Cordaid's experience with impact evaluation

Lenfant, F.; Rutten, Rens

Publication date
2013

Published in
Development in Practice

Development in Practice

ISSN: 0961-4524 (Print) 1364-9213 (Online) Journal homepage: http://www.tandfonline.com/loi/cdip20

Cordaid's experience with impact evaluation

Francois Lenfant & Rens Rutten

To cite this article: Francois Lenfant & Rens Rutten (2013) Cordaid's experience with impact evaluation, Development in Practice, 23:1, 46-56, DOI: 10.1080/09614524.2013.753412

To link to this article: https://doi.org/10.1080/09614524.2013.753412

Published online: 18 Feb 2013.

Development in Practice, 2013
Vol. 23, No. 1, 46–56, http://dx.doi.org/10.1080/09614524.2013.753412

Routledge
Taylor & Francis Group

Cordaid's experience with impact evaluation

Francois Lenfant and Rens Rutten

In 2007, Cordaid started a pilot on participatory impact assessment, intended to enhance accountability and to improve learning. The methodology was based on quasi-experimental design, complemented with qualitative research. This case study illustrates the challenges INGOs and their partners face in their attempt to find a rigorous yet, relevant, useful, and socially acceptable methodology for evaluation and impact assessment purposes. While most local NGOs participating in this pilot consider (parts of) this methodology useful for their learning, this approach proves unsuitable, costly, and inappropriate for an INGO such as Cordaid since it does not respond to its own accountability and learning needs.

Cortaid et leur expérience avec l'évaluation d'impact
En 2007, Cordaid a lancé un programme pilote sur l'évaluation participative de l'impact, dont l'objectif était d'améliorer la redevabilité et l'apprentissage. La méthodologie se basait sur la conception quasi-expérimentale, assortie de recherches qualitatives. Cette étude de cas illustre les défis que les ONGI et leurs partenaires doivent relever au moment de trouver une méthodologie rigoureuse, mais aussi pertinente, utile et socialement acceptable aux fins des évaluations en général et de celles de l'impact en particulier. Si la plupart des ONG qui prennent part à ce programme pilote considèrent que (certaines parties de) cette méthodologie est (sont) utile(s) pour leur apprentissage, cette approche se révèle inadaptée, coûteuse et peu appropriée pour une ONGI comme Cordaid, car elle ne répond pas à ses propres besoins en matière de redevabilité et d'apprentissage.

Cortaid y su experiencia con la evaluación del impacto
En 2007, Cordaid inició una experiencia piloto sobre evaluaciones de impacto participativas, cuyo objetivo consistió en mejorar la rendición de cuentas y el aprendizaje. La metodología utilizada se fundamentó en un diseño semiexperimental, que fue complementado con investigaciones cualitativas. El presente estudio de caso ilustra los retos enfrentados por las ONGI y por sus aliados en su intento por encontrar una metodología rigurosa, pertinente, útil y socialmente aceptable para fines de evaluación y de valoración de impacto. Si bien la mayoría de las ONG locales participantes en esta experiencia piloto consideró que esta metodología, o partes de la misma, era útil para su aprendizaje, el método demostró ser inadecuado, costoso e inapropiado para una ONGI como Cordaid, debido a que no atendió sus propias necesidades en las áreas de rendición de cuentas y de aprendizaje.

Cortaid e sua experiência com avaliação de impacto
Em 2007, a Cordaid iniciou um projeto-piloto sobre avaliação participativa de impacto destinado a melhorar a prestação de contas e aperfeiçoar o aprendizado. A metodologia foi baseada em um modelo quase-experimental, complementado com pesquisa qualitativa. Este estudo de caso ilustra os desafios que as ONGIs e seus parceiros enfrentam em sua tentativa de encontrar uma metodologia rigorosa, porém relevante, útil e socialmente aceitável para fins de avaliação e análise de impacto. Embora a maioria das ONGs locais que participam deste projeto-piloto considere parte dessa metodologia útil para seu aprendizado, essa abordagem mostra-se insustentável, cara e inapropriada para uma ONGI como a Cordaid pois não atende às suas próprias necessidades de prestação de contas e aprendizado.

