Operationalising impact evaluation: from theory to practice

Divya Parmar and Manuela De Allegri
11.1. Introduction

Impact evaluations in the field of microinsurance can employ different research designs. Impact evaluation designs can be experimental, quasi-experimental and non-experimental, qualitative, or a combination of these. Increasingly, quantitative approaches, whether experimental or quasi-experimental, are often combined with qualitative data collection and analysis processes. The resulting designs are described as mixed methods. Irrespective of the specific design adopted, the common objective across all impact evaluations is to assess the “causal” relationship between the programme and a set outcome of interest. This objective clearly distinguishes impact evaluations from other forms of evaluations, like implementation evaluation, which focuses on how an intervention is being implemented, or performance evaluation, which focuses on assessing whether a programme achieved the objectives it originally set to achieve.

All research designs used to study impact evaluations have specific strengths and weaknesses. For example, propensity score matching can only reduce selection bias due to time invariant indicators. Similarly, although randomisation is often considered to be the best available option to estimate impact, it may not always be the most appropriate design to apply in a particular situation [Shadish et al. 2002]. For instance, it may not be politically feasible to offer microinsurance to some individuals whilst denying it to others who have similar needs and could equally benefit from the intervention. The lesson we have learned over the years is that the most appropriate design balances theoretical considerations with field circumstances.

Focusing now on the elements that ought to be considered when translating the theory of impact evaluation into practical fieldwork, this chapter unfolds over five sections. Each section covers one specific aspect relevant to designing, planning, or conducting an impact evaluation in the field of microinsurance. Firstly, we introduce and discuss the elements that need to be considered when choosing an adequate research design, depending on field circumstances. Secondly, we describe the resource requirements for conducting an impact evaluation and how those requirements may reflect on the choice of the research design. Thirdly, we address pragmatic considerations on sample size, tool development, and actual data collection and analysis designs can only control for selection bias due to time invariant indicators.
processes. Fourthly, we address the ethical implications of the various designs and briefly review procedures for ethical clearance in impact evaluation. Lastly, we address the interface and relationship between the research team and the concerned policy makers. Across sections, we draw from direct field experience to complement conceptual models with practical illustrations.

11.2. Selecting an appropriate impact evaluation design

Purpose of the impact evaluation

The ultimate aim of any impact evaluation is not only to determine whether, and by how much, the microinsurance programme yields an impact on a defined outcome through the estimation of causal relationships, but also to provide the necessary evidence base to inform further policy developments. If a microinsurance programme yields beneficial impacts, it is important to understand what factors in the environment enabled change to take place and to assess whether the programme can be scaled up or replicated elsewhere. Conversely, if the programme does not yield the expected impacts, it is important to identify potential bottlenecks and barriers to change that can be removed. Hence, good impact evaluations should employ mixed methods: experimental or quasi-experimental methods focus on capturing the impact, whilst qualitative methods can reveal how change took place (White 2009) in the light of the context in which the programme is embedded (Hintz 2010).

Besides its ultimate purpose to inform policy, there are several field-related factors that need to be considered when selecting an adequate impact evaluation design. The most important factors are briefly described and discussed below. These factors are not listed in terms of their relative priority or importance. They are also not mutually exclusive and, by and large, will depend on the same underlying conditions. Their role and the extent of their influence on the design of the impact evaluation largely depends on a case-by-case basis.

The launch of the microinsurance programme and its targeting strategies

We start by looking at how the microinsurance programme is launched.
Two initial important elements to be considered:

1) How the microinsurance programme is launched: whether it is phased in over time, or launched everywhere at the same time
2) How beneficiaries are targeted and enrolled

**The rollout strategy**

Phasing the rollout of a microinsurance programme over time presents two advantages. On one side, it eases the early administrative burden on the programme management. On the other side, it offers those concerned with the impact evaluation the possibility of recruiting the control sample from areas/districts/subportions of the target population not yet affected by the insurance launch. In technical terms, the element of an impact evaluation that relies on a phased implementation of the microinsurance programme is referred to as *stepped wedge* (Brown and Lilford 2006).

The robustness of the analysis increases substantially if the phasing can take place according to a randomised design, meaning that areas/districts/subportions of the target population are first stratified on the basis of a variable [i.e., characteristic] thought to be important, for example literacy rates or level of urbanisation. Then areas/districts/subportions of the target population are randomly selected from the composed strata so that the final sample includes representatives of all strata.

Selecting villages on the basis of political decisions or feasibility concerns represents the least ideal scenario for a phased rollout strategy. The lack of random assignment to the intervention, in fact, imposes an important threat to the validity of the overall study because it introduces a source of bias in the estimation of the impact effect. In situations where no randomisation is possible, however, a phased rollout of the intervention according to political or feasibility criteria...
is still preferred over a situation with no phased rollout at all. Even in the absence of randomisation, a phased rollout strategy leaves open the possibility of identifying comparable controls, whilst interventions that affect entire regions/districts/target populations at once make it much more difficult for the researcher to identify comparable controls.

