



Timothy Ogden [Follow](#)

Managing Director, FAI and USFD. Senior Fellow, Aspen FSP. Co-Founder, Sona Partners. Co-Auth...
Oct 13, 2015 · 16 min read

Experimental Conversations: Angus Deaton

*This is a chapter from the forthcoming book **Experimental Conversations**, to be published by MIT Press in 2016. The book collects interviews with academic and policy leaders on the use of randomized evaluations and field experiments in development economics. To be notified when the book is released, please [sign up here](#).*

Note: the text is heavily annotated and cited. Unfortunately Medium recently gutted the ability to display annotations or citations. Click on the asterisk in the sidebar, and often “see more”, for relevant annotations.

Tim Ogden: I want to start by understanding if there is a critique of evidence from RCTs independent of evidence from localized studies.

Angus Deaton: I am not sure I understand the question. Is a localized study one that uses 20 million observations from a census? I would have thought that a randomized control trial is usually a localized one.

TO: Yes, but there is lots of work that is localized but not randomized.

AD: I think one of the big issues is that these RCTs are typically small and localized. And a lot of the observational studies use nationally representative data sets.

TO: That’s what I’m trying to home in on: the difference between a critique of a method and a critique of a sample.

AD: Maybe I’ll come at that indirectly. If you go back 50 or 60 years when economists started playing with regression analysis, they thought they had a magic tool that would reveal just about everything. They would run multi-variable regressions on all sorts of things and interpret that within, in a way completely unjustified by today’s standards, a causal framework. Then over the years economists and other people

learned that there were all sorts of problems with that. If you go to an econometrics course now, they're not teaching the magic regression machine. It's more like the regression diseases and what's wrong with regression. I think economists, especially development economists, are sort of like economists in the 50's with regressions. They have a magic tool but they don't yet have much of an idea of the problems with that magic tool. And there are a lot of them. I think it's just like any other method of estimation, it has its advantages and disadvantages. I think RCTs rarely meet the hype. People turned to RCTs because they got tired of all the arguments over observational studies about exogeneity and instruments and sample selectivity and all the rest of it. But all of those problems come back in somewhat different forms in RCTs. So I don't see a difference in terms of quality of evidence or usefulness. There are bad studies of all sorts.

I think it depends a lot on the details. I also think what strikes me as very odd about a lot of the development work, is there was a huge amount of experimentation done in economics 30–40 years ago. There were many lessons from there, and a lot have been forgotten. There are people still around like Chuck Manski and Jim Heckman who understand very well what those problems are.^[i] In his recent book, Manski has some terrific stuff on RCTs, particularly on how many assumptions they implicitly make.^[ii] [The *randomistas*] like to argue that RCTs don't need assumptions but they're loaded with assumptions at least if you're going to use them for anything.

People tend to split the issues into internal and external validity. There are a lot of problems with that distinction but it is a way of thinking about some of the issues. For instance if you go back to the 70s and 80s and you read what was written then, people thought quite hard about how you take the result from one experiment and how it would apply somewhere else. I see much too little of that in the development literature today. Maybe I'm missing something, but my reading of the J-PAL webpage makes me think that when they list estimates they seem to suggest you can use them pretty much anywhere.

Which is pretty weird when you think about it. Causality can change locally too. Even if you've uncovered a causal effect that doesn't mean

that causality will work that way somewhere else. It's not just the size of the effect.

TO: Taking this out of the development context, as you mentioned, RCTs have been done for a long time outside of a development context. There are these touchstone studies—the social policy experiments in the United States—that get referred to often. What is the difference there? Is there one?

AD: First there were hundreds of them and they still go on today. But I think that many of those studies were very high quality. I think those people thought very hard about the strengths and limitations of RCTs, and that much of that seems to have been lost today, which is a pity.

For instance, in a newspaper story about economists' experiments that I read today, a reporter wrote that an RCT allows you to establish causality for sure. But that statement is absurd. There's a standard error, for a start, and there are lots of cases where it is hard to get the standard errors right. And even if we have causality, we need an argument that causality will work in the same way somewhere else, let alone in general.

I think we are in something of a mess on this right now. There's just a lot of stuff that's not right. There is this sort of belief in magic, that RCTs are attributed with properties that they do not possess. For example, RCTs are supposed to automatically guarantee balance between treatment and controls. And there is an almost routine confusion that RCTs are somehow reliable, or that unbiasedness implies reliability.