KEY WORDS: Aid – Monitoring and Evaluation; Methods

Introduction

In the past decade, there have been repeated calls to the development community in general and INGOs in particular to show the results of their work. Given the often significant budgets managed by some INGOs as well as the public nature of their funds, these calls appear more than legitimate. Furthermore, in the midst of declining budgets, especially in the Netherlands, and aid fatigue, there is increasing pressure to search for innovative and compelling ways to show and prove one's interventions are directly related to tangible results. In the Netherlands for instance, funds channelled for international cooperation through INGOs declined from €2.2 billion for the period 2007–2010 to €1.8 billion for the period 2011–2015, and are likely to suffer additional cuts during this period. The Dutch co-financing agreement for the period 2011–2015 requires a rigorous evaluation methodology including baseline, control group/reference group, and a representative sample of projects to be evaluated. The Dutch Ministry and the Policy and Operations Evaluation Department assume that this will give strong evidence of the level of effectiveness of the development programmes supported through Dutch civil society organisations. The demand from the Dutch government to use more rigorous evaluation methodologies coincides with a current trend in international evaluation circles whereby a hierarchy of methods has emerged, placing experimental and quasi-experimental options for impact evaluation highest on the ladder. According to this method, "*an impact evaluation is a study which tackles the issue of attribution by identifying the counterfactual value of Y (Y0) in a rigorous manner*" (White 2009: 4). This definition of impact is thus different than the more generally used definition of OECD-DAC which states that "*impacts are positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended"* (OECD/DAC 2002: 24).

It is in this context that Cordaid made an attempt to improve its M&E systems by experimenting with a participatory impact assessment methodology. The experiment consisted of implementing a rigorous method for impact assessment for six projects in three different countries. The methodology was based on a difference-in-difference analysis, comparing the outcomes after and before the intervention, and comparing the "treatment" group with a "control" group, a so-called quasi-experimental design. An experimental design was not feasible, since the start of the pilot did not coincide with the start of the projects. It would have been possible in one or two projects, but the set-up of the pilot was to test the same methodology in the six different projects. This article reflects the results and lessons learnt from this pilot.

Cordaid and the pilot of Participatory Impact Assessment (PIA)

Objectives of the pilot

In May 2007, Cordaid and CIDIN (Centre for International Development Issues Nijmegen) started a pilot for PIA. Cordaid is a Dutch INGO working on international cooperation in different thematic areas (health, income generation, conflict transformation, disaster risk reduction, and emergency aid) in more than 30 countries with over 800 local partner organisations. CIDIN is a Dutch interdisciplinary academic institute addressing issues of inequality, poverty, development, and empowerment. Cordaid, which has been working together with CIDIN on researching particular development issues for a few decades, felt the need to develop a new and innovative methodology for impact evaluation for two main reasons. First, to address the issue of scarcity of information on outcome and impacts found in M&E reports of partner organisations. The lack of comparison (baseline data or triangulation) was considered a limitation which hampered the gathering of reliable data on whether development interventions set out by partners and supported by Cordaid were effective or not. The second reason was related to the growing political and public pressure for evidence-based results. CIDIN offered to support Cordaid (and the NGO sector in general) in conducting scientifically-based impact assessment to make an unbiased appraisal of the aggregate value of the standard welfare measures. The experimental and quasi-experimental designs for impact evaluation had so far mainly been used for large scale interventions with rather simple, measurable indicators; for an application in the NGO context, more complex elements, such as the analysis of attitudinal response[1] and group learning were included. Other sub-objectives of this pilot were to enhance the capacities of participating partner organisations in using impact assessment as an instrument for monitoring, learning, accountability, and innovation; and to strengthen relationships between local development NGOs and local research institutes, building closer links between practice and research. More generally, Cordaid's expectations with this pilot were to develop a new methodology for the assessment of impact, and to use the outcomes for accountability purposes, as well as for learning how development interventions supported by an international actor can lead to impact at the level of people's lives.