Impact evaluations of microinsurance programmes often set the unit to be randomised at a group level (e.g., a village, an agricultural cooperative, a loan group, etc.), rather than at an individual level. This randomisation strategy is known as cluster randomisation. There are two strong arguments that support cluster randomisation in the context of microinsurance programmes. Firstly, although the benefits of the microinsurance are for those who buy the microinsurance, others may experience the impacts within that cluster. This may be because the programme “leaks”, contaminating those who are not supposed to receive it, thereby weakening any estimate of treatment difference (Flory 2011). This is particularly relevant for microinsurance programmes as they operate in areas where people are already commonly engaged in informal risk sharing. Randomising insurance access at the individual or household level may actually lead to underestimating the true effects of insurance due to the potential risk shifting from those without access to the microinsurance to those with access. Such a case was observed by Flory (2011) in Malawi, where the poorest benefitted from a microinsurance programme—even though they had not enrolled—because of interhousehold transfer practices. Secondly, individual randomisation requires that the implementing organisations have a complete listing of all potential beneficiaries and are able to manage and control the process of enrolling single members in a randomised manner over an extended period of time. These conditions can rarely be met by organisations operating in low-income settings, especially in rural areas, where much of the bookkeeping is still done exclusively manually. In addition, individual randomisation does pose some ethical challenges since it diverts from the communal orientation of traditional societies. Conflict may arise when two neighbouring individuals or households
are enrolled into the same programme at very different time points.

Cluster randomisation is better suited to reduce the risk of contamination across the intervention and the control arms, since it uses exiting barriers (e.g., geographical, group, etc.) to define the very same allocation of the intervention (De Allegri et al. 2008). Continuous monitoring is required on the part of the impact evaluation team to ensure that the intervention is correctly allocated to the selected individuals or clusters. Mistakes in the allocation of the intervention can jeopardise the estimation of the impact at a later stage. The implications that cluster randomisation bears on the sample size are discussed later in this chapter.

A stepped wedge cluster randomised design was applied to evaluate the impact on access to health-care services and financial protection of a health microinsurance programme in the Nouna Health District in rural Burkina Faso. In this study, the 41 villages and seven urban sectors of the target region were formed into 33 clusters. Small neighbouring villages that shared common ethnic and demographic characteristics were grouped together to form a single cluster. Health insurance was offered in a phased manner, so that each year households in an additional randomly selected 11 clusters received the offer to enrol in the scheme. By year three, the entire target region had received the offer to enter the health insurance programme (De Allegri et al. 2008). The design allowed for minimal contamination across clusters and a robust estimation of the impact of the scheme on the outcomes of interest. At the same time, the stepped wedge design facilitated the scheme implementation, by allowing the implementing agency to focus their social marketing campaign on a restricted number of villages per year.

**Targeting strategies**

Another important feature of a microinsurance programme that needs to be considered when selecting an adequate impact evaluation design is the targeting of the beneficiaries. Programmes that target the beneficiaries according to an arbitrary cutoff point open the possibility for the impact evaluation team to apply regression discontinuity designs (RDD). For instance, this may be the case when implementing agencies decide to offer microinsurance only to the poorest 20% of the entire population. For RDD, it is assumed that the households who marginally miss the cutoff (for instance, households that lie in the 22% poorest category) are very similar to the targeted households and therefore can be used as controls. RDD represents a viable
design option when enough people fall in the category just above the cutoff point. A commonly used approach is to look at clusters of settlements that are divided by administrative boundaries. The administrative boundary is seen as a random cutoff, where seemingly similar households are offered microinsurance only if they fall on one side of the boundary. In such situations, the remaining households that stay close to the intervention households but are not offered the microinsurance can be used as controls. There are not many examples where RDD has been used to evaluate microinsurance programmes, but there are few examples from similar interventions, like the publically-funded health insurance programme in Colombia called Régimen Subsidiado (Miller et al. 2009). In the programme, only people below a certain threshold of the poverty index were eligible to receive full public subsidy to purchase health insurance. By comparing the eligible people with those that narrowly missed this threshold, impact of this programme was evaluated.

DID methods can be applied if the microinsurance programme is launched according to a cluster randomised design or if neighbouring areas are similar enough to the intervention area to be used as controls. Quimbo et al. (2010) applied a DID approach to evaluate the impact of a health insurance programme for poor children in Philippines. The evaluation study included 30 public hospitals. These hospitals were divided into intervention and control groups, matched according to demand-and-supply characteristics like household income, number of beds, average costs, etc. Both exit and follow-up home interviews were collected at baseline (round one) and then again after the programme was implemented for two years (round two). Identical baseline and round two data were collected from the intervention and control groups for the DID method. After the launch of the microinsurance programme, applying DID can be often impossible or tricky, as the researcher has to depend on existing surveys for baseline data.

