-being. Broadly, financial well-being measures individuals' ability to pay bills and other financial obligations during the period of transfers, though the specific questions vary widely between studies. OpenResearch uses four different family indexes measuring financial well-being. Their first measure, financial hardship, does not significantly change. The index has two components—they do find a small significant increase in experiencing financial shocks, but that is offset by a small (insignificant) decline in running out of money between paychecks. Their next index component measures savings as a fraction of income, which significantly improves, as covered in the prior section. Next, they have a large number of questions about financial health. Questions pertaining to respondents' current financial situation see significant improvements, including ability to handle a major unexpected expense, and a decrease in reliance on financial help from friends and family. Items that measured longer-term financial circumstances, such as level of confidence in retirement savings, did not show significant improvement. However, even for the items that did improve, effect sizes were small, moving no more than a tenth of a point on a four point scale. Finally, they find a small improvement in their last index measure—financial resilience. This is driven by small decreases in

the extent to which the respondent relies on financial help from others and an increase in whether the respondent has at least $100 in savings.

Chelsea Eats has a “Financial Well-being Index” that shows a statistically significant, though modest, improvement. Specifically, cash recipients are about three percentage points more likely to respond positively rather than negatively to a battery of questions on financial well-being. The specific questions concern the recipient’s being better off financially than six months ago, not having any bills, expenses, or needs that they were unable to pay, ability to pay full utilities last month, ability to pay full rent last month, delay of health care or prescriptions because of costs, and ability to pay for a $300 emergency expense with money on hand. Most of the improvement on the index was being better off financially than six months ago (when transfers began). Other questions, including being unable to pay bills, full rent, delaying health care or prescriptions because of cost, show minimal signs of improvements. Being able to pay full utilities and pay for a $300 expense with money on hand show statistically insignificant three and four percentage point improvements, respectively.

The Compton Pledge uses a financial security index, which consists of whether the household could pay for a $400 emergency bill with current resources without going into debt, could pay all bills in the past 30 days, put money aside for the future in the past 30 days, could pay down debt in the past 30 days, had to ever forgo medical care over the past six months because of the expense, and whether the respondent has health insurance. Baby’s First Years uses a “non-food economic hardship index” that assesses the following items: missing a rent or mortgage payment, missing a payment for oil, gas, water, or electricity, being forced to leave or being evicted from your home, needing medical or dental care for yourself or your child but not receiving it, and missing a payment for phone, internet, cable, or streaming services. Both studies find no significant changes, but it is likely that the small improvements in the OpenResearch and Chelsea Eats studies would not be detectable in their smaller samples.

Takeaway: Self-reported financial well-being shows modest improvement in larger studies, with some reductions in financial hardship and increases in resilience measures like emergency savings.

Time use

Across studies, there is little evidence that cash recipients make large-scale shifts in how they use their time. All studies apart from Chelsea Eats have some measures of time use. Baby's First Years focuses on time use related to child care. They find no detectable change in use of center-based child care, ruling out changes of about three percentage points. They also find a non-significant 3.5 percentage point increase in mother's participation in education and training. This could reflect a true small increase, as the result is fairly imprecise—they cannot rule out increases of eight percentage points. Their last time-use related measure is a parent-child activities index. Here, they find a statistically significant 11 minute increase per week, a five percent increase relative to the control group.

The Compton Pledge has two measures of time use. They find that cash recipients spend a statistically significant 25 more minutes sleeping, an increase to about six hours from five and a half hours in the control group. Second, they find no change in time spent on unpaid elder or childcare, ruling out changes of about an hour relative to about six hours in the control group.

OpenResearch has by far the most comprehensive measure of time use. Rather than asking questions about specific activities, they collect comprehensive time diaries from study participants, which comprehensively show changes in time use across categories. This gives the most holistic picture of time use, but does come with tradeoffs. For one, since they ask respondents to categorize all their time, they do not get specific measures of time use; for instance, the parent-child activities that Baby's First Years asks about are collapsed into one category of "childcare." Second, because they test for changes across many specific time uses, adjusting for the chance that some differences emerge by random chance makes each estimate fairly imprecise. When adjusting for multiple comparisons, the only change that is statistically significant is a five minute per day increase in non-computing transportation. Across other items, decreases in time use of three or more minutes per day include nine minutes for market work, eight minutes for time spent sleeping (the opposite direction of the Compton Pledge), four minutes for child care, and three minutes in other income generating activities. Increases in time use include six minutes for social leisure, five minutes for non commuting transportation, four minutes for solidarity leisure, and three minutes for home production.

Takeaway: Transfers appear to have limited influence on time use, with no major shifts in time spent on work, childcare, education, or leisure activities.

Entrepreneurship and human capital investment

OpenResearch has the most comprehensive measures of entrepreneurship and human capital investment. As their study has the largest transfer, longest timeframe, and youngest sample ($1,000 per month for three years for individuals aged 21–40), it is also the most plausible setting for such effects to materialize. They use a mix of survey data and administrative data (88% of participants consented to the administrative data linkage) to look at changes in completion and enrollment in post-secondary education as well as high-school degrees/GEDs and informal education. Overall, they do not see a significant change in their index measuring a combination of these outcomes. They estimate a one percentage point non-significant increase in enrollment in a postsecondary program; 15% of the control group is enrolled in such a program, and they can rule out increases of greater than 3 percentage points. Similarly, participation in informal education increases one percentage point, when 10% of the control group is in some kind of informal education program, and they can rule out increases of more than three percentage points. They also show that increases are larger among participants under age 30, a pre-specified sub-group analysis based on the idea that younger participants are more likely to invest in education in response to the transfer.

On entrepreneurship, OpenResearch has mixed findings. They measured the entrepreneurship domain in three ways: entrepreneurial orientation, entrepreneurial intention, and entrepreneurial activity. The first two domains measure precursors to entrepreneurship; entrepreneurial orientation captures willingness to take financial risks, and entrepreneurial intention measures interest in starting a business. Both these domains see significant increases from cash transfers, for instance, cash recipients are about three percentage points more likely to have an idea for a business. However, there is little evidence this translates into actual increases in starting a business—the experiment finds no significant effect for this domain, and can rule out effects as large as 2.8 percentage points.

Other studies only have a few measures of entrepreneurship and human capital investment. As covered in the time-use section, Baby's First Years finds a non-significant increase in participation in education and training, and

cannot rule out increases of less than eight percentage points. On entrepreneurship, they estimate an insignificant one percentage point increase in self-employment, but cannot rule out increases of more than six percentage points. The Compton Pledge does not see any significant change in spending on education. Given the short timeline, small monetary amounts, and large minimally detectable effect sizes, any increases in educational attainment and entrepreneurship would have been quite remarkable. The OpenResearch study provides the strongest evidence—and rejects the idea there is large pent-up demand for further educational or business investment that's just short a bit of cash. But their small effects are perhaps suggestive of significant improvements in a permanent program.