Methodological approach

The first step of the pilot was to choose a methodology. CIDIN proposed a methodology based on difference-in-difference analysis, in which one compares the differences in outcomes after and before an intervention of the control and the treatment group. In the next step, the selection of countries and projects, it was decided to focus on countries that are relatively stable and projects that consist of interventions with measurable indicators.[2] The selected countries were Peru, Ghana, and India. The selection of partner organisations in these countries was based on their thematic focus and on the type of interventions. For each country we chose a specific thematic focus to test the methodology. Partner organisations with direct interventions at community level were invited to express their interest in joining this pilot. Each of the invited partner organisations mentioned interest in participating, which led Cordaid and CIDIN to proceed with a selection based on practical criteria, such as accessibility and coverage. The availability of control groups was not a criterion for the selection of projects. The projects implemented by the selected partners, i.e. the object of the pilot, were not in all cases directly linked to Cordaid financing, since Cordaid generally supports capacity building and civil society strengthening activities rather than direct poverty alleviation interventions. Selected partner organisations were: two partners in Ghana working in health care, two partners in India with income generating projects and self help groups for women, and two partners in Peru

implementing water and sanitation projects in urban areas. Subsequent workshops were held together with partner organisations and local research institutes (which were eventually invited on board) to explain the methodology and review its overall design. While CIDIN developed the methodological framework, partners were consulted through workshops to define indicators for the surveys, based on the intervention logics of the specific projects, as well as to select appropriate control groups.[3]

After these workshops, surveys were developed and carried out among 400 households (200 control and 200 treatment) in each project location (which included 2,400 households in total). The approach was further supplemented with in-depth interviews and experimental methods, such as a variation on the public goods games to measure trust. The number of surveys was predetermined as part of the methodology, which was driven by the need to collect sufficient data in order to do a sound statistical analysis. Propensity score matching[4] was also used to make the results more reliable. The methods and samples applied were similar in the six project locations.

The results of the surveys were discussed individually with each partner organisation. In one project area in Ghana, the results of the surveys were discussed with the treatment and control groups, in order to understand the differences found, such as the higher incidence of malaria in the catchment area of the hospital compared to the control group. The pilot was labelled "participatory" because partner organisations were involved in the design of the surveys, in the selection of control groups, and in the discussion of the findings. It was also suggested to strengthen the capacities of the partner organisations in using impact evaluation as an instrument for monitoring, learning, accountability, and innovation. The final decision on the selection of the indicators and the control group, however, was the responsibility of CIDIN. In addition, partner organisations did not conduct the surveys nor did they process the data. In some locations however, partner organisations' staff participated actively in the qualitative fieldwork, which led to additional insights into why an intervention was successful in a specific context. In 2010, two and a half years after starting the pilot, a meeting was organised with all partner organisations, research institutes, Cordaid, and CIDIN to discuss the results thus far and share experiences. The pilot was planned for a five year period and has also been used for PhD research, which resulted in a doctoral thesis on how context matters for the effectiveness of development programmes (see De Hoop 2011).

Insights into the methodologies applied and outcomes for partner organisations

Typically, the monitoring practice of partner organisations consists of gathering general process and output data such as activities realised, number of people reached, or services provided. For evaluation purposes, in-depth interviews are sometimes done at beneficiary level, through focus group discussions for instance. For partner organisations, using control groups and conducting a large number of longitudinal surveys at beneficiary level was a new way of dealing with M&E. Conducting household surveys based on sampling was also new to most partner organisations. More generally, the pilot helped generate information on outcome at household level and assisted partners in using baseline data systematically before starting a project. For instance, in the last year of the pilot, one Peruvian partner used a baseline study for a new project which included a health survey among 400 randomly selected households as well as health measurements taken from 250 pre-school children.

