

Michael Kremer has recently been named the new [Chief Economist of the World Bank](#). In honor of this appointment, we wanted to rerun this conversation between him and [Timothy Ogden](#) in which he discusses his approach to economics.

The below originally appeared in [Experimental Conversations](#), published in 2017 by MIT Press. Since it originally ran, Michael became the University Professor in Economics and the College and the Harris School of Public Policy, the Director of the Development Innovation Lab, as well as the Faculty Director, Development Economics Center at the University of Chicago. He also received the 2019 Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel.

Tim: What was on your mind when you started doing randomized evaluations in Kenya? Why were you thinking about randomization and what impact did you think it would have at the time?

Michael: In the early 1990s, academic economists were paying increased attention to the issue of how to get reliable econometric identification. In other words, how to separate out the causal impact of a specific policy or factor from potential confounding factors. For example, researchers were using instrumental variables techniques.

I had gone on vacation to the rural community in Kenya where I had lived and taught high school after college and was speaking to Paul Lipeyah, a Kenyan friend of mine. He had gotten a job with an NGO and told me he needed to pick seven schools for a new child sponsorship program. It occurred to me that maybe it would be possible to choose 14 schools and randomize across them to evaluate the causal impact of the program. My friend took the idea to his boss [Chip Bury of International Child Support Africa] and they decided to go for it.¹

Based on this initial work, it became clear to me that randomization could be made a practical tool for development economists, and that collaborations between researchers and NGOs could make it possible to test a variety of different approaches. Randomization was not just for very large-scale, multimillion-dollar evaluations of government programs with research questions determined by the government, but could be implemented through collaborations between academics and NGOs. It could shed light not just on child sponsorship programs but on broader questions. Within education, for example, it could shed light on drivers of human capital acquisition. How much of the poor learning outcomes in developing countries was the result of low levels of resources in schools, for example? What if we could use the fact that NGOs were putting large amounts of funding into some schools and not others, to try to understand whether low learning levels were due to lack of inputs, or whether it was due to teacher training, or poor child health or any number of other possible causes? It soon became clear that many important questions in development economics could be answered this way.

Getting involved in programs on the ground and tying one's hands by using randomized evaluations forces researchers to confront realities of human behavior, even if they don't correspond to our models. This can lead to the development of better models over time. When Ted Miguel and I did an evaluation of the deworming program of ICS, we found that deworming provided very substantial benefits—eventually

¹ This was published as Glewwe, Paul, Michael Kremer, and Sylvie Moulin. "Many Children Left Behind? Textbooks and Test Scores in Kenya." *American Economic Journal: Applied Economics* 1 (1) (2009): 112–35. doi:10.1257/app.1.1.112.. While this work was seminal in launching the RCT movement, the paper was not officially published until 2009, well after many other influential RCT papers were published.

leading to increased productivity in the labor force.² But despite these benefits, demand for deworming pills fell away sharply when the NGO imposed even a small cost-sharing requirement. A series of RCTs later confirmed a high sensitivity of demand for non-acute health technologies with price.³ This accumulation of evidence helped us to refine our theories of health demand, showing that the human capital theory of investment in health needed to be adjusted. There's been a very productive interaction between RCTs in development and behavioral economics. These series of results have also had an important impact on policy. The World Bank had been an important advocate of user fees for health, but the most recent World Development Report on behavioral economics signaled an important reversal, making the point that fees can deter usage more than might be expected under a rational model.⁴

Tim: I'm interested in the ideas that were circulating around you when you started working on field experiments. Was that the critiques of IV and looking for better identification? Were you thinking about lab experiments from Kahneman and others? Or the large-scale experiments in the US?

Michael: In 1994 when I started the work in Kenya, I was very much influenced by the movement for better identification in labor economics and public finance, but not by lab experiments. I see these traditions as independent, although there is now some convergence of the lab experiment and field experiment traditions. For instance, Nava Ashraf's work combines elements of each in interesting ways.

I also was not reacting to the critics of instrumental variables. Indeed, I think those working on instrumental variables and those of us working on RCTs were motivated by the same impulse, the concern that a lot of empirical work in economics at the time was potentially subject to confounders and required a lot of fairly strong assumptions. That being said, it's not like IV makes all the problems disappear, and neither does an RCT. I don't think anybody thinks that RCTs are magical, but they are a really useful tool for getting at causal impact. So I would say I was trying to get at causal impact in a way that was part of a broader movement in the economics profession to get better identification.

My main impulse was practical—to get more believable answers to real-world questions. I have always been mainly interested in the underlying questions of what policies can address poverty and I realized that RCTs were a tool that could be adapted to help answer this question. I was motivated to make RCTs a more flexible and useful tool.

