

# THE UNFINISHED BUSINESS OF EVIDENCE BUILDING

## DIRECTIONS FOR THE NEXT GENERATION

**BRIAN SCHOLL**

**T**his chapter covers a few odds and ends about evidence. Evidence is a curious business, in some cases too much business and in other cases not enough business. Much is yet to be tapped on the evidence front, and issues of who does the building and how the evidence is used once built are critical.

The content of this chapter comes from a mix of my own experiences and views of evidence, policy, and implementation that I have developed over a period spanning decades in government, the private sector, and academic research settings. I have worked from the very micro or direct-provider level with organizations so small they cannot rightly be called organizations, on up to the most ivory tower levels of policy, and, really, almost every level in between. Those experiences have enabled me to see an enormous amount of variation in the way organizations organize their work and the obstacles they face, and to develop perspectives on why evidence works for them or why it does not.

When we in the evidence community talk about building evidence, so often our conversation goes to the math and the statistics of it all: experiments, treatment effects, causal estimates, randomization protocols, and

so on. Those are so important in so many ways, but also so unimportant in so many other ways. In my mind, it is the organizations, the institutions, and the people that really matter. The wrong people at the top (leadership) can dead-end any efforts to generate evidence. Wrong people generating evidence get to all the wrong questions and all the wrong answers using all the wrong methods. Wronged people at the bottom (beneficiaries) bear the consequences of getting policies and programs wrong, and those beneficiaries might rally with torches and pitchforks if they feel it is their right that someone gets a policy or program to work (even if they are not articulating their concerns as stemming from a lack of evidence). Organizations and institutions can be set up by good leaders to carry the torch even when bad leaders come along. Evidence is critical to getting our work to work and keeping our democracy democratic, but it cannot help if folks have their fingers in their ears.

In this chapter, I am not conveying a read of the evidence on evidence but, rather, a view of informed experience, and sketching something of an ideal that I think organizations can work toward. Hopefully, some of the lessons I have learned over the years are helpful, or will at least help you the reader feel you are not alone.

### THE PROMISE OF EVIDENCE

These days, “evidence” is something of a buzzword. Evidence, though, is not so much a fad as a set of techniques for advancing knowledge about particular questions.

Evidence generation is an investment. Think about a government agency or a nonprofit that is trying to make something good happen for a number of people: feeding the hungry, sheltering the homeless, teaching students reading and math skills, or reducing unemployment. As shorthand, we can say the outcomes related to this goal are *Y*. For example, maybe *Y* is the number of people who have found homes or the number of people who are not hungry or the number of people who have a job.

The agency has a set of policy levers it can pull to accomplish its goal, like providing food, shelter, or education, and, perhaps, it has the ability to invent new levers to try new ideas. Let’s call those policy levers “*p*.” The overarching problems are: Which levers does the organization pull? How much should it pull each one? How does it think up new levers to construct?

Since the outcome  $Y$  changes when we pull different levers in different amounts, we can think of  $Y$  being a function of  $p$ ; change the mix of levers and we get different outcomes for our beneficiaries. We can denote this relationship as  $Y(p)$ , with  $Y$  being a function of  $p$ . (Don't worry, we are not going to do any real math here; we just need a way to communicate that our outcomes  $Y$  change when our policy mix  $p$  changes.)

There are only so many potential beneficiaries out there, and usually a policy or social sector organization has only limited resources with which to reach them. We can call the best possible value of the outcomes  $Y^*$ . This is the best we can ever do; so  $Y^*$  might be the maximum achievable level of employment (or the minimum level of unemployment).  $Y^*$  is the best outcome the organization can facilitate with any potential combination of activities  $p$ .

Achieving  $Y^*$  is tricky. We will get the best outcome  $Y^*$  using the ideal combination of policies  $p^*$ . This is where the evidence comes in. Evidence can help us get closer to figuring out  $Y^*$  and  $p^*$ . More important than those optimum values, evidence can help us understand the relationship between  $Y$  and  $p$ . The evidence does not just land in our lap. To build this evidence, we need to invest effort and time and resources. If we have not been building the evidence, we have no idea what the relationship between  $Y$  and  $p$  really is. We also do not know what  $Y^*$  and  $p^*$  are. Heck, in many cases, we may not even have data that tells us what our current levels of  $Y$  are, or what the current mix of activities are that are in  $p$ . Without evidence, we are really just stumbling around in the dark without any idea of whether we are helping people or hurting people, whether we are doing our best or less or can improve. The more evidence we build and the better we get at building evidence, the better ideas we will have for trying things that are new.

What we have above is a sketch of what evidence can mean: evidence gives us hope. It is a flashlight when we are wandering in the dark. The further we get from the optimum ( $Y^*$ ), the worse off constituents, clients, and beneficiaries will be, and the more likely we are to be squandering our own resources. For example, if the Fed chair sets the wrong interest rate, or if Congress sets a bad tax policy, they can create distortions in the economy. Those distortions can motivate people to make dumb moves and cause other problems. Even a small social sector organization can inadvertently distort peoples' choices, incentivizing people to put their effort into the wrong thing; for example, by wasting their time in a classroom training that is not working, just because tuition is free. We need evidence of our impact to

make sure we are providing the right services for people. Organizational ignorance is a distortion, one that pushes the costs onto beneficiaries (through a lack of an appropriate solution to their problem) and potentially creates other negative consequences that hurt innocent bystanders.

