

Comment



ABIR ABDULLAH/EPA/SHUTTERSTOCK

An aid programme helped rural labourers in Bangladesh to get work as rickshaw drivers by migrating to dense cities such as Dhaka.

Assessing social aid: the scale-up process needs evidence, too

Ahmed Mushfiq Mobarak

When programmes expand, new complexities and indirect consequences must be studied.

Landless agricultural workers and their families often go hungry between planting and harvest, the 'lean season' when the labour demand falls. In northern Bangladesh, my colleagues and I tested a way to ease this hunger. Instead of trying to force job creation in rural areas, we helped labourers to move temporarily to nearby cities, where construction and other jobs existed.

Our pilot study, which included 1,900 households, was evaluated through a randomized controlled trial (RCT) in 2008, and it seemed to be successful. Small subsidies of US\$11.50 – enough to pay for the round-trip bus fare plus a few days of food – boosted the percentage of

agricultural workers heading to cities during the lean season from 36% to 58%. The families of the migrants consumed more than 600 extra calories per person each day – essentially, they were eating three meals instead of two. Moreover, about half of those who moved chose to migrate again without subsidy during subsequent lean seasons, and many found work with the same employer that they had connected with in 2008.

We scaled up the programme in stages, each time expanding the observations we made: these included risk of divorce, changes in prices of goods and the costs of family separation. These data helped us to capture the

unintended consequences of more migrants leaving their villages and entering urban labour markets. Results continued to look promising, and a large microcredit organization in Bangladesh received philanthropic support to offer seasonal-migration loans to hundreds of thousands of households. But the outcome was disappointing – subsidies mainly reached those who would have migrated anyway, and the programme was promptly discontinued. Although this was disheartening, I remain proud of collecting that decision-aiding information: it prevented waste and meant that the limited money for anti-poverty programmes was better spent.

When programmes enter a ‘scaling stage’, the focus often immediately shifts to solving the practical issues of broader implementation of the programme (such as how to teach government staff about an innovation, distribute subsidies to tens of thousands of people, instead of hundreds, or integrate a programme across government systems). All that work, although essential, overlooks the crucial question of whether exciting pilot results still hold. Many – if not most – development programmes encounter uncertainties and complexities that emerge only at scale. These are rarely observed – and therefore cannot be analysed – during the initial pilots. Simply repeating interventions on the same scale at multiple locales is not enough.

I have spent more than a decade trying to systematically understand how scaling complexities arise, and the methodological tools and data that we need to analyse them. In 2017, I co-founded the Yale Research Initiative on Innovation and Scale in New Haven, Connecticut, with the aim of formalizing informative, systematic evaluations.

‘Evidence-based’ philanthropy and policy-making have become important buzzwords in economic-development and global-health circles. The effective altruism movement – famously described by newspaper *The Economist* as “trying to bring scientific rigour to philanthropy” – directs hundreds of millions of charity dollars each year. Those interested in evidence-based policymaking and philanthropy should recognize that rigorous standards of evidence are needed in the process of scaling up pilot programmes. These standards are the only way to inspire confidence in the results that excited supporters in the first place.

Lessons learnt

Here is what the migration programme taught us about how the set of research questions should expand as an intervention is scaled.

First, consider effects beyond those reaching direct beneficiaries. At scale, programmes often lead to feedback loops that can create market- or city-level changes. In our example, encouraging rural Bangladeshis to migrate affects others competing in the same labour markets as the migrants (both at the origin and destination). Wages could rise in the village and fall in the city. Increased outmigration could affect the informal insurance networks that operate among agricultural labourers, a system for sharing risk in which a family will lend money to or share food with another, knowing that the favour might one day be returned. Perhaps these networks would be strengthened by the migrants’ increased income, or weakened by their prolonged absences.

Second, pay attention to broader social changes beyond the outcome that the original programme targeted. In our example, the migrants’ spouses and children ate more reliably but might have faced new risks of divorce,

“If we care about whether a programme improves lives, we must analyse issues that arise at scale.”

domestic violence or communicable diseases brought back from the city.

Third, anticipate political and operational risks as new players get involved with a programme. For example, landowners, who are forced to pay higher wages when many workers migrate away, are politically powerful and could organize to undermine the programme. Sustaining the programme might require appeasing their concerns, perhaps by offering them labour-saving technologies, such as herbicide sprayers. But those, in turn, could harm the environment.

Simply delegating others to extend the programme also carries risk: those charged with recruiting more participants might focus on the most reachable people, not those most likely to benefit. A common (and not unreasonable) success metric for microcredit organizations is maximizing the number of loans given, but successful interventions often need to target harder-to-reach groups. Addressing this risk requires a thorough understanding of incentives in partner organizations, including the career incentives of their field staff, and how those might differ from the performance incentives in the original researcher-driven pilot trial. Each partnering organization should be considered as a new variable that

can change outcomes.

