

A new generation of economists is trying to transform global development policy through the power of randomized controlled trials.

REVOLT OF THE RANDOMISTAS

BY JEFF TOLLEFSON

In 70 local health clinics run by the Indian state of Haryana, the parents of a child who starts the standard series of vaccinations can walk away with a free kilogram of sugar. And if the parents make sure that the child finishes the injections, they also get to take home a free litre of cooking oil.

These simple gifts are part of massive trial testing whether rewards can boost the stubbornly low immunization rates for poor children in the region. Following the model of the randomized controlled trials (RCTs) that are commonly used to test the effectiveness of drugs, scientists randomly assigned clinics in the seven districts with the lowest immunization rates to either give the gifts or not. Initial results are expected next year. But

smaller-scale experiments suggest that the incentives have a good chance of working. In a pilot study conducted in India and published in 2010, the establishment of monthly medical camps saw vaccination rates triple, and adding on incentives that offered families a kilogram of lentils and a set of plates increased completion rates by more than sixfold¹.

"We have learned something about why immunization rates are low," says Esther Duflo, an economist at the Massachusetts Institute of Technology (MIT) in Cambridge, who was involved in the 2010 experiment and is working with Haryana on its latest venture. The problem is not necessarily that people are opposed to immunization, she says. It is that certain obstacles, such as lack of time or money, are making it difficult for them to attend the clinics. "And you can balance that difficulty with a little incentive," she says.

ESTHER HAVENS/SABIN VACCINE INSTITUTE



This is one of a flood of insights from researchers who are revolutionizing the field of economics with experiments designed to rigorously test how well social programmes work. Their targets range from education programmes to the prevention of traffic accidents. Their preferred method is the randomized trial. And so they have come to be known as the ‘randomistas’.

The randomistas have been particularly welcomed in the global development arena. Despite some US\$16 trillion in aid having flowed to the developing world since the Second World War, there are little empirical data on whether that money improves the recipients’ lives (see page 144). The randomistas see their experiments as a way to generate such data and to give governments tools to promote development, relieve poverty and focus money on things that work.

Trials are showing that offering incentives can boost attendance at vaccination clinics.

Not everyone is convinced. Sceptics argue that the randomistas’ focus on evaluating specific aid programmes can lead them to lose sight of things such as energy, infrastructure, trade and corruption — macroeconomic issues that are central to a country’s ability to prosper, but that are effectively impossible to randomize. “Development is ultimately about politics,” says Angus Deaton, an economist at Princeton University in New Jersey.

Nonetheless, the randomista movement is gaining momentum (see ‘Scale the heights’). Universities are pumping out more economics graduate students with experience in RCTs every year. Organizations ranging from the UK Department for International Development to the Bill & Melinda Gates Foundation in Seattle, Washington, are throwing their financial support behind the technique. “There are hundreds and hundreds of randomized trials going on, and ten years ago that just wasn’t the case,” says economist Dean Karlan at Yale University in New Haven, Connecticut, who is at the forefront of the movement. “We’ve changed the conversation.”

Demand is only rising. This September, governments will gather in New York under the auspices of the United Nations to approve a new set of Sustainable Development Goals, which are intended to guide investments over the coming decade. And in December, questions about financial aid will be high on the agenda at the UN climate summit in Paris, where governments expect to sign a new climate agreement that will probably include commitments by industrialized nations to funnel money into sustainable development in poorer countries. In both cases, the effectiveness of the programmes is likely to be a key concern.

“This is front and centre on a lot of people’s agenda,” says Ann Mei Chang, who is executive director of the Global Development Lab at the US Agency for International Development (USAID) in Washington DC. “Where do we get the biggest bang for our buck?”

PROGRESS AND OPPORTUNITIES

RCTs have been used to test the effectiveness of social programmes at least since the 1960s. But the modern era began in 1997, when one of the most famous and influential RCTs in public policy began in Mexico.

The experiment had its origins three years earlier, when Mexican President Ernesto Zedillo assumed office in the middle of an economic crisis and assigned economist Santiago Levy to devise a programme to help poor people. Sceptical of the conventional approach — subsidies for products such as tortillas and energy — Levy designed a system that would provide cash payments to poor families if they met certain requirements, such as visiting health clinics and keeping their children in school. “And because people were very critical about what I was doing,” says Levy, who now leads strategic development planning at the Inter-American Development Bank in Washington DC, “I wanted to ensure that we had numbers so that we could have an informed debate.”

As it happened, Levy had a natural control group for his experiment. The government was rolling out its payment programme in stages, so he could collect data on families in villages that were included in the initial roll-out, and in comparable villages that were not. Within a few years, his team had data suggesting that the programme, dubbed *PROGRESA*, was working remarkably well. Visitation to health clinics was 60% higher in participating communities than in the control group. Children in those communities also had a 23% reduction in illness and an 18% reduction in anaemia. Overnight hospital visits halved across several age ranges.