Findings of the six projects are presented in the Appendix. In India, for example, the women empowerment project showed that self-help group membership leads to higher degrees of autonomy, but that enhanced autonomy has not – on average – resulted in higher subjective well-being. Qualitative research revealed that this was due to social sanctioning, especially in conservative villages. In Ghana, healthcare utilisation by the poor increased while their

reliance on home remedies, self-medication, and traditional healers reduced. In Peru, community participation in water supply works fostered social cohesion. On three subjects, results have been published in international peer-reviewed journals.

Critical analysis of the methodological principles of the PIA pilot

The use of control groups was one of the major discussion points before starting the pilot, mostly because of ethical concerns. In preliminary discussions, Cordaid and a few partner organisations expressed their reluctance to work with control groups, but this was a non-negotiable element of the methodology. It was therefore decided to consider the control group, if possible, as a future treatment group. In addition, Cordaid and partner organisations always have to make choices on where to intervene, i.e. who the beneficiaries are. These choices are typically based on a situation analysis where elements such as needs assessments, partner's track record, conformity with policy choices, and other factors, are carefully weighed. With regards to this pilot, the control groups were not purposely denied the interventions. The control groups were households in adjacent areas in which the partner organisations were not working. These people may become future beneficiaries of partner organisations and, in many cases, are reached by other organisations through other interventions. The fact that these control groups benefit from other development interventions reduces the ethical concerns, but might raise questions regarding the rigour of this method. As Barahona (2010: 12) points out:

> "... the way in which development agencies balance the use of their resources depending on a range of factors (e.g., their knowledge or perception of what is going on in specific locations, the history of their interventions, their local contacts, the presence and activity of other agencies) makes the use of control groups as counter factual problematic. The argument is even weaker where a development intervention contributes to, but does not cover all, the activities that an implementing agency carries out in a particular area or with a specific population. In these cases, the possibility of establishing a counter factual using a control group is almost nonexistent and in the context of development interventions the assumption that "if my intervention had not happened nothing else would have happened" is unrealistic and arrogant."

The impact assessment focused on interventions at community and household level, in order to make the use of a control group possible, while these interventions were not the only, nor the most important, interventions carried out by partner organisations. Spillover was mentioned by one of the Indian partners as a problem in the selection of control groups. It argued that the success of its intervention resulted in control villages copying its technological innovations and products.[5] The control groups for the Ghana cases were households that live further away from the diocesan health services (and therefore may not be a good comparison) and/or make use of other health care services. Nevertheless, the comparison with control groups did reveal some surprising findings, such as higher incidence of malaria in treatment groups in Ghana, which could partly be attributed to a reduced incentive for prevention when access to care is better. Partner organisations recognised the use of control groups as an interesting method, which may have potential for future use in M&E. From a learning perspective, it is an important exercise to think about one's reference group or reference activity, and bad choices of control group can be stimulating in this regard.

The importance of using mixed methods in impact evaluation is now strongly recommended by the International Initiative for Impact Evaluation (White 2008). In the PIA pilot, a quantitative methodology, consisting of surveys with 400 households, was combined with qualitative focus group discussions. However, since surveys were the dominant method and focus group

discussions were only done in a few cases, the mixed nature of PIA can be questioned. It is only in the course of the pilot that more focus was placed on qualitative research. The use of surveys for 400 households per project was the predefined method which allowed for sufficient data collection, but was not based on specific research questions directly related to the type of interventions under scrutiny. This raises a central issue in relation to the methodology used in this pilot. PIA does not respond to specific evaluation questions – not from the perspective of partner organisation or other stakeholders involved in the intervention or project, nor from the perspective of Cordaid. The design of M&E frameworks should match the nature, needs, and constraints of the intervention. In the mid-term evaluation with all PIA stakeholders, it was concluded that in order to avoid misleading results, it is key to start out with an in-depth study of the intervention plan and the intervention area. Cordaid should have based its approach on the evaluation questions, which may have resulted in a different methodological framework.

