MILK Brief #6: Research designs for measuring the client value of microinsurance[1]

Karlijn Morsink[2] and Peter Geurts,[3] Institute of Governance Studies (IGS) University of Twente, The Netherlands

The value of microinsurance products can be studied in a variety of different ways. Experiments, specifically randomized control trials (RCTs), are viewed by parts of the academic community as the "gold standard" technique in research designs. Sometimes, however, the type of product being studied, the questions a study seeks to answer, or other practical considerations make it difficult, impractical, or even impossible to use an RCT. In other cases, even if an RCT is feasible, another technique may be better suited to the study's objectives. This brief provides a structure to support decisions about appropriate research designs to answer questions about the value microinsurance has to clients.

We begin by describing what a "perfect" research design seeks to achieve, and what it requires. We then discuss practical considerations that can limit researchers' ability to implement perfect studies, and end with a discussion of alternative techniques in research designs. A brief glossary of important research terms is attached.

The impossibility of perfect research

To provide rigorous evidence of the impact of microinsurance on clients, a study must have internal validity (Dalal, Bauchet and Morduch, 2011:8), as well as external, construct and statistical conclusion validity (Shadish, Cook and Campbell, 2002:38). The box to the right defines these four validities and provides examples. It is generally acknowledged that well-designed double-blind experiments (RCTs) are the best design to establish internal validity in social and behavioral research, including research on the impact of microinsurance. However, for reasons discussed below, they may not always be the best way to establish the other types of validity, which are required for 'perfect' research designs.

Four types of validity: Examples

A study concludes that enrollment in a health microinsurance program causes female clients in rural Nigeria to visit the doctor sooner when they are sick. The study's:

Internal validity is the confidence we have that the impact (visiting the doctor sooner) is caused by the insurance and not by some other factor (e.g., that those who enrolled were generally more cautious about their health)

External validity is the extent to which this finding can be generalized to other areas, clients, and products (e.g. men in Latin American cities)

Construct validity is the confidence we have that the indicator (the response to a survey question about how long the respondent waited to visit the doctor) accurately represents the intended concept (the actual amount of time)

Statistical conclusion validity is the confidence we have that the study's statistical methods were used correctly to draw the conclusion

Contextual, qualitative factors should also inform research design and interpretation of results. *Portfolios of the Poor* (Collins et al., 2009) demonstrated the complexity of seemingly straightforward questions about the financial lives of the poor and the importance of gaining an in-depth understanding of these questions. While RCTs *can* be designed in a way that takes qualitative and contextual considerations into account, they do not lend themselves easily to answering these types of questions, and in fact often fail to do so.

[1] This brief summarizes issues discussed in Morsink, K., and Geurts, P.A.T.M. (forthcoming) Research designs for measuring the impact of microinsurance on low income households.

[2] Corresponding author. Tel. + 31634140055. E-mail address: k.morsink@utwente.nl

[3] With Barbara Magnoni and Emily Zimmerman of the MILK Project

Practical and methodological considerations

The questions asked

The first step in developing a research design for studying the impact of microinsurance on clients is to define the study's broad objectives and research questions, grounded in theory from existing literature. Focusing narrowly on the RCT as the only option for conducting rigorous research before considering what we seek to learn from the study can lead us to exclude important questions because they cannot easily be answered with an RCT. The type of product, the questions we have about it, and the context in which it is studied should guide decisions about which impact indicators are appropriate and how they can best be measured.

These considerations are especially important in ensuring that a study has construct validity; drawing an internally valid conclusion about the wrong concept will only blur our understanding of client value and impact. For example, a common measure for the impact of insurance on economic vulnerability is a household's out of pocket payments. For one household these payments may come out of household income or savings, while another household may have to sell a productive asset such as a cow. We might expect the latter to have a greater impact on a family's vulnerability than the former, but the measure of out of pocket payments would not pick up the distinction. In such a case, a measure of out of pocket payments has low construct validity in relation to the concept of economic vulnerability.