But this is a reinvention of statistics. Reliable is something to do with precision. And RCTs in and of themselves don't do anything about precision just by being RCTs. But when you read the literature, the applied literature, there are claims about precision which are often false. There's nothing, *nothing*, in an unbiased estimator that tells you about reliability. If you know what you're doing you get a credible standard error, and that, of course, tells you something about precision. One of the first things one learns in statistics is that unbiasedness is something you might want, but it's not as important as being close to

the truth. So a lexicographic preference for randomized control trials—the “gold standard” argument—is sort of like saying we’ll elevate unbiasedness over all other statistical considerations. Which you’re taught in your first statistics course not to do. In the literature, and it happens in medicine too, people use this gold standard argument, and say we’re only going to look at estimates that are randomized control trials, or at least prioritize them. We often find a randomized control trial with only a handful of observations in each arm and with enormous standard errors. But that’s preferred to a potentially biased study that uses 100 million observations. That just makes no sense. Each study has to be considered on its own. RCTs are fine, but they are just one of the techniques in the armory that one would use to try to discover things. Gold standard thinking is magical thinking.

I think that right now the literature is claiming way too much for RCTs.

TO: Would it be accurate, then, to say that you have less of an issue with the method than the claims being made with and for the method?

AD: Yes. Except that sounds like the method’s OK and it’s just being misapplied. I suppose to some extent that’s true. But it’s not just misapplied, it’s a belief that this method can do something which it can’t, the replacement of statistics with magical thinking.

TO: Given the history of 40 or 50 years of people thinking hard about these issues of selection, identification, bias and validity, why do you think the movement gained such traction in development when it did?

AD: That’s a good question. There are certainly lots of problems with observational data. I think people got very tired of dealing with them. I don’t actually think those problems are avoidable; they have to be faced in one way or another, whatever method you use. But you can see why I think some of it is a youth over old age sort of idea. Here’s a new tool, we can rethink the world with it.

I think the rhetoric was very enticing though I don’t think it delivers much. All this stuff about how policy makers can understand it, it’s just the difference between two means and there is no room for controversy. But that is just a hope. A very good example is what’s

happened with the Kremer and Miguel worms study.^[iii] That study was replicated in India in a paper by Miguel and Bobonis.^[iv] And that is enough to put it into practice through Deworm the World followed by Evidence Action to scale up deworming. But there's a 150 page Cochrane Review, which includes the Kremer Miguel study, which says there is no consistent or obvious effect.^[v] I'm not a fan of the Cochrane Collaborations or of meta-analysis, and I have no special insight into this case, but it illustrates that RCTs don't eliminate controversy.

Now I have no idea who's right and that is not what I am talking about here. But when you think about it for a minute you may realize there might not *be* any right. What works in one place may not work in another place especially for something as complicated as deworming when there are social interactions and it depends on the environment and on sanitation and on whether kids wear shoes and on prevalence and all that sort of thing. Maybe the Cochrane Collaboration review is chasing something that doesn't exist. And I know that Michael and Ted are contesting the Cochrane Collaboration's analysis.

But this is exactly the situation we were in before these sort of studies started. Different studies gave different results and no one could really resolve the discrepancies. I think it's just a really good example that suggests that as we get more results there ain't going to be a clean resolution in all cases, because the results will sometimes be all over the place, even when they are correctly and precisely done. Discrepancies between studies have to do with much more than bias!

I think [RCT advocates] thought they were going to solve a problem which I don't think is solvable. There is no magic bullet. That's the truth of it. It would be interesting to get some of the advocates to explain why they don't talk more about the stuff that was done in the 60's and 70s, why that has not changed the world, and why it lost momentum, at least among academics?

TO: Working on the US Financial Diaries, I've experienced some of the frustration in the trade-off between doing a project that's data- and time-intensive and one that's nationally representative.

AD: Wait a minute. Look at the US Census data or the American Community Survey, which I was working with this morning. I've got 20 million observations in those data sets. There's hundreds and hundreds of questions. I'm not sure why you make that trade-off.

I very much like the financial diaries work and I've learned a lot from them, and to me they are more useful than a series of randomized trials on the topic because they have lots of broadly useful information. I can make my own allowance for how selected they are and I'm not blinkered by the craziness that if it's not a randomized control trial I shouldn't pay any attention to it. Which I've heard more times than I can count.

TO: There's a trade-off that's about cost ultimately.

AD: I'm not sure there is such a trade-off. Governments spend a lot of money on surveys, how expensive they are depends a lot on the questions you ask and how you ask them.

TO: But you can't add anything to something like the American Social Survey. Once you have those questions, it's impossible to learn anything different because of the bureaucracy involved to make any changes.