Analysis of the usefulness of PIA for the partner organisations

It must be stressed that partner organisations started this pilot somewhat reluctantly, although they all volunteered to join. However, the mid-term evaluation showed a change in attitude. Partners indicated a great deal of interest and willingness to continue this pilot, because the research did not lead to an overall judgement, an "all or nothing" assessment of the project, but revealed to what extent an intervention works well or does not. In this respect, it is actually more a research than an evaluation approach. It is useful for learning about specific interventions, and can be used for upward accountability for donors funding these specific interventions. Partner organisations indicated that PIA helped them adjust their interventions (e.g. relate health education to seasonal issues such as malaria), since it gave them insight into the results of their interventions in general and for specific groups of households. The general results, such as increased autonomy for women in self-help groups, or increased cash crop income as a result of the activities with the farmers' federation, were not surprising for partner organisations.

Nevertheless, some partners found the unit of analysis too "micro" (limited to a few communities), in the sense that it did not give them sufficient insight in the broader programme area in which they work. Other partners however felt that the unit of analysis was not sufficiently micro to unpack the impact of specific elements in an intervention programme. PIA casts a rather wide net to detect impacts of an entire intervention package in a specific geographical area. Yet partners are less interested in proving that they do something worthwhile (they are likely convinced of this), and are more interested in "how- to" questions. For instance, in the case of a health service provision, partners are less interested in knowing exactly by *how much* health indicators improved in a specific community as a consequence of health service delivery, but rather whether maternal health has positively responded to the specific interventions of training traditional birth attendants, or whether arranging emergency transport for pregnant women in remote villages would do more to improve maternal health. In any case, it must be stated that all partner organisations (with one exception) said they found the pilot useful and have made concrete statements about how it could assist them in adjusting their interventions in order to improve and show impact. But only in Peru did partner organisations apply a PIA-like methodology on their own initiative. Others have not (yet) done so. It also remains to be seen to what extent their reported interest will translate into action (beyond adjustments to project design). Looking at the objective related to the capacities of partner organisations in using impact assessment for monitoring, learning, accountability and innovation, and their relations with local research institutes, some results have been attained. However the increased capacity of partner organisations is not so much related to the use of the method, but to an

increased knowledge of the importance of using different methodologies and in gathering more systematic information at beneficiary level.

Usefulness of PIA from Cordaid's perspective

From the perspective of Cordaid, the usefulness of PIA turned out to be very limited. PIA was supposed to provide Cordaid with a methodology for impact evaluation that could be used for learning, accountability, and innovation. While this methodology is to some extent useful for partner organisations to evaluate certain interventions at household level, it is difficult to translate these lessons to a broader level. Typically, Cordaid's support is less directed towards project interventions at household level, but rather towards interventions at system level, and/or multi stakeholder processes in complex and fragile societies. For the sake of testing the methodology, the interventions selected for this pilot were those with a direct impact at household level, and were not necessarily interventions financed by Cordaid. In Ghana, for instance, Cordaid supports health organisations in strengthening their management and staff capacities, in order to improve the quality and accessibility of health care. In the Peruvian case, the water and sanitation projects were implemented with other donor's funds, while Cordaid strengthens community participation in local planning committees and encourages cooperation between local government and community organisations. To evaluate the effectiveness of the Cordaid supported activities, a different, broader evaluation framework should have been applied. Since the selection of the projects or interventions for this pilot was primarily based on the applicability of the methodology and not on the relation with Cordaid funding, these findings were to be anticipated. The expectation was that elements of the methodology could be adjusted for more complex interventions after the testing phase, which proved to be wrong. The lack of usefulness for complicated and complex processes (see Rogers 2008) is inherent to this methodology. Social change processes typically involve different sets of complex interventions between various actors while RCTs require discrete and separable interventions. Furthermore, a standard impact evaluation, "*which only reveals net counterfactual mean outcomes for those treated*" (Ravallion 2009: 35), does not give a better understanding of heterogeneity in impacts and the role of contextual factors. Other studies also recommend taking the broader (development) context into account (Ashman 2001; Kolk, Van Tulder, and Kostwinder 2008). Another limitation is that too often, impact evaluations are a black box, which "*say little or nothing about the economic and social processes leading to that impact … only by understanding those processes can we draw valid lessons for scaling up, or for taking the same project to other settings*" (Ravallion 2009: 38). Understanding how the context influences the outcomes of a specific intervention increases external validity (Ravallion 2009). Nevertheless, from a broader, more general development perspective, PIA did provide some general indication of contextual factors contributing to the effectiveness of development interventions. In five of the six cases, the impact assessment included quantitative and qualitative research on social norms, such as trust in health providers, in-group trust, gender norms, and social norms to cooperate. The study found for instance that trust in health providers increases the impact of health education regarding ownership and use of insecticide-treated bednets; that in-group trust is strongly related to crop adoption; and that subjective well-being for women participating in women's self-help groups sharply declines for those women who are surrounded by relatively conservative gender norms (De Hoop 2011).

