

Consuming and Producing Knowledge:
The Dual Role of Policymakers in Advancing Guaranteed Income Research

MPP Professional Paper

In Partial Fulfillment of the Master of Public Policy Degree Requirements
The Hubert H. Humphrey School of Public Affairs
The University of Minnesota

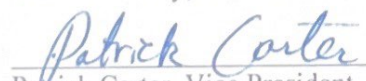
Alex Meyers

April 24th, 2024

Signature below of Paper Supervisor certifies successful completion of oral presentation and completion of final written version:



Maria Hanratty, Assoc. Prof., Paper Supervisor Date, oral presentation 5/2/24 Date, paper completion



Patrick Carter, Vice President, Results for America, Second Committee Member Date
Signature of Second Committee Member, certifying successful completion of professional paper 4/24/24

Introduction

Writing for the Washington Post in 2022, journalist Megan Greenwell wrote that, “If empirical evidence ruled the world, guaranteed income would be available to every poor person in America, and many of those people would no longer be poor. But empirical evidence does not rule the world” (Greenwell, 2022). As more U.S. cities and non-profits have launched guaranteed income pilots, this perspective seems likely to grow: if we can prove that this type of intervention works well, why do we keep testing it on a small scale, instead of adopting widely?

A key challenge lies in “proving” that a program has worked, even when it has been the subject of a well-designed evaluation. In recent decades, there has been an increasing focus in policymaking circles to rely on evidence when making decisions. However, both empirical results and subjective evaluations of the quality of evidence influence policymaker decisionmaking, especially when interpreting results is not straightforward. Further, policymakers embody two roles vis-à-vis the evidence: as *consumers* of existing evidence, and as *producers* of new evidence. Advocates of evidence-based policymaking have taken steps to describe how policymakers interpret evidence and rely on it for decision making (the consumer role). Less has been said about how policymakers influence the quality of evidence that is produced through their decisions in policy design and implementation (the producer role). As a result, there are fewer established best practices for using policy to produce high quality knowledge, and implemented policies may fail to achieve intended results to this end.

Combining policy evaluation frameworks, I describe the process through which policymakers interpret existing evidence, noting both objective and subjective evaluations which influence the impact of data. I suggest five best practices for writing policy which will produce evidence to influence action. I then consider the existing evidence on guaranteed basic income

(GBI) programs in the United States, identifying gaps in policymaker knowledge based on the identified evaluation framework. Finally, I review the language of the 2024 Minnesota Guaranteed Income Grant Program proposal currently making its way through the Minnesota House of Representatives, recommending enhancements to improve its evidence-production capacity based on the best practices. It is my intention that this research outlines clear best practices for writing policy which generates the knowledge needed to advance impactful programs.

Statement of Positionality

I come to this research with expertise, opinions, and potential biases developed through ten years working on empirical research methods, including randomized control trials, for Fortune 100 firms in the private sector. As a leader in marketing performance measurement, I supported clients in designing and implementing testing programs to inform and optimize budget strategies. This work involved developing analytical models, including media mix models relying primarily on Bayesian regression, as well as designing and executing in-market trials relying on randomized controls at both the geographic (e.g. city v. city) and individual level. Importantly, my role focused on supporting senior leaders in interpreting the results of these analyses and determining how to use information to implement significant budget shifts, often exceeding tens of millions of dollars in spending. Through their decisions, these leaders produced evidence of the efficacy of different strategies, and acted as both consumers and producers of evidence.

My clients often viewed themselves as consumers of evidence without recognizing their impact on evidence production. These leaders, seeking better information to decide how to allocate scarce resources to the most effective actions, would happily invest in robust testing

efforts. However, they often had little appreciation for how much influence they held over the production of the evidence they were waiting for. They would defer to partners on aspects of evaluation design while dictating other components, building studies that were a hodgepodge of best practices and leadership preferences. These decisions about what metrics to evaluate, for how long, during what time frame, and so on can and do change results and interpretations. What's more, leaders and their peers usually have a pre-existing conscious or unconscious idea of results that would indicate a need to act. These dynamics often came into conflict, with studies failing to produce actionable results because they were not designed effectively to deliver the needed data. As a result, despite quality evidence, leaders were often unable to take transformative action.

While it was straightforward to make the case for more measurement, and it was often possible to allocate some small portion of budget to innovative tests, it was substantially more difficult to build the case for a rollout of learnings—to move from pilots to full scale strategy change. Leaders faced questions about the applicability of test findings to the broader market; challenges in interpreting analytical results clearly; skepticism of the motives and objectivity of analysis partners, and more. As a result, organizations would dump hundreds of thousands or millions of dollars into technology and testing programs that, while producing insights with potential for significant impact, would go unacted upon, leaving organizations unable to move past the status quo.

While my clients were not working in a public capacity, their relationship to evidence shares many similarities with that of policymakers. This includes challenges such as identifying which types of questions can be answered effectively through testing, determining the level of effect required to choose to move an effective project forward (i.e. is statistical significance

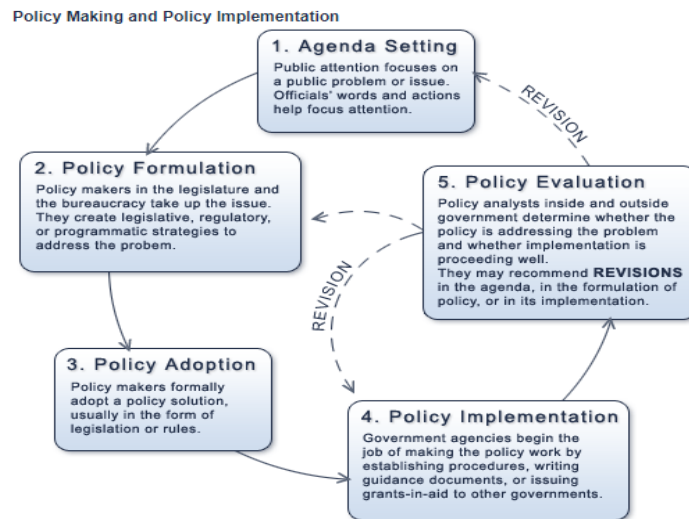
sufficient?), and a whole host of issues associated with the replicability, credibility, and comprehensiveness of insights. Through these experiences, I have seen how often even the best designed tests can be insufficient to drive action, regardless of the measured results.

Throughout this paper, I apply insights from my private sector work to public sector workings. I assume the decision-making processes in these two arenas share many similarities, while recognizing that there are differences in goals and priorities. I further recognize that empirical evidence is only one factor influencing decision making, and in many cases will not be the most important consideration. Nevertheless, I contend that policymakers play these dual roles of evidence producers and consumers regardless of their intentions, and therefore recognizing best practices which can be considered within the specific context at play is a useful practice.

The policymaking process: consuming and producing evidence

Policymaking is generally framed as a circular process entailing agenda setting, policy formulation, policy adoption, policy implementation, and policy evaluation. (Howlett & Geist, 2015) Figure 1 shows this process, highlighting the way evaluation is generally expected to influence policy formulation and implementation through revisions based on data and learnings. This view implies a fairly linear process; realistically, these different steps are likely to overlap, switch order, or be skipped entirely. This messiness creates significant uncertainty about how to advance policy concepts. In response to this uncertainty, policymakers look to evidence to guide policy development (Benson & Jordan, 2015).

Figure 1. The policymaking process



Source: Texas Politics, n.d.

Evidence, in its many different forms, can contribute to each of the different phases, not just evaluation. Indeed, Head (2015, 284 - 285) proposes at least eight roles that researchers and the evidence they produce can play in policy making, ranging from “raising public awareness of key problems and issues” (i.e. agenda setting) to “providing answers to specific questions through modeling and evaluations” and “providing independent scrutiny of government initiatives and developments” (i.e. policy evaluation). Given these many roles of evidence, all types of policymakers, from legislators to agency bureaucrats to policy advocates, are likely to be impacted by it through their work. Whether listening to constituent demands to ascertain policy priority or reviewing scientific studies to determine a program’s impact, these many types of policymakers are constantly identifying relevant evidence and determining how to proceed based on what it offers. In each policymaking phase, the policymaker plays two roles in relation to evidence: **consumer** of existing evidence and **producer** of new evidence.

As **consumers** of evidence, policymakers consider the existing body of evidence and what conclusions can be drawn from it. These conclusions include defining what is known and not known, as well as subjective confidence in the information. Consuming evidence is not an

apolitical or objective process, and different policymakers have different relationships to evidence and the appropriate ways to understand and use it. The political aspects of policy consumption must be noted, as they can contradict the general idea of the objectivity of science-based knowledge. Moreover, as Head notes, “cultural and organizational differences between governmental and academic institutions” result in differences in incentives and approaches to evidence production which influence the types of data produced (Nutley et al., 2007, as cited in Head, 2015, p. 285; Bogenschneider & Corbett, 2010) Policymakers may not value evidence in their own process or may ignore evidence that is misaligned with their personal views on an issue or goals for the policymaking process. On the other hand, policymakers may misinterpret data (intentionally or unintentionally) to make a particular point. Even when interpreted correctly, an over-reliance on evidence can be problematic. Evidence-based policymaking can “prefer scientific expertise and quantitative precision over other forms of knowledge” (Head, 2015, p. 283). It can also contribute to an emphasis on “efficiency...rather than...innovation and social benefit...” Evidence’s impact cannot be understood outside the context of the policymakers who consume it, and therefore it is essential to understand how policymakers operate in this role.

As policymakers consume evidence, they also adopt the language and practices of evidence-based policymaking in their own policy writing activities. In this way, policy makers increasingly serve as **producers** of evidence, a role that is increasingly institutionalized. Christina Ciocca Eller, writing for evidence-based policy advocacy organization Results for America, notes the success since 2012 of the Federal government creating evidence-based policy functions across many different agencies, including the eventual passage of the Evidence Act (2018), which “mandated that each of the 24 agencies covered by the Chief Financial Officers

Act (1990) appoint an Evaluation Officer and...that all agencies appoint a Chief Data Officer” to oversee evidence building, evidence use, and data driven decision making activities (Ciocca Eller, 2024, p. 17). This growth in federal attention has also conditioned a growing community of nonprofit and academic organizations focusing their evidence-building expertise on policy work. However, the efficacy of these evidence-focused bureaucrats may be limited if all policymakers (including everyone from elected officials to agency staff to advocates) do not recognize their roles in the evidence-producing process.

