

Measurement & Evaluation

Experimental Conversations: Perspectives on Randomized Trials in Development Economics

A dialogue with University of Michigan professor Dean Yang on what we can—and can't—learn from randomized controlled trials.

By **Timothy Ogden** | May 8, 2017

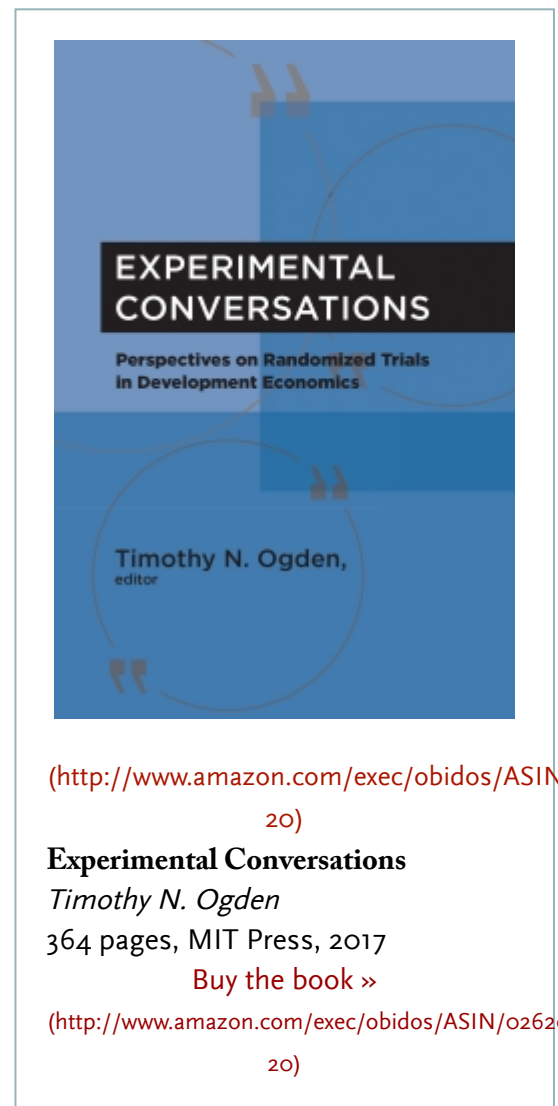
Dean Yang is a professor in the Department of Economics and the Ford School of Public Policy at the University of Michigan. Dean's doctoral research was based on a natural experiment: variation in rainfall amounts across Indonesia. The fact that the country maintained both detailed records of weather and census-type data allowed him to show how Indonesian families prioritized the needs of male children during hard times.

While continuing to exploit natural experiments, Dean's research now includes many randomized controlled trials (RCTs). That research looks at microfinance, risk and insurance in poor households, and corruption. He also has done extensive research on migration and remittances, in particular looking at interventions that might increase the positive impact of remittances.—Timothy Ogden

Timothy Ogden: Why don't we start with natural experiments versus field experiments. Tell me about how the **Indonesia rainfall paper** (<https://www.aeaweb.org/articles?id=10.1257/aer.99.3.1006>) came about.

Dean Yang: I had actually worked on exploiting rainfall variation and weather variation for a couple of years before I got the idea for that paper. I am far from the first person

Subscribe
Sign In



to use weather variations to achieve exogenous variation. At Harvard it was an idea that was floating around. **Michael Kremer** (<https://scholar.harvard.edu/kremer>) was my advisor and **Ted Miguel** (<http://emiguel.econ.berkeley.edu>) was wrapping up his graduate work and they were thinking about things like this. Weather was one category of exogenous variation that people tried to exploit in their dissertation work.

One of my dissertation chapters was looking at weather variation and how it affected international migration from the Philippines and how households responded by sending migrants overseas. Weather variations were very much on my mind at that time. But how the Indonesia paper emerged was kind of idiosyncratic. I was preparing a lecture for my master's level course in development economics here at Michigan. And I read an article by Bob Fogel whose **work** (<http://www.sciencedirect.com/science/article/pii/S1574003X97800268>) was about how early life conditions, nutrition investments in early life in particular, influenced people's outcomes over a lifetime. I discussed the paper with my wife who is a health economist, she's my co-author on that paper, and somehow we both had the brainstorm that it would be great to bring a strong causal identification to these claims Fogel was making about how early life conditions are important for people's long-run outcomes. Over the course of his career, Fogel accumulated a lot of compelling material, but we both felt that a strong causal identification was missing. So we decided to work together on trying to find a context in which we would be able to link weather events around the time of one's birth to later life outcomes.