In addition, the issue of costs is a very critical point, since PIA turned out to be quite expensive. The financial contribution for the six partner organisations was around €575,000 per year (funds that are not in all projects directly contributing to the evaluated interventions), while the yearly contribution of Cordaid to PIA was more than €250,000 per year.[6] Costs fell into three

main categories: costs covering CIDIN for methodology design and overall training; costs made locally to pay/train local researchers, organise meetings, and gather data; and costs incurred by Cordaid in terms of supervision, travel, and attendance. Taking into account the limited usefulness of the overall findings, one can question whether PIA represents an efficient use of resources. It could be justifiable to spend a large amount of money for testing an intervention with this methodology, if the intervention is to be replicated on a large scale. In addition, using rigorous evaluation methods at all times and at all costs may result in the reduction of development to simple interventions in order to facilitate its measurement.

Some reflections on the 'P' of PIA

It was observed in the mid-term evaluation that the 'P' in PIA had not been properly articulated. Stakeholders argued that participation should have been better defined especially in terms of level, intensity, and purpose. The general feeling was that a broader stakeholder dialogue should have been the first step in the process, which could have improved the quality of the treatment/control group selection, as well as the definition of impact indicators and survey design.[7] In the literature, participation in M&E mainly refers to knowledge generated and constructed through the lived experience of participants, rather than through social science (Vanderplaat 1995). Furthermore, as Lennie states (2006: 28), "*the scientific ideal of objectivity is usually rejected in favour of a holistic approach that incorporates the diverse perspectives, values, agendas and interpretations of participants and evaluation professionals*". Scholars such as Dick (1992, 1999), and Guba and Lincoln (1989), argue that such an approach can also be reconciled with scientific rigor, while Mayoux and Chambers (2005: 272) contend that, when used well, "*participatory methods generate not only qualitative insights but also quantitative data which are generally more accurate than those from conventional survey approaches and methods*".

In the case of PIA, participation was mainly envisaged as organising design workshops with partner organisations, and discussing the findings with them. Incorporating values, perspectives, and lived experiences of different stakeholders, including beneficiaries, was not the explicit intention of this pilot. The rigour of data collection and statistical analysis prevailed over the participatory approach. Partner organisations did have a say in determining the impact indicators, and they vetoed several research proposals, as well as field experiments. One of the partners mentioned a preference for a more rigorous evaluation methodology which would have included more quantitative indicators. Both local research institutes and partners expressed a desire to be more involved in the data analysis of the surveys. This would definitely have improved ownership of the project and stimulated organisational learning. However, the statistical data analysis requires substantial capacity development and implies a stronger supporting role to be played by local research institutes. Over the course of the project, such a relationship was built in the case of Peru. In fact, one indirect benefit of PIA is that stronger links between partners and local research institutes have been forged. Apart from the Peru case, one of the partners in Ghana is currently collaborating with the university that was involved in PIA, both in terms of capacity building and joint research-oriented projects.