These data helped to solidify support for the programme. Now known as *Prospera*, it covers almost all of Mexico’s poorest citizens and has inspired similar initiatives across Latin America and into Africa.

“*PROGRESA* was one of the first major national programmes of its kind to get a rigorous evaluation,” says William Savedoff, who works on aid effectiveness and health policy at the Center for Global Development, a think tank in Washington DC. “Today conditional cash-transfer programmes are some of the most heavily evaluated programmes in the world, and that is I think a direct consequence of the Mexican experience.”

The idea of developing hard evidence to test public policies was

bubbling up in parallel in the United States. One of the first trials began in 1994 with a small initiative to analyse the effect of supplying textbooks and uniforms as well as basic classroom improvements to a group of schools in Kenya. Economist Michael Kremer at Harvard University in Cambridge had taught in Kenya years earlier. A friend of his who worked for a non-profit group was initiating the programme, and Kremer suggested that the group roll it out as an experiment. “I didn’t necessarily expect anything to come of this,” he says.

Working with the group, Kremer collected data on students in 14 schools, half of which received the intervention. School attendance increased, but test scores did not. Similar results came from an experiment in 1995 that involved 100 schools. That trial suggested that providing textbooks had little effect on average test scores⁷, owing perhaps to language challenges — the textbooks were in English, which was not the native language for many students. Students who were already scoring higher than their peers, however, pulled further ahead if they had the books.

Kremer continued to run RCTs of other programmes, but it was Duflo — then a student of his — who pushed the idea into the mainstream. Duflo’s 1999 dissertation looked in part at an education initiative in Indonesia that had built 61,000 primary schools over 6 years in the 1970s. She wanted to test a common concern that such a rapid expansion would lead to a decline in the quality of education, thereby offsetting any gains. Running an experiment was impossible, but Duflo was able to use data on the differences across regions to show that the programme had, in fact, increased educational opportunities as well as wages.

This and other early work inspired Duflo to look at RCTs as a way to generate data and definitively measure the effectiveness of policies and programmes. “As soon as I had a longer time horizon and some money I started working on setting some up,” she says.

One of Duflo’s early papers³, published in 2004, capitalized on a 1993 amendment to India’s constitution that devolved more power over public investments to local councils and reserved the leadership of one-third of those councils, to be chosen at random, for women. Duflo realized that this effectively created a RCT that could test the effect of having women-led councils. In analysing the data, she found that councils led by women boosted political engagement by other women and directed investment towards issues raised by them. In some areas, women are in charge of obtaining drinking water, for instance, and councils led by women typically invested more in water infrastructure than did those run by men. “The scale of the policy and the topic were at the time unusual,” Duflo says. “It gave me a sense of the range of things that the tool could possibly cover.”

By the early 2000s, the randomistas were on the upswing. In 2002, Karlan, one of Duflo’s students, joined with her and other researchers to form Development Innovations — now known as Innovations for Poverty Action — in New Haven. The following year, Duflo co-founded what is now known as the Abdul Latif Jameel Poverty Action Lab (J-PAL) in Cambridge with fellow MIT economists Abhijit Banerjee and Sendhil Mullainathan.

The work quickly expanded, and J-PAL has now run nearly 600 evaluations in 62 countries, and trained more than 6,600 people. One of Duflo’s latest projects will revisit her dissertation on education in Indonesia, only this time with secondary schools and randomized control groups. “We will have a randomized version of a paper on the benefits to education soon I hope,” Duflo says.

VENTURE CAPITAL

One enthusiastic convert to the randomista philosophy is Rajiv Shah, a Gates Foundation official who became head of USAID in 2010. Once there he created a fund called Development Innovation Ventures (DIV) to test and scale up solutions to development problems, and he enlisted Kremer as its scientific director. The goal, Shah said, was to

“move development into a new realm” through the use of evidence.

Since then DIV has invested in more than 100 development projects, and nearly half involve RCTs. One, conducted in Kenya by a pair of researchers from Georgetown University in Washington DC, tested a simple method for reducing traffic accidents that involve minibuses — collisions that Kremer calls major and increasing killers. “Two of them crash into each other, and 40 people die,” he says.

In 2008, the researchers worked with more than 1,000 drivers to place stickers on buses that urged passengers to speak up about reckless driving⁴. They then collected information from four major insurance companies and found that claims for serious accidents had dropped by 50% on buses with stickers compared with those without. DIV provided a grant to conduct a larger trial — which found that claims dropped by 25–33% — and a second grant of nearly \$3 million to help to scale up the project throughout Kenya.