AD: It's not just bureaucracy. There are a 1000 people who would like to add a question and it would get totally out of hand if you opened them up to changes. These surveys in the US, especially if they're done by phone, it's very hard to keep people on the phone for more than about 20 minutes. So there are real constraints there. Those are much less severe in a place like India or Kenya though, especially if you're spending dollars and can benefit from the discrepancy between the PPP and the exchange rate.

TO: I guess I often feel discouraged, because of those constraints, about getting answers...

AD: Oh, but that's the beginning of wisdom. It's very hard to do science. If it was easy or there was a magic machine out there we'd all be a lot wiser. It's just very very hard. Things like the financial diaries

and extended case studies are enormously important. Most of the great ideas in the social sciences over the last 100 years came out of case studies like that. Because people are open to hearing things that they didn't necessarily plan, for one thing.

TO: Another of the common critiques of the RCT movement is a lack of a theory of policy change.

AD: I think that's a very complicated thing. These things are slow often, but there is a big political element and there should be. Something I read the other day that I didn't know, David Greenberg and Mark Shroder, who have a book, *The Digest of Social Experiments*, claim that 75 percent of the experiments they looked at in 1999, of which there were hundreds, is an experiment done by rich people on poor people. Since then, there have been many more experiments, relatively, launched in the developing world, so that percentage can only have gotten worse. [\[vi\]](#) I find that very troubling.

If the implicit theory of policy change underlying RCTs is paternalism, which is what I fear, I'm very much against it.

I think policy change very much depends on the context. I don't know if you've read Judy Gueron's book. [\[vii\]](#) I learned a lot from that. What do these MDRC things do? They've gone on and on and continue to this day. Many academic economists were involved in the early days but much less since but MDRC and Abt and Mathematica and so on have gone on doing these experiments ever since. For the Federal government, state governments, and some in Canada. So I'm kind of curious about how they function in the policy space.

I don't think the results from these experiments have had much of an impact on academic knowledge, but that may be wrong. I don't know. I think what the experiments did was to settle disputes between competing political views. There would be a new administration and they would say, "These policies should all be abolished," or, "If we make people go to work before we give them any welfare that will cause them to earn their own incomes and it will reduce costs for the government," for example. The interesting thing is that in the US such arguments have to be costed by the CBO which actually has to estimate

if the financial projections that come out of those proposed policy changes make any sense. When the Reagan people came in, they were not keen on doing any experiments at all but when the CBO didn't agree with their estimates, they became supporters of experiments because they believed it would show that they were right. And sometimes indeed they were, at least as far as RCTs can tell.

Those experiments are mostly about what policy changes did to the budgets of state and federal authorities. They have a case load, and they often care less about the well-being of poor people than they care about the state budget. An RCT is good for that because it gives you the average cost and the average in that context is exactly what you want to know. It resolves the dispute. But that average is not generally useful elsewhere, at least without understanding the mechanisms. And MDRC wrestled with that problem of finding mechanisms from the very beginning but they never resolved it. They thought that by going into the details they could find mechanisms that would generalize or transport and they never managed to do that. You *can't* do that with RCTs. You've got to combine them with theory and observational data so you're right back where you were.

But before you need a theory of policy change you need a theory of transportability. Meaning, it works here, what arguments do you have that it works there? And it often seems that those running RCTs simply assume that these numbers will apply with little discussion of how to move the results from one location to another.

TO: Where do you think the development economic field needs to go from here?

AD: Economics is a very open profession. Young people who come along with bright ideas get a lot of attention, in comparison to a lot of academic fields that are dominated by old people. I think [the RCT movement] will likely fade in the same way that it faded 30 years ago and for much the same reasons. There will certainly be consulting houses that do RCTs for particular purposes, such as *ex post* fiducial evaluation. I think the academic interest will fade, as the problems are better understood, though I think that RCTs will have more of a place in the economist's toolbox than was true twenty years ago, and as with

other methods, we will have a well thought out view of when and where they are useful. More tools are always welcome, as long as we don't think one of them is a magic tool, or that it is the only tool we need. People will go on doing RCTs along with other things too. There's a lot of competition out there, a lot of people thinking about development in lots of ways. I don't think any long term solutions are going to come out of RCTs. We're certainly not going to abolish world poverty this way.

TO: Do you think there are promising leads in abolishing world poverty?

AD: From RCTs?

TO: From anywhere.

