The disruptive power of RCTs[1]

July 30, 2019

Jonathan Morduch
New York University

Abstract

Two very different kinds of RCTs are used by economists, although they often get lumped together. The first kind is evaluative, used to assess whether a policy or intervention worked or not. Critics worry that privileging these RCTs over other evaluation methods can narrow knowledge. The second kind of RCT is exploratory, asking how behavior, institutions, and markets react to changing prices, contracts, and other economic features. By disrupting status quo economic conditions through experimental design, these exploratory RCTs open new questions for empirical micro-economics in ways that other methods cannot. One can be ambivalent about putting evaluative RCTs on a pedestal while also encouraging exploratory RCTs. Examples from RCTs of insurance, microcredit and digital money illustrate the arguments.

Key words: Randomized controlled trial, selection bias, experiment, microcredit, insurance, liquidity, interest elasticity, mobile banking

[1] I'm grateful for comments from Isabelle Guerin, Florent Bédécarrats, François Roubaud, Tim Ogden, and Martin Ravallion and for discussions with Tim Ogden, Michael Kremer, Lant Pritchett and participants at the conference on RCTs in Development at Agence Française de Développement, March 19 and 20, 2019. Views and errors are mine alone.

The disruptive power of RCTs[2]

July 30, 2019

Jonathan Morduch
New York University

There are two distinct ways that RCTs are used in development economics. In the first, RCTs are used to measure impact. In the second, RCTs are used to explore the nature of economic contracts, behaviors, and institutions. The two kinds of RCT are often lumped together by critics, but the two strands speak to very different questions and serve different purposes. Understanding the power of RCTs, and disentangling debates around RCTs, requires first separating the two modes.

Critics are especially uncomfortable with elevating RCTs as the favored tool for evaluation, but one can accept their criticisms – in whole or in part – and still embrace the importance of RCTs (and want to encourage far more RCTs) in the cause of experimentation. Should the randomistas rule (Ravallion 2018)? No. Are RCTs a gold standard (Bédécarrats et al 2017)? No. In practice, however, RCTs have been—and will continue to be—particularly useful exploratory tools.

The first use for RCTs (and the focus of the heaviest criticism) is the promotion of impact evaluation through randomized methods. The criticism is less often about RCTs *per se* than about putting them on a pedestal, with a special status that accords them more credibility than other evaluation methods. These RCTs focus on evaluating government or NGO programs and policies, and the hope of proponents is that having more credible measures of impact through randomization will mean better investments and interventions (Glennerster and Takavarasha 2013; Kremer 2003, Banerjee and

[2] I'm grateful for comments from Isabelle Guerin, Florent Bédécarrats, François Roubaud, Tim Ogden, and Martin Ravallion and for discussions with Tim Ogden, Michael Kremer, Lant Pritchett and participants at the conference on RCTs in Development at Agence Française de Développement, March 19 and 20, 2019. Views and errors are mine alone.

Duflo 2009). The questions usually focus on “what works.” Older studies include RCTs of government programs like Mexico’s Progresa conditional cash transfer program (Levy 2006) and the US Job Training Partnership Act (Lalonde 1986), and, in the most recent wave, evaluations of NGO programs like the microcredit RCTs rounded up in Banerjee et al (2015). For the most part, researchers design the evaluations but not the interventions. Much of the debate in this book tackles whether and how such RCT evidence should matter.

The second kind of RCT has a different character. It aligns with the experimental mindset increasingly adopted by development economists, with RCTs as a critical methodological innovation. While some economic experiments involve lab-based hypothetical scenarios (e.g., Davis and Holt 1993), this strand of RCTs involves experiments in real settings. The studies are based on experimentally-controlled manipulations of price structures, contracts, teaching methods, healthcare protocols, bureaucratic processes, and the like. Here, researchers participate actively in the design of the actual programs and policies, usually together with a government agency, business, or NGO. The questions asked are exploratory, theory-driven, and motivated by the desire to understand economic possibilities and constraints. The contexts are often limited-scale pilots or limited-time trials. The questions are less often about “what works” than “how and why?” or “what could be?” While RCTs for evaluation are criticized for saying little about “why” – why impact is small or large or appears for some people but not others – these studies center on explanation. They ultimately ask whether the world works in the ways that economic theory says it should. The power of these RCTs lies in how they disrupt business-as-usual by manipulating economic environments and thereby allowing vision into what would otherwise remain unseen or untried.

The line between the two kinds of RCTs can be fuzzy, burred by both RCT advocates and critics, and the aim of this essay is to clarify the modes and illustrate the experimental mindset in development economics. The view I put forward is not necessarily that would be made by a full-throated randomista, but it aligns with how RCTs are often used in practice.[3]

The first section of this essay describes the rise of the experimental mindset coupled with RCTs. The second section gives three examples of RCTs that probe questions related to prices, contracts, and the use of financial services in poor communities. The third locates the focus of RCTs on poverty-reducing interventions and the provision of private goods, and considers the argument that RCTs push focus away from studying the systemic forces that shape economies.

Expanding knowledge by creating variation

The primacy accorded to RCTs for evaluation – along with related methods like natural experiments and regression discontinuity designs – leads to the fear that the rise of RCTs for evaluation unduly and unhelpfully downgrades other ways of assessing what works (e.g., linear regression, conventional instrumental variables, ethnography and qualitative evaluation, and machine learning with Big Data). More worrying, giving primacy to these kinds of RCTs risks restricting attention to the set of economic interventions that are most amenable to randomized trials. The fear, at the extreme, is that giving RCTs a special status in determining "what works" could lead to a loss of knowledge, especially relative to what learning from a diversity of approaches could deliver (Ruhm 2018).

Detractors also worry that advocates exaggerate the precision and the ease of generalizability of RCTs (Deaton and Cartwright 2018). They worry that the kind of evaluative information generated by RCTs is often of limited political and practical value (Drèze 2018, Pritchett 2014), and is vulnerable to misinterpretation for lack of context (Morvant-Roux et al 2014). Like other evaluation methods, RCTs have difficulty providing crisp answers, especially when, as is often the case, it is necessary to extrapolate from a study in one place to a policy environment in another (Cartwright and Hardie 2012, Pritchett and Sandefur 2015, Bisbee et al 2017).[4]

[3] See Ogden (2017) for a view from academics and practitioners engaged with RCTs, with a theme around differing theories of change.
[4] As Imbens (2018) notes, however, scholars using RCTs are aware of the limits and are responding with expanded approaches (e.g., Bates and Glennerster 2017).