Any action to implement, expand, continue, change, cancel, or ignore a policy produces new evidence, which further impacts the policymaking process. Policymakers play a significant role in deciding whether their policy will contribute to evidence explicitly (e.g. through designing programs as random control trials) or implicitly (e.g. by extending a benefit to a larger sector of the population), and the extent to which evidence will inform program design (e.g. through requiring certain performance measures). Although this paper discusses the evidence-producing merits of a proposed piece of legislation (and therefore focuses on legislative activity), the frameworks discussed here are likely to have relevance throughout this process.

In addition to occupying different functional roles, policymakers can also have highly divergent goals for policy and its capacity to produce evidence. Evidence production may not be priority as policymakers focus on implementing programs which benefit constituents as quickly and efficiently as possible. For programs already in place, policymakers may oppose evidence production if there’s a concern that results could be detrimental to the status quo. Even for policymakers who are focused on measurement, longer-term goals to produce clear impact results can conflict with short-term demands for measurable outcomes, “lead[ing] to a focus on visible projects rather than on building foundations for sustainable benefits” (Head, 2015, p.

286). In other words, policymakers may focus on being able to show program implementation success (counting how many people received a benefit) as opposed to showing program outcome success (proving the benefit had its intended effect). Although the policymaker relationship with evidence may range from support to indifference to opposition, all must contend with the fact that perception of policy impact will always influence the policymaking process. Thoughtful consideration to this end lends itself to best practices that can influence the policy drafting process.

Consuming and producing evidence: the rise of “Evidence-based Policymaking”

The rise of “evidence-based policymaking” as an explicit priority of many federal, state, and local governments increases the importance of understanding policymakers’ process for interpreting evidence. “Evidence” of one form or another is used in all policy-making decisions—whether in the form of subjective assessments based on the individual’s experience or expectations of how the world works, or more objective evidence arising from designed studies and analysis. However, the concept of “evidence-based policymaking” (EBP) is increasingly associated with more robust analytical methods (Baron, 2018). Jon Baron defines “evidence-based policy” as including “the application of rigorous research methods, particularly randomized control trials (RCTs)...and the use of such evidence to focus public and private resources on effective interventions (Baron, 2018). Xu et al. offer a slightly broader definition, where EBP “refers to ‘the systematic use of findings from program evaluations and outcome analyses (“evidence”) to guide policy and funding decisions’ (The Pew Charitable Trust, 2017, as cited in Xu et al., 2024, p. 3 - 4). Bowen & Zwi note that the evidenced-based policymaking *literature* “largely relates to only one type of evidence—research” (Bowen & Zwi, 2005, p.

0601). They define research as including “empirical evidence from randomized control trials and other trials,” “time series analyses,” “qualitative studies,” and other forms of analytical studies. They suggest that other forms of evidence used in policymaking include “knowledge and information,” such as the results of consultations with networks and groups; “ideas and interests,” including expert knowledge/opinions; “politics”; and “economics.” For the purpose of this paper, a broad definition of evidence is appropriate. As will be shown, a broad definition works in large part because policymakers already have a sense of how to interpret and weigh different types of evidence.

The idea of measuring what works and using that information to design future policies may seem straightforward, but government’s heavy focus on systematized approaches to this process is a recent trend. The fundamentals of evidence-based policymaking are primarily borrowed from the health sciences, where the use of randomized control studies extends back to at least the 1946 streptomycin trials (Bhatt, 2010). In the 1980s and 90s, large-scale random control trials (RCTs) began to become popular for social policy, especially for evaluations of welfare and employment efforts. This trend has also been noted at the state level: as of 2017, “all U.S. state governments had at least one instance of EBP at a program, agency, or state level” (The Pew Charitable Trust, 2017, as cited in Xu et al., 2024, p. 4). In 2013, evidence clearinghouses built and run by the federal government collected evidence of effectiveness from several hundred programs; in 2023, over 20,000 were evaluated (Ciocca Eller, 2024). While it is challenging to determine how often evidence is meaningfully impacting policy decisions, there has certainly been an explosive change in the reliance on evidence as a justification for choices. In 2013 “3 federal agencies invested \$660 million in 6 federal grant programs that defined,

prioritized, or encouraged evidence of effectiveness when allocating funds. In 2023, 11 agencies invested \$30 billion” in such programs (Ciocca Eller, 2024, p. 8).

Baron suggests that this rise in attention on evidence-based policymaking is due to an increasing need for rigorous evaluation methods to measure the effects of interventions, arguing that, “...many countries have already exploited the kinds of interventions that have blockbuster effects. To continue to make progress, evaluation methods need to detect effects that are more modest but still quite important....To determine whether the intervention caused the outcomes we observe, we need an evaluation method that controls for confounding factors—such as a [random control trial (RCT)] that randomizes a sizable number of [subjects]” (Baron, 2018). However, this hypothesis carries a highly cynical undertone—suggesting the primary reason for utilizing evidence in policymaking is that the potential effect of all remaining policy opportunities is so marginal as to require advance study designs in order to measure the effects. Xu et al. suggest instead that evidence-based policymaking is a tool which policymakers can use to “predict effectiveness of a similar program in their own contexts” (Xu et al., 2024, p. 6). In other words, while it is true that evidence helps to determine whether a policy has been effective in the past, it is also important for evaluating whether the same policy would be effective in a different context in the future. In this way, evidence-based policymaking helps “improve government effectiveness by developing and utilizing a more rigorous base of information and scientific evidence to guide decisions about program design, funding, implementation, and management,” not just to allow for measurement of the impact of relatively less impactful programs (Xu et al., 2024, p. 4).

Xu et al. suggest two primary mechanisms which influence policymakers’ use of evidence, and which contribute to the heavier reliance on quantitative methods: “heuristics

related to [evidence's] credibility" which policymakers apply when evaluating evidence, and the institutionalization of EBP norms, which "encourage civil servants to focus on causal evidence" (Xu et al., 2024, p. 5). These two features are connected to one another: the heuristics tend to favor more analytical methods and are codified to encourage other policymakers to focus on evidence which fares well in relation to the heuristics. In other words, policymakers who believe RCTs are the "gold standard" for evaluating program efficacy write policies which refer to RCTs as the gold standard, reinforcing this belief and the reliance on these types of tools, for better or worse.

The heuristics which Xu et al. outline fit into three broad buckets: the degree to which the evidence is likely to predict a future program's effect in the policymaker's unique context, the confidence policymakers can have in the authors of the evidence (especially in terms of bias), and the comprehensiveness of the evidence, especially in terms of demographic subgroup impacts (Xu et al., 2024). Each of these factors, and especially the predictive nature of evidence, results in "EBP initiatives favor[ing] a hierarchy of evidence which ranks the validity of evidence based on its empirical strategy....On such a "pyramid," evidence from RCTs and natural/quasi experiments are generally considered at the top" (Newman, 2022 and Doleac, 2019, as cited in Xu et al., 2024, p. 6). Individual policymakers will prefer these empirical methods over other potential sources of evidence because of their perception that insights from such studies will have more predictive validity for future policy designs.

The confidence policymakers have in RCTs and experimental design then makes its way into formal policy, creating the "EBP norms" which Xu et al. note. A 2017 report of the federal Commission on Evidence-Based Policymaking, while acknowledging that "'evidence' can be defined broadly as information that aids the generation of a conclusion," was intentional in more

narrowly scoping evidence for their purposes as “information produced by ‘statistical activities’ with a ‘statistical purpose’.” (Commission on Evidence Based Policymaking, 2017, p. 8). In describing the types of ‘evidence’ the Commission would consider, they noted that “the result [of a statistical activity] summarizes information about a group rather than a single individual or organization.” Indeed, any ‘evidence’ which assesses impact on an individual ‘means the information is not being used for statistical activities and thus, would not be termed ‘evidence’ in the Commission’s definition.” Since the late 1990s, presidential administrations have increasingly established policy requiring federal agencies to consider evidence, including definitions that have increasingly emphasized more robust analytical methods: see E.O. 12866 (1993)¹; the Evidence-Based Policymaking Commission Act of 2016²; the Foundations for Evidence-Based Policymaking Act of 2018³; and Presidential Memorandum of January 27th, 2021⁴.

What is clear is from this history is that policymakers have connected to the idea of rigorous statistical methods as a useful tool for identifying potentially effective policy and have institutionalized this view. As Cartwright and Hardie write, “Over the last twenty years...it has become standard to require policymakers to base their recommendations on evidence. That is now uncontroversial to the point of triviality—of course, policy should be based on the facts” (Cartwright & Hardie, 2012, p. 3 - 13). As uncontroversial as it may be to desire policy based on facts, it is not so clear that policymakers possess a consistent framework for consuming

¹ Requiring cost-benefit analysis for all federal agency regulations.

² Establishing a commission to develop a strategy for increasing the availability and use of data to build evidence about government programs.

³ Requiring agencies to plan to develop statistical evidence to support policymaking.

⁴ Explicitly identifying methods including the “use of pilot projects, randomized control trials, quantitative-survey research and statistical analysis, qualitative research...and other approaches that may be informed by the social and behavioral sciences and data science.”

evidence. The heuristics outlined by Xu et al. focus primarily on attributes of studies (how they are designed, who conducted them) as opposed to the actual outcomes reported. In other words, while the literature shows that policymakers do rely on evidence, and that they have methods for determining whether they trust the evidence they review, it is not so easy to determine whether or when policymakers are effectively interpreting the evidence in front of them.

Consuming Evidence: A General Framework

During my time in the private sector, a key tactic I used for designing effective evaluations was to determine the impact of potential outcomes prior to conducting the evaluation. Often, clients would approach the evaluation process linearly: identify a question, design and implement an analysis, interpret the results to identify an answer, and formulate a response based on what was learned. However, usually the potential answers to a question reflect a small universe of outcomes. Generally, an analysis can only do one of three things: it can support or refute a pre-existing hypothesis, or it can bring to light new factors which change the hypothesis entirely. Of course, the strength of the learning (i.e. how confident we can be that a hypothesis has been supported or refuted) matters for action, but I often found it valuable to ask “How will you act if your hypothesis is supported or refuted?” By considering this question in advance, clients were able to determine whether an evaluation would truly change their behavior, and if so, what would be necessary to take the next step. Through this contemplation, we often adjusted the evaluation design to ensure we answered the most important question for motivating change. To apply this approach to policy making in the public sphere, it is essential to understand how other policymakers will interpret produced evidence.