AD: I know what I think which is that we should be thinking much more about politics than about micro-detailed studies. So I'm basically in the same boat as Daron Acemoglu and Jim Robinson.^[viii] I think, to a first approximation, it really is all about politics. And as I say in my book, I think aid is making it worse, not better.^[ix] It's fine to say we discovered this marvelous new delivery system and here's how you should deliver aid. And it might make things better locally. You might save lives, you might get people educated but you're not going to abolish world poverty because that has to do with politics, it's not to do with money. Certainly knowledge can help, but once again it's a question of transportable knowledge. There's got to be some theory of how you can take it from one place to another and that requires theory and generalizations and structural models of some sort. They don't have to be intertemporal dynamic programs, the sort of thing that is thought these days to be structural modeling in economics. You have to think about why things are happening and you can't get that out of an RCT in and of itself.

TO: It seems to me somewhat ironic that the most well-known RCT results are about microcredit's lack of impact. I think much of the philosophical appeal of RCTs is the hope that small local actions can matter. On the politics front it's easy to feel that there's nothing to be done.

AD: Well, I'm not sure that's right. I think there are lots of things than can be done on the politics side. Not propping up dictatorships for example. Or not encouraging them to come into existence by supporting governments that have no need to raise taxes. And as far as individuals doing things, I do believe in that too, very much so. But it has to be local. You can't get a team of people from MIT or NYU to fly in and setup a something-or-other, or to do an experiment in one place, and then hand it over to the World Bank for implementation somewhere else. It's certainly true that teams from MIT or NYU or anywhere else can help provide understanding of mechanisms. I always give the example that between France and the US they figured out that HIV was a sexually transmitted disease and how it worked. That's immensely valuable information for individuals all around the world and especially in east Africa and places where the epidemic has been really bad. That's doing a lot. There are a lot of things like that we could be doing. We could stop selling arms to those countries. We're doing a lot of harm. When students come to me and ask me, "How should I help the poor of the world, should I go to Bangladesh, should I go to Africa?" And I say, "No you should go to Washington. That's where you can do the most good." Of course, I don't mean for American poor people[laughter], but for poor people around the world.

E *xperimental Conversations* collects interviews with 20 academic and policy leaders on the use of randomized evaluations and field experiments in development economics. In addition to Deaton, interviewees include Abhijit Banerjee and Esther Duflo, Michael Kremer, Dean Karlan, Rachel Glennerster, Jonathan Morduch, Lant Pritchett, Nancy Birdsall, Tyler Cowen and Judy Gueron. To be notified when the book is released, [click here](#).

. . .

[i] For examples of Heckman's work on this topic, see:

Heckman, J. and Hotz, J. (1989) "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social

Programs: The Case of Manpower Training: Rejoinder” *Journal of the American Statistical Association*, 84(408)

Heckman, J. (1992) “Randomization and Social Policy Evaluation.” In *Evaluating Welfare and Training Programs*, ed. Charles Manski and Irwin Garfinkel, Harvard University Press

Heckman, J. (1992) “Basic Knowledge—Not Black Box Evaluations.” *Focus* 14(1)

Heckman, J. and Smith, J. (1997), “The Sensitivity of Experimental Impact Estimates: Evidence from the National JPTA Study”, NBER Working Paper 6105

[ii] Manski, C. (2013) *Public Policy in an Uncertain World: Analysis and Decisions*, Harvard University Press.

[iii] Kremer, M. and Miguel, E. (2004) “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities”, *Econometrica*, 72 (1)

[iv] Bobonis, G., Miguel, E. and Puri-Sharma, C (2006) “Anemia and School Participation”, *Journal of Human Resources*, XLI(4)

[v] Taylor-Robinson, DC, et al. (2012) “Deworming Drugs for Treating Soil-Transmitted Intestinal Worms in Children: Effects on Nutrition and School Performance”, *Cochrane Review*,
http://www.cochrane.org/CD000371/INFECTN_deworming-drugs-for-treating-soil-transmitted-intestinal-worms-in-children-effects-on-nutrition-and-school-performance

[vi] Greenberg, D. and Shroder, M. (1999) “The Social Experiment Market”, *Journal of Economic Perspectives*, 13(3)

Greenberg, D and Shroder, M (2004), *The Digest of Social Experiments*, Third Edition, Urban Institute Press

[vii] Gueron, J. and Rolston, H. (2013) *Fighting for Reliable Evidence*, Russell Sage Foundation

[viii] Acemoglu, D. and Robinson, J. (2012), *Why Nations Fail: The Origins of Power, Prosperity and Poverty*, Random House

[ix] Deaton, A. (2013) *The Great Escape: Health, Wealth, and the Origins of Inequality*, Princeton University Press