It was an unusual way for a research project to emerge because we had this idea about how to achieve causal identification, but we didn't know where we were going to do it or what data set we might be able to use. I tell my students this all the time—you really need to have a lot of pots on the fire because you never know when all the pieces are going to come together for you to write a good paper. You not only need to have a good idea of how to achieve causal identification, but you also have to find a context and a data set with which to work, if you're not running your own experiments that is.

So we basically set out and scoured all the data that we could find that had the specific set of elements that we needed to have. It turned out that Indonesia was just about the only place that we found where we could do this because the **Indonesian Family Life Survey** (<http://www.rand.org/labor/FLS/IFLS.html>) (IFLS) turned out to have all of the data items that we needed. We then just needed to buy some supplemental weather data from the Indonesian government.

It was actually a surprise to us, in retrospect, that there was only one country and one data set that we could have used to write this paper. If the IFLS didn't exist, or if they hadn't collected very detailed data on birth year and place, we wouldn't have been able to write this paper. There was one other candidate data set, a **Russian living standards survey** (<http://goo.gl/NRxCY>) from the 1990s, I think. It had a similar structure with

data on date and location of birth. But we didn't think it was as promising a context as Indonesia, mostly because we didn't have a strong prior that weather shocks around the time of birth would have a strong effect on people's later life outcomes. We thought if there was, we'd be more likely to see it in Indonesia.

Ogden: There have been a lot of studies and natural experiment papers using Indonesian data, presumably because it has these very good data sets. There are a few other data sets that crop up over and over. People tend to use them because they are accessible and a lot cheaper than running an experiment or collecting new data. Do you worry about external validity issues because of that? That there is something systematically different about Indonesia reflected in the fact that they do collect good data? Or, alternatively, that so many people are running so many tests against that data that there are going to be false positives?

Yang: You could frame your question about development research as a whole. What percentage of results that are published are spurious or can't be replicated? That's a broader concern about the production of knowledge, and what we can learn from articles that are published in journals. I don't think it's anything but a good thing that in Indonesia there is high-quality socioeconomic data collection by the government and by independent researchers like IFLS.

But to address one of the things that you mentioned, yes, Indonesia is a very specific context. But it is a large and important country. So, even if one made the argument, which I'm not making, that findings from Indonesian data only applied to Indonesia, they can have very important human consequences. The point is right that you have to worry about external validity with any research project. Lots of folks ask me about whether my Indonesia findings are likely to apply in sub-Saharan Africa, and I can't do anything but speculate. I certainly wouldn't want to make any strong statements about how it would apply there.

The other question was about specification searching, or not necessarily specification searching but related to publication bias.

Ogden: It's not any particular individual running tests until they find something, but the fact is that with the public data sets you get a lot of tests being run, and no one knows how many tests are being run.

Yang: I wouldn't necessarily frame the issue as being related to any particular data set, but you're absolutely right, there's a limited set of publicly available data sets out there. And researchers are behind the scenes, unobserved by the larger research community, running tests and looking to see what statistical relationships exist in the data. Then the articles that end up being published are these subsets that end up with statistically

significant and publishable results. I think it's widely appreciated that there is publication bias. Some fraction of these results is going to emerge by chance.

Ogden: What was the first RCT you were involved in and what attracted you about the method?

Yang: I'd been exposed to a number of RCTs in graduate school. My adviser was Michael Kremer who really started off the whole wave of RCTs. It was something that was definitely in the air as a methodology, and I got to witness firsthand from the working paper stage all the way to successful publication how to use the methodology, what the challenges were in being convincing and answering more fundamental economic questions. I found it very appealing, and perhaps had some early insight into how these things were done even though I didn't actually work on any of them as a graduate student.

I graduated in 2003, and I was actually exclusively working on non-RCT projects but as I got out of the phase of getting my dissertation chapters published and started to think about new things to work on, I naturally gravitated toward the RCT methodology. I think I also probably felt that I had the raw skills that could make these things work, that perhaps this was something where I might have a comparative advantage.

I worked in business for three years before going to grad school so I felt like I had some ability to see things from the standpoint of practitioners and business people and to frame a research study as something that could also be of business interest to, say, a microfinance institution. And that's turned out to be the case. I feel like I've been able to find studies, and set up studies, that have been of mutual interest to the research community and to the business community, that could provide important bottom-line contributions to MFIs or remitters. I also think I had more of a taste, or less distaste, for management of fieldwork than other academics, at least on average.