Tim: What, of the very many efforts you've been involved in—helping begin the RCT movement, Advance Market Commitments, Deworm the World, Development Innovation Ventures, the Global Innovation Fund—are you most proud of? What do you think will have had the largest impact looking back 20 years from now?

Michael: I see all these initiatives as very closely related. Advance Market Commitments are about finding new ways to promote innovation for development. The others are also about innovation for development. Demonstrating that RCTs could be done, working out the practicalities for how to do this work, training others in the technique, raising funding for others to do it, and helping governments and others use the

² Miguel, Edward, and Michael Kremer. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72 (1) (2004): 159–217. doi:10.1111/j.1468-0262.2004.00481.x.

³ Pascaline Dupas has been involved in much of this work, and some of it is discussed in her interview in *Experimental Conversations*. *J-PAL* has an overview of the topic with links to many papers [here](#).

⁴ This included not only the original economics papers, but a book (*Strong Medicine*) and a working group at the Center for Global Development before an organization was created and funded to manage AMCs.

results in developing their policies are all part of a package. Of course, DIV and GIF support not just RCTs but innovations in development more broadly.

People often underestimate the huge amount of practical R&D that went into finding ways to make it feasible to run randomized evaluations that answer both practical and important theoretical questions in developing countries. The field experiments in the US⁵ were being done with budgets of \$40 million dollars or more. Such evaluations can clearly be very important, as the example of PROGRESA⁶ demonstrates.⁷ And by the way, as far as I know, PROGRESA, even though it was going on around the same time I was working on those first randomized evaluations, was also something that was separate. I don't think they knew what I was doing and I don't think I knew what they were doing. The initial randomized evaluations with ICS in Kenya were done on very small budgets, and we had to work things out from scratch. Working on small budgets in developing countries led to a lot of innovation about how to maximize power from limited samples, how to measure outcomes, and how to randomize. Another difference is that we were working with NGOs, as opposed to the large-scale government evaluations done in the US. This was true both in Kenya where I started, and then with Abhijit in India a few years later.

Working with NGOs, as opposed to big government evaluations, opened up the ability to answer a much wider range of questions, because NGOs are more nimble than governments and they're used to trying different things. They're used to not being able to serve everyone, so it's more natural for them to be willing to try randomizing the order of phase-in. I helped set up a long-term partnership with ICS in Kenya, and the work Abhijit and I did with Seva Mandir⁸ also turned into a long-term partnership. Within those partnerships we were able to test all sorts of different ideas from education, to health, to women's empowerment, to agriculture. Graduate students were able to come and join those partnerships and explore new ideas, and those graduate students became junior faculty and then senior faculty, transforming the field.

I'm also happy that I have been able to work with policy makers to use these results to inform policy. Sometimes this involves scaling of particular innovations—such as deworming or chlorine dispensers—and sometimes it involves developing more general lessons, such as on the impact of price on use of preventative health products or the importance of matching teaching to children's current learning level.

It takes a ton of effort to scale. That became very clear in the experience of deworming. We presented our results to policy makers and I think they were very genuinely excited about them. But a Permanent

⁵ These US field experiments of federal programs are discussed in the interview with Judy Gueron in [Experimental Conversations](#).

⁶ PROGRESA is arguably the progenitor of modern conditional cash transfer programs, which provide social welfare payments conditional on recipients taking specific actions like keeping their children in school or getting vaccinations. PROGRESA was evaluated using a randomized trial which exploited the need to roll the program out over the course of several years and found significant impact on many measures of interest. The success of PROGRESA inspired the adoption of CCT programs in many countries around the world.

⁷ For an overview of the PROGRESA program and its impact, see: Skoufias, Emmanuel, and Bonnie McClafferty. "Is PROGRESA Working? Summary of the results of an evaluation by IFPRI." FCND Discussion Paper 118. Washington, DC, 2001.
<http://ebrary.ifpri.org/cdm/ref/collection/p15738coll2/id/77118>.

⁸ Seva Mandir is an NGO based in Rajasthan, India. It was involved in a number of the early education RCTs as well as the immunization promotion experiment (discussed in the interview with Esther Duflo and Abhijit Banerjee) .

Secretary in a Ministry of Education will have many things to deal with, so sustained engagement and support are needed. In the case of deworming, for instance, we co-founded an NGO, Deworm the World, to provide technical assistance to governments to introduce mass school-based deworming programs. It took a lot of work to get the NGO started and to get some large-scale programs going in Kenya and Bihar. But now Evidence Action, which took over the work of Deworm the World, is successfully supporting national programs in Kenya, Ethiopia, and India. They have already reached 140 million children in the first half of 2015 alone.