Perhaps most worrying, critics argue that the interventions most amenable to evaluation by RCTs are too small, too limited, and too particular. Within economics, RCTs are an easier fit for studies involving private goods than public goods. Moreover, RCTs often focus on marginal impact and on impact on marginal subpopulations (Wydick 2016). They can be used to measure short-term impact when microcredit enters a new region, for example, but not to evaluate how the original customers have fared since the microcredit organization's start (Cull and Morduch 2018).

From a broader vantage, by focusing on small steps to improve the implementation of existing ideas, evidence of impact from RCTs tends to only speak indirectly about the broader structures that perpetuate poverty and inequality. By this view, giving complete primacy to RCTs for evaluation would restrict admissible evidence on "what works" and ultimately narrow understandings of complex economic and social phenomena (Bédécarrats et al 2017).

In contrast, the RCTs for exploration (the second kind of RCT above) more clearly expand knowledge, and most RCTs published by development economists take this direction. The experimental mindset responds to the fact that key variables may not move much in the natural course of things, so experiments are needed to create relevant variation. Prices may not change much in a given moment or sample, nor contracts. Governments, clinics, schools may all act uniformly in a given range. The result is that, while researchers can explore theoretical predictions, they have little hope to take them to the data. Without experimentation, there is too little to observe and thus too little to analyze.

These exploratory RCTs have limits too: it is tempting to draw overly-strong policy conclusions from the trials and pilots, rather than taking them for what they are: informative and provocative but contingent. Yet, at the same time, criticizing these RCTs for being pilots or trials risks missing how they can aggregate to create sharper, more expansive understandings of constraints and possibilities. Although Angus Deaton and Nancy Cartwright argue against giving evidence from RCTs a special status, they note,

> RCTs are often convenient ways to introduce experimentally-controlled variance—if you want to see what happens, then kick it and see, twist the lion's tail… (Deaton and Cartwright 2018, 17).

From the perspective of economic knowledge, twisting the lion's tail with the help of RCTs has pushed researchers to better understand economic theory and question assumptions that were once considered settled.

Consider the case of crop insurance, a product with much potential given the risks of rain-fed agriculture. In practice, however, crop insurance (and its newer variant: index-based rainfall insurance) has been particularly difficult product to sell to farmers. Casaburi and Willis (2018), for example, show that only 5% of Kenyan sugar-cane farmers in their sample purchased rainfall insurance, a finding that reinforces the sense that potential customers are wary of these products, might not understand or trust them, are content to rely on informal mechanisms, and/or find the products too poorly designed or too expensive. Casaburi and Willis, however, use an RCT to experiment with the timing of when the insurance is sold. They ask whether the problem is not mainly the price nor the understanding of customers. Instead, could the low take-up rate occur because insurers ask for the premium to be paid in a lump sum before the planting season, a time when most money is being invested in crops? By randomizing the timing of payment, pushing it to harvest-time (when farmers have liquidity) for a sample of customers, they show an increase in the take-up rate to 72%. In contrast, reducing the cost of the insurance by 30% (but not delaying the timing of payment) only increased demand by one percentage point. The RCT allowed everything else to be kept the same, and, while the finding is not revolutionary, it helps expand perceptions of the problem. Whether the exact parameter is transportable or not is less important than that the study highlights timing and liquidity as constraints to insurance demand to consider seriously in other settings (in addition to highlighting a practical response to the problem).[5]

Casaburi and Willis's experiment in Kenya informs the work of Belissa et al (2019) in Ethiopia. They too investigate the role of liquidity on the take-up of insurance, again asking whether demand is greater when farmers can pay after the harvest when

[5] Similarly, Jonathan Bauchet and I investigate the demand for a life insurance product in Mexico sold to poor women. Using a natural experiment, we find that demand rises by over 59% when customers are allowed to pay in small weekly installments rather than in an upfront sum (Bauchet and Morduch 2019).

liquidity is greater. They additionally explore the role of promoting insurance through Iddirs, local informal risk-sharing mechanisms used by farmers. The Belissa et al (2019) design involves 8579 individuals and 144 Iddirs. The RCT has six treatment arms. The first is a control group that is offered a standard index-based rainfall insurance contract that requires payment before the insurance takes effect. The second group is similar but the product is promoted by a local leader. The third group is also like the first, but delayed payments are allowed. The fourth is similar to the third, but the purchaser is asked to formally sign a binding contract committing to pay the premium after the harvest. The fifth group gets the insurance product promoted through the Iddir (with the possibility of delayed payment), and the sixth gets everything – the possibility of delayed payment, the requirement to sign a binding contract, and promotion of insurance through the Iddir.

Although less dramatic than in the Casaburi and Willis study, delaying the timing of the payment turns out to be substantial for the farmers in Ethiopia, increasing take-up from 8% to 24%. Combining the delayed payment with promotion through the Iddirs intensifies the impact, bringing take-up rates to 43%. Promoting insurance via Iddirs not only helps bring credibility to the insurance product, it also facilitates the collective purchase of insurance against an explicit background of informal insurance. The study, though, shows that about 15% of farmers who agreed to pay after harvest in fact defaulted on their commitments to pay, a level high enough to threaten the economic viability of the insurance product.

The two insurance studies illustrate the fundamental distinction between RCTs for exploration – researcher-designed experiments that open the box to probe mechanisms – versus RCTs for evaluating the impact of established programs. Neither study here measures the impact of insurance on farmers. The main aim is not to evaluate whether insurance "works," and, in line with that, neither study has a pure control group with no intervention. Instead, in both studies the control group has the chance to buy a standard insurance product. Both studies then explore what happens when the products are re-designed in systematic ways to gauge farmer behavior and the viability of the products. The specific results of neither experiment can be extrapolated to other contexts, but the nature of the innovations (the delayed timing of payments, marketing through local groups) and broad concerns (illiquidity, the risk of post-harvest default) can be.