Takeaway: The largest and longest-running study finds suggestive increases in entrepreneurship and educational investment, perhaps indicative of a larger effect in a permanent program.

Intimate partner violence and relationship quality

Two studies assess the frequency of intimate partner violence (IPV)—the Compton Pledge and Baby's First Years. Baby's First Years asked an index of questions assessing support and trust in mothers' co-parenting relationship. Second, they asked if mothers were ever cut, bruised, or seriously hurt in a fight, how often they argued with their partner on important matters, and a ten item scale measuring relationship quality. For this second batch of questions, questions were administered via audio computer-assisted self-interviewing, which allows mothers to record their answers directly into a programmed computer rather than answering questions live with the interviewer. Across the board, there were no statistically significant changes for any of these items for the group that received the transfer. There are two important caveats to these results. First, if a mother reported that she was not currently in a romantic relationship during the age-1 follow-up, she was asked to report on the quality of the relationship with her most recent partner. This relationship could have ended before the start of the study and could not possibly be affected by receipt of cash. Second, the estimates appear to be imprecise. The one question where assessing precision is straightforward, the study cannot rule out as large as seven percentage point reduction in IPV, when control group mean prevalence is 8%. In other words, the effect would have only been detectable if the transfer almost wholly eliminated IPV.

The Compton Pledge has three questions about intimate partner violence. They directly asked if the respondent was physically abused by a partner, and whether the respondent was forced into physical intimacy by a partner. They also ask about the presence of intimate partner violence using a list experiment, a technique designed to elicit more honest responses to sensitive questions (the survey was conducted online and by phone). Combining the three items into an index, they find no significant impact—however, the question asked as a list experiment finds a twenty percentage point reduction in IPV that is statistically significant. It's possible that the list experiment question is more reliable than the direct report, and that cash caused large-scale reductions in IPV. However, the difference in baseline reports of IPV between the standard question and the list experiment in the control group is just two percentage points, which suggests there is fairly minimal reluctance to admit to experiencing IPV. This suggests the large-scale reduction in IPV in the list experiment result likely reflects random chance.

Takeaway: Two studies find no definitive impact on various metrics of relationship quality and intimate partner violence, but are not large enough to detect potentially important effects.

Health

All the pilots evaluate measures of health with varying degrees of specificity. Across the board, self-reported holistic measures of physical health do not significantly improve. Some studies include more detailed health-related outcome data. Baby's First Years finds no change in babies' sleep disturbances or a measure of poor bedtime routines, and they find a suggestive 3 percentage point increase in diagnoses of health condition or disability. Interpretation of this outcome is a bit tricky, as an increase in diagnoses could come from a real increase in prevalence or an increase in parental health advocacy. OpenResearch finds no impact on sleep and perhaps a five-minute decline in time spent "time in physical activity/recreation." They show that a number of physical health biomarkers, including measures of diabetes risk, cholesterol, obesity, blood pressure, inflammation, and a cardiovascular health index, do not significantly change. Finally, the Compton Pledge finds a statistically significant 25 minutes increase in sleep duration.

Studies also assess various components of mental health, mostly finding no significant changes. Chelsea Eats finds no change in reports of being "anxious or depressed," ruling out changes of more than five percentage points. The

Compton Pledge has a battery of questions they combine into an index that includes frequency of stress, depression, a scale of questions on psychological distress, as well as questions about life satisfaction and happiness, finding no significant change. Baby's First Years measures mother's perceived stress, parenting stress, maternal depression (PHQ8), and maternal anxiety. They even measure physiological stress via Hair Cortisol levels, though the sample for this is small due to stopping in-person data collection during the pandemic. OpenResearch has a scale of questions they combine into a "mental health index" that includes prompts on emotional problems that interfere with daily life, mental distress, perceived stress, generalized anxiety, depression, and number of good mental health days out of the last 30. Some of these scales are difficult to provide easy to understand effect sizes, but their null results are generally quite precise. For instance, they can rule out an increase in one more good mental health day out of the last 30.

Several studies also measured consumption of health care via things like doctors visits. In Baby's First Years, there is no change in parents bringing their child to the doctor for illness or injury. OpenResearch finds no increase in receiving a flu shot, cholesterol test, or for women, receiving a PAP, and a four percentage point decline in receiving the Covid-19 vaccine (although this was a non-prespecified exploratory outcome). They do find small increases in office based medical care—the largest increase is for dental visits in the last 12 months, which increased five percentage points.

The one positive health-related improvement comes from the Chelsea Eats study. They link participants to administrative records of health usage, and find a large and statistically significant reduction in ER visits (27%), their only pre-specified primary outcome in their health analysis. This contrasts with the OpenResearch study, which finds an increase in ER visits from their survey responses, and Baby's First Years, which finds no significant change in ER visits.

What could drive these differences in ER use? It's difficult to compare across studies, as they vary on any number of dimensions. Simple averaging of the three results would predict that the typical impact of cash on recipients is no effect. However, the specific nuances of the Chelsea Eats study deserve further attention. Since they rely on administrative health data rather than survey reports like the other studies, their measure should be more reliable. Surveys may not be accurate because of both non-response and misreporting (people may forget or feel stigmatized about admitting to ER usage). Survey non-response turns out to be important in the Chelsea Eats setting—they

found that declines in ER visits were concentrated in the group that didn't respond to surveys. OpenResearch and Baby's First Years had a very high response rate to their study, making non-response a highly unlikely explanation for their findings. However, for OpenResearch, this was in part a product of their design, which only selected people to further participate in the study who consistently responded to surveys. While they found their sample was representative of their targeting criteria on various demographic measures, if people who are more inclined to respond to surveys vary on some important unmeasured dimensions, it raises questions about the findings of all the guaranteed income pilots. While speculative, the specific dynamics of improvements in health for participants in the Chelsea Easts study who did not respond to surveys is very well-evidenced, and there has been no further evidence to either validate or debunk whether this dynamic applies more broadly.

Takeaway: Overall, studies show limited effects on physical or mental health, with the notable exception of reduced emergency room visits in one study with unique administrative data.

Housing

All the pilots have some housing related outcomes measures—though the specifics measures used vary. Baby's First Years finds no change in self-reported housing quality, excessive residential instability, perceptions of neighborhood safety, or change in a neighborhood level "childhood opportunity Index." The Compton Pledge finds an improvement in a scale measuring housing security, which has little relation to the housing outcomes in Baby's First Years. The specific questions in the Housing Security Index include whether the household can pay their rent or mortgage, their likelihood of eviction, and the number of months behind on rent or mortgage payments. Improvement is driven by a large decrease in the perceived likelihood of eviction, perhaps driven by the increase in financial security afforded by the transfers. Finally, Chelsea Eats finds a reduction in residential moves, while OpenResearch's longer-term study finds significant increases in residential mobility; recipients were four percentage points more likely to move to a different neighborhood than the control group.