### THE PATH TO EVIDENCE BUILDING

Buying into the promise of evidence is the first step in the journey. The next step along the path is figuring out how to succeed in building insightful and impactful evidence.

#### *The Power of Working Backward*

Let's discuss a problem that has plagued some evidence-driven organizations: the "dead-end study." Before we even begin: do not panic. There is a solution to the dead-end study: developing a credible theory of change by working backward from desired outcomes in the design process.

What is a dead-end study? Some organizations have built infrastructure, collected data, hired the right folks, engaged consultants, and so forth. In some cases, they have thought their efforts to be fruitless because big, careful, and sometimes expensive studies turned up point estimates of zero or close to it, or repeated tests have yielded inconsistent results, or their tests were simply too scattered to add up to a clear direction.

To be clear, the problem many organizations face is not, in fact, the null result. "Null results" are much maligned, but they can often be informative. Null results are shorthand for: "We tried a couple of different things and either nothing worked or our best ideas didn't show any incremental benefit." Null results feel like a failure because people may have put effort and resources into developing and testing an innovative new idea only to find out that it is a dud. Yet, solid evidence that something does not work is informative. It tells you what not to do (we will return to this point in another way). The problem is really not the null result itself; the problem is when the null result has left the organization without insight on what to do next. This happens when evidence is generated without the conditions to succeed.

I would estimate that at least 80 percent of my time working with organizations to develop evidence is spent on programmatic issues, and less than 20 percent is spent on methodological issues. That is, I spend most of my time asking *Why is the program doing what it is doing?*—a programmatic question—and only a small fraction of time asking *How do I generate evidence*

to tell if what it is doing actually works?—an evaluation question. The reason for this is that many organizations—from the small nonprofit to the large government policy institution—do not have a strong, coherent, and *credible* theory of change that links their activities (p) to a set of outcomes (Y) that they care about. *Credible* is the key word here because it is not sufficient to postulate wildly implausible causal links between activities and outcomes that are not justifiable using existing social science theory and evidence.

Without the theory of change and without some soul-searching to identify potential alternative activities, tests typically can comprise only a limited set of questions of the form: *Does our current activity work to effect change in our outcomes of interest?* We can design a test to determine if what the organization is *currently* doing works, but if the answer is no (it does not work), then we have gained little understanding of the factors that can affect Y, and we have little to guide us on what to do next from either an implementation perspective or an evaluation perspective.<sup>1</sup> An organization or program's theory of change is, in my experience, the most overlooked component of successful and meaningful evidence generation.

How do we construct a theory of change? Truthfully, it is not an easy task either methodologically, institutionally, or emotionally. We need to put our pride aside and open ourselves to hard and uncomfortable questions. Then we have to work backward.

Working backward means starting our theory of change at the end: identifying the outcomes of interest (Y), and developing the causal chain by working the pathway backward through outputs and intermediate outcomes and causal mechanisms on through to a set of activities that can credibly produce change in those outcomes.

Working backward not only forces critical thinking about the assumptions needed to connect current or proposed activities to intended outcomes but also can help organizations identify alternative activities that also can lead to the desired proximate and distal outcomes. In contrast, working forward (starting with the current activity mix) often has the potential to push everyone to contrive assumptions and explanations for how and why current activities lead to outputs that cause changes in behaviors that ultimately lead to the outcomes. In a sense, working forward basically *assumes* that the current activities work, whereas, with working backward, one might not ultimately even situate current activities in the set for consideration. In my experience, the assumptions embedded in working forward simply tend to be unreasonable. In working backward from desired outcomes, there

tends to be stronger footing, perhaps because ignoring current activities forces everyone to think critically about actions that can result in the stage of the causal chain they are focused on.

For example, suppose our goal is to improve student classroom outcomes and our current activity is to engage students in arts and crafts. In working forward, we might need to make a lot of very tenuous assumptions about the immediate effects of the crafts program to draw a link to better student performance. If we were to work backward from the premise that we are trying to help student performance, we might have to be critical of ourselves and identify the main obstacles to student performance and, in time, come to a recognition that there are better options than an arts and crafts program. To be uber-clear, such an arts and crafts program might be valuable for reasons other than student performance, or it may be an important complement to other student-centric programs, so it is possible that we just chose the wrong rationale for the program, but if student performance is the right outcome, we might need to consider alternative activities.

Working backward can lead to a more critical assessment of whether the activities make sense, since they have to fit within the path models described at more distal stages rather than forcing a path from activity to outcomes. To be fair, some organizations will be limited by cultural, capacity, mandate, funding, and other constraints that will narrow the practical range of activities they can implement; an association of school teachers providing after-school tutoring is unlikely to hop into providing basic income support for local families. Yet, in my experience, even in those organizations, working backward can force a much broader conversation about considering alternative (yet feasible) activity options than working forward can do. Evidence generation developed based on that causal chain also can provide a better basis for understanding the factors that affect the desired outcomes. For larger or policy-oriented organizations, working backward using existing social science theory and evidence is critical to help the organization think outside the box and critically assess activities for which alternatives and alternative methods of implementation are possible. These alternatives, if tested, can provide insights that curtail dead-end research.