When it comes to impact evaluation, RCTs are often promoted for reducing selection bias due to nonrandom program access, but the two insurance examples show that selection bias is just one of several big challenges in empirical development economics. Here, a main problem is the lack of relevant variation in insurance contracts (especially the lack of observed contracts offering post-harvest payments), a problem exposed via experimentation through the RCT. Neither study had to be an RCT, but both had to involve experimentation and product re-design. Both had to "twist the lion's tale." The fact that both sets of researchers chose to use RCTs stems from the practicality of joining experimentation with randomization in an exploratory mode.

While Ravallion (2018) traces the history of RCTs in economics to experiments in the 1950s and 1960s (see also Gueron 2017), the notable rise of RCTs in development economics started in the 1990s, following a period of methodological ferment that, among other outcomes, led to focuses on natural experiments (Angrist and Krueger 1999). The move from natural experiments to RCTs was not a large one conceptually, pioneered by Harvard's Michael Kremer in Kenya, and solidified later by the establishment of MIT's J-PAL (Kremer 2003, Banerjee and Duflo 2009; see Ogden 2017 for descriptions of process and motivations from Banerjee and Duflo, and Kremer). Kremer and his colleagues too part in designing the interventions, unlike the previous evaluation-based RCTs that tested government-designed interventions. Kremer (2003) summarizes a series of early experiments to improve schooling outcomes in Kenya, including providing free breakfasts, supplying school uniforms, adding textbooks, de-worming children, and introducing more teachers. Several of the interventions increased school participation substantially at relatively low cost.

The examples show where confusion arises about the types of RCTs. Kremer (2003) describes the RCTs as evaluations of the "what works" sort (in the sense above). Yet, without diminishing their value, they are in essence exploratory. They are largely pilot programs, not large-scale public programs. They usefully document possibilities and constraints, providing an important opening or next step rather than the last word.

The ubiquity of sub-optimality and the potential for innovation

Deaton and Cartwright (2018) are careful to distinguish "what works" RCTs from exploratory "how and why" RCTs.[6] In this context, they consider "when RCTs speak for themselves" and situations with "no extrapolation or generalization required":

> For some things we want to learn, an RCT is enough by itself. An RCT may provide a counter-example to a general theoretical proposition, either to the proposition itself (a simple refutation test) or to some consequence of it (a complex refutation test). An RCT may also confirm a prediction of a theory, and although this does not confirm the theory, it is evidence in its favor, especially if the prediction seems inherently unlikely in advance.

What's at stake in most exploratory RCTs is seldom refuting theory in the sense of Deaton and Cartwright. The two insurance examples, for example, center on well-known ideas (illiquidity, lack of trust), and their importance is unsurprising (in the sense that they are both likely to be somewhere on the list of challenges to providing insurance). Instead, what's ultimately at issue is how much faith to place in constrained optimization. A fundamental tenet of neoclassical economics is the idea that markets yield optimal institutions, goods and services, and prices. In theory, the disciplining function of the market should weed out sub-optimal forms. This tenet holds even in second-best or third-best worlds with constraints like asymmetric information and imperfect contract enforcement (Stiglitz and Weiss 1981). In essence, modern economic theory says that what we see is not necessarily perfect, but it is as good as it can get.[7] In other words, existing insurance processes and products should already incorporate ways to deal with problems of liquidity and trust to the extent feasible.

---

[6] Deaton and Cartwright (2018) anchor a special issue of *Social Science and Medicine* focused on "Randomized Controlled Trials and Evidence-based Policy: A Multidisciplinary Dialogue," edited by Ichiro Kawachi, S.V. Subramanian, and Ryan Mowat and featuring 19 responses from leading statisticians and social scientists.

[7] A fundamental result in the economics of information is that equilibria may not even be constrained efficient (Stiglitz 1986). The RCT research program can be seen as showing a far wider range of circumstances with inefficient outcomes.

But is that generally true? Muhammad Yunus's experimentation with credit contracts in the 1970s, which led to the development of microcredit, illustrates a case where tinkering and re-design created a genuine improvement over what the market had delivered. New contracts led loan default rates to drop sharply and profitability to become possible even when lenders charged relatively modest interest rates (Armendàriz and Morduch 2010). What economists thought had been a constrained-optimal outcome turned out not to be. And even Yunus's tinkering was not the last word in microfinance innovation (e.g., Rai and Sjöström 2004, Field et al. 2013).

The exploratory RCTs carry on in this spirit, driven by experimentation, helping to map how far existing institutions and choices are from what could be possible. Increasingly, the RCTs also map why innovation has not happened (for example, fear about the relatively high rate of default documented by Belissa et al.) and, often, test practical steps to mitigate problems. The contribution of exploratory RCTs is seldom a test of a specific theoretical proposition (like "are individuals rational?") but is a demonstration of an innovation or experimental manipulation that exposes (or deepens understandings of) sub-optimality.

Why RCTs?

Writing about testing theory, Deaton and Cartwright note that generalizability is not always the major concern. They continue,

> [Theory-testing] is all familiar territory, and there is nothing unique about an RCT; it is simply one among many testing procedures. (Deaton and Cartwright 2018, 12)

At a high level of generality, it must be true that "there is nothing unique about an RCT" here. There are, of course, other methods that can prod theory, demonstrate sub-optimality, disrupt, surprise, and expand economic frameworks. Methodologists are now building on the experimental mindset, in some cases improving on bread-and-butter RCTs (e.g., Kasy and Sautman 2019) and in other cases integrating randomized assignment with ethnography (e.g., Duncan et al 2007). There are also non-randomized methods that can be used to analyze exogenously-created disruptions. But RCTs, coupled with an experimental mindset, have been particularly helpful in practice.

Part of the case for RCTs when used for exploration echoes the case for using RCTs for evaluation: Selection bias is a constant worry, and RCTs can help rein it in (while, admittedly, creating other issues). But another part of the case is that when you are *already* manipulating the economic environment in an experimental mode, randomization appears as a smaller stretch.

Researchers using RCTs wonder why one would want to use an alternative method to study their given question in their given place. Why study price elasticities for insecticide-treated bednets, for example, using a non-randomized approach when randomizing prices is feasible? In this line, Deaton and Cartwright (2018) note that a frequent response to their critique of RCTs is: "OK, you have highlighted some of the problems with RCTs, but other methods have all of those problems, plus problems of their own" (Deaton and Cartwright 2018, 16). Deaton and Cartwright refuse to accept that retort because they find that reliance on RCTs substitutes one set of problems for another set. As Ravallion (2018) notes, for example, problems in RCTs arise with selective non-compliance and "essential heterogeneity."

