

Evaluation Fundamentals¹

Aparna **Dalal**
New York University

Jonathan **Bauchet**
New York University

Jonathan **Morduch**
New York University

The Financial Access Initiative is a research center based at New York University, focused on finding answers to how financial sectors can better meet the needs of poor households.

1. This chapter has been prepared for the ILO *Microinsurance Impact Evaluation Practical Guide*.

Introduction

At age 40, Feizal was supporting his family in rural India. He earned a living selling aluminum 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.²

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.

2. The story of Feizal and his family is part of the financial diaries project collected by Orlanda Ruthven and described in Collins, et al. (2009). Faizal's story is just one story from *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. During the research year, the households surveyed experienced 167 financial emergencies related to injury, illness, death, and losses to income and property. Household usually coped with these emergencies by patching together resources from a mix of formal and informal savings, loans and insurance products, as well as by getting help from friends and relatives.

Establishing that causal link with confidence requires careful planning. It is fundamental, however, to allow policymakers and practitioners to design more effective programs, projects and solutions.

Impact evaluations try to measure 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* of the intervention. The focus on outcomes, rather than inputs, distinguishes impact evaluations from "process evaluations." Process evaluations measure program processes and inputs. 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.

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 neighbors 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.

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

The Challenges

Addressing Selection Bias

The ultimate goal of all impact evaluations, and their greatest challenge, is to establish credibly that the intervention caused a difference in the lives of the participants. The rough notion of "making a difference" can be translated into a precise question that is at the heart of every impact

3. 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.

study: “How have outcomes changed with the intervention *relative to what would have occurred without the intervention?*”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). 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 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 eliminating the selection bias can lead researchers to overestimate the impact by 100 percent.

In microinsurance, not separating the impact of the intervention from that of other confounding factors could lead to under- or overestimating the impact of the insurance, depending on the situation. It is possible, for instance, that households who suffer from a pre-existing 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.

In microinsurance, not separating the impact of the intervention from that of other confounding factors could lead to under- or overestimating the impact of the insurance, depending on the situation.

4. This approach to impact evaluation, based on a comparison with hypothetical outcomes, is often called the Rubin Causal Model after its originator, Donald Rubin.

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.

The big challenge arises with unobservable factors. Attributes like an individual's propensity to fall sick, organizational 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 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.”

A Unique Challenge for Microinsurance Evaluations

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 farmer to take more risk, altering their investment and employment strategies, which could have an impact on their well-being. In markets that offer no insurance, households instead mitigate risk by adopting strategies such as crop and labor diversification or under-investing in new technologies that could improve productivity, but incur upfront costs (Foster and Rosenzweig 1995). These risk mitigation strategies diminish the potential returns of their activities (Morduch 1995). Having access to insurance and signing up for an insurance contract can potentially alter these strategies and encourage households to turn to riskier but more productive activities and technologies. 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.

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 *don't* have insurance. This is the “counterfactual,” and it is the key to credible evaluations.

Getting Credible Answers (“Internal Validity”)

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 *don't* 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 can't 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.

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 external to the intervention—an “exogenous” event—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,

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.

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

the adoption of a national policy of large-scale school building, has been used to estimate the impact of education on wages (Duflo 2006). 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 randomized 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.

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. But more subtle threats exist.

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 benefit from being acquainted with households in the treatment group, since it is likely that friends would help each other in times of need. This spillover of benefits from the treatment to the control group “contaminates” the assignment. Spillovers and contamination are technical terms (also revealing the medical origin of randomized evaluation designs) referring to situations in which members of the control group receive benefits from the intervention, thus cancelling the virtues of the random assignment to treatment and control groups.

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,

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.

7. See Bauchet and Morduch 2010 for a more detailed exposition of the theory.

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. Recognizing this possibility, the evaluator might need to assign entire communities to treatment or control.⁸

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 generalized 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 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 minimized 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.⁹ Statistical power generally improves with larger sample sizes, but it is

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.

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

9. 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).

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 emphasized 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.

Whether evaluating microinsurance as a whole or a specific feature of a micro-insurance product, evaluators must carefully define their intervention and place it in a larger context.

Interpreting results

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 cost effective is the intervention at producing these impacts? How confidently can the evidence from one evaluation be extrapolated to other contexts?

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 semi-formal organizations 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.

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 benefited from rainfall insurance during a drought.

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.

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.

The effects on social outcomes in health, education, and women's empowerment were negligible. The study's relatively short time frame, however, limits the scope of the results and their implications to the short-term.

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.¹⁰ 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 the impact of *using* an intervention, the treatment group must be made of individuals assigned to use the intervention.¹¹

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

Evaluations must be designed specifically to measure the impact of access to, or use of, the intervention.

Intuitively, the method of assignment to treatment or control groups must mirror the type of impacts in which the evaluator is interested.

10. 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.

11. 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.

individuals or households can choose whether to participate or not, which re-introduce a selection bias.

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.¹² In many cases, however, organizations care about the distributional impacts of an intervention and not just the average impact.

Imagine an organization 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.