There are two useful frameworks that policymakers can rely on to understand how produced evidence is likely to be interpreted. Checkland et. al (2021, p. 467) suggest that policy

be evaluated both in terms of “program success” (i.e. the extent to which the tested program concept is shown to be successful) and “political success” (i.e. the extent to which the testing of the program created any political consequences). While their framework is especially oriented toward pilots (i.e. small scale, intentional tests of policy concepts), their insights can be generally applied to any form of research which tackles these two types of success. This framework highlights the possibility that a pilot could demonstrate that a program is effective at meeting its stated goals while simultaneously failing to generate political consequences leading to lasting change. As noted, under this logic, we can extend the framing to other forms of evidence—a passionate op-ed with little scientific evidence of program efficacy could nevertheless generate significant public attention, thereby creating political impact without proving program impact.

In my experience, much emphasis is placed on ensuring a pilot will demonstrate program success, while the political success of the pilot is assumed. In one example of this, I was working with a client with a significant television marketing budget. Leadership was skeptical of the impact some ads were having, and measurement tools we were using suggested they were overspending on television advertising. We proposed a randomized control test, suspending television ads in some geographic regions while maintaining them in other, similar markets. By the end of the test, the markets without television had performed nearly identically to those with; indeed, there was no statistically discernible change in sales. However, when we used this data to propose reducing the television budget, we faced pushback that television was too big a part of the marketing program to cut—the risk of lost sales outweighed the potential benefit of reduced cost. So, although we were able to design a pilot which delivered statistically valid and impactful results, we were not able to achieve the political impact needed to drive change. We had

incorrectly assumed that decision makers would rely on evidence once it was presented, as opposed to acknowledging the political effort required to ensure action on pilot outcomes.

Checkland et. al’s framework asks six questions which policymakers can apply to various forms of evidence. The questions touch consider the pilot process (impacts of having conducted a pilot, regardless of outcomes), outcomes (the impacts of the program itself on participants), and effects (the results of the program after pilot completion). These features are further divided into program (relating to the program itself) and political (the broader context within the program was piloted) dynamics. Taken together, this results in six questions as structured in Figure 2.

Figure 2: Six Measures of Pilot Impact

	Program Success	Political Success
Process	Was the program successfully developed and initiated?	What was the political impact of the initiation of the program?
Outcome	Did the pilot program achieve its stated goals? If not, do these negative results provide information about where to focus next?	Did the outcomes of the program have political impact?
Effect	What happened next?	What has been the political effect of the pilot?

Source: Checkland et al., 2021

The first three questions are narrowly pointed at the subject of the evaluation itself. Asking these questions helps policymakers to determine whether a particular program is working. The last three questions put the results of the research into a broader context: given this positive or negative result (or supported or refuted hypothesis), how has the political environment responded? In this way, the framework answers both “what happened?” and “what

was done because of what happened?” As I have suggested, these questions can be asked in advance of executing a program, and the evaluation can then support or refute expectations.

However, as already noted, evidence does not operate outside the context of those who interpret and utilize it. Checkland et al.’s framework generally leans toward the objective, highlighting the observed effects of a program. However, their political success questions fail to explain why an initiative with program success would not have had meaningful impact. This objective evidence of program success will inevitably interact with the subjective process of interpretation by any policymaker engaging the data, and this interpretation will contribute to political outcomes.

To study the subjective factors most influencing policy decisions, Xu et al. surveyed over 1,000 civil servants with the “authority to make decisions on budget, grantmaking, and legislative/policy work” (Xu et al., 2024, p. 10). These policymakers were asked to review pairs of program proposals with variation in the evidence on empirical strategy, recency, location, author affiliation, and effectiveness in demographic subgroups. By varying the features of different programs across the population, the researchers were able to identify which specific features of evidence are most influential in policymaker choices to recommend a program. In this way, the researchers provide at least a partial explanation for the political success of programs (“what was done because of what happened?”), demonstrating why a particular program’s evidence basis may have been more or less trusted beyond the reported results.

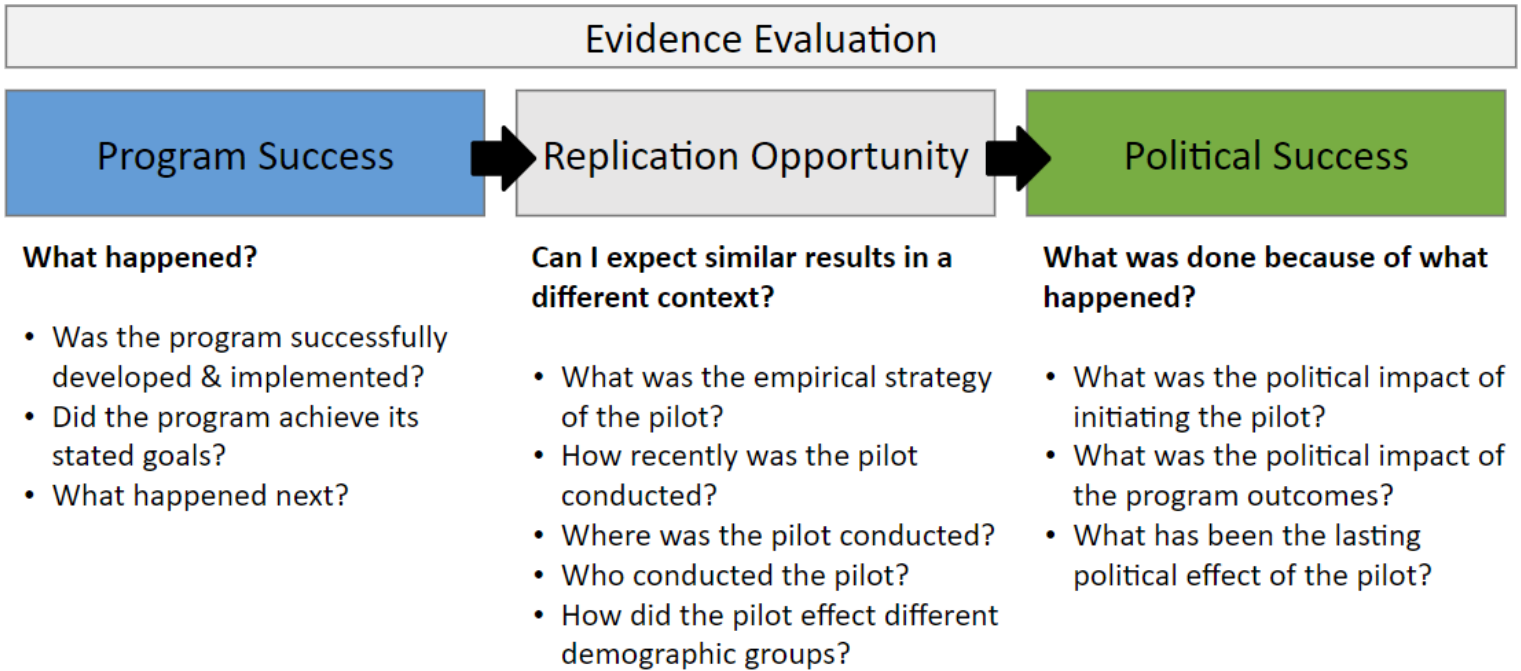
Xu et. al find five preferences policymakers express when assessing evidence, each of which contributes to an overarching goal service of determining whether a program is likely to generate similar results in their unique context. In general, policymakers will prefer evidence with:

- 1) Robust empirical methodologies, especially RCTs or natural/quasi-experimental studies
- 2) More recency
- 3) Greater geographic proximity to the policymaker (especially from within their own state)
- 4) Independent and/or third party evaluators
- 5) Demonstrated positive impacts on demographic subgroups

By combining the Checkland et al. framework for program objective success with the Xu et al. framework for subjective interpretation of evidence, we capture a more complete view of how policymakers incorporate evidence into their decision making. First, program success (the objective measures of program impact) will be considered. Second, policymakers will interpret this evidence, using heuristics to consider whether they believe the outcomes could be replicated in their environment. Finally, there will be political outcomes, both resulting from policymaker interpretations and, in the future, providing another layer of evidence for the program's potential impact.

Figure 3 summarizes this theoretical approach, orienting to the three key questions offered by Checkland et al. and Xu et al.: What happened? Can I expect similar results in a different context? And what was done because of what happened?

Figure 3. Suggested approach to evaluating research for policymakers



Producing Evidence: Applying What Policymakers Know About Themselves

It has so far been demonstrated that policymakers are increasingly emphasizing evidence in their decision-making processes, and that there are both best practices for evaluating evidence for policymaking as well as general heuristics upon which policymakers rely in their subjective interpretation of data. However, despite positive signs that the focus on evidence-based policy is increasing policymaker awareness of the importance of evidence generation, there remains

significant opportunity to improve policymakers' capabilities in this space. Ciocca Eller (2024, p. 51) suggests that “the knowledge generated by a decade of evaluation and data work is not being used in critical decision-making contexts,” including policy development. In response to this, she suggests that resources be prepared for policy leaders to support their awareness of best practices for incorporating evidence into the policy-making process.

One valuable source of information to consider when generating best practices is the knowledge we have about how policymakers interpret evidence. By designing policy which produces evidence that aligns to the heuristics identified so far, policymakers can strategically contribute to the body of evidence in a way that smooths the path to political success. Based on the combined Checkland-Xu framework discussed above, I propose five best practices:

- 1) Define the program explicitly to ensure program features are clear.
- 2) State specific outcomes of interest that the policy is intended to affect.
- 3) Incorporate high quality measurement methods, especially RCT when appropriate.
- 4) Utilize third party evaluation partners to design and execute measurement.
- 5) Design studies with scale and/or repeatability sufficient to respond to questions of geographic, temporal, and demographic differences.