I understand some of the reasons theories of change often are absent. Organizations may be sensitive to opening up the theory of change discussion. Theories of change are fundamentally an element of program design rather than evaluation, so even though they are critical to generating meaningful evidence, they can be viewed as outside of the evaluator's

domain. Organizations have activities, but they often do not have a credible theory of change for why they are doing what they are doing—or they figured it out so long ago that it may be deep in the recesses of their memory. Evaluators can either work within existing activities to develop tests of efficacy, or they can push the very difficult conversation that involves developing and vetting a credible theory of change. It often may be practical for the evaluator to postpone the difficult theory of change conversation and work on evaluations of existing activities to build trust with the programmatic people, which could facilitate tougher conversations in the future. The problem is that postponing those discussions can lead us to the dead-end study.

The theory of change does not eliminate the dead-end study entirely; it simply lays the foundation for meaningful evaluation. A credible theory of change is a necessary but not sufficient condition for meaningful evidence generation. Evidence still needs to be developed strategically, and in a methodologically sound way, because dead-end studies also can arise with a non-strategic approach or an ill-conceived methodology.<sup>2</sup>

### *Pragmatic Parameters: Leadership and Resources*

#### **Management Integration**

One would hope that the push for evidence would come from within the organization and resonate top and bottom through leadership and the rank and file, yet it often comes from external pressure. Management-centric evidence initiatives—ones that feed helpful perspectives directly into the decisions management encounters—will tend to find a warmer reception. Evidence programs should seek to tie into management goals as much as possible so the evidence can help optimize activities along the dimensions management cares about. At the same time, evidence programs can help steer management to a more suitable dashboard if it is not already looking at the right indicators of progress.

Management integration need not be confined to the evidence itself. In some circumstances, it could be creating data tools that support management's direct objectives. For example, consider that the first step in evidence generation often is taking stock of the activities the organization is engaged in, perhaps compiling a dataset of such activities. Instead of conducting this inventory as a one-off research activity, is there a way to create a data collection system that regularly reports out to management? Are

there data collection tools that can be delivered to staff that both captures data needed for evidence generation and helps staff perform the tasks they normally do, making their jobs easier and allowing management to capture productivity gains? Evidence needs friendly management to flourish, but it also will find management and the organization more friendly when it makes the organization's work easier. Good evidence programs will seek to align and integrate with an organization's operations and to keep focus on the organization's overall goals.

### **The Economics of Evidence**

Does evidence cost a fortune? In recent years, too many folks have gotten overzealous and proposed large and time consuming studies with costly data collection and so forth. There is definitely a role for that sort of evaluation, and in some settings, such investments may be necessary for big, expensive programs, broad policy issues, difficult to quantify and study issues, and initiatives that will affect a large number of people. These all deserve careful study and attention.

Yet, the economics of evidence is not really about cost control; it is about adapting research to the institutional incentives, constraints, and other realities that each organization and its leadership faces. What can an organization do? What are they required to do? To what measures are management held accountable? What are their major operational problems?

Unfortunately, the reality is that time and money tend to be focal points of those institutional incentives. One thing I have endeavored to do in organizations with which I have worked is to lower the costs and shorten the length of the evidence life cycle. Decision makers who are told that research will take three to five years and cost several millions of dollars will be unenthusiastic when their job tenure may last only a few years. Reducing the costs and time may be essential.

I cannot say I have a one-size-fits-all solution to money and time problems. Each situation should be examined based on the organization's own resources, opportunities, and constraints. In almost every organization, I find that focusing on better recordkeeping is a first step to supporting operations while also providing administrative data that can be organized by researchers to study the work of the organization. In large and complex organizations, I have found it often is important to make investments to build standing capacity for data collection, and to build internal technical

expertise, which will lower the marginal cost of evidence generation and make individual projects more easy to approve.<sup>3</sup> Smaller organizations will need to be creative but can do some things to generate good evidence without breaking the bank: partner with graduate students who hunger for interesting problems and unique datasets; start out with small qualitative research programs; leverage outreach networks to conduct data collection. Evidence programs can start out with baby steps.

### **Valuing Evidence Investments**

While evidence building does not need to bust the bank, it does need a reasonable amount of support. That is always an uphill battle, especially when decision makers are not researchers. Many professionals tend to see only the perspective of their own profession and cannot appreciate what it takes to generate informative evidence. Unfortunately, many non-researchers in leadership positions seem to believe that evidence generation can be free and instantaneous. The view may arise from ignorance about the costs and benefits of evidence generation, cynicism about the value of evidence, a lack of resources available for evidence, or a view arising because “research” in the form of a Google search feels so fast and free and easy that all evidence generation must be similarly quick and costless. Whatever the cause, this view is obviously unrealistic. Evidence is an R&D-like investment, which can be viewed and evaluated through the lens of cost per outcome.

