

Evaluating Development NGOs¹

Jan Willem GUNNING

→ JAN WILLEM GUNNING is Professor of development economics and director of the Amsterdam Institute for International Development (AIID). He also has been a staff member of the World Bank and Professor at the University of Oxford where he directed the Centre for the Study of African Economies (CSAE). His research interests include poverty dynamics, impact evaluation, and the effect of risk on growth in rural societies.

1. Introduction

How do we know that a development NGO is effective? Consultants writing evaluation reports typically describe what an NGO has done and how outcomes have changed over time, *implying* that such changes can be attributed to the NGO. This was never convincing but only recently has the disenchantment with the before-after comparisons in most evaluation reports become widespread. There now is a surge of interest in using rigorous impact evaluation methods such as randomized control trials (RCTs) to assess what works in development.



LA FERDI EST UNE FONDATION RECONNUE D'UTILITÉ PUBLIQUE.
ELLE MET EN ŒUVRE AVEC L'IDDRI L'INITIATIVE POUR LE DÉVELOPPEMENT ET LA GOUVERNANCE MONDIALE (IDGM).
ELLE COORDONNE LE LABEX IDGM+ QUI L'ASSOCIE AU CERDI ET À L'IDDRI.



1. For a longer and more technical version of this paper see J.W. Gunning, "How Can Development NGOs Be Evaluated?", FERDI Working Paper 51, 2012.

.../... These methods are widely applied but are sometimes resisted by NGOs. This may reflect no more than an irrational aversion to rigour and quantitative methods. But the objection can also be that some NGO activities do not lend themselves to impact evaluation methodologies such as RCTs and this position deserves to be taken seriously.

We consider two situations in which impact evaluation for an NGO indeed requires special methods. First, many NGOs try to achieve their goals indirectly, notably through advocacy. Their theory of change then involves two steps: advocacy leads to collective action and this (possibly via induced changes in policy) to the intended development outcomes. When various NGOs address the same policies or the policies cover wide areas there is little scope for RCTs and in the case of national policies there is none: since everybody is affected the impact of the policy cannot be identified. Secondly, in many NGOs local staff can use their discretion in selecting communities or individuals for participation in, say, a sanitation training program. A standard RCT will then be misleading since it cannot mimic the use of private knowledge of local circumstances in such targeting. We consider these two cases in turn.

► 2. Advocacy: Achieving Change Indirectly

If advocacy was only aimed at changing a national policy then all the evaluator can do is to note that if the policy was not changed there is a *prima facie* case for saying that the NGO was not effective. He can go further if the aim of the NGO is to change, through advocacy, the information available at the local level, to trigger through that information collective action and thereby to improve development outcomes. This theory of change involves three steps: (1) advocacy (e.g. about how little is learned in schools) is picked

up in villages (e.g. through a radio program), (2) this leads to collective action (e.g. complaints by parents to the school principal) and (3) eventually to better learning.

The first step can be tested only in a limited way: if no such information reaches the village the theory is obviously rejected, but if it does arrive then one cannot confidently attribute this to the NGO, notably because many other organizations may be active in a similar way. The second and third step, however, can be tested in regressions, using observational data: high-frequency data (e.g. collected in cell phone interviews) on information and collective action and low-frequency data (e.g. collected through household surveys) on development outcomes.

Note that communities may differ systematically in the information they receive and in how they respond to that information. For example, the communication strategy adopted by an NGO may imply that information is more likely to arrive in those villages where it can trigger effective collective action. This non-random allocation is part of the program's effectiveness and it would be wrong to eliminate it, as would an RCT.

Clearly, NGO activities may not be suitable for RCTs. However, a substantial part of the theory of change in the example (the effect of information on collective action and the effect of the latter on development outcomes) does lend itself to rigorous analysis, contrary to what is often suggested in similar contexts. In this situation the perfect (testing the entire chain) should not become the enemy of the good.

► 3. Imperfect Control

In many NGOs the policy maker has limited control in the sense that decisions on program participation are often left to program officers in the field who can use private information on

who will benefit most. This is, of course, entirely sensible but presents a problem for an evaluation. In the jargon of the econometrician: it introduces endogeneity resulting from a correlation between treatment effects and assignment. In non-technical terms: an RCT would be invalid since it would choose participants (the treatment group) randomly whereas in actual practice they would be chosen non-randomly. This is a situation where indeed standard evaluation methods are not appropriate for an NGO.

Fortunately there is an alternative, proposed by Elbers and Gunning.² This involves collecting observational data³ on changes in outcomes (e.g. health), participation (e.g. in a sanitation program) and other determinants of the outcomes (e.g. income). These data can be used in a regression to estimate the effect of the program in a way which avoids the endogeneity problem. In effect this method produces a weighted average of the effect of the program for different individuals with the weights reflecting how participants are actually chosen. This is, of course, precisely what one would like to know, but it cannot be identified in a standard experiment.

What does this mean in practice? The evaluator should first study how the NGO works. If program officers have discretion to determine whether someone receives “treatment” (and how much) and they base that decision in part on the differences they see between beneficiaries in the effect of the treatment then RCTs are inappropriate. The evaluation should then use observational data and employ the Elbers-Gunning method to estimate the effectiveness of the NGO.

▶ 4. Conclusion

We have discussed two cases where RCTs are not appropriate for evaluating an NGO. In both cases regression methods can be used to test at least part of the NGO’s claim to effectiveness.

2. Elbers, Chris and Jan Willem Gunning (2012), ‘Evaluation of Development Programs: Using Regressions to Assess the Impact of Complex Interventions’, *Tinbergen Institute Discussion Paper 12-069/2*.

3. That is: administrative or survey data rather than experimental data.



Créée en 2003, la **Fondation pour les études et recherches sur le développement international** vise à favoriser la compréhension du développement économique international et des politiques qui l'influencent.



Contact

www.ferdi.fr

contact@ferdi.fr

+33 (0)4 73 17 75 30