Differences in study contexts and outcome measures make it challenging to generalize findings across studies. For instance, while OpenResearch and Chelsea Eats both measure residential mobility, their different results likely

spring from their contextual differences. Both studies interpret their findings positively; Chelsea Eats was a short-term emergency cash transfer during the height of Covid which enabled recipients to stay in place in the presence of a large negative income shock. OpenResearch provided a large transfer over three years (ending well-past the height of the Covid pandemic), better enabling people to move who wished to. OpenResearch further measures desire to move and finds it increases sharply, which further validates that the changes they find there are positive rather than negative. The lack of improvement in housing and neighborhood quality in Baby's First Years appears consistent with the relatively small changes in moving in the OpenResearch Study—a four percentage point increase in living in a different neighborhood would be too small to be detectable for Baby's First Years, where the transfer and sample were considerably smaller ($1,000 vs $333 and n = 3,000 vs n = 1,000). Lastly, the improvement in housing security in the Compton Pledge is not tested in other studies. While the overall picture is a bit muddied given the lack of uniformity in outcomes tested, indications are positive—three of four studies find positive changes in housing-related outcomes.

Takeaway: Most studies find an improvement in housing-related outcomes.

Food security and nutrition

All the pilots have measures of food security, and most show a fairly limited impact. The Compton Pledge, OpenResearch, and the Chelsea Eats studies all find no impact near the conclusion of transfers, while Baby's First Years finds if anything an *increase* in food insecurity (that is not statistically significant). However, there are some positive results. Although Chelsea Eats finds no effect on food security, they do detect significant improvements in food satisfaction, which are further supported by an increase in reports of eating fresh vegetables, meat, and fish. Similarly, Baby's First Years finds increases in the number of healthy foods babies consume each day, and OpenResearch finds small improvements in healthy eating and a reduction in drinking sugary soda. Baby's First Years, OpenResearch, and Chelsea Eats all find significant increases in food-related spending. Finally, Chelsea Eats and OpenResearch also find large improvements in food security in early parts of their study that fade out near the end of their study. For OpenResearch, the large positive effects come one year in, fading out in year two and going to

zero by year three. For Chelsea Eats, the positive impacts are one and three months after recipients started to receive the transfer.

Takeaway: Any improvements in food security decay shortly after the start of transfers, but two studies that measure food satisfaction find durable improvements, and most studies find significant increases in healthy eating and spending on food.

Discussion: What have pilots taught us about guaranteed income?

Debunking cash's harms

Traditionally, the debate around the efficacy of cash assistance has been defined by its potential to trade-off with earnings. Giving people money definitionally makes people better off. But if people work less in response, cash is less effective in increasing the total amount of resources to its recipients, as some increase in income is offset by a decrease in earnings from work.

Of course, there are many reasons to reject the framing of the desirability of cash assistance as a tradeoff with work. Reducing reliance on exploitative, precarious, and low-wage jobs to meet basic needs is often a central justification for basic income. When people are afforded some financial freedom and make a choice to work less, assuming that choice is harmful to their well-being can be both wrong and disrespectful to their autonomy.[24] Even from a narrow efficiency standpoint, time away from work can also mean more time to invest in other productive activities with long-run economic benefits.

This incomplete list of caveats aside, the impact of cash on labor force participation is important to the evaluation of guaranteed income. Many critics' primary objection to guaranteed income centers on its effects on work. For instance, Leslie Price, a fellow at the American Enterprise Institute, writes that the "central concern" behind guaranteed income programs is that they "harm lower-income Americans by disincentivizing work."[25] These critics also reject the logic that freely made choices to work less are in individuals' best interest.[26] While prior evidence makes a strong case that impacts are fairly

minimal, much of that research base is dated, tests the effects of smaller amounts of cash, and comes via interventions that are not as clear-cut as the latest pilots.[27]

Across pilots, impacts on labor force participation are small and inconsistent. Analyzing the average effect across studies, the typical cash transfer results in a decline in hours-worked of just half an hour per week. While it seems likely that large unrestricted cash transfers do result in some marginal decline in work, this effect is so small that it should not be a central concern when considering the merits of such a program.

Beyond labor force participation, critics of guaranteed income often point to cash transfers as inhibiting upward mobility. Working less at one point in time might be undesirable, but a far more damaging threat is that unrestricted cash can knock people off the ladder to the middle class. Writ-large, critics posit that there is a tradeoff between the generosity of the social safety net and upward mobility.[28] To them, caring about the poor should mean more emphasis on policy that can boost people in poverty into the middle class and beyond. Disbursing cash may move people above the poverty line, but that comes with a tradeoff in that it reduces the probability of broader upward mobility, principally by decreasing work.

Here again, pilot results refute critics' worries. Even when a large cash transfer appears to decrease labor force participation—as in the OpenResearch Study—these effects are small in comparison to cash recipients' income growth over time. Broader economic forces that determine economic mobility (like tight labor markets) appear far more impactful than even large cash transfers.

Finally, some critics claim that flexible cash assistance will be used for unproductive ends. Senator Joe Manchin, whose opposition to the expanded Child Tax Credit was instrumental in its downfall—reportedly told fellow Senators that parents would use their Child Tax Credit payments to buy drugs rather than providing for their children.[29] If the marginal use of unconditional cash supported spending on drugs, alcohol, or other products that were actively harmful for recipients, it would undermine the case for no-strings-attached transfers for all but the most radical opponents of paternalism.

Measuring these spending effects is one of the biggest benefits of guaranteed income pilots. High-resolution data on individuals' spending is only possible

to collect with detailed survey data—prior evidence from natural experiments largely does not capture consumption decisions. Across the board, pilots find virtually no evidence that cash is spent on "temptation goods." Instead, increases in consumption seem concentrated on various basic needs like housing and food.

Inconsistent evidence for downstream benefits

One of the biggest strengths of the guaranteed income pilots relative to past research is the ability to measure the impact of cash transfers in unprecedented detail. In a market economy, people with more money are almost definitionally better off—certainly when the potential harms of cash assistance outlined in the prior section are ruled out. But many advocates hope, and pilots test, that being better off will materialize in various specific measures of well-being that are downstream of having more money. There is little evidence of consistent benefits in many of the domains pilots study that are not the direct effect of the transfer.