If all the beneficiaries are targeted at once, then propensity score matching (PSM), with all its limitations, may be the only available option left as DID or RDD will not be applicable. The researchers have to depend on the “eligible but not enrolled” group to be potential controls for those that enrol for the microinsurance programme (Trujillo et al. 2005; Gnawali et al. 2009).

An intrinsic challenge in the selection of the most appropriate study design rests in the difficulty of reconciling the highest research standards (to which a sound impact evaluation ought to abide) with field requirements and
conditions. Those directly concerned with the implementation of a micro-insurance scheme may be more concerned with programme start-up and survival than with producing evidence on expected and unexpected impacts to inform the wider policy arena. Therefore, they may be reluctant to agree to randomisation, stepped wedge designs, and RDD.

In our experience, continuous and open dialogue with those in charge of implementation is the only strategy for reconciling implementation needs with those of a sound impact evaluation. At times, when this fails, research teams in charge of the impact evaluation have no other choice if not selecting the one study design that, although not perfect, best exploits existing field conditions. The objective is always to select the design that presents the fewest threats to validity in the light of existing field conditions. During this process of negotiations between impact evaluators and implementing agencies, the former may often find themselves explaining over and over the concept of impact evaluation, as this is not so clearly differentiated from other forms of evaluation amongst people working away from research. Likewise, evaluators have to be open to working within the framework of an emergent design, meaning they must be willing to readjust design decisions as field conditions unravel. Emerging designs are proper for qualitative studies. Therefore, they do represent a fundamental epistemological challenge for the positivist impact evaluator, who learned to rely exclusively on experimental and semi-experimental models. Nevertheless, years of reconciling research rigor with field needs have convinced us that often no other way is possible than letting the mind be open to adjust design decisions, within the limits of rigorous research, to emerging field conditions.

Who is commissioning and conducting the impact evaluation?

The ability to negotiate with the implementing agency the most robust study design largely depends on who is
commissioning and who is conducting the impact evaluation. Impact evaluation teams may find the dialogue over research designs to proceed better when the need for an impact evaluation has been identified by the agency implementing the microinsurance program and/or when agency itself commissions the impact evaluation. In such cases, implementing agencies are very open to adjust their implementation design to allow for robust estimation procedures because they have a primary interest in the use of the results that will emerge from the evaluation. Large international organisations, such as the International Labour Office (ILO) or the World Bank, often enjoy the privilege of commissioning, and, in some instances, even managing directly both the implementation and the impact evaluation of interventions. This allows them to shape the rollout of programmes in such a way as to derive sound evidence on their impacts afterwards.

Researchers from academic institutions often find themselves facing the exact opposite scenario, being called in by a third party, frequently a ministry or another high governmental institution, to assess the impact of interventions run by local and international non-governmental organisations (NGOs) and/or consultancy groups. In such instances, the risk that the implementing agency perceives the impact evaluation, and, in turn, the research team conducting it, as a potential threat is great, since its reputation and its ability to acquire future contracts may be at stake. As mentioned above, continuous dialogue is the only means to overcome initial scepticism towards the impact evaluation and the team conducting it. Still, at times this initial resistance cannot be fully overcome, the result being that the research team will have to find its analytical way around field conditions far from enabling an optimal impact evaluation design.

The number of resources that can be mobilised to conduct an impact evaluation also plays an important role in shaping design decisions. Impact evaluations that can count on substantial ad hoc funding can engage teams of expert researchers who can guide proper design decisions. This is often not the case when, due to lack of adequate funding, the impact evaluation is planned and conducted directly by the implementing agency. At times, this may result in impact evaluations that rely almost exclusively on routine monitoring and evaluation data to estimate effects.

Timing of the impact evaluation

Dialogue between the implementation team and the impact evaluation team should take place as early as possible, preferably during the conceptualisation
of the microinsurance programme. Early exchange allows for the impact evaluation to be factored into the implementation design from its very onset, ensuring that adequate conditions are in place to allow for a robust estimation of the effects of the intervention. A sound impact evaluation is one that relies on data collection that takes place both in the intervention and control areas at both baseline and follow-up (multiple follow-ups are possible). If the impact evaluation is not factored into the implementation design early enough, it may become impossible to produce a robust evaluation at a later point in time, either because the programme is launched in a way that there are no appropriate controls (e.g., the microinsurance programme is launched in the entire target region simultaneously) or because relevant data has not been collected from relevant controls already at baseline.