Still, the most prominent alternative approaches to causal inference (especially applications of instrumental variables) are subject to well-known limits. If nothing else, the history of empirical development economics has established that (1) selection bias often matters a lot and (2) plausible instrumental variables and natural experiments are hard to find. This is true across economics, but particularly so in development economics.

One illustration is offered by Beaman et al (2018), who construct an experiment to measure selection into borrowing in a sample of farmers in Mali. Their aim is to measure returns to capital and the impact of microcredit for agriculture, paying attention to the possibility that the most promising farmers are more likely to borrow than others. With no exogenous, excludable variation of prices and other external factors, estimation with instrumental variables is not feasible. So, instead, they construct a two-stage RCT. To get insight into the extent of selection bias, Beaman et al. randomly select 88 out of 198 villages in Mali in which to offer loans through a local microfinance provider. They then randomize the allocation of capital grants to a sample within the 110 villages without the loans and a sample of non-borrowers in the 88 villages that received the loans. They can then measure returns to capital for both borrowers and non-borrowers.

On average, returns to capital were large and positive with clear evidence of liquidity constraints. Recipients of the capital grants (which were worth about $140) in the 110 villages without the loans increased land under cultivation by 8%, use of fertilizer by 16%, and total input value by 15%. As a consequence, net revenue increased by 13%. Similar results were found for borrowers in the 88 villages that received microcredit (counter to well-publicized negative results summarized by Banerjee et al 2015). But farmers who chose not to borrow but who had access to microcredit had essentially zero returns to capital at the margin. Thus, comparing the returns of borrowers to non-borrowers—without accounting for the endogeneity of borrowing and the heterogeneity of returns—would greatly overstate the net returns to microcredit access.

The lack of plausible instrumental variables tends to be greater in micro-economic studies of development because market failures drive interlinkages between household choices and between markets, especially in informal settings (e.g., Stiglitz 1986, Bardhan 1984). It is thus harder to find plausible excludable variables because, without complete markets, more elements of the economy prove to be endogenous. Empiricists working on the canonical agricultural household model (Singh et al 1986), for example, have exploited a recursive property that justified analysis of production independent of consumption variables, but the reverse does not hold, effectively ruling out the use of any production variables as instruments when analyzing consumption choices of producer-consumer households (including farmers and small-scale entrepreneurs). And even the recursiveness property depends on strong assumptions about the completeness of markets, including insurance markets.

The result is many good ideas but far fewer convincing ways to challenge and test the ideas – even when it is possible to observe naturally-occurring variation in the economic environment. Moreover, using instrumental variables often leads to situations in which instruments may not be fully convincing but nonetheless estimated parameters are substantially affected by IV estimation.[8] The Local Average Treatment Effects (LATE) framework helps to explain why: with heterogeneous treatment effects, OLS and IV essentially estimate different parameters (Imbens and Angrist 1994, Imbens 2009). Like the understanding of weak instruments (Staiger and Stock 1997), the LATE framework challenged what can be learned from instrumental variables strategies. It became clear that differences between OLS and IV estimates could not be assumed to result solely (or even mostly) from removing bias (a natural interpretation only under the assumption of homogeneous treatment effects). Instead, IV generates particular parameters that are specific to the interaction of the instrument and the endogenous variable when treatment effects are heterogeneous (Heckman and Urzua 2010). While RCTs also generate parameters that are local and specific, their interpretation can be read through the experimental design. They thus offer a mode of interpretation that is often clearer than in a typical LATE from an IV regression, especially one that does not draw on a natural experiment and one with multiple continuously-defined instruments (Samii 2016).

Three examples

To illustrate RCTs for exploration, I describe three examples of experiments with contracts, prices, and access to financial markets and products:

Microcredit contracts

The typical microcredit contract takes an unexpected form for a business loan. Although described as a loan for investment in small-scale enterprise, loans look more like consumer loans, with contracts that require repayment in regular installments starting shortly after the loan has been disbursed. In loans from Grameen Bank, for example, the installments are weekly and start the week after disbursement. In effect, the loan size is diminished since part of the loan must be returned to the lender nearly immediately.

---

[8] This is the case, for example, with Pitt and Khandker (1998), a well-known non-randomized evaluation of microcredit in Bangladesh that relied on the assumption of treatment effect homogeneity and the use of particular functional forms for identification, and which ultimately proved not to be robust even on its own terms. For a critical discussion, see Roodman and Morduch 2015.

This structure, however, helps minimize the size of installments and has been promoted as a way to maintain high loan repayment rates (Armendàriz and Morduch 2010).

Might this structure, though, discourage investment and reduce profits for customers (and, possibly, local economic growth)? Might borrowers do better if they had more time to invest before repayments start? Field et al (2013) designed an RCT to test that proposition, asking whether the "classic" microfinance contract inhibits investment in high-return business opportunities? They worked with an NGO that served with women in low-income neighborhoods of Kolkata, India. Each client received an individual-liability loan varying in from Rs. 4,000 ($90) to Rs. 10,000 ($225), with a modal loan amount of Rs. 8,000.

After group formation and loan approval (but prior to loan disbursement), groups were randomized into two contracts. In the control group, 85 groups were assigned to the regular debt contract with repayment in fixed installments starting 2 weeks after loan disbursement. In the treatment group, 84 groups were assigned to a contract that included a grace period of two months. Other features of the loan contract were held constant. The total interest paid was identical, and once repayment began, all groups repaid every 2 weeks over 44 weeks, at a group meeting.[9]

Three years later, the new contract looked like a success: borrowers in the treatment group had 57% higher profit levels on average. They were also using 81% more capital and taking greater risks as they invested more. The problem, from the lender's vantage, however, was that repayment problems increased by three times: 52 weeks after the loan should have been fully repaid, 6% in the treatment group had not fully repaid compared to under 2% of the control group. The repayment problems were large enough that the contract was not profitable at feasible interest rates.

The study is not an impact study that asks: "does it work?" Instead, it investigates the nature of contracts and constraints, comparing one kind of contract against another. In the course of the study, a measure of returns to capital could be estimated (11-13% per month), suggesting that access to more capital would be welfare-enhancing, but that was not the main aim of the study. Instead, the RCT helps to get at a persistent question:

[9] Because the treatment group had loans with a longer debt maturity (55 as opposed to 44 weeks before the full loan amount was due), they faced a slightly lower effective interest rate on the loan.

why do the measured impact of microcredit appear to be so modest (Banerjee et al. 2015)? Are lending methods part of the problem? Can they be improved?