Often there are pockets of more positive results. For instance, while many holistic measures of health don't improve—there is strong evidence that Chelsea Eats caused a large decrease in hospitalizations that other studies could not have identified. Similarly, Baby's First Years found an array of benefits for children—from more quality time with parents to more child-focused consumption that other studies did not measure. There are various potential explanations for why benefits show up in some contexts but not others. Maybe the studies did not include enough participants, did not last long enough, or were thrown off by the pandemic. These are all valid critiques. However, if cash had tremendously powerful benefits, they should materialize even given the many limitations of these studies.

The lack of improvement for some outcomes is particularly dispiriting. Across pilots, there was little evidence that cash improved household food security or various measures of economic related stress. These are outcomes that seem particularly closely related to a lack of money and improvable even when a transfer lasts a relatively short time, which makes the lack of consistent improvement especially confounding.

Other guaranteed income pilots do find more consistently positive results. For instance, the very first quantitative guaranteed income pilot evaluation in Stockton found increases in full-time employment and improvements in both physical and mental health for cash recipients. None of those effects were

broadly replicated in the studies I review, but several other pilot studies have results more similar to Stockton.

Idiosyncratic contextual factors might cause the pilot studies I review to have fewer identifiable downstream benefits than other research. However, my review was not a random collection of four pilot projects, but studies that had particularly strong methodological practices that helped ensure any effects they identify are attributable to cash. While no study is immune from critique, especially for findings that are consistent across all four studies, it is more likely that methodological differences, to a greater degree than context differences, explain the different results. Even weighing the evidence in each study equally, if positive effects of guaranteed income appear in only some studies, it reinforces the notion that cash's benefits in these specific domains are inconsistent.

Some explanations for the lack of downstream effects

Grounding expectations: what kind of improvements can we expect?

When summarizing results, I have tried to contextualize the precision of each study's findings. For some domains, studies could not rule out large positive effects of cash, like for reductions in debt or intimate partner violence. Other results were fairly precisely estimated. However, this element of precision only captures how well the study can distinguish between the control and treatment group. A study can have very precise null results, but if treated participants are only given $10 a month, it conveys little information about cash's potential benefits.

Of course, none of the studies gave participants just $10 a month. However, there may be some reason to think the transfers given may not be large enough to have produced important effects. For instance, transfers were given at the household level, while most UBI proposals would give transfers at the individual level. This is an important distinction. For instance, in the Compton Pledge, the average transfer is $438 per month, but the average household size is 4.4 people, making the transfer per person per month just about $100. Relative to the control group, treated households got an income boost of about 13 percent from the transfers.[30] Combine that with a sample

size of less than 350 households receiving cash, and it's unclear how disappointing the lack of significant positive effects on psychological well-being, financial security, and food security should be.

What would help contextualize the results is knowing the correlation between income and various outcome measures, and comparing them to the studies results. For example, in the Compton Pledge, control group households earned $40,000 annually, while treatment group households received an additional $5,000 per year in transfers. Outside the context of any experiment, we know from a large body of research that, on average, households with $45,000 of income are somewhat better off than households with $40,000. A huge number of measures of well-being are correlated with income. However, it's important to know exactly how much better off they are, and if that amount is large enough that the experiment's sample size can detect it.

Moreover, it's unclear if expecting a temporary transfer should be expected to have the same effect as the descriptive correlation between income and well-being. People who have more income are better off because of that extra income, but also likely have other advantages that have an independent positive impact on well-being. To make this concrete, the Yale Budget Lab's modeling of the Child Tax Credit assumes that just 20% of the correlation between childhood household income and adult earnings is causal.[31] In other words, extrapolating their logic, they would expect the Compton Pledge transfer to make a household with $40,000 of income as well off as a household with $41,000, rather than $45,000.[32]

Among the studies reviewed, OpenResearch is the most likely to detect meaningful effects due to its large sample size, significant transfer amount, and extended duration. It also analyzes the descriptive correlation between household well-being in its sample for all the health related impacts that can help contextualize its results. In most cases, its estimates are much smaller than what would be expected if the relationship between income and health was fully causal. However, health is a domain where you might expect outcomes to be particularly resistant to change from short-term cash. Comparing the descriptive correlation between income and well-being to experimental estimates across all outcomes and studies would be incredibly useful for contextualizing the results.[33]

Cash as an untargeted intervention

Another explanation for the lack of more positive downstream results may stem from the fundamentally untargeted nature of guaranteed income interventions. Proponents hope (and research tests) for impacts as diverse as improved health, catalyzed entrepreneurship, and human capital investment—cash does not explicitly target any particular domain. The lack of targeting makes it far more likely that positive effects are very diffuse as different people direct their increase in resources to different goals. Given this dynamic, comparing guaranteed income to interventions that specifically target a certain domain may inevitably find cash's results disappointing.

Take research on health outcomes. The OpenResearch paper on health outcomes concludes that, "more targeted interventions may be more effective at reducing health inequality between high- and low-income individuals." Programs that target health directly have a far higher probability of improving health because all their effort goes towards that particular domain. Not every health related treatment is successful on every individual it is applied to, but because the treatment is fully directed towards health, it has a far better chance of showing significant improvement in aggregate. In contrast, many guaranteed income recipients may use their funds in ways that are totally unrelated to their health. The few that do direct their funds toward health may have the same rate of success improving their health as those who get a more targeted intervention. However, because few direct their funds for this purpose, the overall impact on health may not be detectable.[34] The same logic can explain null results across a variety of domains. Not every person who receives guaranteed income wants to start a business, and many who try to start a business fail. Therefore, the overall impact of guaranteed income on new business creation could be too small to be detectable in these experiments, but potentially sizable among the small fraction of people who have goals around entrepreneurship. Future work may be able to better understand guaranteed income effects by explicitly asking what people's goals for using the transfer are, and linking those answers to quantitative measures of that specific domain.

Lack of substitution from work

Much of the discussion around guaranteed income's other positive impacts comes from an assumption that it will reduce paid employment. Many advocates accept the premise that people will work less, but point to how that time can be productively reinvested in other activities with long-run returns.[35] However, across studies, there was very little reduction in work in

the first place. By and large, cash recipients do not have more time to "reinvest" in other activities.

The time use results from OpenResearch are particularly demonstrative of this phenomenon. They find that cash participants spend an average of eight minutes less per day on "market work" —leaving a minuscule amount of time to establish new commitments and activities. Moreover, the effect on time spent on market work itself isn't statistically significant, so even if this time is fully "reinvested" in other productive activities, the study is not large enough to detect them.

This dynamic applies to more than daily uses of time measured by OpenResearch. The lack of impact on outcomes like educational investment, time with children, and entrepreneurship (among others) all may stem from the lack of extra time to devote to these activities.[36] The OpenResearch time use results are not an outlier—it is the largest study with the strongest evidence that participants work slightly less, and the time use results are consistent with their other employment-related results.