Field circumstances are such that, at times, impact evaluations are decided only once a programme is already in operation. In such situations, information may be available on the areas/districts/population targeted by the microinsurance programme, but most likely not on any comparable controls. To overcome the estimation problems resulting from the lack of data on controls, researchers are advised to search for alternative data sources (e.g., household surveys, census, demographic health surveys, facility-based data, and programme data) applicable to the control areas/districts/population. If no alternative data source is available, then the researcher has to derive an estimation of the effect by relying on cross-sectional data collected both from the intervention and control individuals and/or households. Propensity score matching, with all its limitations, is probably the only viable option in these situations (Gnawali et al. 2009).

11.3. Resource consumption

General remarks

The different impact evaluation designs all imply different resource consumption requirements. Therefore, selecting the most appropriate design also requires an assessment of the specific resources available. Resources are usually more abundant when funds are made available from a third party, either in the form of a research grant or in the form of a contract established directly between the impact evaluation team and the agency commissioning the study. Such contracts, however, are very rarely directly managed by the agency implementing the microinsurance programme, since such agencies usually operate in conditions of financial limitations and would be unlikely to divert funds from implementation towards research. Such contracts are
more common when the impact evaluation is commissioned by a governmental or supra-governmental institution to an external evaluation team (either a research institution or a consulting agency). Therefore, if on one side, earmarked contracts increase the resources available for field research, on the other side, they limit ownership of the impact evaluation on the part of the implementing agency. The implications of having an ad hoc research team conduct an impact evaluation have already been discussed in detail earlier.

**Research team**

Team requirements vary substantially depending on the design and on the scope of the single impact evaluations.

Experimental designs that rely on randomisation are very demanding on the organisations implementing microinsurance, as they require a constant managerial and administrative engagement to ensure the correct allocation of the intervention.

Quasi-experimental and non-experimental designs are less burdensome from this point of view. Irrespective of the design that is applied, all quantitative impact evaluations require that people with specific expertise in statistics are an integral part of the research team. Decisions on design, randomisation, and sampling can only be achieved with the support of expert statisticians. This is especially important, as discussed in detail later, given the frequent absence of formal and complete population data to serve as a sampling frame in most settings where microinsurance is implemented. Likewise, irrespective of the design, the support of statisticians and econometricians is essential during the analytical phases. The need for support during the analytical phases is greater in case of complex quasi- and non-experimental designs, combining multiple designs into one to compensate for lack of randomisation and/or baseline data (Shadish et al. 2002).

Impact evaluations applying mixed methods require an even larger set of expertise to come together in the design and analytical phases. Mixed methods designs, in fact, rely on the application of multiple methods at once and, as
such, multiple people have to engage in decision making throughout the length of the evaluation. This means that multiple traditions of conducting research are merged into one single study. The process can be extremely challenging, as it entails open dialogue across people with radically different scientific mind-sets. In the experience of the authors, the key success factor when conducting a mixed methods impact evaluation is the mediating capacity of the team leader. Ideally, the team leader herself should be well acquainted with multiple scientific traditions and respect them all equally to be able to facilitate dialogue in teams where ethnographers and econometrists need to reconcile diverging scientific paradigms into a single study.

Likewise, the scope of the impact evaluation determines the range of expertise that will need to be brought into the research team. The scope of an evaluation is determined by the number and typology of outcome indicators that need to be assessed. Complex impact evaluations that aim to capture changes on multiple dimensions (e.g., agriculture, health, poverty) often require large multidisciplinary teams to be involved already at the stage of study design and tool development. We, the authors, have been involved in impact evaluations, bringing together fifteen scientists from at least ten different disciplines. Working on an impact evaluation with a broad scope of analysis may entail having to facilitate dialogue between clinicians, psychologists, economists, political scientists, geographers, and agricultural experts. In our experience, this dialogue is complex and not always easy, but very rewarding and enriching in the end.

Field team and data collection processes

The human resources needed to design and plan an impact evaluation and to analyse the data emerging from the field are only a minimal part of overall human resources needed to conduct such an effort successfully. The people actually in charge of data collection in the field are key to any impact evaluation. Data collection teams vary substantially across impact evaluations. Much of the variation can be explained in relation to the nature and scope of the single evaluations. For instance, assessing the impact of a health microinsurance scheme on health indicators may require that people with some clinical training handle the data collection processes, whilst this may be totally irrelevant when assessing the impact on poverty reduction of crop microinsurance. Likewise, purely quantitative versus mixed methods designs imply reliance on very different sets of expertise.
The overall educational levels in a given country also influence the selection and composition of the field data collection team. We have worked in settings where field interviewers could, at the very best, be expected to hold a high school degree, as well as in settings where a college degree was the minimum prerequisite to being enlisted as a potential field interviewer. It goes without saying that the initial educational level of the staff recruited to work on field data collection plays a very important role in the overall process. Interviewers with lower education may require longer training sessions. Yet, these may be the people who really know the communities where data collection is taking place and may therefore be much better placed to engage their respondents than their better-educated, often urban, counterparts. In general, it is easier to identify and recruit interviewers to work on quantitative data collection tools than to identify interviewers with sufficient experience to be able to conduct qualitative in-depth interviews and/or facilitate focus group discussions (FGD).