For example, if status quo intervention A costs \$500,000 and intervention B costs \$100,000, then, in principle, there is a large gain from finding out that both interventions are equally effective. There are many ways to look at the value in this setting and how much one should be willing to spend on an evaluation, but a framework with the flavor of Return on Investment (ROI) is often reasonable. In fact, thinking in terms of ROI gives us a different context for the null results discussed above. In this example, the null hypothesis is that intervention A and B are equally effective. Failing to reject the null here feels disappointing from a programmatic and evidence-generation perspective; you did not come up with a better mouse-trap. Actually, though, you have achieved an impressive win. We can implement intervention B and save \$400,000 for other projects. Even if A and B have similar costs and identical impacts, externalities may differ considerably. One intervention might create distortions or adverse investment incentives or adversely affect local markets in other ways even if intentions are noble.<sup>4</sup>

While it is fair to think in terms of long-run ROI, evidence programs have other benefits beyond their direct focus of study. They can develop as an early warning system to better understand emerging risks and how to respond to them; they can develop internal expertise to be able to identify and address other problems; and they can inform a host of management decisions in a variety of contexts. In the case study by the Camden Coalition in this volume, a Health Information Exchange launched in 2010 provided benefits in the fight against COVID-19.<sup>5</sup>

The point here is that evidence should and can have value, but it will do so only if we are crafting it and evaluating it in the right way. If, for example, the academic value of evidence for researchers or the compliance value of evidence for funders is prioritized over more actionable evidence with practical value to those providing and receiving the services being evaluated, this can contribute to the sense that evidence is not worth what it costs, since it is not answering the questions that matter to these stakeholders and it is taking resources from other priorities.

The question that arises often enough is: Could we be over-investing in evidence? Are we doing too much? In my experience, that concern most frequently arises within organizations that are doing almost nothing, and often comes out of a fear of change. While it is clearly conceptually possible to be over-investing in evidence, I can think of no organization that is actually doing so. In my view, while this issue is often fretted about, we are nowhere close to a world where opportunities for evidence generation have been over-exploited.

### **Building an Infrastructure for Evidence**

Building credible evidence often requires technical skills: economics, statistics, other social sciences, econometrics, experimental design, qualitative research. All these skill areas can come into play when building a program of evidence or particular studies.

Most organizations do not have these capacities lying around (statistician in the cupboard?). Often enough, organizations reach out to consultants to augment their capacities. Consultants can bring expertise and experience and a fresh perspective on the work of the organization. After all, consultants have not been in the trenches trying to deliver the goods and services and policies, and they can ask “smart dumb questions.” Smart dumb questions are naïve questions about the organization, its work process,

its goals and motivation, and so forth that are asked until there is mutual clarity about activities and rationales. In my mind, smart dumb questions are essential, and organizations need to have the patience for these questions if they are to succeed in building evidence. These probes can be uncomfortable for the organization because they not only will include questions like: *What are you doing?* There also will be questions like: *Why are you doing this? Why do you think this works? What made you draw that connection?*

Those questions can be unsettling because organizations have spent a whole lot of time figuring out how to do what they do and may not remember the original rationale, or they may feel their decisions or motives are being challenged, or that they are being told they do not know anything. I often have needed to ask the same questions over and over again to make sure I have gotten it right and that I have understood the motivations and the details. That is, I ask a lot of smart dumb questions, which I find helps me if the questions are dumb enough.

Relying on consultants alone may not solve the evidence problem. They may leave the organization with an interesting research result or fresh perspectives, but rarely do they build the organization's internal capacity along the way. I have personally spent a lot of time worrying about the gap in knowledge and expertise between an organization and external help and the role of expertise in helping an organization achieve its goals. The issue really is the disparity in knowledge between the consultant and the client, and the propensity for that disparity to end in a whole lot of nothing.

Maybe an example will illustrate better than pontificating. When the term "impact evaluation" was all the rage, there were organizations that were externally bullied or forced to march down Impact Evaluation Highway. In many circumstances, consultants were hired to conduct impact evaluations, which sometimes generated reports that had the words "impact evaluation" written on the cover page. Report complete, external pressure eased, life went on. The problem was that some client organizations did not have the capacity to know what an impact evaluation was. Sure, some people knew a few of the basics, but by-and-large, if a client organization did not know much about impact evaluation, a consultant could pass along a report with "impact evaluation" in the title and with contents consisting of a bunch of nonsense, and no one would be the wiser. That is because the client organization did not have the skills to distinguish an impact evaluation from gobbledygook, or even to know the difference between an expert consultant and a charlatan.

One key protective factor in this realm is how evidence-fluent staff are and how much the organization has bought into evidence. Organizations with a lot of buy-in typically want to have the right people to make sure they get the questions and answers right. Organizations lacking the buy-in typically just want to check a box and move on.

Does every organization need to build its own evidence shop or have a team of PhDs at the top? Clearly no. After all, building an evidence group takes energy and resources that may be inefficient for really small organizations. Yet, in my view, most organizations should at minimum develop enough expertise to be good partners and informed consumers of the products they pay for or depend on. Social sector entities should establish relationships with organizations that have good track records of being honest brokers, particularly ones that have been not only tirelessly working with individual organizations to help them build and manage evidence programs but also trying to change the ecosystem so that evidence is more valued, valuable, and strategic. I expect that, over time, there will be more options for cooperative learning systems where small organizations can work together as an association or consortium to finance collective evidence-generation capacities and share learnings.