From my experience working on scaling up deworming, it became clear that there was a need for more institutional support to scale up the lessons coming out of RCTs. When Raj Shah became USAID Administrator and asked me to get involved, this is what I told him I wanted to work on. He was excited by the idea and suggested that I work with Maura O'Neill, who has a background in entrepreneurship. Together we co-founded Development Innovation Ventures within USAID to finance early-stage piloting of new ideas in development, rigorous testing of those ideas, and scaling of those that proved most successful. DIV has funded a lot of RCTs around the world and has helped organizations scale those that work. This is really quite a different model to much of development aid. Instead of aid agencies making top-down decisions about what to invest in and then issuing calls for proposals to implement the vision, this approach involves an open call for innovative ideas, provides resources to support rigorous testing, and then makes further support conditional on results. The DIV approach has influenced the creation of related funds in Peru and Tamil Nadu. DFID was interested in what we were doing, and so over the past few years I have been involved in setting up a new international venture, the Global Innovation Fund, which is supported by the US, the UK, Sweden, Australia, and the Omidyar Network. We just launched it this year, and I am very excited about it.

The modern movement for RCTs in development economics often gets put in the evaluation category, but in fact the movement is about innovation, as well as evaluation. It's a dynamic process of learning about a context through painstaking on-the-ground work, trying out different approaches, collecting good data with good causal identification, finding out that results do not fit pre-conceived theoretical ideas, working on a better theoretical understanding that fits the facts on the ground, and developing new ideas and approaches based on theory and then testing the new approaches. The idea for DIV and GIF is very much about innovation, so if you need support to pilot your idea before you're ready to rigorously evaluate impact, DIV and GIF will both pay for that. Then they'll pay for a rigorous impact evaluation, and if it works, they also go on to the next stage, which is trying to help transition innovations to scale.

So I see getting the modern movement of randomized evaluations started, showing they were possible, working with NGOs, working to try to scale successful development approaches like deworming, drawing out lessons like the sensitivity of preventive healthcare to fees, and then trying to build institutions to keep doing these things as a package that I hope will have lots of impact.

Tim: There's an implicit theory of change there about the problem being the lack of institutions to generate and use evidence.

Michael: I think that's right. You can see a lot of what I have worked on as creating an institutional framework to allow ideas in development to be rigorously tested and scaled. Even my early work in Kenya was building the local infrastructure to allow people to run high-quality field experiments—from trained enumerators to systems people could run grants through. Many others have helped set up that infrastructure too and the world has come a long way in the last 20 years. There's J-PAL, there's IPA, there's 3ie, there's the SIEF program at the World Bank, there's DIME, so tremendous progress has been made.

Tim: There's a bit of irony in the "institutions matter" critique that RCTs are paying attention to things that are too small, but the movement has created institutions.

Michael: RCTs can't be used to answer every issue, but they can shed light on many issues. Deworming may seem like a small question but evaluations of deworming shed light on the interrelation of health and education, people's sensitivity to small copays for non-acute health, responsiveness of behavior to health information, and information flow through social networks. It also helped shed light on methodological issues about measuring externalities.

There are now many good RCTs on political economy questions. Ben Olken has done a lot of good work on corruption using RCTs⁹ and there is a lot of work on making bureaucracies more responsive.¹⁰

Some of my work on education that initially started as an evaluation of a textbook program wound up being about institutions. As I mentioned earlier, one of the great things about randomized evaluations is that it forces researchers to get very in touch with the reality of what's on the ground. Academics are not just sitting around theorizing about the challenges of development; they are getting their hands dirty. Very often in that process they realize their theories are wrong, and they come up with new theories. And then when you test those theories with an RCT, you are often surprised by the result. There is no room to tweak your result to fit your preconceived ideas, it is what it is. RCTs force you to confront reality.

I had taught in a school in Kenya, and seeing how few resources the schools had made me think it must be good to have more resources, including more textbooks. But we found that textbooks only improved test scores for those who were already performing well. That made me think about the issues of fit between the curriculum and where the students currently were.

Textbooks were written at a level far from the level of many students. One of the big lessons from RCTs in education is the mismatch between where curricula are and where teaching is oriented and where the typical student is. There are institutional reasons for that. There are political economy reasons for that. I think we understand a lot more about one of the key institutional problems in education because of randomized evaluations. In that case it didn't start out with an analysis of the institutional reasons for the mismatch, but the data brought us to it.

We've identified a number of very useful interventions, such as remedial education, to address that problem, but it remains a big political economy problem to get those interventions implemented. Those are useful policies to help address this problem, but we probably also need curricular reform, and that's a harder task. But there's no reason why if we do have curricular reform, we can't understand the impact of that with randomized evaluations.