Field data collection often absorbs most of the financial resources available for an impact evaluation. In recent years, the use of digital devices to collect data has shown potential to reduce quantitative data collection costs substantially, whilst also ensuring greater quality of the data being collected. Digital devices, in fact, allow the research team to insert very accurate filters and logical checks in the structure of the questionnaires being used, limiting the possibility of error by the single interviewers. In our experience, two rounds of data collection on a sample of approximately 1200 households are sufficient to recover the costs of purchasing 15 tablets.
Digital devices can speed up the process of obtaining data.

Through the use of mobile phone lines and digital devices also facilitate immediate data sharing between the field and the impact evaluation team. This process allows the impact evaluation team to carry out constant quality checks and to feed preliminary results to the implementing and commissioning agency much faster than traditional pen and paper surveys. Field data entry represents an alternative to digital data collection, which also allows using time more efficiently and controlling for completeness and quality of data. In this model, a data entry clerk accompanies the interviewers to the field to input data in the computer one day after the actual data collection activity has taken place, at the latest. This allows the data entry clerk to identify missing and inconsistent information in time to be able to verify it directly in the field.

Digital solutions to facilitate qualitative data collection are, unfortunately, not as advanced as those supporting quantitative data collection. In-depth interviews and FGD are easily recorded on tape or using MP3 recorders, but skilled transcribers are still needed to report the recorded text on paper. A number of software solutions are available to automate this process, but in our experience, none of this software provides a valuable alternative to the traditional method of transcribing interviews. The available software, in fact, mostly operates only in few languages (English, French, and Spanish) and largely relies on voice recognition. Therefore, it is of little or no use at all to researchers working with communities in low-income settings where these languages are not spoken. Similarly, automated translation devices are not sufficiently advanced to afford researchers a faster, yet accurate, translation of transcribed text from a local language into the language of analysis.

Irrespective of whether digital devices or traditional pen and paper surveys are used for field data collection, interviewers are usually organised in microteams of two to three interviewers and one supervisor. When the scope of the impact evaluation entails collecting data from multiple sources (for instance in a health facility and at the household level), it is usually more effective to create specialised microteams than to train all interviewers on all relevant data collection tools. Microteams can be composed of quantitative and qualitative interviewers, targeting different respondent constituencies and data collection needs at the same time. We have widely worked with this system of multiple specialised microteams, targeting communities at once (and thus keeping transport costs low) to maximise the amount and quality of data collected in a day from multiple sources.
Before embarking in the actual data collection, field interviewers receive training (normally two to seven days, depending on the complexity of the data collection tool) and assist the research team in piloting the tool. The pilot phase is essential for identifying potential pitfalls in the data collection tool in time for the research team to modify questions and adjust them to the specific setting where the impact evaluation will take place. Research teams wishing to rely on digital data collection may need to consider one to two additional days of training. The cost for these additional days of training is quickly recovered through faster field processes, higher-quality data (requiring minimal cleaning), and the absence of data entry at a later stage.

Costs for field data collection vary dramatically across settings and it would be unfair for us to pretend that we can provide the reader with realistic price estimates applicable across settings. In general, in low- and middle-income countries, the cost of an interviewer (and similarly of a data entry clerk) varies from US$5 to US$50 per day. In most countries, interviewers will expect to receive additional compensation if the data collection takes place in a location that does not allow them to return to their homes in the evening. Accommodation and per diem costs can, at times, be higher than the actual compensation received as interviewers. On average, supervisors cost 20% to 40% more than simple interviewers. In many resource-poor countries, transport costs are an important cost driver when conducting an impact evaluation. This is due to the poor conditions of the roads which require data collection teams to hire high-quality vehicles, and to the extreme high price of fuel in some settings.

Data requirements also tend to differ depending on the research design. Simple randomisation and RDD designs generally rely on cross-sectional data and data are collected from both the intervention and the control groups. PSM is generally applied to cross-sectional data, but need larger sample of eligible but non-enrollees to ensure that researchers can find enough controls that closely match with the enrollees. Stepped wedge cluster randomisation, on the other hand, has the largest data requirement, as data needs to be collected from the intervention and control groups at every step of the cluster randomisation. For example, if the phased rollout takes three years to implement, like the health microinsurance programme in Nouna District discussed before, then data needs to be collected for at least three years. Whilst designing the impact evaluation study, data requirements and duration of the study need to be taken into consideration. These factors will further
have financial implications for the impact evaluation study.