The larger the organization and the more complex its work, the more urgent it becomes to build an internal evidence team. In the case of large social sector organizations or government entities that develop wide-ranging policies, it often is critical to build internal expertise with technical specialists who can really understand the organization; integrate into work processes; ask important questions of colleagues; help the organization learn; and liaise with senior academics studying the domain. As the size and costs of programs increase and the need for evidence-generation activities grows, it becomes more important to avoid mismatches between the research objectives and the organization's goals because the consequences of bad programs or policies can be enormous. The fact that so many large policy entities around the world still operate without this internal capacity to generate evidence, and instead pass on the costs and distortions of ignorance to their beneficiaries and stakeholders, is, in my mind, unforgivable; the world is too complex, the policy challenges are too great, and resources are too scarce to be used unwisely.

Internal expertise has another advantage: the organization can potentially in-source aspects of the evidence production process that are costly and time-consuming to outsource. It can put people who care about the re-

search in the driver's seat. Of course, evaluator independence is often important, too, so in-sourcing has to be done in a way to preserve that independence so the research is kept honest.

### **Creating a Culture of Evidence**

In parliamentary debate, Winston Churchill once rebuffed a critic's chiding of a proposal's multitude of changes: "To improve is to change; to be perfect is to change often."<sup>6</sup> This could very well be the motto of the evidence community. Evidence, for the most part, is an exercise in innovation: how to make processes work better, how to develop better products or combinations of services. At its best, it really is about continuous improvement. Yet organizational change is tough unless there is enough buy-in from all the parts of the organization that are involved in the undertaking. A more evidence-friendly culture provides fertile ground for evidence to be developed, to prosper, and to be applied. Building a culture of evidence—a learning organization—is an important ingredient for change.

Learning organizations are organizations that have achieved a heightened state of awareness about improvement. These organizations are "skilled at creating, acquiring, and transferring knowledge, and at modifying its behavior to reflect new knowledge and insights."<sup>7</sup> They have a culture of learning that permeates the organization and provides structures that support the learning agenda. Learning has to have the commitment and space to grow. A culture of evidence means that almost everyone has drunk the Kool-Aid and buys into the importance of evidence. Evidence is part of the ecosystem.

A continuous improvement model, undergirded by a culture of evidence and learning, becomes more essential and pressing when organizations are larger, more complex, and have different activities. In that context, it is hard for one-off studies to cover everything. For example, large, policy-oriented organizations operate in a sea of complexity and often do not have direct levers of control over their outcomes; their actions are moderated by the vagaries of human behavior. In these circumstances, the need for a continuous improvement culture seems essential to keep policy on target. "One and done" studies are tough to consider as an evidence program on their own.

A culture of evidence means getting uncomfortable. Evidence should challenge your assumptions, your closely held beliefs, and all of your opinions. The deeper your beliefs, the more the scrutiny of the process

of evidence generation and the evidence itself has the potential to unsettle you. You have to go into it all with an open mind, or you will never learn anything. It is always going to be uncomfortable; you just need to learn to live with it. Relax, let it go, and get uncomfortable.

Leadership is a key ingredient in the culture equation. Leadership can set a tone for, reinforce, dismantle, or circumvent attempts at building evidence, guiding the ship to open sea or into the rocky shallows. If management does not embrace a learning culture, the learning agenda will persistently toss on turbulent waters. Strong, visionary leadership can prioritize evidence and ensure that learnings are used. Myopic leadership can stone-wall an evidence program, divert an evidence initiative to inconsequential points of inquiry, or banish the results to a basement repository. If key players in the organization are brainstorming reasons why evidence generation cannot happen instead of brainstorming ways to make it happen more easily, only strong leadership can get everyone on board. With time, good leadership can foster strong culture, and strong culture can keep the organization on the path even if bad leadership comes along.

## **TROLLS AND TRAPS ON THE PATH TO EVIDENCE BUILDING**

### *Methodological Fundamentalism*

In my research, I do a lot of experiments in which participants are randomized into different conditions. This allows me to estimate the effects of the different alternatives I study with a high degree of precision and little concern that alternative explanations might be driving the difference in results for the different groups. Randomized control trials can be highly informative in building evidence in a number of contexts. Regrettably, RCTs still are underutilized, and there often still is considerable sensitivity and resistance to using them in some contexts. In the early days, we spent lots of time explaining to various stakeholders the ethics of withholding treatment to a control group, but I do not personally find myself debating the ethical merits as much these days. I think people better understand that withholding a project from beneficiaries is bad only if it has demonstrably positive benefits, and that most programs do not have enough resources to reach everyone in the first place. As Jim Manzi highlights in his chapter in this volume, RCTs have an important role to play.