Research designs must also consider the distinction between ex-ante and ex-post effects of insurance.[4] Ex ante effects are those felt regardless of whether a shock occurs (for example the peace of mind a breadwinner feels by knowing that her life insurance policy will provide for her family if she dies, or a farmer's willingness to invest in fertilizer because he is partially protected by insurance from the consequences of a drought). To observe ex-ante effects, only insurance coverage, and not the occurrence of insured events or the payment of claims, needs to be adequately represented in the sample (Cai, Chen, Fang, & Zhou, 2010). The ex-post effects of insurance (for example its impact on utilization of healthcare services or its effectiveness in protecting a family from the financial consequences of a breadwinner's death) can only be observed after the occurrence of the insured event. One of the most significant challenges in measuring ex-post effects is finding enough people, both insured and uninsured, who have experienced the risk event. This challenge is especially significant when studying less frequent risk events such as disasters, death, disability, and property damage (e.g. Morsink, Geurts & Kooijman-van Dijk, 2011).

Effectiveness and efficiency of designs

Effectiveness and efficiency of a research design must always be taken into account. In practice, research always requires tradeoffs between methodological rigor and feasibility in terms of cost, logistical capability, and time, keeping in mind the practical applicability of the research results. Products in the early stages of design and implementation can be expected to undergo many large and small changes that will alter their value to clients. For evaluations of non-mature products, costly and time-intensive rigorous studies can be of scientific interest but are, from a practical perspective, less appropriate for those implementing the program. In addition, research designs using RCTs are often prohibitively costly, particularly where sample sizes must be very large (in many cases up to several million dollars). Simpler, less expensive RCTs using only administrative data rather than surveys may cost less, but provide less depth of information. Other, less expensive research designs, relying on quasi-experimental or qualitative approaches, may be a more practical way to achieve a study's objectives while maintaining methodological rigor.

Ethical concerns

Microinsurance providers interested in studying the impact of their products may find randomization of the

[4] MILK separates value into three types: **expected** value (behavioral incentives and peace of mind, which are ex-ante effects), **financial** value (ex-post effects) and **service quality** value (through access to product-related services, which are generally also ex-post effects). For a description of MILK's approach to understanding these three types of value, see MILK's Client Value Summary, available here.

insurance treatment (which experiments require) in their potential client market unethical or in conflict with institutional objectives, especially as it may affect their trustworthiness in the field. Even when ethical considerations do not exclude the possibility of an RCT entirely, they limit the effectiveness of its design. The ideal design of an RCT would be double-blind, keeping both the researchers and the subjects unaware of who receives the treatment. However, a double-blind design is not feasible for any study of the effects of having insurance (versus no insurance) because the ethical rule of informed consent requires participants to be aware of coverage, and because as a practical matter participants must be aware of their insurance coverage to use it. Therefore most RCTs studying insurance are still not "perfect" experiments (Radermacher et al., 2009). When a study is not double-blind, researchers may be more inclined to interpret uncertainties in accordance with their expectations (i.e. to support of the study's hypotheses) (Shadish, Cook and Campbell, 2002:441).

Alternative research designs and methodological solutions

While RCTs remain the "gold standard" in rigorous research design in many instances, they can be infeasible or undesirable due to the considerations outlined above. Alternative research designs can be employed to study the products, answer the questions, and observe the contexts that RCTs cannot. This section describes several such alternatives and outlines the methodological considerations needed to maintain rigor in these alternative studies.

Quasi-experiments

In the case of rare risk events, resource constraints or other limitations, quasi-experiments may be the most appropriate research design. Quasi-experiments are similar to experiments in that they also take a sample from the population and have treatment and control groups but are different because the treatment and control conditions are not randomly assigned. Careful **sampling** of control and treatment subjects is important.[5] When studying low-frequency risk events, random assignment is rarely feasible. Stratified random sampling of respondents from the population can provide sufficient control and treatment study subjects, and the results can be generalized to the general population by taking the ratios of sample versus population into account. **Controls for self-selection and endogeneity** are also particularly important in quasi-experiments.[6] When these problems are adequately addressed, the study's internal validity comes close to that of an RCT. An advantage of quasi-experiments is that they do not rely on artificial choice circumstances, which allows for inclusion of **context factors** from different levels through multi-level analysis (Snijders and Bosker, 1999) (e.g. showing effects of variables at district, community, and household levels on individual respondents) and for fewer threats to external validity (applicability in other contexts).