⁹ This includes work in Indonesia on extortion of truck drivers and on corruption in road building and the gap between perceptions of corruption and actual corruption. For an overview, see Olken and Pande, "Corruption in Developing Countries" *Annual Review of Economics* (2012), Olken, Benjamin A., and Patrick Barron. "The Simple Economics of Extortion: Evidence from Trucking in Aceh." *Journal of Political Economy* 117 (3) (2009): 417–52. doi: 10.1086/599707, Olken, Benjamin A. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115 (2) (2007): 200–49. doi: 10.1086/517935.

¹⁰ See, for instance, Banerjee et al.'s work with the Rajasthan police, and Olken's work on performance pay for tax collectors in India.

Some other critiques of RCTs, for example, about whether results from one context will generalize to others, are general points about empirical work, not about randomized evaluations.

Tim: I want to go back to the issue of AMCs that we set aside at the beginning of this discussion. There's a reasonable argument that AMCs increase the number of vaccines and, combined with GAVI, gets a lot more kids vaccinated—and that one of the most important things to happen in development is keeping a bunch of kids from dying or being permanently handicapped physically or cognitively. How would you compare the potential impact of building institutions to generate evidence, versus “we saved a bunch of kids”?

Michael: Institutions to encourage R&D on health products for the developed world include both public biomedical research funding and intellectual property rights to encourage private sector R&D. The idea of AMCs was to expand the set of tools we had available for encouraging R&D. Intellectual property rights systems create R&D incentives but also involve some static distortions, and those can be quite costly. There can also be some dynamic distortions. I had written a paper on the idea of buying out patents—firms could voluntarily sell their patents to governments, which could put them in the public domain.¹¹ That preserves the dynamic R&D incentives created by intellectual property rights but avoids some of the static distortions associated with intellectual property rights, as well as some of the dynamic distortions, like the incentive to develop “me-too” drugs just to get around the patent, and the disincentives to develop follow-on drugs.

AMCs for vaccines have a lot of the same properties as buying out a patent, but it seemed like they were institutionally and politically easier to do.¹² There was a huge amount of work to move from an academic idea to something that could be implemented in policy.¹³ It was great that was done for the pneumococcus vaccine, and then that vaccine was developed.

I think the biggest potential benefit of AMCs would be to encourage R&D for diseases that are a little more distant. In the case of pneumococcus, there was already a vaccine for the strains that were common in the rich world, but there wasn't one that covered all of the strains that are relevant to a lot of poorer countries. It was a technological challenge that I don't want to minimize, but it was less of a technological challenge than developing a completely new vaccine. I haven't spent time on this for a while, though I am writing an academic paper on vaccines and drugs right now. But, in general, I think there's scope for institutional innovation to come up with new mechanisms to encourage R&D and to make the products accessible. I hope I'll go back to doing some work on that at some point.

Tim: Do you think about your role in training either directly or indirectly so many of the people involved in doing randomized evaluations? At the risk of goading you into saying some not very humble things, what's the counterfactual of a world without Michael Kremer?

Michael: It's been great to work with some amazing graduate students and co-authors over the years, many of whom are now doing incredible things. One of the great things about setting up the operation at

¹¹ Kremer, Michael. “Patent Buyouts: A Mechanism for Encouraging Innovation.” *Quarterly Journal of Economics* 113 (4) (1998): 1137–67. doi:10.1162/003355398555865.

¹² Kremer, Michael, and Rachel Glennerster. *Strong Medicine: Creating Incentives for Pharmaceutical Research on Neglected Diseases*. Princeton: Princeton University Press, 2004.

¹³ This included not only the original economics papers, but a book (*Strong Medicine*) and a working group at the Center for Global Development before an organization was created and funded to manage AMCs. See: Levine, Ruth, Michael Kremer, and Alice Albright. *Making Markets for Vaccines: Ideas to Action*. Washington, DC: Center for Global Development, 2005.

ICS in Busia has been seeing these incredibly talented people who are getting training in top-notch institutions and bringing that training face to face with life in Kenya and generating fantastic ideas.

I think that was very valuable for the students, but it's also valuable for the world insofar as a lot of those students went on to do great things. I'm very happy about that.

Tim: One version of the impact of the field experiment movement is we've transitioned from a situation that, in general, serious economists didn't go to the field to one where they do. Does that resonate with you?

Michael: The world needs all sorts of research.

We need field researchers, theorists, macroeconomists. There's fantastic work going on in economic history. I think the idea of spending time in the field and being involved in randomized trials and other fieldwork is great, and that is one very important set of techniques and tools. I'm also open to other approaches. I do think that, in general, it is good for development economists to have spent time living and working in developing countries, but that does not mean that every project should be based on fieldwork.