Fourth, scale up in reasonable increments. We chose to expand the migration programme in stages and over several years. At each stage, we developed new experimental designs to better understand the broader range of benefits and costs, and the new complexities and risks. With 1,900 households, we focused on indirect or unintended consequences at the household level, such as the risk of divorce and changes in health outcomes. We started studying spillover effects on the villages that the migrants left behind (such as changes in wages, food prices or labour-market outcomes among non-migrants) only after we expanded our sample to 35,000 households. When the programme was expanded to 150,000 (a point at which effects might have been discernible in cities), we looked at urban spillover effects. At each stage, our goal was to rigorously examine the complexities likely to arise at that scale.

Fifth, expand methodologies to track the full range of welfare effects. My team knew that mass migration would bring income into the village, and that could raise the prices of food and other goods. Tracking prices and wages required an experimental design in which we varied the fraction of villagers who received transport subsidies. Other welfare effects are less quantitative: the cost of family separation or the discomfort of living in an urban slum cannot be valued directly through standard economic measurement in an RCT. We had to borrow methods from adjacent fields, such as macroeconomics. Because the harm of family separation is not directly observable, we mathematically modelled the concept of disutility (an economic term capturing harmful or adverse effects), and inferred its magnitude by calibrating the model to experimental data.

Growing pains

Promoting migration to mitigate seasonal hunger continued to look sensible, even after all these layers of research tracking the broader costs and benefits. However, the change in organizational incentives as the programme moved to large-scale implementation through a microlender proved to be the Achilles heel. The intervention worked when it induced new people to migrate – but not when it was subsidizing migration that would happen anyway.

The complexities we investigated are not unique to this intervention. For example, the benefits of job-training programmes might not scale if many skilled workers compete for a limited number of vacancies. Large programmes that improve agricultural productivity might reduce crop prices. And increased



MAJORITY WORLD/SHUTTERSTOCK

The migration programme collapsed after expanding: subsidies mostly reached those who would have migrated anyway.

productivity could lead farmers to expand their crop fields, resulting in land conversion and deforestation. Such unintended consequences abound. However, effects at scale can look even more promising than in pilots owing to macroeconomic multipliers, in which one change affects many factors: the money migrants send home to their villages means that more students can attend better-funded schools and that shopkeepers have more customers, benefits that accrue beyond participating households. And scale up can sometimes be achieved more easily than expected: scientists might face less resistance when expanding vaccination programmes to remote populations, because hesitancy and misinformation often make it difficult to persuade city-dwelling social-media users to get vaccinated.

Often, researchers are tempted to declare victory after a successful pilot and to jump to an ‘implementation phase’, assuming that the pilot results will hold. But if we truly care about whether an expanded programme improves lives, we must continue to ask questions and analyse issues that arise only at scale.

When I advocate evaluation of scaling, people often respond that this wastes time: any lifesaving programmes should be expanded as rapidly as possible. But performing such assessments does not necessitate delays; in fact, broader implementation is often a prerequisite for generating evidence at scale. As a bonus, the implementing organization can use this experimentation to iterate and improve

the programme being tested. Still, there is no obvious stakeholder – global organization, government or philanthropic group – to demand that research be done for informed scaling.

Perhaps the largest barriers to evidence-based scaling are institutional. Once a group has coalesced and an organization has formed around a specific cause, momentum and baked-in assumptions make it difficult to change course. That’s especially true if the process for gathering evidence during scale up is not built in. Yes, there is a risk that collecting data and expanding questions might overturn gratifying pilot results. And discussing what remains to be assessed could undermine support for a programme because philanthropists and policymakers are more interested in ‘sure bets’ than in evaluating uncertainties and risks, and the complexities of doing that well.

This uncertainty aversion can create a bias towards supporting simpler public-health programmes, such as deworming campaigns to tackle childhood parasites, for which the potential for unintended consequences is smaller. But settling for simplicity might mean missing out on more ambitious innovations with greater potential.

Doing scaling research is tough. It requires multi-year initiatives and multi-site trials, plus a gathering in of diverse researchers using a variety of analytical tools to study problems, interventions and outcomes from multiple angles and to synthesize evidence to judge whether and how an intervention should grow.

Despite these headwinds, scaling is now an emerging area of scientific inquiry. In public health, it goes by the moniker ‘implementation science’; medicine distinguishes between ‘efficacy trials’ in the laboratory and ‘effectiveness trials’ in the field. In the social sciences, the Abdul Latif Jameel Poverty Action Lab at the Massachusetts Institute of Technology in Cambridge has launched an ‘evidence to policy’ team to facilitate scale up of effective pilots, and the University of Chicago in Illinois has published books that serve as toolkits for successful scaling.

Efforts to support the scaling of promising innovations are essential, but insufficient: it is time to institutionalize the practice of investigating complexities and testing confidence in pilot-scale evidence. That can be a constructive exercise, if analytical rigour improves the likelihood that programmes succeed at scale.

Yes, there will be disappointments and disillusionment. Overall, however, this will elevate effective ideas and improve millions, or even billions, of lives.

The author

Ahmed Mushfiq Mobarak is a professor of economics at Yale University in New Haven, Connecticut. Twitter: @mushfiq_econ
e-mail: ahmed.mobarak@yale.edu

The author declares no competing interests.