Learning about Measuring Advocacy and Policy Change: Are Baselines always Feasible and Desirable?

This paper captures some recent challenges that emerged from establishing a baseline for an empowerment and accountability fund. It is widely accepted that producing a baseline is logical and largely uncontested – with the recent increased investment in baselines being largely something to be welcomed. This paper is therefore not a challenge to convention, but rather a note of caution: where adaptive programming is necessary, and there are multiple pathways to success, then the ‘baseline-endline’ survey tradition has its limitations. This is particularly so for interventions which seek to alter complex political-economic dynamics, such as between citizens and those in power.

There is an increasing pressure in development practice to establish baselines; in part driven by donor pressure to better demonstrate results and attribute impact. Certainly, the conventions around impact evaluation state that it is necessary to have a baseline in order to provide a benchmark against which to measure impact. This seems logical and is largely uncontested. But are baselines always feasible, and what happens if it does not seem possible (or even appropriate) to undertake one? This paper was prompted by recent challenges in setting up a baseline for an impact evaluation of a multi-donor Empowerment and Accountability (E&A) fund – as well as other experiences of similar programmes in sub-Saharan Africa. The common thread in these programmes is that they all seek to alter the relationship between citizens and the State. This includes the use of evidence to improve accountability (e.g. budget tracking) and advocating for policy change (i.e. research-based advocacy). These are themes that are also of interest to the Impact and Learning Team (ILT).

The paper starts by defining the type of E&A programme that is the subject of this Practice Paper In Brief, highlighting both the key attributes of this intervention type and its potential implications for monitoring and evaluation. The paper then goes on to consider what baselines are, and some of the constraints and challenges about what is possible in this particular case. Finally, the paper concludes that while baselines are generally useful, the assumption that a baseline can always be done, and that it provides the only basis against which to measure change, does not always hold true. This may be particularly so for only a relatively small number of interventions, such as those that seek to influence government and alter policy – i.e., interventions where more adaptive approaches to design and management are often necessary, as the intervention responds to the particularities of the political dynamics as well as opportunities that emerge in any given policy context at any specific point in time.
The rise of Empowerment and Accountability funds

Over the past decade there has been an increasing emphasis on ‘demand-side’ governance interventions, and although the term is itself not always helpful, these are broadly programmes that aim to support civil society in achieving democratic and development outcomes. These have been variously termed ‘transparency accountability initiatives’, ‘voice and accountability’ or ‘empowerment and accountability’ interventions. These initiatives are increasingly being supported through indirect institutional arrangements, which introduce new dynamics around the way by which civil society is supported and how results are captured. There are several potential advantages to managing donor funds in this manner, including the role performed by the intermediary agent which may be better placed to increase efficiency, disbursement, communications and managing for results. There may also be advantages for the donor in terms of the transfer of transaction costs to the intermediary, and for multi-donor programmes, the potential for greater harmonisation between donors (Disch and Moberg 2008).

Certainly, in sub-Saharan Africa, the past few years have seen the emergence of a number of large multi-donor mechanisms that often (although not always) provide support to civil society through grant-making funds. A recent workshop brought together eight such programmes ranging in size from USD 8 to 50 million per programme, and working in countries as diverse as Democratic Republic of the Congo, Ethiopia, Ghana, Kenya, Nigeria, Sierra Leone, Tanzania and Zambia (Fletcher and Nkonkolimba 2012). Together these eight programmes alone handle funds of approximately USD 272 million (GBP 175 million), and there are others on the continent that follow a similar modus operandi.

Programme attributes

The lessons summarised in this paper are drawn from a particular subset of Empowerment and Accountability (E&A) programmes. The programmes are far from identical, but they share a number of common characteristics. They are not individual non-governmental projects, but programmes (often multi-donor) that are designed to support a range of civil society activities. A study on behalf of a number of donor agencies (Disch and Moberg 2008) provides a useful way in which to characterise these modalities; by using three dichotomous choices that tend to dominate donor thinking when structuring development assistance. These are:

- Whether funding is for core or project activities;
- Whether it is provided directly or through intermediaries (indirectly);
- Whether it is provided unilaterally or through joint arrangements.