Justifications for guaranteed income from the pilot results

Where do these results leave justifications for a guaranteed income? For a guaranteed income that includes all low-income adults layered on top of the existing safety net, the current crop of studies does not support the idea that a guaranteed income can have a "return on investment" from the perspective of government finances.[37] The consequences of entrenched poverty are certainly expensive for the federal government. However, there are scant results that would support the idea that guaranteed income significantly reduces these costs. For instance, there is little evidence that a cash infusion reduces barriers to finding higher paying employment or enables high-value entrepreneurship. Nor is there much evidence that cash improves health and thus reduces health care costs.

The pilot studies thus far do not test all the possible channels of reduced government spending—like reduced criminal justice involvement. And it's possible a permanent guaranteed income would have different results. A more conservative conclusion would be to say that the studies I review show no *new* evidence that a guaranteed income could pay for itself. That said, the notion that a guaranteed income could provide a return on investment would

require dramatic differences between the impacts of permanent and temporary transfers, or other large benefits in domains the pilots failed to capture. Writ-large, this kind of large-scale positive impact for an untargeted intervention applied at scale would be unprecedented in research on social policy. The case for a return on investment is much stronger around unconditional cash for children—a topic I return to later.

Evidence of a return on investment would be an especially compelling reason for action, helping negate difficult questions of financing and, at least in principle, ideological opposition to redistribution. However, guaranteed income should not need to pay for itself to be justified. Congress does not serve as an investment committee, and a larger cash-based safety net that eases the financial pressures of low-income Americans can be justified on its own terms. The pilot results clarify that rather than specific claims about the efficacy of cash, the case for guaranteed income should rest on broader ethical grounds. The US is a rich country that can do more to support the lives of its lowest-income residents. While the pilot results do not show that recipients' lives were radically transformed by a cash transfer, it unquestionably reduced poverty and material hardship during the span of the program. Pilots also allowed recipients the freedom to make purchases to support their basic needs. These are fairly straightforward direct impacts of cash, but that makes them no less important to justifying guaranteed income.

The pilot results that may be less intuitive to some show that unconditional cash assistance is not harmful. The two principal risks—that recipients work significantly less and that money is spent irresponsibly—are strongly rejected by the pilot results.

While a broad guaranteed income policy cannot “pay for itself”, policy designs that modify the existing safety net have a strong case for cost-effectiveness.[38] Getting better results than status-quo programs with the same level of expenditure is a far lower bar to clear than creating a new program that pays for itself on top of the status quo.

Critically, all the studies thus far have tested the effects of additional cash on top of the status-quo programs, many of which were made more generous during the pandemic. However, the evidence they produce still has relevance for assessing a more cash-first safety net. The prime economic justification for in-kind assistance is that cash will be put to poor use.[39] This is soundly rejected by all the studies under review. If there is no negative impact of

cash, the policy case for in-kind assistance, with its well-documented administrative burdens and inefficiencies, is tenuous.

Another justification for guaranteed income is the potential to overhaul the existing safety net. Current policy creates a patchwork of different programs with varying eligibility criteria, which can create significant challenges for low-income households enrolled in multiple programs. As earnings increase, these households may face steep benefit reductions—or even lose support entirely—at so-called "benefits cliffs." These fast phase-outs and abrupt cutoffs can unfairly and arbitrarily penalize households. Additionally, the complexity of the system may discourage earning more income, potentially trapping people in poverty. A guaranteed income that replaces existing programs could eliminate these steep phase-outs and benefits cliffs, creating a fairer safety net and potentially large increases in labor force participation independent of the effects of replacing in-kind benefits with cash.

While this kind of wholesale reform holds promise, pilot studies contribute little to advancing this argument. There is limited empirical evidence to support claims about the disincentivizing effects of phase-outs and benefits cliffs, and designing such a reform will inevitably run into challenging policy design problems.[40] While pilots demonstrate the effectiveness of cash transfers, they offer no insight into the work disincentives of benefits cliffs or the practicalities of implementing a broader overhaul.

Pilot results and the expanded child tax credit

Discussions about universal basic income often feel detached from political reality. A proposal for a universal cash benefit or a targeted guaranteed income inclusive of all adults has not even been introduced as a bill in Congress. However, the temporary 2021 expanded Child Tax Credit was a de facto guaranteed income for children, and bringing it back remains a plausible near-term path forward for large-scale guaranteed income policy in the US.

Unfortunately, the pilot results will do little to resolve empirical debates over the expanded CTC. The CTC can be justified by the same terms as a broad-based guaranteed income—more cash resources to support the material needs of low-income children with minimal negative side effects is an unalloyed good. However, there is a strong case that the CTC can cause more transformative effects as children age into adulthood. For both sets of

arguments, the pilot results only make a marginal contribution to the existing evidence.

To critics, the main potential threat to the positive impacts of cash on children is if it makes their parents work less. For instance, asked to respond to a study claiming every dollar of investment into the Child Tax Credit yields ten dollars in benefits, American Enterprise Institute senior fellow Scott Winship said, "you're more likely to find that a policy is worthwhile if you simply assume the largest potential costs don't exist (in this case, worsened child outcomes in the long run from reduced parental work and increased single parenthood)."[41]

As reviewed in the prior section, aggregating results across studies finds a minimal decline in labor force participation. Moreover, the studies that break down labor supply results between parents and non-parents find parents do not work significantly less or even work more. OpenResearch finds no impact on time at work among parents in their time-use survey, and generally small and insignificant impacts on other income measures. The Compton Pledge specifically focuses on single mothers, finding no decrease in labor force participation and a large (30%, $p<.01$) *increase* in hours worked.

While guaranteed income studies should ease worries about labor force participation, critics of the CTC may point out that the guaranteed income studies also do not precisely rebut their criticisms.[42] For instance, while there was virtually no evidence the temporary CTC expansion reduced earnings, a permanent policy could have a larger impact. Temporary guaranteed income pilots do not test the effect of permanency—but the two longer running programs show no strong pattern of larger effects over time.[43]

More importantly, the CTC (theoretically) disincentivizes work in two distinct ways. First, it increases recipients' income regardless of how much they work (an income effect). Second, it reduces what economists call the "return to work." Before the CTC was expanded, only parents with substantial earnings were eligible for the credit, providing a kind of tax bonus for working. The expanded CTC gave most parents the full credit regardless of whether they were working, taking away this tax bonus. This second mechanism (called a substitution effect) was far more important in predicting the relatively large decreases in work in models from its chief critics.[44] However, guaranteed income only tests the first mechanism (the income effect), rather than the substitution effect, which is about how much income is taxed away as parents

earn more. Thus, it does not speak to the core debate about the disincentivizing effect of an expanded CTC.