At the same time, there has been something of a cult of RCTs emerging in the evidence community: methodological fundamentalism. This religious

zeal for RCTs—this rejection of all non-RCT evidence—seems almost as problematic as the apprehension some have had about using RCTs. Good evidence can come in many forms, and methodologies to create it must be faithfully followed, attuned to the circumstances and the research questions. Process evaluations, qualitative research, observational studies, and other approaches all can have a role, *if executed using strong methodological standards*. Moreover, an absolute focus on RCT evaluations can undermine research and learning on topics that are hard to study with these methods, such as studies where treatment cannot be withheld for ethical reasons or studies of issues that cannot be randomized in a practical way (e.g., historical political events). An emphasis on RCT-only research undermines credible evidence that can be generated in a variety of contexts. Much evidence generation can be obtained through other methods leading up to an RCT. Qualitative research can provide important background and contextualize quantitative work, and if you have the data to do an observational study, often it will provide important results that may inform a future RCT. The fields of econometrics and statistics have developed robust tools for dealing with data for which random assignment was not available. At the same time, while causal inference methods often provide a powerful set of tools, the focus on causal identification can at times create a “research bias” analogous to the well-known publication bias (the biased perspective that researchers gain when academic editors are biased against publishing null results).<sup>8</sup> Such research bias deters researchers from taking up the mantle of important research questions for which strong causal identification methods are not available, limiting research in important areas where knowledge generation is desperately needed.

The bottom line is that evidence comes in many forms. We should strive to find the highest quality research design appropriate to the question, circumstances, and problem, and to apply methodologies rigorously, but not shy away from tackling questions that add value even if the research design does not conform to some religious view on what evidence is about.

### *The Dark Side’s Abuse of Evidence*

There is a dark side of evidence: it may be co-opted in a way that seeks to deceive rather than inform. For most of my career, I believed that co-opting evidence was not possible because bad evidence could be critiqued openly in public debate in order to debunk bad methodologies or faulty claims.

In recent years, I have had experiences that changed my mind. These were experiences for which the cake was already baked; evidence, rather than being pursued to discover and inform, was curated to justify decisions that had already been made. In these cases, the curation of evidence was intended to mislead or misdirect the public about the decisions being undertaken. If the public raised questions about projected outcomes, projections were generated using contrived assumptions to demonstrate the impact that would occur or that concerns raised were unfounded.

Much of this “evidence” was not evidence at all but, rather, just numerical tricks in the guise of evidence. In some cases, actual programs of evidence generation existed, but in those cases, spokespeople abused the programs.

Why did any of this happen? I am not sure I know all the reasons, but I will provide a few observations. In some cases, decision makers used processes to quell any questioning of the results that were presented. They ruled questions out of order, or ignored them, or asked the critics to discuss the issue in a sidebar conversation that never materialized. This failed to give critics a venue to raise legitimate objections. Where critics voiced concerns, the proponents of policies could easily marginalize or bury queries in a mountain of paper and talk, weaponizing process control.

Crucially, the lack of an independent watchdog to call foul created the conditions for such abuse to take place. This may be a byproduct of modern media, the busy lives of ordinary citizens, and a lack of the requisite head-space to fully understand the implications of thousands of pages of policies and proposals. A lack of local coverage—with all media attention grabbed by sensational national headlines—in this new information order, a deep dive into complex policies or local issues does not gather much attention. The lack of local issue coverage is of extreme concern. Many decisions that can make people’s lives better or worse happen at the local level, but the demise of local press over the past few decades means little is monitored or dug into deeply. More generally, the truth often is in the details, but it is hard to communicate deep truths in a 280-character limit Twitterverse.

More evidence sorcery lies in crafting questions to curate proof points for a desired position. This includes survey questions like: “Do you feel this project is: a) a great project; b) the best project ever; or c) all of the above!” Decision makers can control the evidence-generation process to forestall asking any questions that might be meaningful or challenge a course already decided upon.

This is not a full list; other tricks abound. The important thing to note is that all these efforts are evidence in name only. Attempts to deceive rather than inform fly in the face of any acceptable standard. The evidence community should pay attention, as this tarnishes us all, whether or not we personally sully our hands. These approaches use the banner of evidence to deceive or misinform, often waving this banner in front of folks who are not evidence gurus and may not be sufficiently trained to debunk details of sham findings. They may notice something smells wrong with the “evidence” but may not have the snout to ferret out what is rotten.

It is the evidence community that needs to find a way to police such shenanigans and help regular people understand that this is not what evidence is intended for. We may not want to be the beat cop, but it might be a role we cannot forgo. The evidence community may need to develop ethical standards that weed out those that distort evidence and evidence methods to the detriment of constituents.

In my view, and the view of the true professionals in the field, evidence is supposed to be for the people. Our tools are intended to find new ways of helping make lives better. That is true whether or not the folks we are trying to help have the background to understand regression analysis or causal chains. It has become more and more obvious that the evidence community has not been doing enough to democratize the evidence we ourselves generate, by bringing stakeholder voices into each step of the process and ensuring we listen to them as well as communicate our research findings. That is a tall order, for certain. Generating evidence is hard enough, and some of the masters are neither skilled at nor feel they have time to listen to participant input and circle back with findings. But they should.

We need to go even further. We need to hold ourselves and others accountable for the evidence (or “evidence”) they generate. We need to fight against the dark side to avoid the tyranny of fake “evidence.”

### *The No-Evidence Trap*

Entities that do not have robust evidence programs may have, in some circumstances, fallen into a “no-evidence trap,” where it is difficult to build evidence. Legislators, external watchdogs, funders, and others may decry the lack of evidence yet find that, no matter how hard they press the organization, there is little movement toward examining effectiveness.