“The really big win is when developing countries, or firms or NGOs [non-governmental organizations] change their policies,” Kremer says. But one question now facing DIV is whether such a strategy — or indeed any project that proves effective in one setting — can be repackaged and deployed in other countries, where different cultural factors are at play (see *Nature* 523, 516–518; 2015).

SCALE UP

Effecting policy change is the precise aim of the Global Innovation Fund, which was launched in September 2014 with \$200 million over 5 years from the UK Department for International Development, USAID and others, and which follows the DIV model of rigorous testing. Interim director Jeffrey Brown, who is on loan from USAID, says that the fund has already received more than 1,800 applications for projects in 110 different countries and will be announcing its first suite of grants later this year. “We are essentially trying to become a bridge over the valley of death for good development ideas,” he says.

But such organizations still provide only a tiny fraction of the billions of dollars that are spent each year on development aid, let alone the trillions of dollars that are spent by governments on domestic social programmes. Even at lending institutions that have taken this evidence-based framework on board, the portion of investments that is covered by rigorous evaluations is small.

At the World Bank, which started a Development Impact Evaluation division in 2005, the number of projects receiving formal impact evaluations — through RCTs or other means — rose from fewer than 20 in 2003 to 193 in 2014, mostly covering things such as agriculture, health

and education. But that still represents just 15% of the bank’s projects, says evaluation-division head Arianna Legovini, who leads a team of 23 full-time staff and has an annual budget of roughly \$18 million. Although many of these evaluations more than pay for themselves over the long term, one constraint is the up-front cost: the average price of an impact evaluation is around \$500,000. “If I did not have donor funding,” she says, “these studies just would not happen.”

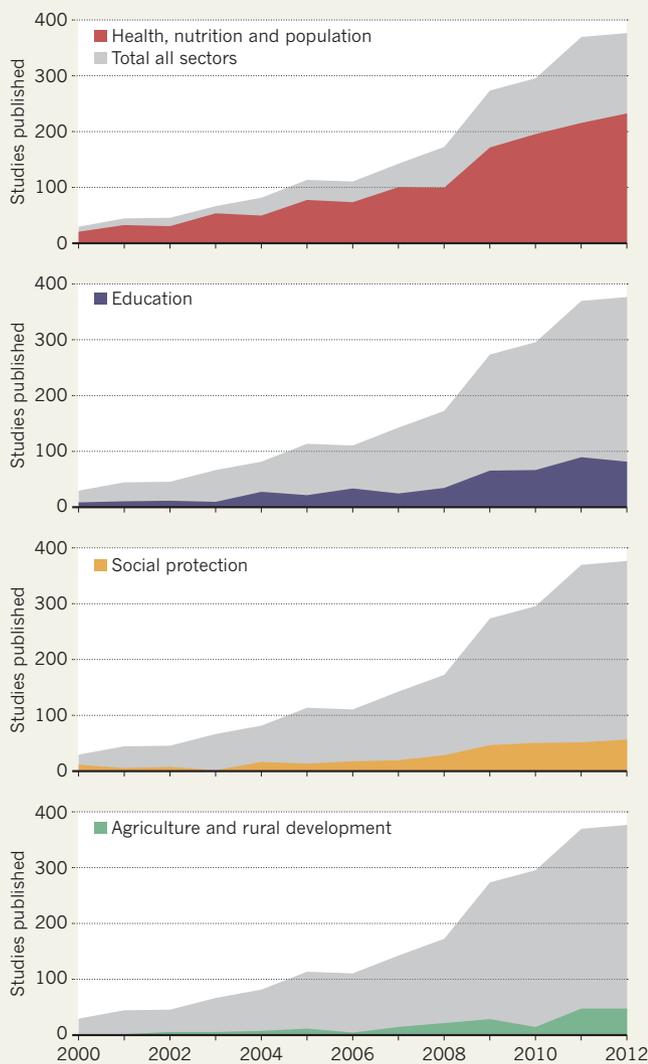
The World Bank is trying to make the most of its resources by working directly with developing countries on implementation. More than 3,000 people have attended its workshops and training sessions since 2005, most of whom were government officials in developing countries that are receiving funds from the bank.

The bank is also making efforts to assess the impact-evaluation programme itself — although the analysis is based largely on whether payments for projects are made on time as a proxy for implementation of the initiatives. An analysis by Legovini and two of her team suggests that development projects that undergo a formal impact analysis are more likely to be implemented on time than are those that do not have evaluations, probably because of the extra attention

**“THE FAD NOW IS LET’S
PILOT IT, AND IF IT WORKS
WE’LL TAKE IT TO SCALE.”**

SCALE THE HEIGHTS

The growing influence of the randomized controlled trial in economic spheres can be seen in the number of studies published each year. Most of the increase is in four sectors — although many studies overlap.