The first two best practices support measurement of “program success” from the Checkland framework. Under the framework, to demonstrate efficacy, a program should produce evidence that it (1) was successfully implemented, (2) achieved its stated goals, and (3) had an ongoing effect. While I use language of “success” and “achieve” here, an unsuccessful program will still effectively contribute to the evidence, so long as it can be demonstrated that the program did not achieve its goals in fact, as opposed to not producing evidence one way or the other. In other words, a funded program which is unable to deliver its intended service because

of practical issues tells policymakers something about the path forward; a funded program which may or may not have delivered the intended service because no data was collected tells little to nothing. In suggesting best practices, I focus on demonstrating implementation and stated goal success; the ongoing effect is likely to be knowable regardless of the policymakers' actions.

Successfully writing policy which generates this information requires at least two things of policymakers: first, they must define the program explicitly to ensure it is clear what the features of the implemented program are; second, they must state specific goals and/or outcomes of interest that the policy is intended to affect. Both these recommendations are best practices associated with the writing of program evaluations (U.S. Department of Health and Human Services Centers for Disease Control and Prevention, 2011). Moreover, excluding these features is a common cause of policy failing to generate useful data. Joellen Killion, in "Why Evaluations Fail," identifies "clarity of outcomes" and "clarity of evaluation purpose" as two of the three key risks to evaluations' ability to generate quality data (alongside "selection of appropriate methodology") (Killion, 2017). And while it is unreasonable to expect policymakers to design a complete program evaluation for every proposed policy, programs should be clear enough for evaluators to know if they were implemented as intended and articulate specific, measurable goals.

While these may sound like an achievable floor for policy development, the consistent application of such tools varies. When evaluation design and execution is required through legislation, with responsibility for determining the specifics deferred, there can be mixed results. Consider Minnesota's Office of Grants Management (OGM), which in 2008 implemented 13 requirements for state agencies managing state-funded grants. One of these requires state agencies to "collect regular progress reports to demonstrate how the grantee has used grant

funds...” (Office of Legislative Auditor, 2023). However, a 2023 report of the Office of the Legislative Auditor found that “...15 of 16 Minnesota Department of Education (MDE) grantees either submitted at least one report late or did not submit required reports at all.” Additionally, although staff are required to review these reports, “MDE staff documented that they had reviewed only 62% of the submitted progress reports.” In this scenario, leadership expectation that evaluation would be conducted, let alone used for program improvement, was not borne out in the actions of those to whom responsibility was deferred. In this scenario, a lack of specificity in the measurement approach is likely to have contributed to the eventual challenges in implementing a consistent process that could be actively monitored.

The third best practice is to incorporate high quality measurement methods, especially RCT when appropriate. For better or worse, RCTs are considered the “gold standard” of evidence-based policymaking (Ciocca Eller, 2024). They provide the most scientific approach to measuring the impact of a particular policy. On the other hand, they can lack external validity, meaning their findings cannot be extrapolated to different contexts. This is because they measure effects under highly controlled scenarios (Tomlinson, Ward, and Marlow, 2015). They can also be time- and energy-intensive, expensive, or simply impossible for the program under evaluation. Regardless of their strengths and weaknesses, Xu et al. demonstrated that policymakers have internalized the preference for these types of studies when reviewing evidence (Xu et al., 2024). There are many reasons policymakers may choose not to use RCTs, but those decisions should be made with consideration for the fact that evidence produced through other means is likely to be viewed with more skepticism. As such, where possible, policymakers should use the most robust measurement methods available.

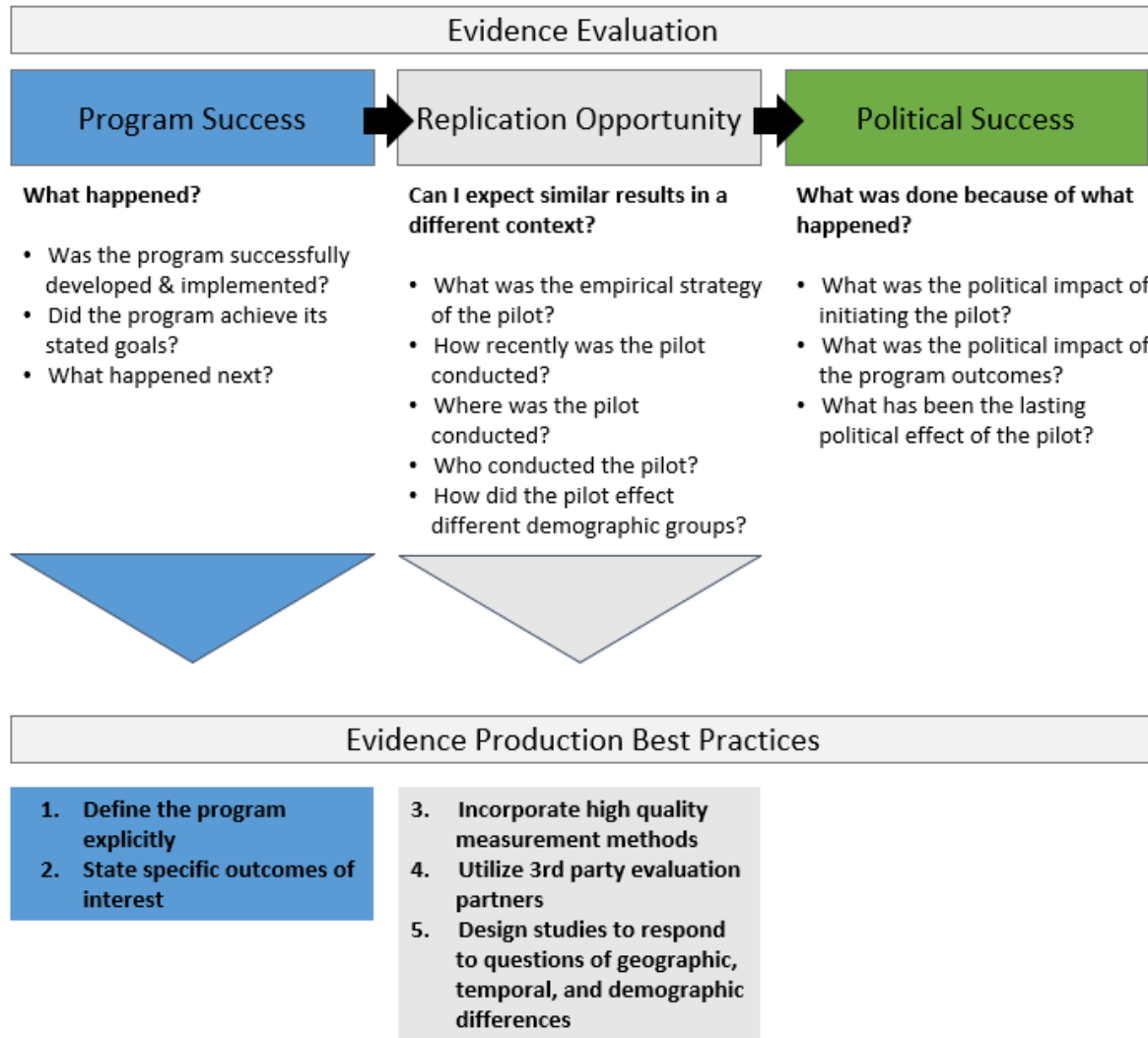
Use of robust methods goes hand-in-hand with the fourth best practice, relying on third party evaluation partners for the design and execution of evaluations. Demand for more support in evidence-based policymaking has resulted in a significant increase in the number and types of organizations available for this work. Ciocca Eller notes that from 2015 to 2022, the number of nonprofit and academic organizations focused on evidence-based policymaking grew from just a few to many dozens (Ciocca Eller, 2024). These organizations provide support not only in the effective design of evaluations, but also in funding to ensure evaluations are prioritized. In addition to these explicitly external partners, many layers of government are expanding their internal capabilities through establishing or expanding teams dedicated to evaluation work. In either case, policymakers have demonstrated a preference for evidence which is produced by seemingly expert and unbiased partners (Xu et al., 2024). Of course, working with these partners can be challenging or at least inconvenient for those directly responsible for program administration. In my experience, external evaluators may lack program context, especially if the evaluation is treated as a separate exercise conducted alongside the program, as opposed to an inherent part of the program itself. Their suggestions for how to effectively conduct evaluation can cause challenges for program design if introduced too late in the process. External evaluators can also struggle to gain the trust of program administrators and participants, especially if it is not made clear how their findings will be used. However, each of these challenges can be overcome by embedding evaluation in program design from the start. When this is done, the value of the external evaluators' expertise, coupled with the additional confidence they garner from policy consumers, outweigh the potential challenges.

The final best practice is to design studies which will stand up to policymaker preferences for evidence from their own state, which has been conducted recently, and which provides

information about key demographics. These demands are challenging to meet; obviously a policymaker in California can do nothing about a colleague in New York's preference for New York-based evidence, and creating quality evidence across a wide array of demographic groups requires larger studies and, as such, higher costs. However, policymakers can consider designing studies which will generate evidence that is convincing despite these biases. Best practices already outlined will support this, such as ensuring high quality methods, but policymakers should also consider factors like scale and duration to produce evidence which is able to incorporate a larger section of their constituency.

By applying these five best practices, policymakers can design policy proactively to answer the questions future consumers will ask and to overcome their biases to build trust. In this way, intentional policymakers claim their role as producers of evidence, and more effectively influence outcomes beyond the specific program at hand. Figure 4 expands on Figure 3, showing how these evidence production best practices align to the evidence consumption behaviors that inform them.

Figure 4. Evidence Production Best Practices to Support Policymaker Consumption



In the next section, I apply these frameworks to the case of guaranteed basic income (GBI) and the 2024 Minnesota Guaranteed Income Grant Program, which advances GBI research through the first state-administered and largest US-based GBI program ever run. I apply the Checkland and Xu to the body of GBI evidence to systematically identify what is known and what remains to be learned about this type of policy. Finally, I apply the five best practices outlined to the current draft of the Minnesota Guaranteed Income Grant Program bill, identifying opportunities to improve the language to contribute to the evidence most effectively.