The trap may be the by-product of incentives within and external to the organization. If generated, evidence could be used to defund rather

than reform; people within the organization might feel the need to decrease transparency just to protect the organization. As frustration grows on the part of external observers, their calls for evidence may become sharp and aggressive. Such threats further provoke concerns that the evidence will be used to curtail good work rather than to increase organizational effectiveness, perpetuating a cycle where the organization is less willing to develop solid evidence and ask meaningful questions.

One can interpret the intransigence of the rank-and-file staff of no-evidence organizations in a number of ways. On one hand, it can seem that when entrenched individuals within an organization have interests in maintaining the status quo they will do anything to resist change, even if it means blocking their pro-evidence colleagues (who also are trying to serve the organization's interests). In some entities in which I have observed this dynamic, personnel within the entity seemed to have an entrenched anti-evidence culture that I found hard to sympathize with. They seemed so concerned about protecting the organization from the immediate threat of potentially derogatory evidence that they became blind to the longer-term threat of being an organization that is completely ineffective or causing harm. On the other hand, these often are well-trained professionals who have dedicated their lives to a cause, so they may view preserving the institution and its mission as paramount, regardless of effectiveness: you can't win if you don't play. Any way one interprets a reluctance to generate evidence, it is important for key stakeholders (particularly funders) to understand that the path to successful reform involves both carrots and sticks.

Traps of these kinds have arisen in all sorts of organizations. The trap persists in some government entities whose very existence is contested. This also can be the case for social sector nonprofits that are caught pivoting between participant and donor demands, or feeling the squeeze of unrealistic budgets. I also have seen versions of the no-evidence trap in settings where funders have a track record of changing course and failing to update accountability measures in synch with the change in mission.

Reducing the if-then mentality ("if results are not demonstrated, then resources will be cut") will be key to building better evidence in cautious sectors. It certainly is difficult to engage in honest research when so much is on the line. Evidence needs to be built within a partnership between implementors and their funders. It is an exercise in innovation but, at its best, also a joint exercise with stakeholders in discovery.

## **BUILDING THE NEXT GENERATION OF EVIDENCE**

The past decade or two have reinforced the link between evidence and political democracy. Over that period, citizen belief in the existence of objective truth has wavered, and the nation's ability to agree on basic facts has thrown us all into a tailspin. It seems so much of this effect is based on a lack of appreciation for evidence. If we could just tie down facts and evidence, there might be more consensus about the problems we face and the solutions that are feasible. Of course, it is more complicated than that.

The politics of evidence at the national level—in our current political climate—are problematic to say the least. On one side, you have folks who are unwilling to budge an inch on hard-won programs. On the other side, you have folks who want evidence simply so they can dismantle programs and who are equally willing to filter out any inconvenient truths. While there are definitely some heroes who have tried to chart a middle course, there are too few honest partners. An evidence-based approach to the issues recognizes not only that a given problem exists within society (e.g., people are hungry or children are not doing well in school) but also that existing methods can be improved, restructured, reformed, or reorganized to better meet society's goal (e.g., fewer hungry people). The measure of success of a policy or program should not be whether or not a difficult-to-achieve goal (such as complete eradication of world hunger) has been attained without considering the many millions or billions of people who may have been helped out by imperfect interventions. Unfortunately, a balanced view like this increasingly comes across as contradictory in our modern politics: one part heresy for each tribe.

Society's problems do not just disappear when one ignores or manipulates evidence: people are still hungry or unemployed, or students still are falling behind. The only thing that happens when citizens' needs are not met is that people lose faith in society. Ultimately, government and institutions are a reflection of the preferences of people, so a loss of faith in institutions runs the risk of becoming a loss of faith in the entire economic and political system. After all, institutions create the setting in which an economy can prosper and meet the needs of its people, and where individuals have protected rights.

Evidence may not solve the crisis in democracy on its own, but it certainly can provide some fundamental truths on which to latch. The more

decision making can be attuned to actual facts rather than political ideologies, the more likely we are to find common ground. Maybe that is too idealistic. Maybe our ideologies trample all reason and love of country. But I do believe that, if we are able to find the right track, evidence must guide us.

Society faces enormous challenges, from dealing with climate change and its effects to dealing with life-altering consequences of new technologies to health issues; racial and wealth inequities; changes in the structure of work; evolving demands on social programs; persistent challenges of how to best educate the next generation; and so on. As life continually becomes more and more complex, and financial and other resources become more and more pressured, the inefficiencies of ineffective policies and programs become ever more difficult to overlook. Evidence, if thoughtful, strategic, and well executed, can illuminate the path so we can focus on getting from Point A to Point B rather than tripping over ourselves in the dark. Evidence itself seems a necessary though not sufficient condition for democracy to prosper.

The best organizations (and institutions and democracies and economic systems), the ones that are going to be leaders in developing the next generation of evidence, are going to ask real questions and build interesting and stimulating environments for building evidence. Environments that attract people with key skills, environments that those folks relish—not simply because of the paychecks they pocket but because their work—both stimulates their synapses and is valued. These organizations will provide a place where evidence-generation gurus get a voice in how projects and policies are conceived and implemented rather than being told to stay in their lane. These folks will be a key part of the leadership and management of the organization, and their impact on the organization will grow over time as learning leads to deeper and deeper understanding of the ways things work.