D. B. CAMERON, A. MISHRA AND A. N. BROWN, J. DEV. EFF. HTTP://DOI.ORG/GN8 (2015).

that is given to initial set-up, roll-out and monitoring⁵.

This finding is good news for individual projects, but it is also a potential thorn in the side of many RCTs. Positive effects seen in a trial setting may disappear when the programme is scaled up, governments take over and all the extra attention disappears (see *Nature* **523**, 146–148; 2015).

“The fad now is let’s pilot it, and if it works we’ll take it to scale,” says Annette Brown, who heads the Washington DC office of the International Initiative for Impact Evaluation, an organization that funds impact evaluations as well as meta-analyses of existing studies. Brown says that researchers and governments should probably conduct rigorous studies when any programme is scaled up to ensure that the results continue to hold true — just as the government in Haryana is doing now.

RANDOMIZATION BIAS

From a political perspective, the strongest argument in favour of well constructed RCTs — that they do not lie — may also be the biggest factor working against them. Local politicians often want to cut ribbons and release money into communities, whereas international donors, including governments and NGOs, want flagship programmes that show how they are improving the world. They do

not welcome results showing that initiatives are not working. Even in Mexico, Levy says, some of the subsidies that he fought against when he created *PROGRESA* have regained political favour.

But the randomistas have been accused of succumbing to their own biases. Some fear that their insistence on the RCT has skewed research towards smaller policy questions and given short-shrift to larger, macroeconomic questions. One example comes from Martin Ravallion. An economist at Georgetown University and a former research director at the World Bank, he cites an antipoverty programme in China that received \$464 million from the bank in the 1990s. Although the programme involved road construction, housing, education, health and even conditional cash payments for poor families, a study based on data collected in 2005, 4 years after disbursement ended, found minimal average impact on citizens⁶. “That was the only long-term study of integrated rural development, which is the most common form of development assistance,” Ravallion says.

Yet some families did benefit, and by combining statistics with economic modelling, he and his team showed that the difference lay in basic issues, such as education level. For Ravallion, the message is that aid is best targeted at the literate poor, or more broadly at issues such as literacy. “Governments need to know these things,” he says. “They can’t just know about the subset of things that are amenable to randomization.”

To Alexis Diamond, a former student of Duflo’s who manages project evaluations at the International Finance Corporation, the private-sector development arm of the World Bank in Washington DC, the debate between the randomistas and the old-guard economists is in many ways about status and clout. The latter have spent their careers delving into ever more complex and abstract models, he says. And then “the randomistas came along and said ‘We don’t care about any of that. This is about who has a seat at the table.’”

Diamond says that he tries to strike a balance at his organization, where most evaluations still rely on a mixture of quantitative and qualitative data, including expert judgement.

Duflo shrugs off the debate and says that she is merely trying to provide government officials with the information — and tools — that they need to help them spend their money more wisely. “The best use of international aid money should be to generate evidence and lessons for national governments,” she says.

She points to an anti-pollution programme in industrial plants in the Indian state of Gujarat. Partnering with a group of US researchers, the state ran an experiment in 2009 that divided nearly 500 plants into 2 groups. Those in the control group continued with the conventional system, in which industries hire their own auditors to check compliance with pollution regulations. The others tested a scheme in which independent auditors were paid a fixed price from a common pool. The hope was that this would eliminate auditors’ fear of being black-balled for filing honest reports. And it did: independent auditors were 80% less likely to falsely give plants a passing grade, and many of the industrial plants covered by those audits responded by curbing their pollution. In January, regulators rolled out the programme across the state.

“My hope, in a best-case scenario, is that in the next ten years you are going to have many, many of these projects run as a matter of course by governments in the spaces where they want to learn,” Duflo says. ■ [SEE EDITORIAL P.135](#)

Jeff Tollefson is a reporter for *Nature* in New York.

1. Bannerjee, A. V., Duflo, E., Glennerster, R. & Kothari, D. *Br. Med. J.* **340**, c2220 (2010).
2. Glewwe, P., Kremer, M. & Moulin, S. *Am. Econ. J. Appl. Econ.* **1**, 112–135 (2009).
3. Chattopadhyay, R. & Duflo, E. *Econometrica* **72**, 1409–1443 (2004).
4. Habyarimana, J. & Jack, W. *Heckle and Chide: Results of a Randomized Road Safety Intervention in Kenya* (Center for Global Development, 2009); available at go.nature.com/lyewbv
5. Legovini, A., Di Maro, V. & Piza, C. *Impact Evaluation Helps Deliver Development Projects* (World Bank Group, 2015); available at go.nature.com/tv3hqo
6. Chen, S., Mu, R. & Ravallion, M. *J. Publ. Econ.* **93**, 512–528 (2009).