11.4. Sampling

Sampling is probably one of the most challenging operational aspects facing teams conducting an impact evaluation. Different elements ought to be considered and reconciled into one coherent, yet feasible, sampling strategy. This process inevitably rests on an open dialogue between the implementing agency that knows the population it serves with its microinsurance programme, the commissioning agency that often defines the outcomes of interest to be observed, the research team that operationalises the outcomes of interest into measurable variables, and specialized statisticians and econometricians. As mentioned earlier, a rigorous impact evaluation always requires the contribution of experienced statisticians. Even in the absence of a comprehensive research team in situations where the implementing agency conducts its own impact evaluation, the participation of a statistician is essential. This section identifies some of the key aspects to consider in the sampling process, but can by no means replace the contribution of a statistician in the field.

The first step is getting agreement amongst the various stakeholders on which outcome of interest will be observed. Whilst this may appear to be a simple task to accomplish, in reality this is far from being the case. Implementing and commissioning agencies often aim to produce changes that cannot so easily be observed and measured in the field within the timeframe of an impact evaluation (normally two to three years). Let us take, as an example, a health microinsurance scheme that aspires to reduce maternal mortality by including coverage of facility-based delivery amongst its services. Although maternal mortality rates are still very high in many low- and middle-income countries (LMICs), maternal mortality, per se, is a rare event. This makes capturing differences in maternal mortality due to an intervention extremely difficult, unless the impact evaluation team works in a setting where comprehensive population health surveillance data are available. Similarly, microinsurance targeting agricultural production ultimately aims at reducing poverty, but measuring changes in poverty levels is a very challenging task, since the mere definition of what constitutes poverty in a given setting comes into question.

In these situations, the responsibility of the research team is to engage the implementing and commissioning agencies in a dialogue that allows translating the ambitious objectives of the microinsurance programmes into observable and measurable outcome
indicators. For instance, increases in the utilisation of facility-based delivery can be taken as a proxy of expected reductions in maternal mortality. Changes in asset protection can be taken as a proxy of poverty reduction. In managing this dialogue, it is essential that the research team remains open to exploiting all opportunities for data collection available in a given setting. For instance, in some settings, it may actually be possible to measure changes in maternal mortality even in the absence of formal population surveillance systems. In some African countries, due to pressure to achieve the United Nations’ Millennium Development Goal 5 (MDG5), chiefs and community health workers have been prompted to keep record of all maternal deaths, allowing research teams to access community-based data, which can be very helpful in assessing the impact of those microinsurance programmes that target women. Once the outcomes of interest have been identified and agreed upon by all stakeholders, the research team needs to determine at what level these can be observed: institutional level (e.g., health facility, microinsurance implementing agency, agricultural cooperative, etc.), community, household, or individual level. A common mistake in sampling takes place at this stage, which we will illustrate with an example. Let us imagine that a health microinsurance programme seeks to improve access to care for children less than five years of age. Let us further imagine that the outcome of interest has been operationalised as proportion of children less than five years of age with fever who report to a health facility. Then, at this stage, the research team incorrectly samples households or mothers instead of sampling children with fever. The mistake derives from confusing the object of the sample itself (in this case, children less than five years of age with fever) with the strategy needed to identify and/or to collect information on the actual sample (in this case, households or mothers). This mistake can easily jeopardise the analysis of the entire impact evaluation, if, for instance, 400 mothers instead of 400 children with fever were sampled for the interview. This indicates the need to clearly align the outcome of interest with the applied sampling strategy.

Parallel to identifying the level at which the outcomes of interest can
be observed, the research team also needs to define with the implementing agency the size of the change that the microinsurance programme is expected to produce. Together with the size of the target population and the study design, the value of the expected change (i.e., the expected difference between treated microinsurance recipients and untreated microinsurance non-recipients) is fundamental for the calculation of the relevant sample size.