Organizations that are next-generation leaders are going to be the ones that embrace continuous evidence approaches; that take a strategic approach to evidence generation; and find ways to keep those in their field honest. These leaders will include evidence experts throughout all stages of *both* project development and process improvement. They will understand the importance of equity in their activities and give those experts a real voice. These organizations will not only offer their ear to advice from evidence producers but also give their own internal experts a real career path within the organization, one that allows them to rise to the

highest ranks of leadership. No one wants to hit a glass ceiling, which signals they are not valued. At the same time, evidence experts will not truly be impactful within an organization if they are a bunch of folks with hammers and screwdrivers looking for things that look like nails and screws. Rather, these internal experts need to be attuned to the organization's goals and mission. They will need to serve as effective translators between social science, methods, and practitioners, and they will need to be adaptable.

Scale is important in all of this, and the idealized state alluded to in this chapter is not for every entity. Small organizations can use evidence to get better at their work, but it is not reasonable for them to become evidence-first institutions. Yet, as we have discussed, for larger organizations, particularly those that affect large groups of people or enact policies that have wide-ranging effects on people, the economy, or other aspects of society, marginalizing evidence generation can create distortions that hurt people and society and can undermine trust.

For those organizations, it is time for reform, even if it is painful. It is time for action. It is time to shake up old norms and bad habits and take evidence seriously. The challenges and crises we face are too extreme to ignore. The resources are too scarce, making the inefficiencies too glaring to gloss over. It is time to turn on the lights and stop stumbling in the dark. Society, and lives, may depend on it.

## NOTES

1. Even if the answer is yes (it does work), we have probably missed an opportunity to identify something better.
2. Three big-picture methodological misalignments are: omnibus program evaluations, overstating of findings, and failure to take account of heterogeneity. These are three examples of where one could still go wrong even when generating evidence using a credible theory of change. "Omnibus" program evaluations lump all activities under "the program," leading evidence generation to fail to distinguish the effects of different program components (for example, your grants program may be working, but it could be undermined by a counterproductive training program); approaches that do not allow one to assess what parts are working and what parts are not also can lead to a dead end. Researchers and practitioners commonly overstate the generalizability of findings when, in fact, the activities work only in a really specific set of conditions. This can lead to erroneous conclusions that a certain activity

works everywhere, leading to misdirected programmatic effort and inconsistent results. And a failure to account for heterogeneity among beneficiary groups is one of many factors that also can lead to inconsistent results (see, for example: Christopher J. Bryan, Elizabeth Tipton, and David S. Yeager, “Behavioural Science is Unlikely to Change the World without a Heterogeneity Revolution,” *Nature Human Behaviour* 5, no. 8 (2021): 980–989).

3. Standing data-collection capacity often provides the opportunity to evaluate not the total cost of research initiatives (which typically have high fixed start-up costs) but, rather, the marginal cost of a new project. With standing capacity, negotiations with contractors for data collection can be on more favorable cost terms. In my own efforts, standing capacity has reduced timeframes considerably (from timescales in years to timescales in months or even in weeks), and the reduction in time and costs can be enough to take time and cost issues off the table. Internal expertise also can be helpful with cost cutting, by allowing internal experts to in-source key parts of the production process, and also can provide other benefits.

4. For example, tax and investment incentives that have been used to encourage, preserve, or create affordable housing might lead to less housing affordability if higher-end units are put into the development at a higher rate than affordable units (thus diluting the prevalence of affordable units in the community, reducing income diversity, and raising average housing expenses in the community overall). In other contexts, unemployed workers may desperately seek to augment their skill set with free job training, but if workers are acquiring antiquated skills in dying industries, they may invest their time poorly and find little or no benefit in the job market. Initiatives that seek to promote environmental preservation may backfire if they increase negative attitudes toward the cause of interest because they are punitive for households that do not have the option to adjust their lifestyle, say, due to a disability.

5. See chapter in this volume. [AU: Why is this highlighted? Which chapter?]

6. House of Commons, June 23, 1925, [https://api.parliament.uk/historic-hansard/commons/1925/jun/23/finance-bill-1#S5CV0185P0\\_19250623\\_HOC\\_339](https://api.parliament.uk/historic-hansard/commons/1925/jun/23/finance-bill-1#S5CV0185P0_19250623_HOC_339).

7. David A. Garvin, “Building a Learning Organization,” *Harvard Business Review*, July-August 1993, <https://hbr.org/1993/07/building-a-learning-organization>.

8. Chalmers Lain, “Underreporting Research Is Scientific Misconduct,” *JAMA* 263, no. 10 (1990): 1405–1408; Phillipa J. Easterbrook, Ramana Gopalan, J. A. Berlin, and David R. Matthews, “Publication Bias in Clinical Research,” *The Lancet* 337, no. 8746 (1991): 867–872; Annie Franco, Neil Malhotra, and Gabor Simonovits, “Publication Bias in the Social Sciences: Unlocking the File Drawer,” *Science* 345, no. 6203 (2014): 1502–1505.