Applying the framework to GBI evidence

Universal Basic Income (UBI) policies provide “regular cash payment by the government...on a monthly or annual basis.” (Stewart, 2021) While UBI generally refers to policies which provide these payments to everyone, Guaranteed Basic Income (GBI) programs target payments to a specific community—maybe individuals below a certain income threshold, or who are participating in another benefit program. In either case, the defining characteristic of the programs is that they are unconditional, both in terms of not requiring program recipients to take certain actions and not dictating how funds can be utilized. This feature distinguishes guaranteed income programs from other cash transfer programs like Unemployment Insurance (which requires recipients to “actively seek suitable employment each week”) (*Index*, n.d.) and the Supplemental Nutrition Assistance Program (which limits eligible spending to “food for the household”) (*What Can SNAP Buy? | Food and Nutrition Service*, n.d.). Generally, GBI programs are also distinguished from one-time cash transfer programs like the Earned Income Tax Credit, which only provides payment to individuals in a lump sum once per year as a tax refund. For the purpose of this paper, “GBI” will refer to programs that provide unconditional payments on a consistent basis over a defined period of time and would therefore include the more than 30 programs launched in the United States since 2020 and monitored by Mayors for a Guaranteed Income (*Mayors for a Guaranteed Income*, n.d.).

The idea of guaranteed income is not new to the United States. In the 1790s, Thomas Paine and Thomas Spence both advocated unconditional payment programs as a response to land privatization and individuals’ diminishing ability to provide for their own care through hunting, fishing, farming, and so forth (Widerquist, 2024). Dr. Martin Luther King advocated basic income as a response to “inadequate and often demeaning conditional programs,” saying in

Where Do We Go From Here: Chaos or Community?: "...the programs of the past all have another common failing — they are indirect. Each seeks to solve poverty by first solving something else. I am now convinced that the simplest approach will prove to be the most effective — the solution to poverty is to abolish it directly by a now widely discussed measure: the guaranteed income." (Widerquist, 2024; King Jr., 1967) Economists, including Milton Friedman, suggested that basic income "would represent a more effective approach to poverty than existing policies....[Representing an] attempt to simplify and streamline the welfare system while also making it more comprehensive." (Widerquist, 2024) In short, guaranteed income rests on the idea that individuals understand best how to improve their material conditions; by providing unconditional cash payments, governments empower recipients as opposed to attempting to dictate their behavior through paternalistic requirements. Guaranteed income is expected to provide recipients a floor to stand on, making it easier to participate in valuable but nevertheless unpaid activities, such as providing childcare, pursuing job training/education, or even simply taking time to seek more stable, fulfilling, and higher paying work.

Since 2010, "interest in basic income has grown enormously" (Widerquist, 2024). The Great Recession of 2008; increasing political and activist attention on income inequality; labor automation, and climate justice; Andrew Yang's 2020 presidential campaign and his "Freedom Dividend" basic income proposal; and the economic and cultural meltdown of the COVID-19 pandemic have all buoyed interest and experimentation in basic income policies. Along with interest, popular knowledge and support of such policies has increased. In polling released in 2024, 62% of likely voters supported a federal guaranteed income program which would provide monthly payments of \$500 or \$1,000 to all adult citizens making below the community's median income (Lake Research Partners, 2024). Moreover, after these voters were subject to "engaged

debate” over the programs to provide additional information, support increased by 2 points overall, with a 7-point increase in intensity of support (i.e. increase of voters who “strongly support” the proposal).

Skeptics of guaranteed income programs have traditionally leaned on two primary arguments: first, that providing such payments at any scale would be prohibitively expensive; and second, that unconditional payments will reduce participation in the labor force (Minogue, 2018). This first argument is based in simple math—if one advocates a universal income of \$12,000 per year per person and multiplies by the current U.S. population, they find themselves proposing a program cost of nearly \$4 trillion per year, or approximately 65% of total government expenditures in 2022 (\$6.3 trillion). (Ready, D., Salazar, J., & Verboon, C., 2023) Indeed, this figure would exceed the cost even of programs it might theoretically replace: in 2022, the total cost of social security, Medicare, Medicaid, refundable tax credits, and coronavirus relief totaled just short of \$3 trillion. Basic income opponents generally hypothesize this cost would be primarily born through increases in taxes, and that its scale would require those taxes be levied on at least some low- and middle-income individuals (Doar, 2018). Of course, this kind of one-to-one comparison with current program expenditures at an estimated full-scale rollout is premature, as the specifics of a national program have not been outlined. As already discussed, a final adopted program might limit recipients by income, as nearly all U.S. GBI pilots to date have. A universal program could be designed as a negative income tax, as was proposed by Nixon in the 1970s—creating a universal *guarantee* of income, but not universal *payments*. (Bregman, 2018) Or, a program that empowers individuals financially could justify the downsizing or elimination of other social safety net programs while also creating efficiencies and benefits in areas like healthcare (Wright, 2010). In other words, the potential cost of a fully

scaled program should not preclude exploration of its possible benefits, given we cannot yet know how those benefits would stack up against the expenditures. Instead, basic income studies should clearly and intentionally explore both sides of the cost-benefit equation.

The question of labor force participation is both a moral concern and an economic one. Oppositional writing on basic income likens it (pejoratively) to “welfare,” with the added concern that it is distributed “no strings attached.” (Glahn, 2022) Beyond this moral issue, opponents suggest “UBI would destroy—not improve—incentives to work,” causing individuals to be “less driven than ever to earn their way out of poverty” (Doar, 2018). Fortunately, program effects on labor participation is a common performance measure and one that is well suited to evaluation. In an era of evidence-based policymaking, we can treat hypothesized benefits and harms as just that: hypotheses to be tested empirically.

In the next section, we will use the Checkland and Xu frameworks for reviewing evidence as outlined. Using this framework, we will identify gaps in and issues with the current body of evidence that may be resolved by evidence-producing policymakers interested in advancing basic income knowledge and adoption.

Applying the Framework: Reviewing Existing Evidence

Basic income is an interesting policy to apply the consumer / producer of evidence concept to because it has both been tested many times in many different contexts *and* continues to be framed by advocates as something which should be piloted more. Why is it that, having (presumably) consumed the existing evidence of basic income, policymakers continue to support only continued small scale experiments? As producers of evidence, are policymakers designing programs which are likely to shift this conversation meaningfully? Using the framework for evidence consumption, I will review the existing literature to provide a point of view on what is

known or not. Then, applying the best practices for evidence-producing policy, I will evaluate the current draft of the Minnesota Basic Income Grant Program and propose improvements to advance basic income development.

Program Success - What Happened?

The Checkland framework begins with consideration of “program success”—answering to what happened when policymakers attempted to implement the program. This question is broken into three key parts: was the program successfully developed and implemented? Did the program achieve its stated goals? And what happened after the conclusion of the program? Many policy ideas never pass this first test, being successfully developed and implemented. Consider all the proposed legislation which never receives a hearing, is not passed, or fails to be implemented in its intended form. Even for programs which are implemented, Checkland suggests degrees of success, where a more robust program which rolls out faster suggests greater success than a smaller program or one which takes longer to begin.

Basic income programs have been developed and successfully implemented many times, in many different contexts. In 2022, the World Bank released a report documenting 962 cash transfer programs in 203 countries and estimating that “cash transfers were distributed to 1.36 billion people—17% of the world’s population—during the pandemic period” (Richterman et al., 2023, p. 575). Although the largest scale, truly unconditional basic income programs have been run in low- and middle-income nations, pandemic experiments rapidly expanded to developed nations, including the United States. Of the 962 programs identified by the World Bank report, “672 were newly introduced during the pandemic” and “about half of the population...of North America were covered” by unconditional cash transfers (e.g. the American Rescue Plan Act Economic Impact Payments) (Richterman et al., 2023, p. 575; Gentilini, 2022, p. 2). As of the

beginning of 2024, a dashboard of GBI pilots maintained by the Stanford Basic Income Lab, University of Pennsylvania Center for Guaranteed Income Research, and Mayors for a Basic Income reported data for 31 pilots completed or in process by cities and counties since September of 2020 (*Home*, 2023). Together, these U.S. pilots have made payments to over 10,000 individuals and have provided variation in design across multiple factors: size of payment (from \$375 to \$1,000/mo), duration (10 months to 3 years), number of participants (from 25 to 3,200), and eligibility/targeting. Programs throughout the world have been studied extensively, with the Stanford Basic Income Lab publishing an umbrella review summarizing sixteen reviews covering more than 500 papers concerning basic income programs run in U.S. and international contexts dating back to the 1970s (Hasdell, 2020).

While these many evaluations demonstrate that programs were implemented, it is challenging to determine whether they were implemented as originally intended. However, we can look to specific anecdotes for examples of success on this front, such as the Saint Paul, MN, People's Prosperity Pilot (PPP). On March 27th, 2020, the Coronavirus Aid, Relief, and Economic Security Act ("CARES Act") was signed into law and established funds for state and local governments to use to provide emergency financial assistance to individuals and families (Flynn et al., 2023). Throughout that spring and summer, the city of Saint Paul and local nonprofits conducted analysis to evaluate the opportunity for direct payment programs. On September 3rd, 2020 Mayor Melvin Carter signed an executive order directing Saint Paul's Office of Financial Empowerment to develop and implement a guaranteed income program. By October, the program and measurement approach had been designed, participants selected, and payments begun. The program concluded in April of 2022, and in November of 2023, the Center for Guaranteed Income Research (CGIR) at the University of Pennsylvania published their

evaluation of program effect. That report suggests that Saint Paul was able to design and execute this program as intended, although it is possible changes were made throughout. It certainly seems fair to suggest that Saint Paul was able to design, implement, and evaluate a basic income program bearing the key components of such an initiative, and any potential changes are likely to have been minor in the broader context of basic income research.