But what if half of the treated population gains by 100 percent, and the other half loses 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 sub-groups 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 it requires a larger sample—the more subgroups the larger. Subgroups also need to be identified before the evaluation and built into the design; one cannot measure distributional impacts depending on preliminary results of the analysis, which is often tempting.

A clean estimate of impact for specific sub-groups 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.

12. The theory underlying the rigorouslyness 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.

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 organization. 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 policy-makers, funders, and implementing organizations 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 organization.

Care Must Be Taken in Extrapolating to Other Settings (“External Validity”)

When a particular intervention is shown to be particularly effective, it should be replicated in other regions or countries where it could improve households’ well-being. Most impact evaluations aim to provide generalizable ideas about “what works.” The degree of generalizability of the findings from an evaluation is called “external validity.”

Most impact evaluations must trade off internal for external validity. Some statistics-based evaluations exploit data coming from large geographical areas, varied contexts, and/or diversified populations, so that their conclusions are applicable to a wider range of situations (high external validity). But the lack of random assignment casts permanent doubts on their ability to provide an unbiased estimate of impact (low internal validity), and, without that, they are of little use for social planning.

Evaluations based on random assignment into treatment and control group, on the other hand, do have high internal validity, but are necessarily implemented with a specific partner in a particular context, which reduces one’s confidence that impacts it measured would also happen in a different setting. For instance, a randomized evaluation of flip charts

as teacher's aide 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.

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 policy-makers 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 organizations.

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, thereby increasing scale, sustainability, and social impact.

This is an exciting time for the microinsurance industry. The past few years has seen an influx of interest from insurers and investors, and regulators are driving new initiatives to broaden access. As organizations 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.

In microinsurance, impact evaluations can test the effectiveness of two different insurance products or test the effect of specific elements of the products... Understanding the impact of their operations on client participation and well-being can enable practitioners to design better products and services, thereby increasing scale, sustainability, and social impact.

Suggestions for Further Readings

Bauchet, Jonathan and Jonathan Morduch. 2010. "An introduction to Impact Evaluations with Randomized Designs." Financial Access Initiative.

Collins, Daryl, Jonathan Morduch, Stuart Rutherford and Orlanda Ruthven. 2009. *Portfolios of the Poor: How the World's Poor Live on \$2 a Day*, Princeton, NJ: Princeton University Press.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2008. "Using Randomization in Development Economics Research: A Toolkit." In T. Paul Schultz and John Strauss (Eds.), *Handbook of Development Economics* 4: 3895-3962. Amsterdam; New York: North-Holland.

Karlan, Dean, Nathanael Goldberg and James 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.

References

- Armendáriz, Beatriz and Jonathan Morduch.** 2010. *The Economics of Microfinance*. Second edition. Cambridge, MA: MIT Press.
- Bauchet, Jonathan and Jonathan Morduch.** 2010. "An Introduction to Impact Evaluations with Randomized Designs." Financial Access Initiative.
- Churchill, Craig.** 2006. *Protecting the Poor: A Microinsurance Compendium*. Geneva: International Labour Organization.
- Collins, Daryl, Jonathan Morduch, Stuart Rutherford and Orlanda Ruthven.** 2009. *Portfolios of the Poor: How the World's Poor Live on \$2 a Day*. Princeton, NJ: Princeton University Press.
- Duflo, Esther, Rachel Glennerster, and Michael Kremer.** 2008. "Using Randomization in Development Economics Research: A Toolkit." In T. Paul Schultz and John Strauss (Eds.), *Handbook of Development Economics 4*: 3895-3962. Amsterdam; New York: North-Holland.
- Duflo, Esther.** 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review* 91 (4): 795-813.
- Foster, Andrew, Mark Rosenzweig.** 1995. "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture." *Journal of Political Economy* 103 (6): 1175-1209.
- Gine, Xavier, and Dean Yang.** 2007. "Insurance, Credit, and Technology Adoption: Field Experimental Evidence from Malawi." Policy Research Working Paper Series 4425, The World Bank.
- Glewwe, Paul, Michael Kremer, Sylvie Moulin, Eric 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, Dean and Jonathan Zinman.** 2010. "Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts." *Review of Financial Studies* 23 (1): 433-464.
- Karlan, Dean and Jonathan Zinman.** 2009. "Expanding Microenterprise Credit Access: Using Randomized Supply Decisions to Estimate the Impacts in Manila." Yale Economic Growth Center Discussion Paper No. 976.
- Karlan, Dean, Jonathan Morduch and Sendhil Mullainathan.** 2010. "Take-up." Financial Access Initiative Framing Note.
- Karlan, Dean, Nathanael Goldberg and James 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, Signe-Mary.** 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, Jonathan.** 1995. "Income Smoothing and Consumption Smoothing." *Journal of Economic Perspectives* 9 (3): 103-114.
- Rutherford, Stuart.** 2000. *The Poor and Their Money*. New Delhi; New York: Oxford University Press.
- Shadish, William R., Thomas D. Cook and Donald T. Campbell.** 2002. "Experimental and Quasi-Experimental Designs for Generalized Causal Inference." Boston: Houghton Mifflin.