Conclusion

The PIA pilot has generated important lessons concerning the possibilities and limitations of PIA as an approach. The methodology proved useful for those organisations that are directly involved in the implementation of relatively straightforward interventions as it enables them to learn about the effectiveness of such interventions. These organisations may use (elements

of) the methodology to test specific interventions at household level. PIA generated interesting information on outcomes at beneficiary level, and stimulated partners to start using baseline data prior to starting new interventions. It also produced useful findings on certain context characteristics and how they influence development effectiveness such as the importance of trust in health providers for the effectiveness of health education.

For Cordaid, PIA turned out to be of limited value, since it is not possible to use this methodology to give useful and relevant information on the effectiveness of the more complex causal strands of Cordaid's support. It is clear that developing appropriate methods for meaningful impact evaluations of complex, multi-actor interventions remains a challenge. PIA might be useful when simple innovative ideas emerge within Cordaid that could be tested in the field through performing an RCT-like exercise. For more complex interventions, theory-oriented evaluation approaches are more appropriate (Pawson and Tilly 1997).

The evaluation requirements for the Dutch co-financing funds have the same methodological principles as PIA. The Dutch Ministry and the Policy and Operations Evaluation Department expect that this methodology will deliver strong evidence of the level of effectiveness of development programmes financed through Dutch civil society organisations. However, based on the experience with this pilot, it is expected that this evaluation approach will provide us with insight in the effectiveness of some specific development interventions, but it will not give us average values on common indicators, nor irrefutable evidence of the effectiveness of the total co-financing channel for development aid. For a proper assessment of the effectiveness of Dutch civil society organisations, sufficient attention should be paid to how their partners are supported, the quality of the relationships, as well as on what development logics or theories lay behind this support.

Acknowledgements

The authors wish to thank Luuk van Kempen for his critical comments on previous versions of this article.

Notes

1. Analysis of behavioural differences and change over time (i.e. risk attitudes, willingness to invest).
2. For instance, projects dealing with violence against women were not considered appropriate for this pilot.
3. This turned out to be quite a challenge for almost all projects. For instance, for the health projects, control groups are households that live further away from the Diocesan health services, and/or make use of public health services.
4. With propensity score matching, families of the treatment group are matched with families of the control group, making use of an estimation of the chance of participation in the programme, based on observable characteristics before the start of the programme.
5. The researchers do not completely agree with this argument since the surveys reveal that hardly anybody in the control villages knows the NGO and there is little production of the crops introduced by the NGO.
6. This does not include the financial and staff contributions of CIDIN.
7. Partner organisations, however, did not argue the fact that the methodology was a non-negotiable element of this pilot.

References

Ashman, D. 2001. "Civil Society Collaboration with Business: Bringing Empowerment Back in." *World Development* 29 (7): 1097–1113.

Barahona, C. 2010. *Randomised Control Trials for the Impact Evaluation of Development Initiatives: A Statistician's Point of View*. ILAC Working Paper 13. Rome: ILAC.

De Hoop, T. 2011. "How Context Matters for Development Effectiveness. A Study into Social Norms and Heterogeneous Impacts." PhD Diss., Radboud University.

Dick, B. 1992. "Qualitative Action Research: Improving the Rigour and Economy." *Proceedings of the Second World Congress on Action Learning*. Brisbane: University of Queensland, July 1992.

Dick, B. 1999. "Sources of Rigour in Action Research: Addressing the Issues of Trustworthiness and Credibility." Paper Presented at the Association for Qualitative Research Conference, Melbourne, Australia, July 6–10.

Guba, E., and Y. Lincoln. 1989. *Fourth Generation Evaluation*. Newbury Park, CA: Sage.

Kolk, A., R. Van Tulder, and E. Kostwinder. 2008. "Partnerships for Development." *European Management Journal* 26 (4): 262–273.

Lennie, J. 2006. "Increasing the Rigour and Trustworthiness of Participatory Evaluations: Learnings from the Field." *Evaluation Journal of Australasia* 6 (1): 27–35.

Mayoux, L., and R. Chambers. 2005. "Reversing the Paradigm: Quantification, Participatory Methods and Pro-Poor Impact Assessment." *Journal of International Development* 17: 271–298.