Inherent to this discussion is an assumption that any decrease in work is undesirable. Yet parents might substitute away from work to spend more time with their children, ultimately benefiting their long-run development. While pilot results could speak to what parents do if they reduce their time at work, the studies found such small reductions in employment that there was not a significant amount of time to establish other activities.

What about the potential benefits of the child tax credit, rather than its potential harm via reduced parental labor supply? A large body of evidence has found that sustained cash transfers in childhood can have an array of positive impacts, from increased earnings to reduced criminal justice involvement to improved health. Most of the guaranteed income pilot studies thus far do not focus on child-specific results, nor are they long enough to assess impacts in adulthood, so they do little to advance this literature.

One exception is the Baby's First Years pilot. Relative to prior research, Baby's First Years offers more granularity to the specific pathways by which unconditional cash could help children, where prior research typically focuses on measures of well-being in adulthood after receiving cash in childhood. Researchers found that parents devoted more time to learning and enrichment activities with their children, and at least a third of the transfer went to child-specific goods. However, there was no sign that parents were less stressed, a key hypothesized mechanism for how the researchers thought cash would benefit children.[45] Overall, given the large research base that already exists, the study should at most marginally alter the evidence base on the effectiveness of the Child Tax Credit for promoting long-term child development.

Conclusion

Guaranteed income pilot studies indicate that the benefits of cash are straightforward—supporting spending on basic material needs while reducing poverty and income instability. They do not show that cash can mitigate all the ills associated with poverty, nor does it appear, on its own, to set up recipients for upward mobility. However, the findings reinforce previous research showing that unconditional cash has minimal downside risk. Guaranteed income does not cause reckless increases in spending on "temptation goods" or cause large-scale exits from the labor force.

Will these findings catalyze momentum for larger scale policy, as many of the advocates hope? If these particular cash experiments had transformational impacts on downstream outcomes that would help offset their cost, it may have made the case for guaranteed income so compelling that it would overcome traditional ideological divides. However, the studies reviewed find little evidence for these kinds of impacts. Pilots show that the main benefits of cash are to alleviate poverty, support consumption, and reduce the consequences of income instability, while also counteracting the common claim that unconditional cash aid to the poor is counterproductive to their interests. These are compelling reasons to support guaranteed income, but they are not especially novel.[46]

As much as pilots provide high-quality evidence for policy, there is some truth to the cliché that "a true guaranteed income has never been tried." A well-developed literature on the benefits of cash benefits for children has found the consistent positive results from long-running programs, while the impact of shorter temporary programs (similar to the pilots) have been more lackluster. If the same patterns hold for adults, investing resources into longer-running pilot programs that better approximate permanency could provide valuable evidence strengthening the case for guaranteed income.

While the pilots test cash on top of status-quo programs that invoke common ideological divides over the adequacy of the existing safety net, they do point to a path forward for the reform of status-quo programs. In-kind income support programs come laden with inefficient bureaucratic red tape. Though these restrictions make the programs manifestly less beneficial to their recipients and more costly to administer, they routinely are justified on the fear that the simple disbursement of cash would be put to poor use—an idea that is soundly rejected by the pilot evidence. Even if these recent studies permit no sweeping conclusions about the viability of a guaranteed income policy, they should at minimum further encourage policymakers to introduce unconditional cash wherever possible into our existing safety net.

Appendix

Explanation of Inclusion Criteria For Guaranteed Income Pilots

My main discriminating methodological criterion is to focus on randomized controlled trials. Comparing people who receive cash to a comparable control group is crucial for accurately assessing the effects of guaranteed income. For instance, some pilots track outcomes of cash recipients over time, comparing outcomes when they received cash to before transfers began. This is problematic because many households experience low-incomes only temporarily, independent of cash interventions.[47] As a result, any positive changes might not be attributable to guaranteed income. While studies that include a control group are better positioned to assess the impact of cash, without randomization, they risk attributing the effect of cash to incidental differences between the control group and those receiving cash.

Among randomized control trials, I only review studies with minimal differential attrition—when participants in the cash group respond to surveys at roughly the same rate as those in the control group. When response rates differ, it undermines the assumption that the treatment and control groups are comparable except for the cash intervention. For example, it is possible recipients of guaranteed income may be more likely to stay engaged in the study and respond to surveys, even during economic hardship, while non-recipients may drop out. If more disadvantaged individuals disproportionately leave the control group, it can skew results, making the guaranteed income group appear worse off than if both groups had similar response rates.[48]

Lastly, I only review studies with large samples, pre-analysis plans, and clear evidence of successful randomization. Without a large sample, studies may find there was no difference between those who received GI and versus those who did not, not because GI had no effect, but because a small study cannot differentiate random variation from the impact of cash.[49] A pre-analysis plan explains how researchers plan to analyze the data before they receive it. When there are an array of different outcome measures (as is common for GI studies), some may show signs of a significant positive or negative effect that is really just random noise. Pre-analysis plans show what outcomes researchers are prioritizing to help guard against over-interpreting a stray significant result that's a product of random chance. Finally, evidence for successful randomization comes from comparing the baseline characteristics

of GI participants to control group participants before they start receiving cash—which should show minimal differences. This provides an important check that those who received GI really were selected randomly, and thus lends more confidence to interpreting subsequent differences as caused by the cash itself rather than incidental differences between groups. The famous negative income tax experiments which showed small negative labor supply effects and dampened momentum for a national level Guaranteed Income may have actually been a product of randomization failure.[50]

My focus on rigorous quantitative evaluation should not be read as a dismissal of qualitative evaluations of Guaranteed Income. Participants' stories offer some of the most persuasive evidence on the value of cash and can better capture its varying utility across a dazzling array of different domains. Qualitative research can highlight nuances and personal dimensions that multiple choice questionnaires fail to capture. However, it is also difficult to move from qualitative accounts of individual experiences to average effects in aggregate, something randomized controlled trials are specifically designed to assess. When qualitative and quantitative research study the same conceptual outcome in the same study get different results, it could reflect some kind of problem with the quantitative study. However, it is more likely driven by the fact that qualitative evidence is based on a handful of experiences that might not be representative of the broader group of recipients.