It's also worth noting that, in the U.S., this issue of program design standardization is somewhat alleviated by centralization of basic income research through the Center for Guaranteed Income Research (CGIR) at the University of Pennsylvania and their partnership with Mayors for a Guaranteed Income (MGI), a coalition of mayors advocating and testing basic income programs, and which “operates as a centralized clearinghouse that provides funding and technical assistance for cities looking to adopt their own GI payments” (Juras et al., 2023). In order to receive MGI support, cities are required to:

- Provide unrestricted cash payments to recipients of an amount between \$375 and \$1,000/month
- Run the pilot for at least 12 months
- Include at least 110 participants
- Not retest eligibility once initially determined (cities can set their own eligibility requirements, but once found eligible and included in the program, recipients will not be removed from the program until it concludes)
- Utilize a mixed methods evaluation approach
- Conduct public engagement “to effect narrative change”

Furthermore, CGIR, in partnership with Abt Associates, published a “Data Analysis Plan” to outline the approach that will be used for evaluating MGI-supported programs, detailing methods

for sample selection and recruitment, data collection, and data analysis (Juras et al., 2023). This plan covered 21 U.S. pilot studies as of publication in 2023 and will certainly contribute meaningfully to continued understanding of program development and implementation success. As each of these programs is executed and evaluated, policymakers consuming evidence will have greater cause to believe in its efficacy.

The program success framework turns next to “program outcome success”—the extent to which the program achieved its stated goals. Basic income’s success on this front is much more complicated to assess. First, as noted, much of the literature is new, with studies launched during the pandemic and only now producing published evaluations. Second, many of these programs have been conducted in developing nations, limiting (though not eliminating) their relevance for the U.S. context. Third, proponents of GBI believe it can have positive effects on many variables; as such, studies often measure a diverse array of outcomes, and there is inconsistency across studies. For instance, CGIR’s analysis plan dictates that all studies measure 20 variables, including financial, psychological, and physical well-being, housing and food security, parenting practices, subjective sense of self, paid and unpaid work, job quality, and educational attainment (Juras et al., 2023). While this has the potential to create valuable insights on many metrics, it may create challenges when conducting studies using small sample sizes, which all CGIR’s published evaluations do.

Despite these challenges, evidence that has been produced is compelling on many outcome measures. As noted, evidence in low- and middle-income international trials has been very promising; however, for simplicity and comparability, this paper will focus on the U.S. trials conducted by MBI and CGIR in recent years. CGIR has begun releasing reports for the first wave of pilots (as of date, this includes Stockton, CA’s SEED, Saint Paul, MN’s PPP, Paterson,

NJ's Guaranteed Income Pilot Program (GIPP), and Ulster County, NY's Project Resilience). In each of these, between 100 and 130 participants received \$500/month for 12 to 18 months.

Except for the PPP, all were structured as RCTs. The PPP compared metrics over time to identify trends. The studies consider the same primary measures noted in the analysis plan above, with some variation.

Results across metrics vary widely across reports; however, for nearly all metrics, data directionally improves for treatment group participants, with some reports achieving statistical significance and others not. On measures of financial well-being, child well-being, physical well-being, and life attitude measures, at least three of five studies reported statistically significant improvements for basic income recipients. Four of five studies reported increased rates of employment over control, although these differences were not statistically significant. However, no difference on this measure refutes the idea that providing basic income would incentivize individuals to stop working. In Patterson, where a decline in full-time employment was measured, it was associated with “ease[ing] the burden of making ends meet and providing care, allowing recipients to better balance their time” as they tended to child- and elder-care (DeYoung et al., 2023). As one might expect, across all reports the qualitative evidence is very positive, with participants regularly reporting increased feelings of agency and self-worth. As a consumer of this evidence, one is likely to feel optimistic, but not convinced. The struggle of these studies to produce sufficient power to consistently report statistically significant outcomes is a major issue. However, as additional studies are released utilizing consistent outcome measures and RCT structures, meta analysis may be possible to produce greater confidence.

Checkland's framework turns next to “program effects success,” which asks whether the pilot had any lasting effects for its relevant program: was it sustained or more broadly rolled out,

modified, or discontinued? As in the exploration of program effects, this answer varies widely across the individual GBI pilots that have been discussed, with one important consistency: no GBI program in the United States has been sustained and rolled out broadly within the entity conducting the pilot (i.e. none of the cities or counties conducting pandemic pilots have adopted ongoing programs. Most U.S. pilots (SEED, GIPP, and Project Resilience) ended without immediate follow-up after the initial funding period. In 2022, Saint Paul announced CollegeBound Boost, a second iteration of the PPP which again pilots GBI, this time in a true random control trial design (*Guaranteed Income / Saint Paul Minnesota*, n.d.). However, as noted, many other pilot programs have launched throughout the country since these first trials. The argument can certainly be made that, at a national level, these programs have had a meaningful effect of promoting interest and inspiring more pilot action. However, the individual pilots have not been successful in leaning on their results to inspire broader adoption at any scale.

Replication Opportunity - Can One Expect Similar Results In a Different Context?

We can now apply the Xu framework to develop a sense of how policymakers are likely to respond to the evidence. As noted, the primary question policymakers are trying to answer through their interpretation of evidence is whether they can expect to see similar results in their unique context. To predict this, they consider the empirical strategy of the study, the recency and geography of the pilot, the credibility of those conducting the evaluation, and the evaluations ability to report on effects on a variety of demographic groups.

Focusing on these recent U.S. pilots, I would expect mixed responses from lawmakers. While the pilots are designed as “gold standard” RCTs, they utilize small sample sizes which

limit power, likely contributing to low occurrences of statistical significance on outcomes of interest. Pilots were conducted recently, but most occurred during the pandemic or early in pandemic recovery. The pilots were run exclusively in blue states, and mostly run by cities in major metropolitan areas. All these studies were run in partnership with MGI and CGIR, which lends credibility from the outside evaluator perspective, but may not provide as much confidence of objectivity. Finally, the studies generally do not disaggregate results by demographic group, again because the sample sizes are too small to allow this.

Looking at all these features together, a key theme emerges: while studies have used methods best practices designed and evaluated by generally credible organizations, they have lacked the scale necessary to demonstrate effects across geographic and demographic lines or with power sufficient to consistently deliver confident outcomes. This challenge is likely to be alleviated in the future by the growing body of MGI/CGIR pilots, the resulting opportunity for meta-analysis, and the increasing pool of larger scale pilots. However, the impact of any one relatively small pilot is not likely to move the conversation forward meaningfully based on the frameworks proposed here.

Political Success - What Was Done Because of the Pilot?

Given these dynamics, we can now consider the final Checkland question, “beyond the individual program itself, were there any positive political implications of the pilot’s initiation, outcomes, and/or effects?” While Checkland breaks each of these aspects down separately, it is sufficient for this paper’s purposes to combine them and simply explore the political effects of these pilots overall. There have been four significant impacts: first, the initiation of the pilots and their measured impact success has created greater awareness of direct transfer programs as a tool in policymakers’ belts. The creation of MGI, and their subsequent work with CGIR to create

tools like the analysis plan, has empowered interested policymakers to establish programs that have the experience of basic income proponents behind them. Second, the design of the Saint Paul PPP utilizing American Rescue Plan dollars highlighted a funding stream which policymakers may not have considered for these types of programs. These dynamics taken together have encouraged the increased pace of piloting similar programs in other cities and counties. The specific use of ARPA dollars to fund these programs is highlighted in the National League of Cities “Local Action Tracker,” which calls out large projects in Chicago, IL (\$30 million dedicated to providing 5,000 families with \$500 per month for one year) and Seattle, WA (\$25 million in direct cash assistance) (Pine & Geraghty, 2021).

The launch of these programs has, in turn, contributed to the greater awareness of and support for basic income efforts, noted earlier. Finally, this attention has encouraged larger entities (i.e. state and federal lawmakers) to begin considering basic income programs. In 2023, Rep. Bonnie Watson Coleman (D-NJ) reintroduced the Guaranteed Income Pilot Program Act to the federal legislature. This act would establish a program within the U.S. Department of Health & Human Services (HHS) to distribute monthly cash payments to 10,000 individuals for a three-year period (Watson Coleman, B, 2023). During both the 2023 and now 2024 legislative sessions, Minnesota state lawmakers have proposed legislation to establish the Minnesota Guaranteed Income Pilot Program, which would direct the Minnesota Department of Human Services to establish a grant program to provide funding to local governments to establish basic income pilots providing monthly payments of between \$300 and \$1,200 to recipients for at least 18 months. Grant applicants would be responsible for proposing program design, with the legislation providing little direction for recipient targeting, measurement approach, or primary outcome measures.

Although the federal program failed to generate traction, the Minnesota program is currently making its way through committees and receiving popular interest, including a Star Tribune write-up. While the policy is unlikely to pass this year, its advocates have suggested this is an opportunity to start a conversation and consider how the policy could be improved. This paper will now turn to this question, specifically considering how the policymakers can claim their roles as producers of evidence.

Producing Evidence: Best Practices

As discussed, the current body of GBI evidence has had some political success in creating greater interest in similar programs and a willingness from lawmakers to advocate their own experiments. In Minnesota, several programs, including Saint Paul’s CollegeBound Boost (a follow-up to the PPP) are in process. During the 2023 legislative session, lawmakers funded a basic income pilot targeting homeless youth to be administered by the Minnesota Department of Human Services (DHS). Now, the Minnesota Basic Income Grant Program is being advanced through committees in the state House by Rep. Athena Hollins (DFL – Saint Paul). If passed, this bill would allocate \$100 million to a basic income pilot program. The Minnesota Department of Human Services would administer the grant, empowering local government entities, tribal nations, or non-profit organizations (grantees) to provide monthly cash payments to recipients “For the purpose of disrupting poverty, building wealth, advancing equity, and supporting recipients’ basic needs” (Minnesota Basic Income Grant Program, 2024). Grantees would be able to provide payments of at least \$500 per month to eligible recipients for at least 18 months. The bill requires recipients to be receiving public benefits or have a household income less than or equal to 300% of federal poverty guidelines, although grantees can set additional eligibility requirements. The bill also requires various aspects of evaluation: grantees must submit to

evaluation by an entity selected by DHS, must participate in a basic income “community of practice” to share learnings, and must measure:

- Economic status of recipients before/after participation
- Employment status of recipients before/after participation
- Physical & mental health status before/after participation
- Food & housing security before/after participation
- Ability to enroll in further education due to participation
- Any other relevant information identified by the grantee

Using the best practices derived from the Checkland-Xu framework, we can consider whether this bill structure is ideal for advancing the evidence base of GBI.