OECD/DAC. 2002. *OECD Glossary of Key Terms in Evaluation and Results-based Management*. Paris: DAC.

Pawson, R., and N. Tilly. 1997. *Realistic Evaluation*. London: Sage.

Ravallion, M. 2009. "Evaluation in the Practice of Development." *World Bank Research Observer* 24 (1): 29–53.

Rogers, P. 2008. "Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions." *Evaluation* 14 (1): 29–48.

Vanderplaat, M. 1995. "Beyond Technique: Issues in Evaluating for Empowerment." *Evaluation* 1 (1): 81–96.

White, H. 2008. "Of Probits and Participation: The Use of Mixed Methods in Quantitative Impact Evaluation." *IDS Bulletin* 39 (1): 98–109.

White, H. 2009. *Some Reflections on Current Debates in Impact Evaluation*. New Delhi: International Initiative for Impact Evaluation.

The authors

Francois Lenfant (corresponding author) works as an independent consultant and is guest researcher at the University of Amsterdam. His area of research is multi-stakeholder partnerships in conflict countries in Central Africa. He has published a number of articles on the topic in various scientific journals. <f.lenfant17@gmail.com>

Rens Rutten is a human nutritionist with over 23 years experience in international development cooperation. For the last five years she has been working for Cordaid as a manager of programme evaluations, and is coordinating the pilot for Participatory Impact Assessment. She has worked in Latin America and was an advisor for health programmes with Memisa and Cordaid.

Appendix: Selected findings for the six projects

1. Health: Mission hospital in Brong-Ahafo region (Ghana)
   - Healthcare utilisation by the poor increased and their reliance on home remedies, self-medication and traditional healers reduced
   - Incidence of serious illness reduced (self-reported)
   - No reduction in working days lost to illness (disease burden from malaria unaffected)
   - Enhanced access to care (in combination with health insurance) created perverse incentive for malaria prevention, as witnessed in reduced bednet use*

2. Health: Three rural mission health centres in Upper East region (Ghana)
   - Not been able to break reliance of poorest (food insecure) on traditional healers
   - Outreach activities on malaria prevention effective in stimulating bednet use, but only among those who trust modern health providers*
   - Responsiveness to outreach varies across ethnic groups
   - Health-related beliefs improved among women (not among men due to lack of targeting)

3. Women empowerment: Support of women self-help groups (SHGs) in Orissa State (India)
   - SHG membership leads to higher degree of autonomy (freedom of mobility and self-assessed control in life)
   - Enhanced autonomy has not - on average- resulted in higher subjective wellbeing
   - SHG members in relatively conservative villages report lower subjective wellbeing (related to identity loss and social sanctioning)
   - Young married women do not benefit from SHGs (excluded from participation by in-laws)

4. Livelihoods: Linking smallholders to markets through farmers' federation, Uttarakhand State (India)
   - Federation activities increased cash crop income and household savings (mostly reserved for children's education) of participating farmers
   - Farmers belonging to lower castes largely unable to benefit (social discrimination and self-exclusion)
   - Impact of federation on cash crop adoption stronger in communities with high level of interpersonal trust (social learning)
   - Federation raised income aspirations in project area

5. Water & sanitation: Expansion of water supply and promotion of rational use in rural Cuzco (Peru)
   - Availability (number of hours per day) and predictability of water supply increased, resulting in substantial time savings.
   - Educational workshops effective in improving water use practices (treatment before drinking, reduction in waste)
   - Diarrhoea incidence reduced among those who participated in full intervention package (water meter installation and educational activities), attesting to strong complementarity between different interventions.
   - Community participation in water supply works fostered social cohesion (more trust)

6. Water & sanitation: Promotion of hygiene practices in marginal neighbourhood of Lima (Peru)
   - Home visits effective in convincing high-risk households to buy sanitation kit (in-house tap installation and toilet)
   - No impacts (yet) of educational activities on health indicators, since project designed as complement to government programme (*Agua para Todos*), which was severely delayed. Government works completed only in June 2012.
   - Results published in international peer-reviewed journal