Endnotes

1. VerBruggen, R. (2024, July 24). Free money, less income: Study finds no-strings cash leaves the poor worse off. *New York Post.*
2. Jain Family Institute. (2021, August). Position paper on guaranteed income.
3. Widerquist, K. (2018). *A critical analysis of basic income experiments for researchers, policymakers, and citizens*. New York: Springer International Publishing.
4. I use the terms guaranteed income, universal basic income, and basic income interchangeably when discussing the experiments. While these terms can represent different policies, the pilot evidence is relevant to all specific implementations of basic income via cash's effect on low-income households.
5. See this comment from philanthropist John Arnold, "I think the pilots evaluate the benefits but not the costs. There's a difference in one person in a community getting $1k/month vs everyone in the community getting it. Housing price certainly go up in the latter, for instance. The pilots put a ceiling on what the benefit is. The net impact post inflation will certainly be less. How much less is impossible to know, but given the benefits in this study weren't overwhelming to begin with, it does not bode well for the policy." John Arnold[@JohnArnoldFndtn]. (2024, July 23). I think the pilots evaluate the benefits but not the costs. There's a difference in one person in a community getting $1k/month vs everyone in the community getting it… [Tweet]. *X.* Retrieved from https://x.com/JohnArnoldFndtn/status/1683775684843048962
6. For examples of UBI plans that include financing, see Fleischer, M. P., & Hemel, D. (2020). The architecture of a basic income. *U. Chi. L. Rev.*, *87*, 625. And Hartley, R. P., & Garfinkel, I. (2023). Income Guarantee Policy Design: Implications for Poverty, Income Distribution, and Tax Rates. *National Tax Journal*, *76*(2), 317-346.
7. Page, M. E. (2024). New Advances on an Old Question: Does Money Matter for Children's Outcomes?. *Journal of Economic Literature*, *62*(3), 891-947.
8. Jones, D., & Marinescu, I. (2022). The labor market impacts of universal and permanent cash transfers: Evidence from the Alaska Permanent Fund. *American Economic Journal: Economic Policy*, *14*(2), 315-340.
9. Parolin, Z. (2023). *Poverty in the pandemic: Policy lessons from COVID-19*. Russell Sage Foundation.
10. In a particularly instructive example, a journalist from CleanTechnica concluded that UBI is not a good solution to society's problems because it would increase consumption that causes climate change, writing, "instead of asking to have a $1000 bandaid placed on this sucking wound, we instead need to stop participating in the things that fund the rent-seekers and their partners in the political class" encouraging people to "use less power from the grid, stop buying gasoline, ride bikes, and otherwise starve the beast." Sensiba, J. (2024). Amid debate over results of largest UBI study, there may be a better solution that also lowers emissions. *CleanTechnica.*
11. Balakrishnan, S., Lewis, M., & Nuñez, S. (2020). Reweaving the Safety Net: The Best Fit for Guaranteed Income.
12. Marinescu, I. (2018, February). *No strings attached: The behavioral effects of U.S. unconditional cash transfer programs* (Working Paper No. 24337). National Bureau of Economic Research
13. Joint Committee on Taxation. (2022, October). *Macroeconomic analysis of a permanent child tax credit expansion*.

14. When standard errors are small and rounded, I am conservative and assume they are rounded down when constructing confidence intervals (e.g. I assume a standard error of .01 was rounded down from .014, which gives the widest possible confidence interval given the potential values of the underlying unrounded standard error).
15. For example, in studies on one-time cash assistance administered by GiveDirectly, one study initially suggested that cash was more effective for especially low-income households. However, when the study was replicated by the same team of researchers a few months later, they found no evidence of this heterogeneity. See Pilkauskas, Natasha V., Brian A. Jacob, Elizabeth Rhodes, Katherine Richard, and H. Luke Shaefer. "The COVID Cash Transfer Study: The impacts of a one-time unconditional cash transfer on the well-being of families receiving SNAP in twelve states." *Journal of Policy Analysis and Management* 42, no. 3 (2023): 771-795. And Jacob, B., Pilkauskas, N., Rhodes, E., Richard, K., & Shaefer, H. L. (2022). The COVID-19 cash transfer study II: The hardship and mental health impacts of an unconditional cash transfer to low-income individuals. *National Tax Journal*, *75*(3), 597-625.
16. I concentrate on labor force participation rather than income for two reasons. First, income measures appear nosier than measures of labor force participation—OpenResearch reports income differences of 60% depending on the way it is asked. Second, income is not measured the same way (or in some cases at all) across studies, making the same kind of meta-analysis impossible.
17. For those more technically inclined, scaling results to the size of the transfer, the meta-analysis estimate for the elasticity of hours worked is -0.16 (se = .08) and employment is -.11 (se = .05).
18. These results only account for one member of the household–to the extent other household members also decrease how much they would work, these figures would understate the total employment impact of cash. Studies generally do not provide detailed information about other household members to assess this systematically, but the available data points paint a mixed picture. OpenResearch and the Compton Pledge show larger effects on household income than individual income, consistent with small additional employment reductions for other family members. Chelsea Eats and Baby's First Years show the opposite. For Baby's First Years, the annual earnings for the mother has a smaller decline than total household earnings. For Chelsea Eats, the small increases in work hours worked are virtually the same at the individual and household levels.
19. Center for Law and Social Policy (CLASP). (2021, November). *Key findings from national child tax credit survey: CTC monthly payments are helping improve family well-being*.
20. In results pre-specified as exploratory, the Compton Pledge finds a small significant decrease in cash recipients' satisfaction with their current employment situation.
21. The Chelsea Eats consumption number comes from multiplying weekly food expenditures by four.
22. Chelsea Eats was designed to more efficiently support food assistance with cash instead of relying solely on in-kind food distributions—they do have one question on clothing spending among their secondary outcomes, which increases a marginally significant $15, but the time-period this spending is reported for is ambiguous.
23. Baby's First Years estimates a (non-significant) debt decrease of $2,000, and cannot rule out a debt decrease of over $6,000. Similarly, they cannot rule out an increase in car ownership of up to ten percentage points at conventional levels of statistical