1) Define the program explicitly to ensure clarity of implemented program features.

The Minnesota Basic Income Grant Program (MBIG) provides grantees with significant latitude in the design of the program, creating significant risk of generating many highly dissimilar programs throughout the state. The bill allows for a variety of grant administrators (governmental and non-profit), a range of monthly payment amounts and payment durations, with no specified logic for determining these attributes, and only dictates that grantees submit to evaluation by a DHS-selected party, without clearly delineating the evaluation method. The policy does specify broad goals and related performance measures; however, specific metrics to be used in measurement are not dictated, and the breadth of goals is likely to create challenges in measuring outcomes with confidence, as demonstrated in the MGI/CGIR pilots.

Proponents of this bill could argue that, as DHS is the administrator of the grant, they will be able to define evaluation methods more explicitly, and to focus on consistency across grantee

designs should they choose. However, this policy as written does not require DHS to operate in that fashion, and statements from bill authors suggest that this would not be desired—instead, the focus has been on empowering local communities to decide what will work best for them (Hollins, March 14th, 2024). This latitude may be a valuable feature of the bill; however, it must be recognized that it is likely to reduce the evaluative power of the program. DHS could choose to require certain evaluation techniques (e.g. RCTs) or not; grantees could apply for large grants to include sample sizes sufficient to measure statistically significant effects or could design small programs to target a specific subset of the population. Each of these decisions may have merit, but as a vehicle for producing evidence, this flexibility is likely to reduce the MBIG’s efficacy.

To achieve maximum explanatory effect from the program, the authors should design a more narrowly scoped program which considers the other best practices noted below. While this program would reduce local grantee autonomy, it would likely increase the evidence value of the program overall. It would also contribute something unique to the basic income pilot program pool: a state-designed program with significant funding and access to a large, diverse geographic region with a variety of population characteristics. Indeed, by centralizing the program approach, Minnesota would become the largest administrator of a basic income pilot in the country. By funding many disparate studies, Minnesota still contributes to the evidence, but in a much more disjointed way less likely to deliver the kind of learning that policymakers are most likely to respond to.

2) State specific outcomes of interest that the policy is intended to affect.

MBIG follows the CGIR analysis plan’s lead in defining a wide variety of performance measures which grantees must collect data for, while also allowing grantees to collect additional

data as they see fit. However, will CGIR also provides a detailed data dictionary which defines the variables to be collected for quantitative analysis and outlines survey design for qualitative data, MBIG leaves flexibility for DHS, the evaluator, or the grantees to determine how to define the measures outlined and collect data to report on them. As with the CGIR studies, it is possible that including so many variables of interest could muddy the waters on program impact; additionally, the flexibility in program design could lead grantees to design programs in unique ways to intentionally solve for specific outcomes over others. For instance, the 2023 homeless youth program is available only to youth experiencing homelessness in St. Louis and Hennepin Counties. This narrow program is likely to have very different effects on housing stability and many of the other metrics than a program available more broadly to low-income families. However, because the program is designed differently, it may be challenging to determine how effective basic income programs are for the subpopulation of homeless youth relative to the broader population (as the broader population has not been included in a similar program). In other words, by designing a program only for a narrow segment of the population, we lose the ability to compare impacts to the population at large.

An alternative would be to design a larger program which includes these subpopulations, whose results can then be disaggregated. As noted, the funding level of MBIG sets it up to be the largest basic income program in the country. At \$500/month for 18 months, the program could provide payments to over 10,000 families or individuals, more than 3 times larger than the largest program CGIR is currently evaluating (Los Angeles, CA's BIG LEAP is providing \$1,000/month to 3,048 participants for 12 months). At this scale and duration, Minnesota has a significant opportunity to move knowledge of basic income efficacy forward by ensuring consistency of metrics and power of analysis.

3) Incorporate high quality measurement methods, especially RCT when appropriate.

Policymakers are more convinced by evidence arising from RCT studies than other empirical methods. As demonstrated by CGIR’s work, basic income programs can be effectively designed as RCT studies which can be paired with other analysis forms, including qualitative evaluation. However, MBIG does not require the use of RCT in program design. Again, while DHS and/or the selected evaluation partner could implement this requirement, the policy as written opens the door to less analytical designs. As discussed, while there are limitations in RCTs which may represent good reasons for allowing other approaches, policymakers demonstrated preference for this method makes it the best for generating quality evidence. As such, specifying that MBIG grants must be measured utilizing RCTs would be a more effective way to ensure an effective contribution to basic income knowledge.

4) Utilize third party evaluation partners to design and execute measurement.

MBIG does require evaluation by “at least one independent, research-based entity.” (Minnesota Basic Income Grant Program, 2024) This approach aligns to the best practice of utilizing a third-party evaluator with the independence and technical knowledge to produce trustworthy evidence in the eyes of policymakers. When choosing an evaluator, DHS could take this best practice a step further by working with an organization conducting other analyses and with the ability to align evaluation design across studies. Intentionally designing the evaluations in this way would increase the opportunity for meta-analysis, expanding the opportunity for confidence and disaggregation potential of insights.

5) Design studies with scale and/or repeatability sufficient to respond to questions of geographic, temporal, and demographic differences.

Policymakers look to evidence to understand what kind of effects they can expect in their unique contexts. In the case of MBIG, the authors should consider which policymakers they are most trying to convince through their study. It seems most logical to focus on Minnesota lawmakers who could support the continuation of this program should it prove successful; as such, the program should be designed to convince lawmakers throughout the state that basic income will work for their constituents. With this in mind, the program would ideally provide insights for the types of geographic and demographic variability present in the state. As Rep. Hollins notes, the program may have different impacts within different communities; however, instead of relying on a diversity of grant applicants to provide the necessary population to study these differences, the program could be designed intentionally to include a representative sample from throughout the state.

Additionally, by dictating a program design now, Minnesota could begin establishing the infrastructure to run such a program, instead of asking a wide variety of grant partners to design and implement their own programs. The question of repeatability becomes much more challenging when programs might vary on size and duration of payments, eligibility requirements, fund distribution mechanisms, and so on. By designing a consistent program now, Minnesota can borrow from best practices already identified across other basic income programs and iterate to improve the program over time. While this approach might cause less effective methods to be employed from the outset, it avoids a situation where many small programs test a variety of approaches without the scale to determine what works best.

The current MBIG legislation represents a useful starting point for a statewide basic income program. However, it fails to account for evidence-production best practices identified

here. If implemented as currently written, it risks producing basic income evidence which is not conclusive or trusted by policymakers reviewing it in the future. At best, this risks causing continued confusion about what works best in basic income, and ongoing calls for piloting which slow broader rollout. At worst, this approach opens the door to poorly designed programs that could be targets for basic income opponents.

It is possible that the bill's authors' primary motivation is *not* to produce high quality evidence. As currently written, the bill emphasizes goals of poverty reduction and such—not to learn about the efficacy of GBI in achieving these goals. Certainly, the best practices above create some challenges which might reduce the immediate effect of the bill, for instance by limiting local organizations' ability to tailor program design to target communities based on specialized knowledge. However, given the current state of basic income discourse, and noting that the authors have included multiple references to evaluation design while consistently referring to the program as a pilot, it seems reasonable to infer their wish to produce high quality evidence.

By taking account of the best practices outlined here, MBIG can produce high quality evidence while supporting program recipients. This approach has the potential to contribute dramatically to the state's understanding of basic income efficacy, potentially creating a path to the continuation and expansion of a program with the potential to dramatically improve residents' lives.

Conclusion

As evidence-based policy practices become more influential, it is essential for policymakers to understand their role in both consuming and producing useful evidence. By reviewing existing research on evidence-consumption behaviors and best practices, I have

proposed a consistent framework for evaluating both the objective and subjective merits of a body of evidence. I then used this framework to propose five best practices for writing evidence-producing policy which is more likely to be acted on by other policymakers. By following these best practices, I suggest that policymakers will design trials which more efficiently generate the knowledge needed to confidently advance new program concepts.

In the second half of this paper, I applied this approach to the body of evidence surrounding guaranteed income programs. I find that, despite their use of robust empirical methods implemented by independent research organizations, the small scale of pandemic-era pilots in the United States has limited their potential for convincing policymakers of evidence sufficient to adopt programs at scale. Considering this challenge, I use the five evidence-producing policy best practices to review the Minnesota Guaranteed Income Grant Program proposal, noting several enhancements that would support the policy in achieving its stated goal of demonstrating basic income efficacy.

By internalizing and consistently applying the approach outlined here, policymakers seeking to demonstrate the efficacy of programs are likely to have more success. Developing savvy evidence producers among policymakers will contribute to a more useful body of evidence for effective policies and will empower decision makers to allocate resources to those programs most likely to improve the lives of constituents.