Qualitative designs

Qualitative methods such as focus group discussions and key-informant interviews are not well-suited to provide external validity and statistical conclusion validity. They are, however, useful to obtain information about the scope, objectives, and theory of change and thus can contribute to internal and construct validity. They are valuable on their own or as a supplement to quantitative data (Leeuw en Vaessen, 2009: xiv). Qualitative work can be used to increase the construct validity and external validity of experimental and quasi-experimental studies. It can inform the design of studies by helping us to understand what may potentially influence client value, and can also help to explain unexpected findings in quantitative studies and provide context to those findings.

Client math

The MILK Project's Client Math studies combine quantitative and qualitative techniques (mixed methods) to explore the impacts of different microinsurance products. They begin by developing a theory of change (expected impacts of insurance) through an examination of existing literature, product details, staff interviews, and/or focus group discussions with clients. The next step is field studies that consist of ex

[5] This can be achieved through procedures of matching such as propensity score matching on single or double differences (see glossary for definitions of these terms).
[6] See techniques of Heckmann (Heckmann, 1978, 1979) or Altonji (Altonji, Elder and Taber, 2000)

post documentation, through detailed surveys, of the responses to shocks reported by people who had and did not have insurance. Because they do not use random assignment of insurance coverage and because they seek out respondents after a financial shock takes place (rather than waiting for shocks to happen), they can provide insight on how insurance is used for low-frequency risks. The quantitative component of the studies typically does not focus on providing **statistical significance** or **causal adequacy**, but it does offer the ability to measure the size of a financial shock and other non-financial effects, such as externalities on a range of people that are selected randomly. They offer insight into how different types of low income people respond to financial shocks with formal and informal financial tools, and provide a measurable "math" formula that can be compared across population segments, products and countries. While individual Client Math studies tend to have relatively low internal validity, internal validity can be increased by aggregating findings across studies through so-called meta-analysis. Client Math studies also lend themselves to consideration of qualitative, contextual factors and a deeper understanding of how insurance works in practice for low-income policyholders, because they involve small numbers of in-depth interviews. As in quasi-experiments, careful random **sampling** of the treatment and control groups is essential. The importance of careful sampling is intensified in Client Math studies because they involve small sample sizes (50-60 people).

Moving Forward

Any researcher who seeks to study the value of a microinsurance product should begin the process with a careful exploration of theory, the objectives of the study, the context in which it will be carried out, the audience, and the relevant practical constraints, keeping in mind the pros and cons of different available research methods summarized in the table below. This exercise may lead to a determination that an RCT is the most appropriate technique, but it might instead lead to quasi-experiments or qualitative studies, or to a design such as Client Math that combines several techniques. Beginning with the objectives, theories of change, context, and practical considerations rather than the research technique ensures that we don't exclude important questions when RCTs are not sufficient to answer them, and that we use the best tools available to study those questions.

Pros and Cons of Well-designed Alternative Designs

	Pros (+)	Cons (–)
RCT	High internal validity; high statistical conclusion validity	Often low external and construct validity; usually high cost; often infeasible; ethical constraints
Quasi-Experiment	High internal validity; high external validity; high statistical conclusion validity; feasible for most products; incorporation of context factors	Challenges to internal validity without appropriate controls; medium cost (generally much cheaper than RCTs); often low construct validity
Client Math	Low cost; feasible for most products; high construct validity; internal validity (through meta-analysis)	Challenges to internal validity without meta-analysis. Sampling requires precision; low external and statistical conclusion validity
Qualitative	High construct validity; potential for high internal validity within particular settings; can be low cost; feasible for all product types and stages; add context to other studies	Low or no external validity and no statistical conclusion validity

Glossary of Research Terms

Adverse selection or ***self-selection*** – Tendency for riskier individuals to be more likely to purchase insurance, which can skew the measured outcomes; can influence internal validity.

Control group – In an experiment, a comparison group that do not receive a treatment (i.e. those without insurance coverage).

Double-difference – Sampling technique that uses data on project and control observations before and after the insurance intervention to "difference out" heterogeneity in participation in the program (Khandker, Koolwal and Samad, 2010: 71).