significance. Chelsea Eats' only outcome related to financial assets is bank account balance and amount held in cash. They find no significant changes.
24. Working less because of an unconditional cash transfer is distinct from working less because you will start to lose eligibility for a cash transfer at higher income levels; the latter is more clearly an economic harm or an inefficient economic effect, while the former is simply exercising a preference.
25. Ford, L. (2024, March 18). *Evaluating guaranteed income projects*. American Enterprise Institute, Center on Opportunity and Social Mobility.
26. Doar, R. (2018, Spring). Universal basic income would undermine the success of our safety net. *The Catalyst: A Journal of Ideas from the Bush Institute*, (10).
27. Marinescu, I. (2018, February). *No strings attached: The behavioral effects of U.S. unconditional cash transfer programs* (Working Paper No. 24337). National Bureau of Economic Research
28. Orrell, B., & Winship, S. (2021, December 16). Building upwardly mobile families: Highlights from my conversation with Scott Winship. *AEIdeas.*
29. Shabad, R., Haake, G., Thorp, F., & Tsirkin, J. (2021, December 20). Manchin privately raised concerns that parents would use child tax credit checks on drugs. *NBC News.*
30. The 13% income boost comes 18 months into the study at the time outcomes were evaluated. At the start of the study, recipients had lower average incomes, and the transfers represented a 20% increase in income. This phenomena of income growth from the start to the end of the study is common across studies. While studies target households below or close to the poverty line for initial inclusion, household incomes grow over time because experiencing poverty is often a temporary setback rather than a permanent condition. This is another factor that may lessen the pilot's effects—most households appear to be well above the poverty line at the time outcomes are assessed.
31. Yale Budget Lab. (2024, April 12). Simulating the long-term impact of cash assistance to children on future earnings. *The Budget Lab.*
32. On the other hand, the logic of poverty traps might suggest that increasing poor households' material resources might have larger effects than the correlation between income and well-being.
33. Another helpful point of contextualization would be to specifically look at the effects of temporary income increases within individuals over time in panel data.
34. Importantly, this explanation cannot account for certain theories of guaranteed income's impact that are more holistic. For instance, health could improve not because people specifically spend their extra income on health-related consumption, but because of a decrease in chronic money-related stress. Similarly, ideas about guaranteed income as a means of solving a "poverty trap"—where the conditions of scarcity make poverty self-reinforcing—do not find much support.
35. Some advocates do not think reductions in labor force participation must be justified by investments in other productive work, and see one of the virtues of guaranteed income as empowering individuals to make the best choices for themselves.
36. It is certainly possible to reallocate time from other activities, such as returning to school while working, but discussions about guaranteed income have primarily focused on how such effects stem from reducing time spent on work.
37. See, for instance, Finkton, D., Jr. (2023). *End poverty. Make trillions*
38. Many guaranteed income advocates believe cash should add to the existing safety net, not replace it. This is in part a response to conservative proposals to use

guaranteed income as a much smaller replacement to the existing safety net, and in part from a feeling that existing safety net programs are not large enough. However, making the existing safety net more cash-like does not mean it should be made smaller, and does not preclude making it larger. Moreover, a common position in the guaranteed income community is that cash can be more effective than paternalistic and inefficient in-kind assistance, which is in tension with keeping the status-quo safety net untouched.

39. Many safety-net programs also compensate for failed markets or provide services distinct from income support. For more details about how a guaranteed income could best fit into or replace existing safety net programs, see our prior report: Balakrishnan, S., Lewis, M., & Nuñez, S. (2020). Reweaving the Safety Net: The Best Fit for Guaranteed Income. Jain Family Institute.

40. For more on this topic, see Balakrishnan, S., Lewis, M., & Nuñez, S. (2020). Reweaving the Safety Net: The Best Fit for Guaranteed Income. Jain Family Institute. and Balakrishnan, S., Constantino, S., & Nuñez, S. (2020). *Building a helicopter: Pathways for targeting & distributing a US guaranteed income*. The Jain Family Institute.

41. Jarow, O. (2024, September 17). Do the benefits of the expanded child tax credit actually fade with time? *Vox.* None of the studies assess the impact of guaranteed income on family structure.

42. For a more forceful rebuttal of these criticisms, see Landry, J. (2022, February 1). *The expanded child tax credit and parental employment: Tenuous evidence points to work disincentives*. Jain Family Institute. And Fenton, G. (2023, June 14). *Gains from expanded child tax credit outweigh overstated employment worries*. Center on Budget and Policy Priorities.

43. Baby's First Years shows slightly smaller effects over time, while OpenResearch shows slightly larger effects over time.

44. Corinth, K., Meyer, B. D., Stadnicki, M., & Wu, D. (2021). *The anti-poverty, targeting, and labor supply effects of replacing a Child Tax Credit with a child allowance* (No. w29366). National Bureau of Economic Research.

45. Chelsea Eats also pre-specified a primary outcome as children's school attendance and did not find any significant difference. However, the context for this finding shows it has little bearing on the efficacy of cash for children. Despite the study occurring in the midst of Covid-19 (endline data was collected in April 2021), average attendance was very high, with just 12 days missed measured via school administrative data. Improvements in attendance from a school dealing with chronic absenteeism would be beneficial, but some low level of absence is healthy, especially during the height of the Covid pandemic. Moreover, the group not receiving cash may have felt more financial pressure to send a child to school who was sick so they could receive school provided breakfast and lunch.

46. Many of the results echo findings from the Negative Income Tax Experiments, completed some fifty years ago. On labor supply: "The negative income tax plans tested in the experiments were expected to reduce work effort among participants, and they did so. The work reductions were probably smaller than most opponents of a negative income tax had feared, but larger than advocates had hoped." On consumption: "The results suggest no general increases in frivolous or outlandish expenditures. Indeed, expenditures induced by experimental treatments follow (at least in aggregate) the same patterns observed from non-experimental income. In other words, for most expenditure categories such as food, clothing, health expenditures, and so forth, the results show nothing startling or unexpected." On

other downstream effects: “medical care utilization did not increase and health status did not improve as a result of the income maintenance payments. Hence, to the extent that improved health is of particular interest, programs aimed directly at health care have a better chance of success than do cash transfers. In terms of psychological well-being and participation in community life, again the researchers found no effect. Overall, the results suggest that the lives of recipients were not altered dramatically by the payments offered in the experiments.” From Munnell, A. H. (Ed.). (1986). Lessons from the income maintenance experiments: Proceedings of a conference held at Melvin Village, New Hampshire, September 1986. Federal Reserve Bank of Boston.

47. King, B. (2024, July). *Monthly and episodic poverty: 2022* (Current Population Reports, P70BR-196). U.S. Census Bureau.

48. Researchers can try to control for the characteristics that predict why people drop out, but this makes a randomized controlled trial more like an observational study. There are always questions about if the control variables adequately account for all the differences between groups.

49. For instance, an evaluation of Embrace Mothers Birmingham reported that, “The study’s small research sample (N=263) means that GI would need to have very large impacts on participants’ well-being for the evaluation to be able to detect them in Birmingham alone.” Jefferson, A., Juras, R., Yang, H., Cocatre-Zilgien, E., Rosenberg, S., Kappil, T., & Smith Playfair, A. (2024, August). *Mayors for a Guaranteed Income evaluation final report: Embrace Mothers Birmingham*. Abt Associates. This problem is only about interpreting studies in isolation, trials with small samples can (and should) be included in meta-analyses of the effects of Guaranteed Income.

50. For a history of failed attempts to adopt a national guaranteed income, see Steensland, B. (2011). *The failed welfare revolution: America's struggle over guaranteed income policy*. Princeton University Press. For evidence of randomization failure in the negative income tax experiments, see Riddell, C., & Riddell, W. C. (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), 427-467.