The first dimension refers to the contractual relationship between the donor and the CSO; and whether it is earmarked for specific projects or activities/outputs/results, or instead for core support to achieve a strategic objective. So for example, the former is usually tied to a project proposal that usually incorporates a logical framework approach (with a hierarchy of objectives and expected results). The latter (core support) is usually not tied to specific outputs but rather organisational objectives (or a strategy) with more flexibility on the use of the funds such as to improve the organisation’s capacity to attain larger societal objectives.

The second dimension concerns the interrelationship between the donor and the implementing CSO. Simply put, direct support is where funds are transferred directly to the CSO, whereas indirect support is where the funds are transferred through an intermediary or managing agent. This is more commonly a private firm (or consortium of firms) or an umbrella CSO. The exact responsibilities of the intermediary vary from simple grant-making operations, to a range of support from building capacity, communications and knowledge brokering, through to support to the executive and legislative functions of the national government.

The third dimension concerns the relationship between different donors. This is self-explanatory, with ‘unilateral’ being when only one donor is supporting the CSO (or CSOs) and ‘joint’ referring to two or more donors. More often than not, these arrangements make use of pooled funding arrangements, with a steering committee or Board to oversee its management.

Based on the typology of Disch and Moberg (2008), then the interventions covered in this paper are mostly a form of project funding delivered through indirect, joint arrangements (Joint-Indirect-Project, J-I-P), with some providing core support (Joint-Indirect-Core, J-I-C). See Table 1. They are generally large multi-donor mechanisms, usually grant-making funds, and usually managed by a managing agent such as a consortium of private firms, or an umbrella NGO. They provide support (communications, technical, capacity, etc) and funds to mostly CSOs (both international and national), but some also work with the legislative and executive arms of government.
Table 1. A typology of support models to CSOs

<table>
<thead>
<tr>
<th></th>
<th>Direct</th>
<th>Indirect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Project</td>
<td>Core</td>
</tr>
<tr>
<td>Unilateral</td>
<td>U-D-P</td>
<td>U-C-P</td>
</tr>
<tr>
<td>Joint</td>
<td>J-D-P</td>
<td>J-D-P</td>
</tr>
</tbody>
</table>

Source: Disch and Moberg (2008). The highlighted intervention types are the subject of this paper.

**Dealing with complex attributes**

In terms of monitoring and evaluation, this type of CSO support has a number of interesting features which have a bearing on the monitoring and evaluation (M&E) design. Recent writings by Barder and Ramalingam (2012) argue that the characteristics of an intervention have an effect on the design of the M&E arrangements. Indeed, there is often a tension between complex programmes working in complex environments and the increasing pressure to demonstrate simple, easily communicated results. Using the three dimensions of the problem (simple/complex), the intervention (simple/complex) and the context (stable/dynamic), Barder and Ramalingam (2012) argue that simple interventions in stable settings are more amenable to traditional results-based management – such as vaccination programmes, particularly when the vaccinations are supplied through direct delivery rather than say a more dysfunctional government health system. Barder and Ramalingam (2012) contrast such interventions with complex problems, addressed through complex interventions in diverse and dynamic settings (such as interventions in fragile or post-conflict situations). Most governance (and civil society) interventions probably sit somewhere between these two extremes – a somewhat complex intervention in a somewhat dynamic context.

While Stern et al. (2012) look more specifically at ‘impact evaluation’ designs; they similarly highlight the complexity of the intervention (its programme attributes) as a key factor in selecting an appropriate evaluation design.⁵ The concept of the ‘attribute’ used by Stern et al. (2012) differs from that of Barder and Ramalingam (2012) in that it encompasses both the intervention and the context. For instance in their review of six DFID programmes, