Chapter 4:

What can we learn from impact assessments?

Jonathan Bauchet

Aparna Dalal

Jonathan Morduch

http://www.microinsurancenetwork.org/sites/default/files/194603_MIN_BRO_CHAPTER_4_final.pdf

2014

Chapter 4 in *Practical Guide to Impact Assessments of Microinsurance*.
Edited by Ralf Radermacher and Katja Roth

Microinsurance Network

http://www.microinsurancenetwork.org/groups/practical-guide-impact-assessments-microinsurance

What can we learn from impact assessments?

4.1. Introduction

How can we determine that an intervention is making a real difference?

At age 40, Feizal was supporting his family in rural northern India. He earned a living selling aluminium pots, which he strapped on to his bicycle and took from village to village. The sales provided the lion's share of his household's $36 average monthly income. But one day Feizal had a bad fall from his bicycle and broke his leg. Initially, he relied on the care given by traditional doctors – at a cost of $33. After three months, the leg showed no improvement, but Feizal's family could not afford modern treatments. It took several more months, and the resources of his extended family, to pull together $250 to pay for a hospital treatment. The family had to draw on a wage advance from Feizal's son's employer and deplete the family's savings, which had been reserved for Feizal's daughter's wedding. In the end, Feizal's leg was treated in a modern hospital and he recovered. But he had spent eight months with no income, his family's savings were gone, and the family was $100 in debt.[1]

What would Feizal's situation have been if he had access to health insurance? Would Feizal have gone to a modern doctor sooner, thereby receiving better treatment and minimizing recovery time? Could proceeds from an insurance policy have helped him avoid falling into debt? Would Feizal have been able to protect his family's consumption levels? These questions are at the heart of impact evaluations.

With certain assumptions, evaluators can establish that the difference between Feizal's situation and that of insured individuals was *caused by* having a microinsurance policy, i.e., it is the impact of microinsurance. The rough notion of "making a difference" can be translated into a

[1] The story of Feizal and his family is part of the financial diaries project collected by Orlanda Ruthven and described in *Portfolios of the Poor* (Collins et al. 2009), in which authors provide a window into the financial lives of 250 households in India, Bangladesh, and South Africa.

precise question that should be at the heart of every impact study: “How have outcomes changed with the intervention *relative to what would have occurred without the intervention*?”[2]

Yet, even if that question could be answered in Feizal’s community in northern India, the particular answer would likely be different elsewhere—and those other places might be of even greater interest to policymakers or investors. The question about what we can learn from one place to understand another place – in technical jargon, the problem is “external validity” – deserves much more attention, and we return to it at the end. But since we can’t learn from *any* place unless a study is credible and free from important biases, we devote most of our attention to the problem known as “internal validity”: Are we *actually* measuring what we want to?

4.2. What quantitative impact assessments measure

Impact evaluations try to measure and understand the change in a participant’s life that occurred because of an intervention. The “intervention” could be a policy, a project, an insurance product, or a specific feature of a product. For instance, the intervention could relate to a particular product feature, such as the extent of coverage, a change of pricing structure, or variations in the distribution channel.

Impact evaluations focus on the *outcomes* and impacts of the intervention. The focus on outcomes and impacts, rather than inputs and outputs, distinguishes impact evaluations from “process evaluations.” Process evaluations measure program processes, inputs and outputs. They answer questions like: How many insurance education sessions did trainers conduct? How many farmers attended the sessions? How many households purchased a given insurance policy? These indicators are a measure of the efficiency of the intervention. While they are useful in estimating the outreach of the program, they give little information about how the program affected household well-being.

[2] This approach to impact evaluation, based on a comparison with hypothetical outcomes, is often called the Rubin Causal Model after its originator, Donald Rubin, a statistician at Harvard. See Rubin (1974) for the very origins of this model.

Impact evaluations, on the other hand, look to answer questions like: Did farmers invest in high-grade seeds because they had insurance coverage? Did the change in investment result in higher income? Did insurance make a notable difference in coping with a drought?[3] Sometimes the likely answers seem obvious, but well-designed evaluations have the power to surprise. During droughts, for example, families often get help from neighbours or relatives, and sometimes from the state. Families borrow, draw from their own savings, and many even migrate. The impact of insurance might be large for some families but not for others. Impact evaluations are critical in quantifying the intervention's true value and understanding the underlying mechanisms.

This chapter is mostly focused on quantitative impact evaluations, estimating the amount of change caused by an intervention for a population of interest. Qualitative impact evaluations are also used in some settings and are particularly useful for gaining insight into *how* interventions generate impacts. They proceed from a different logic, however, and it is beyond the scope of this chapter to explain them in detail.

4.2. The greatest challenge of quantitative impact evaluations: addressing selection bias

The ultimate goal of quantitative impact evaluations, and their greatest challenge, is to establish credibly that the intervention *caused* a difference in the lives of the participants.[4]

The challenge is to separate the change that was caused by the intervention from the change that would have happened anyway without the intervention. In other words, how can the evaluator establish that the outcomes have been caused by the intervention, and not by other concurrent events, underlying trends, or characteristics of the participants? For instance, evidence shows that richer and more educated households are more likely to sign up for health microinsurance (Giné and Yang 2007). In other words, these households *selected* themselves into this intervention. If insured households happen to have good health outcomes, is it due to the insurance itself, or to their capacity to afford better hospitals and better understand

---

[3] These questions illustrate individual- or household-level outcomes. Possible indicators for such outcomes include income, asset ownership, nutrition, education levels, health status, or the cost of medical treatment. Microinsurance impact evaluations can also focus on institution-level outcomes.
[4] Impact evaluations can be improved by exploring how differences were caused, sometimes by adopting a mixed method approach and conducting qualitative research in parallel to the quantitative evaluation.

doctors' recommendations, even without health insurance? These characteristics can have a significant effect on impact estimates. In a study of microfinance clients, for example, McKernan (2002) found that not isolating the effect of microcredit from other concurrent effects can lead researchers to overestimate the impact by 100 percent.

When measuring impact in microinsurance, not separating the impact of the intervention from that of other confounding factors could lead to underestimating or overestimating the impact of the insurance, depending on the situation. It is possible, for instance, that households that suffer from a preexisting illness are more likely to buy health insurance - the classic "adverse selection" problem.[5] In this case, an impact evaluation that compares health outcomes of these households with another set of households could find a (mistakenly) negative impact of microinsurance on health condition since the insured individuals appear to be in worse health than uninsured patients. On the other hand, richer farmers are more likely to have enough disposable income to contract a rainfall microinsurance product, so comparing these farmers to those who did not sign up for insurance will make it seem like the insurance was successful at helping them deal with a drought, when part of the measured impact is, in fact, due to the better initial situation of the insured farmers.

This potential bias in the results is called *selection bias*. Disentangling the influence of individuals' characteristics from that of the intervention (i.e., addressing the selection bias) is surprisingly difficult to do. Some individual characteristics can be observed, measured, and controlled in a statistical analysis. For instance, gender, age, and residence location are likely to influence both the decision to contract insurance and the outcome from having insurance. Most of these kinds of factors are easy to measure and their influence on the outcomes can be separated out by statistical means.

The big challenge arises with unobservable factors. Attributes like an individual's propensity to fall sick, organisational ability, or access to social networks, are far harder – and others are impossible – to measure. But they can create big biases. Not all hope is lost: certain evaluation

[5] For further discussion of the adverse selection problem, see Churchill (2006) in the context of microinsurance or Armendáriz and Morduch (2010) in the context of microfinance generally.

methodologies make it possible to recover the net impact of the intervention, free of selection bias. We highlight their principles below in the section titled "Getting credible answers."

Impact evaluations of microinsurance present a specific challenge. While the most fundamental benefit of insurance is that it offers households protection when facing emergencies, having access to insurance can also provide important benefits in the *absence* of adverse events. Economic theory suggests, for example, that having insurance can allow farmers to take more risk, altering their investment and employment strategies, which could have an impact on their well-being. Impact evaluations of microinsurance that capture the impact of these altered strategies can help us understand the full effect of the access to insurance. This is not a methodological consideration, but is a practical challenge when designing microinsurance impact evaluations.

4.3. Getting credible answers ("internal validity")

4.3.1. Control groups are essential

While it might seem that researchers would spend most of their time trying to capture what happens when people have insurance, they end up spending even more time trying to capture what happens when they *do not* have insurance. This is the "counterfactual," and it is the key to credible evaluations. The question is: What would have happened to the participants had they not received the intervention?

Unfortunately, we cannot ever know what would have actually happened to an insured individual had she not had access to insurance because people can only be in one circumstance at a time. But with the right design, an impact evaluation can form a credible estimate of the counterfactual for a group of participants taken together.

The counterfactual is usually estimated by measuring impacts for individuals who do not participate in the intervention, but who are similar to those who do, in as many respects as possible. The group of individuals who participate in the intervention is commonly referred to as the “treatment group,” and the group of those not participating is referred to as the “control group”(Shadish et al. 2002).[6]

Once treatment and control groups are formed correctly, the quantitative impact of the intervention is measured by comparing outcomes in the treatment and control groups. Statistical techniques are typically used to increase the confidence that the results are not spurious – that is, that they would also be likely to occur in other samples. The difference in outcomes between the two groups is a good measure of the causal impact of the program if, and only if, the groups are truly comparable.

4.3.2. Control groups need to be truly comparable

Having a control group is a necessary, but not a sufficient condition, to eliminate the selection bias. The way that treatment and control groups are constituted is fundamental. While having a control group eliminates the effect of general trends such as national macroeconomic conditions on the measured impact, it does not necessarily eliminate the influence of individuals’ characteristics. In fact, the selection bias will always exist when individuals are allowed to self-select to participate in the intervention. This is because their observed and unobserved characteristics influence both their decision to participate and their outcomes. To use the same example as above, richer households will be more likely to be able to afford insurance premiums as well as to cope with unexpected financial shocks.

The only sure way to eliminate the selection bias is to let an event or rule external to the intervention – an “exogenous” event or rule – determine who participates in the intervention. In this situation, individuals are “assigned” to the treatment and control groups, they do not form the groups themselves.

Two main types of exogenous events have been used by evaluators. First, a national or state policy that changes access to programs. For example, the adoption of a national policy of large-scale school building has been used to estimate the impact of education on wages (Duflo 2001).

[6] The methodology has roots in the medical procedures used to test the effectiveness of drugs.

In this methodology, receiving more education was decided by the Indonesia legislature, and, as such, was not related to any personal characteristic of the students benefitting from the policy. We are not aware of any such study on microinsurance. Second, an exogenous event could be a lottery that decides who receives microinsurance and who cannot. Here again, the assignment to treatment and control groups is not related to characteristics of the households or individuals who are being insured – it is random.[7] This randomised controlled trial methodology has become a gold standard in quantitative impact evaluation (Bauchet and Morduch 2010), and can be used to estimate the impact of microinsurance.

In addition to these types of events, in some cases, exogenous rules can be used to eliminate the influence of participants' characteristics on measured impacts. In regression discontinuity designs, for example, an eligibility rule with a clear cut-off point can be used to create credible inference. Some microfinance institutions in Bangladesh, for example, had a rule that they only served households owning under a half-acre of land. A potential study design is to compare the outcomes of households just below the half-acre cut-off (who thus get access to the treatment) to households just above the cut-off. This method requires additional assumptions. The most critical assumption is that participants ranked just above and just below the cut-off are similar in observable and unobservable characteristics.[8]

4.3.3. Control groups need to not have access to the intervention during the evaluation

While great care must be devoted to creating truly comparable groups, even greater care is often necessary to maintain the separation of the treatment and control groups over the course of the evaluation. Obviously, allowing participants to switch group would reintroduce the selection bias that the initial exogenous assignment aimed to eliminate. However, more subtle threats exist.

---

[7] See Bauchet and Morduch (2010) for a more detailed exposition of the theory.

[8] For more details on random assignments please see the chapter on "Experimental Designs" in this volume. Quasi-experimental and other non-experimental methods, described in the chapter on "Non-experimental designs" in this volume, rely on stronger assumptions to attempt to eliminate selection bias.

Households in the control group, for example, might have opportunities to sign up for microinsurance products, maybe from a competing insurer. These households self-select to purchase insurance, which, in addition to reducing the estimated impact, would reintroduce a selection bias.

Households in the control group might also be acquainted with households in the treatment group, and benefit from their relationships, for example, through help in times of need. This spillover of benefits from the treatment to the control group "contaminates" the assignment. The threat of spillovers can be mitigated by implementing specific designs. The level of assignment is the single most powerful way to address the threat of contamination. Rather than assigning individuals to treatment and control groups, families, households, or entire communities can be assigned to each group. In evaluations of microinsurance, for instance, members of the same family should be assigned together to either treatment or control. At the village level, weather insurance could encourage some farmers to adopt riskier and more productive crops and techniques, which in turn would have positive impacts on the entire community. Recognising this possibility, the evaluator might need to assign entire communities to treatment or control.[9]

4.3.4. Studies need to be big enough to reveal the impact of uncommon events ("power")

Since asking all clients how the insurance affected them is generally too costly, a sample of clients is surveyed and statistical methods are used to determine whether conclusions based on the sample can be generalised to all clients.

How big a sample is needed? This question is particularly important for studies investigating risk, such as microinsurance impact evaluations, because most events are uncommon. To observe the effect that microinsurance has on households' ability to cope with adverse events, a sufficient number of these events need to happen in both the treatment and control groups. The need for big samples also arises from the presence of "noise" in all measurements, due to natural variations in the data and measurement errors. This noise might even be particularly

[9] See Bauchet and Morduch (2010) for more details on the level of assignment and other responses to spillovers.

loud when measuring outcomes and indicators of social processes, such as risk, vulnerability, or income. But with a large enough sample, the impact of "noise" can usually be minimised and the impacts of interventions emerge clearly. If the sample is too small, the noise may mask the intervention's real effects: measured impacts may be positive and large, but conventional measures of statistical significance would not be able to establish that the measured impacts are nothing other than noise. Intuitively, the larger the sample, the more confident one is that findings based on that sample are valid for all clients. But, when data collection is required, large samples can be expensive. Evaluators are always trying to balance sample size with budget constraints.

The statistical concept of "power" refers to the ability to detect the impact of an intervention with statistical methods. Power calculations are used to determine the sample size that is required to detect the program's effect.[10] Statistical power generally improves with larger sample sizes, but it is not as simple as that. The design of the evaluation matters as well. Power calculations are based on the size and variation of the impact, the size of the sample that is used to measure the impact, and the desired level of statistical significance.

The important point is that impact evaluations need to consider sample size issues seriously and carefully to ensure that the study is able to capture the impacts while keeping budgets under control.

This section has emphasised the need to adopt rigorous evaluation designs, based on exogenous assignment to treatment and control groups, to estimate the causal impact of an intervention in an unbiased manner. Achieving a high degree of internal validity is necessary for all impact evaluations, and influences the way findings can be interpreted.

[10] This section focuses on how power calculations are used to determine a sample size, pre-study. Power calculations can also be used post-study to estimate the level of power obtained with a given sample size (see Bauchet and Morduch (2010) and Duflo et al. (2008) for more technical introductions and references).

4.4. Interpreting results and understanding the change

Even in evaluations using the design that best establishes internal validity (i.e., the degree of confidence that impacts are caused by the intervention), interpreting results requires stepping back and critically considering the evidence. At least three broad questions should be asked: What impacts are being measured? How did these impacts come to be? How cost effective is the intervention at producing these impacts? How confidently can the evidence from one evaluation be extrapolated to other contexts?

Questions of what intermediate steps have led to these impacts or what pathways theoretically underlie these changes are important as well. As we will explain in detail below, it is important to understand the underlying theory of change, in particular when programs are planned to be transferred to other contexts. Therefore, researchers are not only interested in quantifying the impact, but also in knowing why and how the impact occurred. To get at these kinds of questions, qualitative or participative studies can help probe the underlying mechanisms.

4.4.1. Interventions are not implemented on a blank slate

Books such as *The Poor and Their Money* (Rutherford 2000) and *Portfolios of the Poor* (Collins et al. 2009) reveal how active the financial lives of poor households are. Households use an array of formal and informal saving, loan, and insurance products and maintain financial relationships with their friends and relatives in order to manage their irregular income, finance large expenditures, and smooth consumption.

Whether evaluating microinsurance as a whole or a specific feature of a microinsurance product, evaluators must carefully define their intervention and place it in a larger context. The impact of introducing a microinsurance product in a new market, for example, is a marginal impact, that is the impact of *adding* the product to the mix of informal mechanisms and formal products already available to households. These include the informal strategies described by Collins et al. (2009), as well as insurance products offered by semiformal organisations such as microfinance institutions, social insurance schemes offered by the government, and interventions that other insurers or their partners might be implementing. In most cases, the counterfactual is not the absence of insurance mechanisms. The new insurance product will most likely supplement rather than replace the strategies previously used by households. Thus, the challenge is to parse the net impact of the new product, and, ideally, document its complementarities and exclusivities with existing strategies. Interpreting impact estimates accurately requires an understanding of the intervention's precise effect.

4.4.2. The intervention needs time to produce impacts, but long-term impacts are more difficult to attribute to the intervention

Most impact evaluations measure the outcomes of the intervention one or two years after it was implemented. These (relatively) short-term impacts might be smaller than, or different from, expected impacts, which often require time and multiple exposures to the intervention to emerge. In microfinance, for example, borrowers may not experience improvements in their business and livelihood until after they complete several loan cycles. Similarly, the impact of microinsurance might appear long after households have contracted their first insurance product: households might not adopt new crops, for example, until they have personally benefitted from rainfall insurance during a drought.

Moreover, interest in the results of the evaluation is often high, particularly when the intervention is popular or seems promising, and policymakers and businesses often can't wait. Estimating short-term impacts satisfies a rightful desire to learn how policy and programs can be improved. But budget permitting, additional surveys should be conducted to estimate both short- and long-term impacts.

Ideally, researchers would wait three to five years before measuring the impact of an intervention. In some cases, more time might even be necessary. Waiting that long, however, makes it very difficult to maintain the separation of treatment and control groups and prevent spillovers, which is a fundamental requirement to be able to claim that the intervention caused the observed impacts. In addition, the risk of attrition, i.e., the drop out of participants, is higher the longer researchers wait before following-up to measure post-intervention outcomes. At the least, attrition requires the evaluation to be initiated with a larger sample, but it can also introduce bias in the estimate of impact if participants with specific characteristics drop out of the study.

4.4.3. The impact of access may be as important as the impact of use

Evaluators must decide whether they plan to estimate the impact of *access to* microinsurance, or the *use of* microinsurance when designing the evaluation. This choice influences the design of the evaluation and determines how findings should be interpreted.

Insurers, for example, are certainly interested in evaluating the impact that their products have on the well-being of households who sign up for them. Concern for insured households' well-being and good business practices would also recommend evaluating the impact of adding or modifying specific features of insurance products on households who use these products. Finally, evaluating the impact of using specific products or features is fundamental, since, if they are not effective among those who use them, they should probably not be offered, at least in their current form.

Policymakers and funders, however, are also particularly interested in the impact of offering an intervention such as microinsurance, knowing that not all households who are eligible for it will use it. Many policies, particularly aiming to promote development in a broad sense, are interventions offered to individuals who are not required to participate. The impact of having access to the intervention, rather than actually using it, is, therefore, more relevant to policymakers deciding on which policy to support, or how to improve a given policy.

Evaluations must be designed specifically to measure the impact of access to, or use of, the intervention.[11] Intuitively, the method of assignment to treatment or control groups must mirror the type of impacts in which the evaluator is interested. To measure the impact of *access to* an intervention, the treatment group must be constituted of individuals who are exogenously given access to the intervention. Some of these households will decide not to participate. To measure

[11] Being able to decide which type of impacts is measured is most common in randomized experiments. In many natural experiments, the exploitable source of exogeneity dictates which type of impacts is measured.

the impact of *using* an intervention, the treatment group must be made of individuals assigned to use the intervention.[12]

The impact of having access to the intervention is typically lower than the impact of using the intervention, since some of those offered the intervention do not participate, but must still be considered part of the treatment group. The treatment group here is constituted of those having access to the intervention (i.e., being offered to participate), regardless of whether they use it or not. In an evaluation of the impact of access to an intervention, the internal validity provided by a natural or random experiment does not extend to comparing only those who use the intervention, since individuals or households can choose whether to participate or not. That reintroduces a selection bias.

4.4.4. The distribution of impacts can be (at least) as important as the average impact

Impact evaluations, particularly ones based on exogenous assignment into treatment and control groups, are typically designed to determine the average impact of a program.[13] In many cases, however, organisations care about the distributional impacts of an intervention and not just the average impact.

Imagine an organisation which offers a microinsurance product to a randomly-selected group of households and temporarily denies access to the product to another group. The first group is the treatment group and the second is the control group. The difference between the average outcome of the treated group and the average outcome of the control group is an accurate estimate of the intervention's average impact (see, notably, Bauchet and Morduch 2010 and Duflo et al. 2008 for technical details). This is the causal impact of the microinsurance program. The average impact is an important parameter, and is often what social investors and practitioners want to know.

---

[12] We do not explore these differences further in this introductory chapter, and leave it to more detailed explanations of quantitative evaluation designs later in this volume.
[13] The theory underlying the rigorousness of randomized evaluations applied to a comparison of average outcomes in the treatment and control groups, but does not extend to comparison of medians or other measures of distribution such as percentiles.

But what if half of the treated population gains by 100 percent, and the other half lose by 100 percent? In this case, the average impact is zero. Zero is a clean estimate, but it hides the action. Thus, practitioners and investors might care about who is gaining and losing, so that they can target the programs appropriately.

A clean estimate of impact for specific subgroups can be estimated through clever designs. For example, stratifying the treatment and control groups by gender allows one to estimate the impact for men and for women separately. Stratifying means dividing the sample along one or more observable characteristic (such as gender), and performing the assignment to treatment and control for each subgroup separately rather than for the entire sample at once. One limitation of this method is that dividing into subgroups generally requires a larger sample. To have the greatest credibility, subgroups should be identified before the evaluation is started (based on expectations of the way that impacts are likely to vary in different parts of the population) and built into the survey design.

4.4.5. Cost-benefit calculations are critical companions of impact evaluations

Well-designed impact evaluations will provide evidence about the difference that a given intervention makes in the lives of people and/or the performance of an organisation. Knowing the impact of a specific intervention is not the only guide for future action, however. The costs of producing such impact must be factored in recommendations for replication or scaling-up.

Cost-benefit analyses are widely used tools of public policy and should also be systematic companions to impact evaluation. They allow policymakers, funders, and implementing organisations to compare different interventions, or different features of an intervention, and implement the one(s) that provide the best “bang for the buck.” For example, health microinsurers might want to know whether establishing a cashless payment system would provide additional benefits compared to the current mechanism that reimburses for health expenses incurred. The impact on both households and on the insurer of the change in coverage needs to be evaluated and compared to the increased (or decreased) costs for both insured patients and the insuring organisation.

4.5. Generalising from one place and time to another ("external validity")

Most impact evaluations aim to improve the understanding of "what works," both to determine whether the investments have been effective and to learn about possibilities for other places. The ability to generalise the findings from an evaluation is called "external validity."

Learning from one context to another requires both external validity and internal validity. Some statistics-based evaluations exploit data coming from large geographical areas, varied contexts, and/or diversified populations, so their conclusions may be applicable to a wide range of situations (high external validity). But if those studies lack an exogenous determinant of participation into the intervention, such as a random assignment, they may perform less well in providing unbiased estimates of impact (low internal validity). It is then difficult to draw clear lessons.

Evaluations based on random assignment into treatment and control group, on the other hand, do have high internal validity, but they are nearly always implemented with a specific partner in a particular context, which can reduce confidence that measured impacts would also extend to a different setting. For instance, a randomised evaluation of flip charts as teacher's aides in schools in Kenya (Glewwe et al. 2004) only tells us whether the flip charts helped raise test scores for these students in these schools in this region of Kenya. One could imagine that students or schools in other parts of Kenya, India, or Peru have different educational needs, and would benefit differently (or not at all) from their teachers' using flip charts.

We need to understand the specific context of the evaluation before drawing general conclusions. This means considering three big questions:

1. How does the population studied *there* differ from the population I'm interested in *here*? Are they better educated? Poorer? Healthier? Etc.

2. How do supporting inputs differ? Are there critical government programs in place? Good roads and transport? Community institutions? Effective organizations to deliver the interventions in question?

3. How do alternatives activities differ? Does the studied intervention mostly substitute for existing opportunities? Does it complement them? Morduch et al. (2013), for example, found that a very promising anti-poverty program in South India ended up having no net impact because alternative options were so good (and the control group availed themselves of those options). The same program had bigger impacts in sites with very similar populations, but where, it seems, such good alternatives were lacking.

Some of these questions can be addressed with an eye to understanding how and why the intervention worked or not. Combining qualitative and participatory research designs with rigorous quantitative evaluations can be applied to increase the understanding of the mechanisms that produced impacts and to gain external validity.

4.6. Conclusion: Using evaluations to improve operations

It is tempting to view evaluations as mainly backward-looking assessments. But their greatest power is often as forward-looking guides to innovation and improvement. Businesses, donors, investors, and policymakers often have to select between competing programs when deciding how to allocate scarce resources. Rigorous impact evaluations are an indispensable tool for strategic planning. They inform choices that leaders must make. Knowing what difference a specific intervention makes also calls upon all stakeholders to improve the intervention, try alternative – and potentially better or cheaper – methods, and share the knowledge gained with other individuals and organisations.

Karlan et al. (2009), for example, identify several ways in which rigorous impact evaluations can help microfinance institutions increase both their sustainability and social outreach, including improving their borrower risk assessment techniques and learning about the impact of the price of the loans on demand. In microinsurance, impact evaluations can test the effectiveness of two different insurance products or test the effect of specific elements of the products, such as different marketing techniques, pricing structures, or distribution channels. Understanding the impact of their operations on client participation and well-being can enable practitioners to design better products and services, and thereby increase scale, sustainability, and social impact.

This is an exciting time for the microinsurance industry. The past few years have seen an influx of interest from insurers and investors, and regulators are driving new initiatives to broaden access. As organisations make new investments and test innovations, they should pay attention to whether their products are having the impacts for which they hoped. When done right, impact evaluations are a tool to efficiently direct future allocations, design better products, and improve operations.

Suggestions for Further Readings

Bauchet, J., and J. Morduch. 2010. An introduction to impact evaluations with randomized designs. Financial Access Initiative.

Collins, D., J. Morduch, S. Rutherford, and O. Ruthven. 2009. *Portfolios of the Poor: How the World's Poor Live on $2 a Day*. Princeton, NJ: Princeton University Press.

Duflo, E., R. Glennerster, and M. Kremer. 2008. Using Randomization in Development Economics Research: A Toolkit. In T. Eds. P. Schultz and J. Strauss, *Handbook of Development Economics*. Amsterdam; New York: North-Holland. 4:3895-62.

Karlan, D., N. Goldberg, and J. Copestake. 2009. Crossfire: Randomized control trials are the best way to measure impact of microfinance programmes and improve microfinance product designs. *Enterprise Development and Microfinance* 20(3): 167-176.

Karlan, D., J. Morduch, and S. Mullainathan. 2010. Take-up: Why Microfinance Take-up Rates Are Low & Why It Matters. Financial Access Initiative Framing Note.

References

Armendáriz, B., and J. Morduch. 2010. *The Economics of Microfinance*. 2nd ed. Cambridge, MA: MIT Press.

Bauchet, J., and J. Morduch. 2010. An introduction to impact evaluations with randomized designs. Financial Access Initiative.

Churchill, C. ed. 2006. *Protecting the poor: A microinsurance compendium*. Geneva: International Labour Organization.

Collins, D., J. Morduch, S. Rutherford, and O. Ruthven. 2009. *Portfolios of the Poor: How the World's Poor Live on $2 a Day*. Princeton, NJ: Princeton University Press.

Duflo, E., R. Glennerster, and M. Kremer. 2008. Using Randomization in Development Economics Research: A Toolkit. In T. Eds. P. Schultz and J. Strauss. *Handbook of Development Economics 4*:3895-62. Amsterdam; New York: North-Holland.

Duflo, E. 2001. Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review* 91(4): 795-813.

Giné, X., and D. Yang. 2007. Insurance, Credit, and Technology Adoption: Field Experimental Evidence from Malawi. Policy Research Working Paper Series 4425, The World Bank.

Glewwe, P., M. Kremer, S. Moulin, and E. Zitzewitz. 2004. Retrospective vs. prospective analyses of school inputs: the case of flip charts in Kenya. *Journal of Development Economics* 74(1): 251-268.

Karlan, D., N. Goldberg, and J. Copestake. 2009. Randomized control trials are the best way to measure impact of microfinance programmes and improve microfinance product designs. *Enterprise Development and Microfinance* 20(3): 167-176.

McKernan, S.M. 2002. The impact of microcredit programs on self-employment profits: Do noncredit program aspects matter? *Review of Economics and Statistics* 84(1): 93–115.

Morduch, J., S. Ravi, and J. Bauchet. 2013. "Failure vs. Displacement: Why an innovative anti-poverty program showed no net impact." Indian School of Business and New York University, unpublished manuscript.

Rubin, D. B. 1974. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology* 66(5): 688-701.

Rutherford, S. 2000. *The Poor and Their Money*. New Delhi; New York: Oxford University Press.

Shadish, W. R., T. D. Cook, and D. T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.