The RCT departs from market surveys by testing a real product rather than asking about preferences over hypothetical scenarios. A market survey might reveal a preference for delayed repayments, but would likely say little about the consequences for investment, business outcomes, and loan repayments. A trial could be run without an RCT, of course, but coupling with an RCT is a natural way to clarify comparisons.

Microcredit interest rates

The example above focuses on microcredit contracts, but the most important microcredit innovation was likely the choice to raise interest rates. It was not an obvious move. State-run banks had been created explicitly to provide subsidized credit in poor areas because it was thought that customers could not pay high interest rates. But early leaders in the field felt pressure to cover their basic costs, so interest revenue was imperative. With little more than casual evidence, micro-lenders reasoned that poor households seemed to borrow regularly from moneylenders who charged 5% or 10% per month, so charging 20% or 30% per year did not seem prohibitive. Using the logic of diminishing marginal returns to capital, micro-lenders also reasoned that capital-starved entrepreneurs could have high returns to their first increments of capital (Armendàriz and Morduch 2010).

The mantra soon became: "poor households need access to credit, not cheap credit." Implicit in this conclusion was the assumption that the elasticity of loan demand with respect to interest rates was effectively zero (Morduch 2000). Accordingly, interest rates were raised. Cull et al (2018), for example, show that for a sample of 1330 microfinance institutions between 2005 and 2009, average inflation-adjusted microfinance interest rates were 25 percent per year (21 percent at the median). These interest rates allowed microfinance institutions to reduce dependence on subsidy, although only about a quarter were truly free of subsidy.

Lenders assured themselves that there were limited tradeoffs with outreach. But the essential assumption – that the demand elasticity with respect to interest rates was zero – was untested and largely untestable. With available data, the challenges were: (1) lending institutions seldom changed interest rates, so there was little to analyze; (2)

while different lenders charged different interest rates, so much else differed between institutions that separating out the causal impact of interest rates by comparing borrowing levels across institutions was a non-starter; (3) even when interest rates varied within institutions, the differences were almost always tied to different products serving different kinds of customers; again variation was hard to exploit. (For their part, market surveys always indicate that borrowers want cheaper credit, but it's not clear how strong borrowers' sensitivities are.)

Dehejia et al (2012) made a first attempt to estimate the elasticity of loan demand in a non-randomized difference-in-difference framework, exploiting a quasi-experiment (not an RCT). *Safe*Save, a lender in the slums of Dhaka, had charged its customers 2% per month for loans, but they felt that rates had to be increased to 3% to cover costs.[10] So when new branches opened, *Safe*Save charged 3% there. Eventually, the older branches were brought into conformity with the new branches, giving a chance to see changes in loan demand as interest rates were increased from 2% to 3% per month in the older branches. Loan demand in newer branches could then be used to control for macro shocks and broader conditions in a difference-in-difference framework. The situation was unusual in that prices were raised in some branches but not others, keeping all else the same.

Counter to the assumption that the elasticity would be zero, Dehejia et al (2012) estimate a long-term elasticity over -1.0. In other words, raising the interest rate by 10% led to a greater than 10% drop in demand. Rather than greatly expanding revenues, the interest rate hike slightly undercut net revenues and reduced borrowing. The study directly contradicted expectations – and an important pillar of microfinance thinking. In the first rigorous test, customers were shown to care about interest rates, and they borrowed less as prices rose.

The study rested on strong assumptions. Most important was the assumption that the timing of the move from 2% to 3% was effectively random: that it was independent of demand patterns in the branch. The case relied solely on the recollection of the lender's chairperson. A case also had to be made for comparability across

[10] Disclosure: At the time, I was a member of the *Safe*Save cooperative, effectively serving as a member of its governing board. The institution is now part of the NGO BRAC.

branches in order to interpret the difference-in-difference, relying on a demonstration of the similarity of pre-change trends. In addition, the result was based on evidence on the choices of 5147 members of a particular institution in just one set of branches in the Dhaka slums, and it was not clearly generalizable.

Still, the result mattered because it was plausible and so sharply countered expectations of practitioners (if not of economists, who take it on faith that most often demand curves slope downward). The study laid out an argument that poor households, particularly the poorest, *did* take price increases into account—and reduced loan demand accordingly.

A broader case was provided by an RCT. Karlan and Zinman (2018) describe a similarly-motivated study of Banco Compartamos in Mexico. The bank is the largest lender in Latin America, serving millions of borrowers, rather than the thousands served by *Safe*Save. Compartamos is known as one of the most commercially-focused micro-lenders, charging interest rates around 100 percent per year (Rosenberg 2009). The bank wanted to reduce interest rates, and Karlan and Zinman (2018) saw the chance to estimate interest rate elasticities by convincing Compartamos to reduce interest rates to different levels in different places, creating randomized treatment and control groups in the process (just as the *Safe*Save study also needed heterogeneity across branches).

Randomization at Compartamos proceeded at the branch level, covering branches spread across Mexico. Forty regions were randomly assigned to a "high rate" group: their loans cost about 10 percentage points below existing interest rates. Another forty regions were randomly assigned to a "low rate" group with loans costing about 20 percentage points below existing interest rates. The study assessed elasticities by comparing loan demand across branches.

As with Dehejia et al (2012), borrowers were shown to be sensitive to interest rates. Karlan and Zinman (2018) estimated an interest rate elasticity after the first year of -1.1. By Year 3 it was -2.9. Moreover, as with Dehejia et al (2012), the move did not obviously help profits. After the price change, Compartamos had more borrowers but more costs too.

Had the researchers not intervened, Compartamos would likely have reduced interest rates everywhere at the same time, leaving no control group. And if Compartamos had instead deliberately chosen some branches as first-movers, the risk of selection bias would have arisen. The RCT thus created analytically useful variation. The use of randomization by Karlan and Zinman eliminated the challenge in comparing behavior across branches. It also eliminated the concern that the choice to reduce interest rates (and by how much) in any particular branch was driven by local conditions. Rather than yielding plausible estimates that rely on a chain of assumptions (as in Dehejia 2012), the RCT parameters estimated by Karlan and Zinman (2012) are transparent and tightly measured.

