

Scaling up good ideas is really, really hard — and we're starting to figure out why

Promising programs often fall apart when we try to offer them to more people. Here's why.

By Kelsey Piper | Dec 10, 2018, 2:40pm EST



Refugees receive social relief in Bangladesh | LightRocket via Getty Images

**FUTURE
PERFECT**

Finding the best ways to do good. Made possible by The Rockefeller Foundation.

A couple weeks ago, I wrote about **Evidence Action's No Lean Season**, a program that in its early stages worked incredibly well to help members of the poorest families in Bangladesh move to the city and find jobs — but that stopped working when Evidence Action moved up to larger-scale trials.

The result was disappointing, but it isn't actually rare. Programs often show exceptional results in a small trial, but disappoint at scale. Why does something like that happen? How can the gains from a program be so substantial when it's first attempted and disappear entirely as soon as it's expanded to cover more people?

I asked these questions of Mushfiq Mobarak, an economics professor who led the research into No Lean Season. He works at Yale's **Research Initiative on Innovation and Scale**, which studies the science of scaling and the research methods needed to successfully study program effects. He thinks

the development community isn't worried enough about whether the results from its programs will scale. "We're applying really high rigorous standards to the initial program implementation," he said, "but not applying those same standards to the program at scale."

It turns out that getting sustained results beyond smaller pilot programs is well known to be exceptionally difficult. It's difficult because there are a bunch of contributing factors — correcting just one doesn't necessarily improve anything, and we need new approaches to research to correctly account for all of them.

Here's how programs fail to scale:

1) Aid changes how governments behave

Small aid experiments don't typically affect government policy, but big aid programs do — and that can have consequences. "When we run these programs for an evaluation at pilot scale," Mobarak told me, "the government often doesn't notice or care. But of course if it becomes policy, then the government cares."

A small experiment in one village doesn't merit much notice; an aid program that costs millions of dollars or more is definitely going to feature in a government's planning and policy. In some cases, the government might shift its own aid elsewhere, expecting that the aid program will be all its recipients need. (Imagine that an NGO opened hundreds of food banks in Kansas, and in response, the state government decided to make it harder to get food stamps because it perceived less of a need.)

Decisions like those can be a good thing or a bad thing for the local population, depending on how well the NGO substitutes for the service the government was providing and depending on what the government does with those resources elsewhere. But either way it's unintended — the NGO did not mean to shuffle around access to government services with its own program. In effect, the NGO has "crowded out" the government.

Interventions can also produce the exact opposite effect, "crowding in." That's when governments respond to a program in one region or industry by increasing their investment. For example, a nonprofit might start a skills training program. The government might then think, "We're going to bring in a factory to take advantage of the skilled workers," Mobarak said.

That might seem like good news, but it can cause researchers to vastly overestimate the impacts of their program and to incorrectly expect skills-training elsewhere to have the same effects.

2) Governments make demands of aid programs

When a program is a success, local politicians notice. Often, Mobarak said, they want to claim the credit.

“Governments might react to associate themselves with the program and try to take credit for it,” he said. “When we ran a very large-scale sanitation program in Bangladesh, we were told locally elected politicians want to come and give a speech at your voucher distribution ceremony.” Nonprofits usually can’t refuse these demands, lest their programs get shut down.

But Mobarak has some reservations about the big-picture consequences of letting politicians claim the credit for aid programs. In democratic countries, this may mean lying to voters, who then can’t pick the best candidates. If you give in to too many demands like these, “you’re making it difficult for voters to vote out bad politicians,” he said. “We’re messing up their information environment.” (Making local elections less corrupt and more competitive **can save lives**, so this is a big deal.)

Politicians can intervene in more directly corrupt ways, too: granting permission to scale up the program only in regions that benefit their constituencies, or only if they get to control some of the resources the program distributes.

All of this means that aid programs can have negative political consequences at scale that they don’t have in a small experiment, and may not help people as much as predicted.

3) You run into the laws of supply and demand

When you give one person a goat, the effects on the overall, community-wide (let alone nation-wide) supply of goats will be very small. But “if you give everybody a goat, the prices of goats start to fall,” Mobarak said. So will the price of goat milk, goat manure, and similar products. If your intervention assumed that people could sell goat milk to support their families, it might not work once goats are everywhere.

Economists call these “general equilibrium effects” — when your program has to be evaluated not just in terms of the impact on recipients, but in terms of the impact on markets.

For example, a program that trains skilled workers can drive down the wage for the type of work they do by creating a larger supply of skilled workers. Or **giving some people money might increase food prices, hurting people who don’t receive the money.**

There’s an argument that general equilibrium effects shouldn’t be ignored even in small trials. They are still there on the margin, after all. But in practice, it’d be impossible to detect them, and small-scale studies don’t consider them. At scale, though, they can’t be ignored. **Some research has found** that the spillover and general equilibrium effects of some changes are far bigger than the direct effects. Failing to analyze the spillover means you’ll be badly mistaken about the effects of the program.

4) Scaling up means new people and processes

With a small study, there is intensive support, monitoring, and engagement to keep people on track. Not everything will be run exactly according to the research plan, but the process is on the whole more manageable. Moreover, program officials are often on the ground, seeing out their vision themselves.

When a program scales, it has to hire and train many new people — or in some cases transition to using government resources and civil servants. And that can change a program's effectiveness. "You turn over a program from a highly motivated NGO," Mobarak said, to people who know less about it and are less driven to see it succeed — or informed about what it will take. A lot can be lost in transmission.

Scaling up right is essential to doing what all charity should be doing — actually helping people

Every study in development, Mobarak said, is meant to answer one question: "If this program were to be scaled up as policy, would it affect people's lives?"

In practice, researchers often end up measuring other questions, like, "Was this cost-effective in our small-scale trial?" They often have no choice, because the big questions can't always be studied directly.

But we can't afford to take our eyes off the ball. "Even if it's a real good program trial," Mobarak said, "the answer you get from the trial is just different from the answer to the question you really care about." A program needs to do a lot more research, including meticulously monitoring the program as it scales it up, and being willing to admit bad news like No Lean Season did.

The evidence-based development world has often been too optimistic about how well the evidence they see from program trials translates to policy. But "if you're going to be evidence-based at the first step, you should also care about evidence at the second step," Mobarak said.

That bar may be intimidatingly high. There are very few programs that have had the resources to study all their effects systematically at scale. But ultimately, if we want to do good, it's going to take attention to all of these questions — and the **honesty to face the answers**.

*Do you ever struggle to figure out where to donate that will make the biggest impact? Or which kind of charities to support? Over 5 days, in 5 emails, we'll walk you through research and frameworks that will **help you decide how much and where to give**, and other ways to do good. **Sign up for Future Perfect's new pop-up newsletter.***