At this stage, the statisticians have all the information needed to assist the impact evaluation team in the calculation of an adequate sample size. In LMICs, although not always complete, the data compiled at the institutional level is often sufficient to identify an adequate sampling frame for samples drawn at this level (e.g., health facility, microinsurance implementing agency, agricultural cooperative, etc.). On the contrary, drawing community-based samples (at the household or individual levels) can be very challenging in the absence of a comprehensive population surveillance system. In these situations, researchers lack the needed sampling frame and may need to invest first in enlisting all households and individuals in a given community before being able to draw a representative sample from the same. Alternatively, the impact evaluation team may opt for a more pragmatic strategy and identify the complete sample from a number of selected communities using the “spin the bottle method” or “random walk” (Milligan et al. 2004). The trade-off between the two strategies just described is in terms of accuracy and costs, with the former method being more accurate, but substantially more costly. Experimental designs can rely on smaller samples and can, therefore, substantially reduce data collection costs. In addition, although not ideal, experimental designs can lead to a valid estimation of the effect even in the absence of baseline data, offering yet another opportunity to curb field costs. Experimental designs that rely on cluster rather than individual randomisation, however, require larger samples to allow for analysis to account for intra-cluster correlation (the probability that sampled units within a cluster are more similar to one another than sampled units across clusters). In addition, impact evaluations that rely on cluster randomisation design need to ensure that a sufficient
number of clusters are identified and factored into the design. The power of the estimation is directly proportional to the number of clusters included, as well as to the number of units sampled within each cluster.

Compared to experimental designs, quasi- and non-experimental designs, such as RDD and DID, require larger samples and baseline data. The larger sample is justified by a need to rule out the possibility that factors external to the microinsurance programme are responsible for the observed change and this can only be done by controlling for a large number of possible confounders. Propensity score matching can be applied on cross-sectional data, but due to its need to match intervention and control units at the analytical level, it also requires larger samples.

One last important element to consider is potential attrition over time. Impact evaluations frequently rely on longitudinal data collection methods, interviewing the same communities, households, or individuals several times before and after the launch of the microinsurance programme. If resources are available, it is best to include in the original sample a larger number of communities/households/individuals than the minimum number identified by the statisticians. This strategy safeguards the power of the estimation in case a large number of sampled units should be lost to follow-up. If this strategy is not feasible, the impact evaluation team should be ready to consider replacing households lost to follow-up in a sample. Again, this process needs to be assisted by experienced statisticians and demographers.

11.5. Ethical considerations

In recent years, it has become impossible to conduct an impact evaluation without first obtaining ethical clearance from an ethics committee affiliated with an academic institution or institutional review board (IRB). IRBs are independent ethics committees or ethical review boards, designated to approve, monitor, and review biomedical and behavioural research involving humans. Requirements vary greatly across boards, but, generally, impact evaluation teams are required to compile documentation, including a detailed description of the research protocol, study tools, informed consent forms, and the resumes of all concerned investigators. Most are also required to describe the process of selecting respondents and ensuring that consent is taken from the respondents or from their legal representatives. Depending on the complexity of the research protocol, ethical clearance is obtained in a period of time that ranges from three to six months. Most IRBs charge for their services.
and academic IRBs generally charge a nominal fee ranging from US$2000 to US$3000 per submission. Commercial IRBs can have substantially higher prices.

Impact evaluation teams affiliated with academic institutions are normally required to obtain clearance from the ethical review board at the same institution for which they work. Most universities in Europe and in North America have developed internal structures that work very rigorously to comply with highest international ethical standards. Impact evaluation teams who are not affiliated with an academic institution normally obtain ethical clearance from commercial ethical review boards. This chapter purposely does not include any reference to any commercial IRBs because we do not wish to advertise one service provider over another. We leave it to the interested reader to look for such services online, making his/her own decisions on the rigour of the process proposed to obtain clearance.

Most LMICs have also established their own independent IRB. Application procedures and payment conditions vary greatly across countries. Impact evaluations that are funded and conducted exclusively at the local level can apply directly to country-specific IRBs. Impact evaluations that receive international funding and/or are supported by international teams normally have to apply for ethical clearance both in the country where the impact evaluation team is based and in the country hosting the impact evaluation.

It is often difficult to understand the need for ethical clearance in the case of studies that do not deal with medical products or otherwise potentially toxic substances. The implicit assumption of impact evaluation teams is that their work will result in no harm. Research teams involved in impact evaluation, however, need to reflect carefully on such assumptions to guarantee that their work results in no harm and that, if possible at all, their work yields benefits for those with greater needs. Even impact evaluation rooted in the
social sciences can raise a number of justifiable ethical concerns, especially when some form of randomisation (whether at the community or at the individual level) and/or purposeful targeting takes place. Both randomisation and purposeful targeting into microinsurance entail that a selected group of communities/households/individuals enjoys the benefits of a programme, whilst others do not. Step wedge randomisation represents an effort to reduce potential sources of ethical conflict, since it ensures that the benefits accrued by one group will be shared with all others within a foreseeable period of time. Both randomisation and purposeful targeting have to be communicated very clearly to the concerned communities. Risks and benefits of such designs have to be discussed openly and in simple terms to allow target populations to appreciate the long-term benefits and support the intervention (Marshall 2007).