On the other hand, Compartamos is unusual: the baseline interest rate was very high, and the policy change involved reducing interest rates rather than increasing interest rates. As with the result from *Safe*Save, the estimates are not directly exportable to other settings. Nevertheless, the two studies together can shift priors in a Bayesian sense, and, the experimental mindset behind the Compartamos RCT allows us to see something that would have otherwise been hard to see.[11]

Poverty, migration, and mobile money

Technology is transforming the financial landscape, taking focus away from traditional microcredit, but use of technologies like mobile money (using telephones to make payments and maintain digital wallets) is highly self-selected. The choice to adopt new technology is reinforced by policies by providers to focus on the most lucrative parts of markets. Most often that means that poor households are disproportionately excluded. The corollary is that the poor households that adopt tend to be unusual. How then to assess the possibilities for technology in poor communities?

The rural population of Bangladesh has been steadily drawn to Dhaka, largely driven by the hope of employment in the ready-made garment industry. Factories, large and small, are following China's lead and exporting globally. The jobs are often filled by younger workers who support families in the countryside. This dynamic is in the spirit

[11] The RCT doesn't answer all questions. It appears that much of the increase in lending was due to new borrowing (not substitution from other sources), but there remain questions about impacts on well-being and risks of over-indebtedness. Moreover, the RCT is limited in what it can reveal about context and heterogeneity. The results also say nothing about the ethical questions surrounding charging relatively high interest rates to poor borrowers (Rosenberg 2009).

of the Lewis (1954) model of rural-urban migration and economic growth, and Bangladesh has been growing at about 6-7% per year. But where does this leave households remaining in rural areas? One question is whether technology can help migrant workers in Dhaka send money back to their families? Can the technology lead to increases in levels of remittances from urban migrants to rural families? Can it be a mechanism to reduce poverty and spatial inequality?

Lee et al (2018) use an RCT based on an encouragement design to study how access to mobile banking changes lives in very poor communities in Bangladesh. We started with a sample of households in the rural northwest that were determined to be "ultra-poor", a group that suffers especially during the *monga* (lean) season. The households had participated in a program with a local NGO that helped their adult children move to Dhaka factories. The study follows both sides of the remittance equation, senders and receivers. In Dhaka, we followed urban migrants originally from the northwest. In the northwest, we followed their extended families. In the control group, just 11% had bank accounts and 20% were actively using mobile money.

One reason for low initial adoption of the technology was the hurdle created by English-language menus on the telephone interface used by the mobile money providers. The main experimental intervention, designed by the researchers, involved training randomly-assigned groups in both urban and rural settings about how to use the technology. Participants were given hands-on experience with sending remittances, received translated menus, and got assistance with account sign-ups. (The training cost about $12 per household.) The control group received neither training nor help.

The first result was the finding of a large increase in active mobile money usage, from about 20% in the control group to 70% in the treatment group. Remittances from urban migrants back to their rural families increased by 30% relative to the control group. That flow of money led to a drop in extreme poverty in the rural area. Average consumption increased by 7% on average relative to controls, and gains were particularly notable during the lean season. Migrants, on the other hand, were more likely to report diminished physical and emotional health, consistent with pressures to work longer hours and increase remittances enabled by the mobile banking technology.

The experiment behind the RCT reduced barriers to entry for particularly excluded groups. That might have happened eventually without an RCT, but the experimental intervention allowed a clear comparison to a similar control group at a historical moment when causal inference was possible. By centering on the migration-remittance relationship, the study presents an alternative path to improving rural conditions. Standard responses are to bring resources into rural areas through microcredit and "graduation" programs which aim to raise productivity in rural areas (see the RCT by Bandiera et al 2017). Here, instead, the mechanism involves helping rural workers find more remunerative employment in cities – and then facilitating a mechanism to move resources from the city to rural areas.

While this might seem to be a "what works" evaluation, the study is better seen as an inquiry into spatial inequality and whether intra-household sharing is limited by costs. The point of the study is not to show that illiterate Bangladeshis are deterred by English-language menus required for operating mobile money accounts. That is not a surprise, and would not have been worth studying so intensively. Instead, the point was to use that hurdle (and a training program to overcome it) as a way to induce variation in who uses mobile money and who does not. In other words, the hurdle was the key to forming a treatment and control group (through an "encouragement design") that allowed the mapping of the consequences of access to mobile money for migrants and their families. In the end, the study does not promote a particular solution so much as contribute to understanding the channels of exit from rural poverty.

Market failure and private goods

RCTs by nature are particularly useful in studying discrete interventions. They are particularly well-suited for inquiries around the delivery of private goods. The examples above are in that line. By the same token, RCTs are far weaker in assessing the role of public goods and macro change (Hammer 2017).

Some criticize RCTs for pushing the focus of development economics toward the provision of private goods, but this orientation within development economics and development policy emerged decades before RCTs came to the fore. The 1970s saw a fundamental shift in development economics toward concern with the provision of private goods. This came in the context of a broader shift toward concern with rural development, absolute levels of poverty, undernutrition, high mortality rates, and low educational attainments. The shift can be seen in the "basic needs" literature and criticisms of growth-based development (e.g., Chenery et al 1979), the re-orientation of the World Bank under Robert McNamara, the rise of information economics within development economics (Stiglitz and Weiss 1981, Stiglitz 1986, Bardhan 1984), and a focus on "merit goods" (Musgrave 2008).[12] United Nations Millennium Development Goals and the Sustainable Development Goals—with their focus on poverty, health, education, and basic rights—reinforce the focus. One reason that RCTs took hold is because they are particularly well-matched for inquiries about the delivery of key goods and services.

As Rodrik (2009) and Ravallion (2012), note, this puts the focus on fairly small interventions, not on the larger macro changes that drive poverty, inequality, and economic growth. Restricting attention to interventions that can be studied with RCTs, critics argue, impedes attempts to bring systemic reform in places where systems are badly broken, distorted, and unfair. To put it too sharply, RCTs are particularly good for studying the impact of band-aids, and as a result we will have many studies of band-aids. RCTs are also particularly good at investigating delivery mechanisms ("last mile problems") rather than large, sectoral policy priorities (first mile problems?). Instead, critics argue, we need to tackle the structural inequalities, environmental conditions, political imbalances, and weak infrastructure that generate and reproduce the harms that band-aids can only cover up.