Ogden: I'd like to ask about how you interpret and internalize evidence. You've done [work in Malawi on commitment savings devices](http://www.journals.uchicago.edu/doi/full/10.1086/684014) (<http://www.journals.uchicago.edu/doi/full/10.1086/684014>). Lots of people are working on similar studies, trying to understand the underlying pathways of commitment devices. If those other experiments showed something quite different than yours, what would it take for you to conclude that the Malawi results were the outliers?

Yang: What would it take to move my priors in the opposite direction on a particular finding?

Ogden: Right, but let me broaden it. Do you distinguish between evidence from natural experiments and randomized experiments? Are they essentially equivalent in terms of how we should interpret results from

each of those methods? Or should we think of them differently? More generally, do you think in a Bayesian way? What does it take to change your mind?

Yang: That's a great question. When it comes to methodology, I don't think that evidence from one type of study should necessarily be privileged over another. I don't have a strong stance on that. Most of the issues I would raise are practical and have to do with how easy it is to do a good job with a credible causal identification using one approach versus another. The broad point I would make is that both types of approaches should be in the development economist's toolkit. I don't think it's right to throw out one tool or exclusively privilege another. But the two different tools have real plusses and minuses. The beauty of randomized trials, in my mind, is that you can achieve much stronger claims to causal identification, conditional on the experiment being conducted appropriately and dealing with issues like selection and attrition. The drawback of RCTs is that it's often extremely difficult, if not impossible, to answer certain questions. I think development researchers have been pushing the boundaries on the types of questions that RCTs can answer. I expect that's going to continue because there are great returns to pushing those boundaries. But certain types of questions are probably never going to be answered using RCTs, like most major macroeconomic questions. Practically, there's a limit, a limit we haven't reached yet, to the number of real world organizations or governments that are going to cooperate in designing and implementing an RCT. As researchers, we don't have the ability to actually run a program. So, for the most part, you have to collaborate with someone.

So the ability to ask questions is only partly under the control of researchers. You have to have real world organizations that are willing to randomize. For a long time that limited the ability of researchers to ask questions like "what is the impact of microcredit on household well-being?" until [Abhijit and Esther](http://faculty.wcas.northwestern.edu/~cgk281/MoM_AEJ.pdf) (http://faculty.wcas.northwestern.edu/~cgk281/MoM_AEJ.pdf) and [Dean \[Karlan\] and Jon Zinma](https://academic.oup.com/rfs/article/23/1/433/1579262/Expanding-Credit-Access-Using-Randomized-Supply) (<https://academic.oup.com/rfs/article/23/1/433/1579262/Expanding-Credit-Access-Using-Randomized-Supply>) were able to find MFIs that were willing to randomize in some way.

Those are the downsides of RCTs in my mind. The beauty of the natural experimental approach is, in principle, that there's no limitation on what research questions you can answer. In reality, you never know whether you're going to be able to find a particular measure, but the range of questions is not limited in theory to any particular realm. The challenge is finding causal identification, finding some natural real world variation that will allow you to answer a particular question. The other key downside to the natural experimental approach is that because one is not typically able to work with explicit lottery-based randomization—in the real world there are very few programs that are lottery based—one always has to spend a great deal of time convincing people that you've found plausible exogenous variation. I would say, the typical natural experimental study—though some studies are quite convincing—is much less convincing in

terms of causal identification than an RCT. Even though one can try to provide auxiliary evidence or subsidiary analyses to try to rule out alternative stories, typically it's not possible to dot all the I's and cross all the T's and to rule out with 100 percent certainty all potentially confounding stories about the causation in a particular context.

Excerpted from *Experimental Conversations* by Timothy Ogden, published by the MIT Press. Copyright 2017 All rights reserved.

Notes

¹ [The Indonesian Family Life Survey](#) is an ongoing project of the RAND Institute. Begun in 1993/1994, it has a sample representative of more than 80 percent of the Indonesian Population.



Timothy Ogden ([@timothyogden](#)) is managing director of the [Financial Access Initiative](#) at NYU-Wagner and executive partner at [Sona Partners](#), a thought leadership communications firm. He is the author of [Experimental Conversations](#), a collection of interviews with economists conducting field experiments on poverty alleviation, co-author of [Toyota Under Fire](#), and has developed and edited more than 20 books, including award-winning books on innovation and marketing. He serves as president of the [Bardet Biedl Syndrome Foundation](#), vice-chair of GiveWell, and as a contributing editor for *Alliance* magazine.

Sign In
Subscribe

If you like this article enough to print it, be sure to subscribe to SSIR!

Copyright © 2017 Stanford University.
Designed by Arsenal, developed by Hop Studios