Endogeneity – Independent variable is correlated with the error term in the regression model (commonly because there is an omitted variable that effects the independent variable and dependent variable or because there is a measurement error in the dependent variable); can influence internal validity.

Ex-ante impact – Impact of insurance before the risk event (e.g., increased investment because of higher sense of security).

Ex-post impact – Impact of the insurance after the risk event occurs, often the effect of the payout or of using covered services.

Meta-analysis – A set of quantitative methods for synthesizing a set of related research hypotheses.

Multi-level analysis or ***hierarchical analysis*** – Regression analysis taking into account the different aggregation levels at which

the variables have their effect (e.g., characteristics of villages are likely to influence households in each particular village).

Propensity score matching – Sampling technique that constructs a statistical comparison group based on a model of probability of participating in the treatment, using observed characteristics. Respondents are matched on the basis of this probability, or propensity score, to non-respondents. The average treatment effect of the program is then calculated as the mean difference in outcomes across these two groups (Khandker, Koolwal and Samad, 2010: 53).

Quasi-experiment – A study design that is similar to an RCT but lacks random assignment of the treatment and control.

Randomized control trial (RCT) – Research design in which study subjects are randomly assigned to alternative treatments (in a study of microinsurance, the treatment would be insurance coverage). An RCT is ***double-blind*** if both the researchers and the subjects are unaware of who is receiving the treatment. Double-blind studies of insurance are typically not feasible.

Research design – The entire set up of a study, including the research objective, the theory of change, the construction of the variables, and the selection of a technique (such as an RCT) with which to carry out a study.

Stratified random sampling – Random sampling of respondents within theoretically relevant categories of the population (e.g. example sampling respondents in more or less risky locations).

Treatment group – In an experiment, the group that receives the intervention or treatment (i.e. those who have insurance).

References

Altonji, J.G., Elder, T.E. and Taber, C.R. (2000). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. Paper presented at the Winter meetings of the American Economic Association.

Cai, H., Chen, H., Fang, H., Zhou, L. (2010) The effect of microinsurance on economic activities: Evidence from a randomized natural field experiment.

Collins, D., Morduch, J., Rutherford, S., & Ruthven, O. (2009*). Portfolios of the poor: How the world's poor live on two dollars a day*. Princeton, NJ: Princeton University Press.

Dalal, A., Bouchet, J. and Morduch, J. (2011). Evaluation fundamentals. Chapter prepared for the ILO Microinsurance Impact Evaluation Practical Guide. Financial Access Initiative Research Framing Note.

Heckman, J. (1978). Dummy endogenous variables in a simultaneous equations system. *Econometrica 46,* 931-961.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica 47*, 153-161.

Khandker, Shahidur R., Gayatri Koolwal, and Hussain Samad. (2009). Handbook on impact evaluation: Quantitative methods and practices. Washington, DC: World Bank.

Leeuw, F. and Vaessen, J. (2009). Impact evaluations and development. NONIE guidance on impact evaluation. Washington, DC: World Bank.

Morsink, K. Geurts, P.A.T.M. & Kooijman-van Dijk, A.L. (forthcoming) Impact of micro insurance on vulnerability of low income households in the Philippines: the case of typhoon re-housing insurance.

Morsink, K., and Geurts, P.A.T.M. (forthcoming) Research designs for measuring the impact of microinsurance on low income households.

Radermacher, R. Ashok, S. ,Zabel, K. and Dror,I. (2010). What do we know about the impact of microinsurance? Paper submitted for discussion for the Annual International Microinsurance Conference 2009 in Dakar.

Shadish WR, Cook TD, Campbell DT. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton-Mifflin. Boston.

Snijders, T.A.B., and Bosker R.J (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London : Sage Publishers, 1999.

Microinsurance Learning and Knowledge (MILK) is a project of the MicroInsurance Centre that is working collaboratively to understand client value and business case in microinsurance. Barbara Magnoni leads the client value effort and Rick Koven leads the effort on the business case. Contact Michael J. McCord (mjmccord@microinsurancecentre.org), who directs the project, for more information.