The critics make a fundamentally important point, and perhaps it is well to stop there. But stopping there risks ignoring a broader history, a deeper conflict, and important, unanswered questions about the roles of band-aids and delivery mechanisms, knowledge, and progress.

First, this framing makes explicit that what often takes the form of technical, statistical debates about the appropriate methods to ensure internal validity and external validity is instead most fruitfully recognized as part of a political debate about the scope and nature of intervention. The technical debates can be resolved on their own terms—and are being resolved on their own terms through statistical innovation and

[12] Schooling is included here as a private good because, unlike typical public goods, schooling is largely "rival" and "excludable". Since there are clear externalities for the larger community, schooling is perhaps best thought of as a merit good.

improved research designs—but that cannot resolve the more fundamental political tensions about the scope of intervention.

Second, the theoretical argument for systemic reform is compelling. The massive reductions in global poverty in recent decades, for example, have resulted from broad, systemic change, especially in Asia (Ravallion 2012). Yet, systemic change is not always possible, and sometimes leaves parts of populations behind. Broadening access and service delivery, and expanding the provision of basic goods, remains a fundamental agenda for governments, aid agencies and foundations.

One might reasonably argue that development economics *should* be much more focused on context and on public goods (Hammer 2017), macro interventions, and other kinds of policy, but it is misleading to argue that RCTs are at the root of perceived imbalances. The political economy and history run deeper, and there continue to be justifiable reasons to focus on improving the delivery of private goods and services (even absent RCTs). The RCT results will not spur revolutions, but they can, cumulatively, create necessary steps to better outcomes.

Conclusion

Debates on RCTs are often unsatisfactory. They fail to distinguish between types of RCTs and types of questions. Much of the criticism of RCTs is compelling both on philosophical and technical grounds, and critics rightly argue that RCTs are not a uniquely valuable source of credible impact evaluations. Other methods are useful too, and sometimes superior. We need more description, more qualitative data, more big data, more studies with other empirical strategies.

At the same time, however, the terms of debate fail to emphasize what is truly innovative and exciting about RCTs. First, all imperfect approaches are not equally imperfect. Adding new tools like RCTs broadens the scope of methodological possibilities. Second, often the setting needs to be shaken up in order to see something.

Randomistas emphasize the role of RCTs in determining what works and what does not. I have instead focused on those RCTs that pull economic structures apart. The difference between the two kinds of RCTs above – RCTs for evaluation versus RCTs for exploration – is the difference between studying what exists versus tinkering and rethinking to create different possibilities to study, to push further in exposing theory to reality. Coupled with an experimental mindset, these RCTs create exogenous variation that gives a new way of seeing how important markets, institutions, and processes work.

References

Angrist, Joshua and Alan Krueger. 1999. "Empirical Strategies in Labor Economics." In Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3. Amsterdam: North-Holland.

Armendàriz, Beatriz and Jonathan Morduch. 2010. *The Economics of Microfinance*, Second edition. Cambridge, MA: MIT Press.

Bandiera, Oriana, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. 2017. "Labor Markets and Poverty in Village Economies." *Quarterly Journal of Economics* 132 (2): 811-870.

Banerjee, Abhijit and Esther Duflo. 2009. "The Experimental Approach to Development Economics." *Annual Review of Economics* 1:151–78.

Banerjee, Abhijit, Dean Karlan, and Jonathan Zinman. 2015c. "Six Randomized Evaluations of Microcredit: Introduction and Further Steps." *American Economic Journal: Applied Economics* 7(1): 1–21.

Bardhan, Pranab. 1984. *Land, Labor, and Rural Poverty: Essays in Development Economics*. New York, NY: Columbia University Press.

Bates, Mary Ann and Rachel Glennerster. 2017. "The Generalizability Puzzle." *Stanford Social Innovation Review*, Summer.

Bauchet, Jonathan and Jonathan Morduch (2019). "Paying in Pieces: A natural experiment on demand for life insurance under different payment schemes." *Journal of Development Economics*.

Beaman, Lori, Dean Karlan, Bram Thuysbaert, and Christopher Udry. 2018. "Selection into Credit Markets: Evidence from Agriculture in Mali." Working paper.

Bédécarrats, Florent, Isabelle Guérin, and François Roubaud. 2017. "All that Glitters is not Gold. The Political Economy of Randomized Evaluations in Development." *Development and Change*.

Bisbee, James, Rajeev Dehejia, Cristian Pop-Eleches, and Cyrus Samii. 2017. "Local Instruments, Global Extrapolation: External Validity of the Labor Supply–Fertility Local Average Treatment Effect," *Journal of Labor Economics* 35 (S1): S99-S147.

Belissa, Temesgen, Erwin Bulte, Francesco Cecchi, Shubhashis Gangopadhyay, and Robert Lensink. 2019. "Liquidity constraints, informal institutions, and the adoption of weather insurance: A randomized controlled Trial in Ethiopia." *Journal of Development Economics* 140: 269-278.

Cartwright, Nancy and Jeremy Hardie. 2012. *Evidence-Based Policy: A Practical Guide to Doing it Better*. Oxford, UK: Oxford University Press.

Casaburi, L., & Willis, J. (2018). Time versus State in Insurance: Experimental Evidence from Contract Farming in Kenya. *American Economic Review, 108*(12), 3778-3813. doi:10.1257/aer.20171526

Chenery, Hollis; Ahluwalia, M.; Bell, C.; Duloy, J.; Jolly, R. 1979. *Redistribution with Growth*. New York, NY: Oxford University Press.

Cull, Robert and Jonathan Morduch. 2018. "Microfinance and Economic Development." In Thorsten Beck and Ross Levine, eds. *Handbook of Finance and Development*. Cheltenham, UK: Edward Elgar.

Cull, R., Demirgüç-Kunt, A. and J. Morduch. 2018. "The Microfinance Business Model: Enduring Subsidy and Modest Profit." *The World Bank Economic Review*, Volume 32, Issue 2, 1 June 2018, Pages 221–244.

Davis, Douglas and Charles A. Holt. 1993. *Experimental Economics*. Princeton, NJ: Princeton University Press.

Deaton, Angus. 2010. "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature* 48: 424-455.