Bibliography

- Baron, J. (2018). A Brief History of Evidence-Based Policy. *The ANNALS of the American Academy of Political and Social Science*, Volume 678(Issue 1), 40–50.
<https://doi.org/10.1177/0002716218763128>
- Benson, D., & Jordan, A. (2015). Environmental Policy: Protection and Regulation. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 778–783). Elsevier.
<https://doi.org/10.1016/B978-0-08-097086-8.91014-6>
- Bhatt, A. (2010). Evolution of Clinical Research: A History Before and Beyond James Lind. *Perspectives in Clinical Research*, 1(1), 6–10.
- Bogenschneider, K., Corbett, T.J. (2010) Evidence-Based Policy Making: Insights from Policy-Minded Researchers and Research-Minded Policy Makers. Routledge, London.
- Bowen, S., & Zwi, A. B. (2005). Pathways to “Evidence-Informed” Policy and Practice: A Framework for Action. *PLoS Medicine*, 2(7), e166.
<https://doi.org/10.1371/journal.pmed.0020166>
- Bregman, R. (2018). *Don't believe in a universal basic income? This is why it would work, and how we can pay for it*. World Economic Forum. <https://www.weforum.org/agenda/2018/05/how-we-make-basic-income-reality-Rutger-Bregman/>
- Cartwright, N. & Hardie, J. (2012). Evidence-Based Policy: A Practical Guide to Doing It Better. 10.1093/acprof:osobl/9780199841608.001.0001.
- Checkland, K., Hammond, J., Coleman, A., Macinnes, J., Mikelyte, R., Croke, S., Billings, J., Bailey, S., & Allen, P. (2023). “Success” in policy piloting: Process, programs, and politics. *Public Administration*, 101(2), 463–480. <https://doi.org/10.1111/padm.12790>
- Ciocca Eller, C. (2024). The Power of Evidence to Drive America’s Progress. Results for America. <https://results4america.org/wp-content/uploads/2024/02/The-Power-of-Evidence-to-Drive-Americas-Progress-Results-for-America.pdf>
- Commission on Evidence-based Policymaking. (2017). The Promise of Evidence-based Policymaking. <https://www2.census.gov/adrm/fesac/2017-12-15/Abraham-CEP-final-report.pdf>
- DeYoung, E., Tandon, N., Neves, C., Castro, A., & West, S. (2023). *The American Guaranteed Income Studies: Paterson, New Jersey*. University of Pennsylvania Center for Guaranteed Income Research.
https://static1.squarespace.com/static/5fdc101bc3cfd2dcf0a2244/t/65805bc4127a02409e650544/1702910920285/CGIR+Final+Report_Paterson+NJ_12.2023.pdf
- Doar, R. (2018). Universal Basic Income Would Undermine the Success of Our Safety Net. *The Catalyst*, 10. <https://www.bushcenter.org/catalyst/are-we-ready/doar-universal-basic-income/>

- Doleac, J. L. (2019). 'evidence-based policy' should reflect a hierarchy of evidence. *Journal of Policy Analysis and Management*, 38(2), 517-519
- Evidence-Based Policymaking Commission Act of 2016, Pub. L. No. 114-140, 130 Stat. 327 (2016). Retrieved from <https://www.congress.gov/114/plaws/publ140/PLAW-114publ140.pdf>
- Executive Order 12866, Regulatory Planning and Review (1993)
- Executive Order 14005, Memorandum for the Heads of Executive Departments and Agencies on Restoring Trust in Government Through Scientific Integrity and Evidence-Based Policymaking, President Joseph R. Biden Jr. (2021, January 27). Retrieved from <https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/27/memorandum-on-restoring-trust-in-government-through-scientific-integrity-and-evidence-based-policymaking/>
- Flynn, K., Castro, A., West, S., Tandon, N., Thompson, A., & Cochran, B. (2023). *The American Guaranteed Income Studies: Saint Paul, Minnesota*. University of Pennsylvania Center for Guaranteed Income Research. https://static1.squarespace.com/static/5fdc101bc3cfda2dcf0a2244/t/65805c8bf9ff72725dd4ef13/1702911117175/CGIR+Final+Report_Saint+Paul+PPP_2023.pdf
- Foundations for Evidence-Based Policymaking Act of 2018, Pub. L. No. 115-435, 132 Stat. 5449 (2019). Retrieved from <https://www.congress.gov/bill/115th-congress/house-bill/4174>
- Guaranteed Income | Saint Paul Minnesota*. (n.d.). Saint Paul Minnesota. Retrieved March 14, 2024, from <https://www.stpaul.gov/departments/financial-empowerment/guaranteed-income>
- Gentilini, U. 2022. Cash Transfers in Pandemic Times: Evidence, Practices, and Implications from the Largest Scale Up in History. *World Bank Group*. <https://documents1.worldbank.org/curated/en/099800007112236655/pdf/P17658505ca3820930a254018e229a30bf8.pdf>
- Glahn, B. (2022, January 18). *Money for nothing*. American Experiment. <https://www.americanexperiment.org/money-for-nothing/>
- Greenwell, M. (2022, October 24). Perspective | Universal Basic Income Has Been Tested Repeatedly. It Works. Will America Ever Embrace It? *Washington Post*. <https://www.washingtonpost.com/magazine/2022/10/24/universal-basic-income/>
- Hasdell, R. (2020). What We Know About Universal Basic Income: A Cross-Synthesis of Reviews. Stanford Basic Income Lab. <https://basicincome.stanford.edu/research/papers/what-we-know-about-universal-basic-income/>
- Head, B. W. (2015). Policy Analysis: Evidence Based Policy-Making. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 281–287). Elsevier. <https://doi.org/10.1016/B978-0-08-097086-8.75030-6>

- Hollins, A. (2024, March 14th). Testimony on the Minnesota Basic Income Grant Program [verbal communication, Human Services Finance Committee, 2023 – 2024 Legislative Session]. Minnesota House of Representatives.
- Home. (2023). The Guaranteed Income Pilots Dashboard. <https://guaranteedincome.us/>
- Howlett, M., & Giest, S. (2015). Policy Cycle. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 288–292). Elsevier. <https://doi.org/10.1016/B978-0-08-097086-8.75031-8>
- Index (.). (n.d.). Applicants - Unemployment Insurance Minnesota. Retrieved February 8, 2024, from <https://uimn.org/applicants/needtoknow/eligibility-requirements/index.jsp>
- Juras, R., Jefferson, A., Yang, H., Thomas, H., & Kappil, T. (2023). Data Analysis Plan: Mayors for a Guaranteed Income Pilot Evaluations. *University of Pennsylvania Center for Guaranteed Income Research*.
https://static1.squarespace.com/static/5fdc101bc3cfda2dcf0a2244/t/6576fcabc1a71b16fb4d8d6f/1702296780440/MGI_data_analysis_plan.pdf
- Killion, J. (2017). Why Evaluations Fail. *The Learning Professional*, 38(2), 26 – 30.
<https://learningforward.org/wp-content/uploads/2017/08/why-evaluations-fail.pdf>
- King Jr., M. L. (1967). *Where Do We Go from Here: Chaos or Community?* Beacon Press.
- Lake Research Partners. (2024). National Survey Results Summary.
<https://static1.squarespace.com/static/60ae8e339f75051fd95f792e/t/65d3f264d2a6c70038ca259d/1708388968295/LRP+MGI+national+Survey+Memo+1.5.2024+%282%29.pdf>
- Mayors for a Guaranteed Income*. (n.d.). Mayors for a Guaranteed Income. Retrieved April 21, 2024, from <https://www.mayorsforagi.org>
- Minnesota Basic Income Grant Program, H.F. 2666, Minnesota House of Representatives, Minnesota, 2023-2024 Legislative Session.
- Minogue, R. (2018). Five Problems with Universal Basic Income. *Third Way*.
<https://thirdway.imgix.net/pdfs/five-problems-with-universal-basic-income.pdf>
- Newman, J. (2022). Politics, public administration, and evidence-based policy. In *Handbook on the politics of public administration* (p. 82-92).
- Nutley, S., Walter, I., Davies, H. (2007). *Using Evidence: How Research Can Inform Public Services*. Policy Press, Bristol.
- Office of the Legislative Auditor, State of Minnesota. 2023. Oversight of State-Funded Grants to Nonprofit Organizations. <https://www.auditor.leg.state.mn.us/ped/pedrep/osfgno.pdf>

- The Pew Charitable Trust. (2017). How states engage in evidence-based policymaking: A national assessment. [https://www.pewtrusts.org/~media/assets/2017/01/how-states-engage-in-evidence-based-policymaking.pdf](https://www.pewtrusts.org/~/media/assets/2017/01/how-states-engage-in-evidence-based-policymaking.pdf).
- Pine, J. & Geraghty, L. (2021). Guaranteed Income Pilot Projects with American Rescue Plan Act Funding. *National League of Cities*. <https://www.nlc.org/article/2021/07/21/guaranteed-income-pilot-projects-with-american-rescue-plan-act-funding/#:~:text=Minneapolis%2C%20MN%20is%20planning%20to,months%20to%20approximately%20200%20households>.
- Ready, D., Salazar, J., & Verboon, C. (2023). The Federal Budget in Fiscal Year 2022. *Congressional Budget Office*. <https://www.cbo.gov/system/files/2023-03/58888-Budget-Infographic.pdf>
- Richterman, A., Millien, C., Bair, E. F., Jerome, G., Suffrin, J. C. D., Behrman, J. R., & Thirumurthy, H. (2023). The effects of cash transfers on adult and child mortality in low- and middle-income countries. *Nature*, 618(7965), 575–582. <https://doi.org/10.1038/s41586-023-06116-2>
- Stewart, M. (2021, March 10). *The pros and cons of universal basic income*. College of Arts and Sciences, University of North Carolina. <https://college.unc.edu/2021/03/universal-basic-income/>
- Texas Politics—The Policy Making Process*. (n.d.). Retrieved April 15, 2024, from https://texaspolitics.utexas.edu/archive/html/bur/features/0303_01/policy.html
- Tomlinson, M., Ward, C., & Marlow, M. (2015). Improving the efficiency of evidence-based interventions: The strengths and limitations of randomised controlled trials. *South African Crime Quarterly*, 51(0), 43. <https://doi.org/10.4314/sacq.v51i1.5>
- U.S. Department of Health and Human Services Centers for Disease Control and Prevention. (2011) Office of the Director, Office of Strategy and Innovation. Introduction to program evaluation for public health programs: A self-study guide. Atlanta, GA: Centers for Disease Control and Prevention.
- Watson Coleman, B. (2023). Watson Coleman Leads Introduction of Guaranteed Income Pilot Program. <https://watsoncoleman.house.gov/newsroom/press-releases/watson-coleman-leads-introduction-of-guaranteed-income-pilot-program>
- What Can SNAP Buy? | Food and Nutrition Service*. (n.d.). Retrieved February 8, 2024, from <https://www.fns.usda.gov/snap/eligible-food-items>
- Widerquist, K. (2024, March 7). The Deep and Enduring History of Universal Basic Income. *The MIT Press Reader*. <https://thereader.mitpress.mit.edu/the-deep-and-enduring-history-of-universal-basic-income/>
- Wright, E. O. (2010). *Envisioning Real Utopias*. Verso Books.

Xu, C., Cheng, Y. (Daniel), Wang, S., Merrick, W., & Carter, P. (2024). Evaluating Use of Evidence in U.S. State Governments: A Conjoint Analysis. *SSRN Electronic Journal*.
<https://doi.org/10.2139/ssrn.4693327>