An additional ethical concern arises from the fact that, through the various study components embedded in an impact evaluation, extensive data on socioeconomic profile, health, and economic behaviours of communities/households/individuals will be collected and stored in databases. To allow for the identification of the same sampled units over time, in the case of longitudinal studies, data needs to be stored so that it can be traced back to specific households and/or individuals. This creates a privacy protection concern. To protect privacy, access to the full dataset containing actual names and/or otherwise identifiable ID codes is normally restricted to one or two people in each impact evaluation team. The dataset is then usually cleaned of all personal information before being used for analysis by the rest of the study team. A small compensation, usually in-kind, is frequently offered to households or individuals who participate in long or repeated interviews. In traditional communities with a strong collective orientation, this compensation needs first to be discussed with the community leaders. If community leaders are not consulted, compensation can create conflicts in the community.
11.6. Dialogue with stakeholders and knowledge transfer

At different points throughout our discussion, we have drawn the reader’s attention to the need to engage in an open dialogue with the implementing and commissioning agencies as the only means to ensure a sound, yet feasible, impact evaluation. In the last section, we will draw attention to the overall sociopolitical context within which impact evaluation of microinsurance programmes takes place. We explained at the very beginning that impact evaluations are not conducted with the sole goal of assessing a causal relationship between a given intervention and a set outcome of interest. Through the identification of this causal link, impact evaluations ultimately intend to influence and shape policies. Furthermore, sound impact evaluations respond to the ultimate single objective of improving the living conditions of the concerned communities.

Striving to influence policy requires acknowledging that impact evaluations do not happen in a sociopolitical vacuum and that impact evaluation teams need to engage with a number of politically relevant stakeholders, beyond those directly implementing and/or commissioning the evaluation. Relevant stakeholders include policy makers, national and international agencies involved in neighbouring and overlapping programmes, and, last, but surely not least, the communities themselves. It is important that dialogue with all concerned stakeholders is initiated before the impact evaluation even takes place and is maintained throughout the course of the fieldwork. Consensus can be created by incorporating the concerns of all relevant stakeholders into the design of the impact evaluation and by ensuring that results are fed back at every stage of data collection. Effective knowledge transfer takes place only when communication strategies are adjusted to the specific constituency being addressed at a specific point in time. In addition, open dialogue allows the impact evaluation team to understand what other programmes may take place in the same region, accounting for potential confounding factors when estimating the effect of the microinsurance intervention.
Impact evaluation teams can rely on informal community gatherings to exchange with the communities themselves. For instance, in the Nouna Health District in rural Burkina Faso, the Centre de Recherche en Santé de Nouna (CRSN) holds regular meetings with village chiefs and traditional authorities to ensure that the content of the research activities it leads is properly communicated to the communities and their concerns incorporated into future research. This constant, open dialogue empowers the population, who feel that the work of the CRSN is not only feeding the interest of the political decision makers, but is addressing actual concerns at the community level. The health microinsurance programme, amongst others, was a result of previous studies conducted by CRSN that identified lack of financial resources as one of the leading causes explaining low health-care utilisation. Empowerment translates into community support to research, opening the door to experimental, quasi- and non-experimental designs that are appreciated for their potential to improve the living conditions of the population in the longer term.

Policy briefs, official dissemination meetings, and other formal knowledge brokering activities are more often used to exchange information with policy makers and concerned national and international non-governmental agencies. Scientific publications, either in the form of peer-reviewed articles or discussion papers, represent only a limited source of knowledge transfer from the academic to the policy making community. Scientific publications are largely needed because they allow for information to be shared beyond the borders of the setting where the impact evaluation took place. As such, they may create interest in microinsurance in other settings or countries, but the academic language they use may at times be inadequate to inform policy decisions directly at the local level.

11.7. Conclusions

Impact evaluations should be designed well in advance, ideally already whilst the microinsurance programme is being designed. Early planning allows the impact evaluation team to integrate the evaluation in the launch of the programme and develop the best available methodology to measure impacts. It provides time to plan and collect baseline data prior to the start of the programme.

A team best suited to evaluate impacts must consist of field experts and experienced statisticians. Regular consultations with the microinsurance implementing team will ensure that field realities are taken into consideration in the design of the evaluation. The implementing team should be consulted when the research methodology and the sampling strategy are being
developed, as they should fit with the field setting whilst allowing for the best evaluation design to be implemented.

When designing, conducting, analysing, and disseminating the results of the impact evaluation, it is vital that the team adheres to all ethical guidelines and regulations. Impact evaluations not only provide evidence for the existence of impact, but also provide an opportunity to guide policy and influence the living conditions of the communities. This requires that the impact evaluation results be disseminated to all stakeholders—not only policy makers and international organisations—including communities that are at the centre of these programmes.

References


Results of Impact Assessment Studies