Deaton, Angus and Nancy Cartwright. 2018. "Understanding and misunderstanding randomized controlled trials." *Social Science & Medicine* 210: 2-21. https://doi.org/10.1016/j.socscimed.2017.12.005.

Dehejia, R. Morduch, R. and H. Montgomery (2012) "Do interest rates matter? Credit demand in the Dhaka Slums". *Journal of Development Economics* 47 (2), 437 - 499.

Drèze, Jean. 2018. "Evidence, policy and politics: A commentary on Deaton and Cartwright." *Social Science & Medicine* 210: 45-47.

Duncan, Greg, Aletha Huston and Thomas Weisner. 2007. *Higher Ground: New Hope for the Working Poor and Their Children*. New York: Russell Sage.

Field, Erica, Rohini Pande, John Papp, and Natalia Rigol. 2013. "Does the Classic Microfinance Model Discourage Entrepreneurship among the Poor? Experimental Evidence from India." *American Economic Review*, 103(6): 2196-2226.

Glennerster, Rachel and Kudzai Takavarasha. 2013. *Running Randomized Evaluations: A Practical Guide*. Princeton, NJ: Princeton University Press.

Gueron, Judith. 2017. "Chapter 2 - The Politics and Practice of Social Experiments: Seeds of a Revolution." Chapter 2 in Banerjee, Abhijeet and Esther Duflo. The Handbook of Economic Field Experiments, vol. 1: 27-69.

Hammer, Jeffrey. 2017. "Randomized control trials for development? Three problems." Brookings Institution Future Development Blog. May 11, 2017.

Heckman, James and Sergio Urzua. 2010. "Comparing IV with Structural Models: What Simple IV Can and Cannot Identify." *Journal of Econometrics* 156 (1): 27-37.

Imbens, Guido. 2018. "Understanding and misunderstanding randomized controlled trials: A commentary on Deaton and Cartwright." *Social Science & Medicine* 210: 50-52.

Imbens, Guido and Joshua Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2): 467-475.

Imbens, Guido. 2010. "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)." *Journal of Economic Literature* 48: 399-423.

Karlan, Dean and Jonathan Zinman. 2018. "Long-Run Price Elasticities of Demand for Credit: Evidence from a Countrywide Field Experiment in Mexico." *Review of Economic Studies*, forthcoming 2019.

Kasy, Maximilian and Anja Sautmann. 2019. "Adaptive Treatment Assignment in Experiments for Policy Choice." Preliminary draft, June 2, 2019, Harvard and MIT.

Kremer, Michael. 2003. "Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons." *American Economic Review (Papers and Proceedings)* 93 (2), May: 102-106.

LaLonde, Robert. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *The American Economic Review* 76 (4): 604-620.

Lee, Jean, Jonathan Morduch, Saravana Ravindran, Abu Shonchoy, and Hassan Zaman. 2018. "Poverty and Migration in the Digital Age: Experimental Evidence on Mobile Banking in Bangladesh," NYU working paper.

Levy, Santiago. 2006. *Progress against Poverty: Sustaining Mexico's Progresa-Oportunidades Program*. Washington, DC: Brookings Institution.

Lewis, W. Arthur. 1954. "Economic development with unlimited supplies of labor", *Manchester School*, 22, 139-91.

Morduch, Jonathan. 2000. “The Microfinance Schism,” *World Development* 28 (4): 617 - 629.

Morvant-Roux, S., I. Guérin, M. Roesch, J.-Y. Moisseron. 2014. “Adding value to randomization with qualitative analysis: The case of microcredit in rural Morocco.” *World Development* 56: 302–312.

Musgrave, Richard. 2008. “Merit Goods.” In S. N. Durlauf & L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics: Volume 1 – 8* (pp. 4173-4176). London: Palgrave Macmillan UK.

Ogden, Timothy. 2017. *Experimental Conversations: Perspectives on Randomized Trials in Development Economics*. Cambridge, MA: MIT Press.

Pritchett, Lant. 2014. “An Homage to the Randomistas on the Occasion of the J-PAL 10th Anniversary: Development as a Faith-Based Activity.” Center for Global Development, Washington, DC. https://www.cgdev.org/blog/homage-randomistas-occasion-j-pal-10th-anniversary-development-faith-based-activity

Pritchett, Lant and Justin Sandefur. 2015. “Learning from Experiments When Context Matters.” *American Economic Review: Papers and Proceedings* 105(5): 471-75.

Rai, Ashok and Sjöström, Tomas. 2004. “Is Grameen Lending Efficient? Repayment Incentives and Insurance in Village Economies.” *Review of Economic Studies* 71(1): 217–234.

Ravallion, Martin. 2012. “Fighting Poverty One Experiment at a Time: *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*: Review Essay.” *Journal of Economic Literature* 50 (1): 103-14.

Ravallion, Martin. 2018. “Should the Randomistas (Continue to) Rule?” Center for Global Development working paper 492, Washington, DC.

Roodman, David and Jonathan Morduch. 2014. “The Impact of Microcredit on the Poor in Bangladesh: Revisiting the Evidence.” *Journal of Development Studies* 50 (4): 583-604.

Rosenberg, R. 2009. “The New Moneylenders: Are the Poor Being Exploited by High Microcredit Interest Rates?” CGAP Occasional Paper 15. Washington, DC: Consultative Group to Assist the Poor.

Ruhm, Christopher. 2018. “Shackling the Identification Police?” NBER Working Paper No. 25320, November.

Rodrik, Dani. 2009. “The New Development Economics: We Shall Experiment, but How Shall We Learn?” In *What Works in Development? Thinking Big and Thinking*

Small. Ed. William Easterly and Jessica Cohen. Brookings Institution Press: 24-48.

Singh, Inderjit; Squire, Lyn; Strauss, John, eds., 1986. *Agricultural household models: Extensions, applications, and policy*. Baltimore, MD: The Johns HopkinsUniversity Press.

Staiger, Douglas and James Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3): 557-586.

Stiglitz, Joseph. 1986. "The New Development Economics." *World Development* 14 (2): 257-265.

Stiglitz, Joseph and Andrew Weiss. 1981. "Credit Rationing in Markets with Imperfect Information." *American Economic Review* 71: 393–410.

Wydick, Bruce. 2016. "Microfinance on the Margin: Why Recent Impact Studies May Understate Average Treatment Effects." *Journal of Development Effectiveness* 8(2): 257-265.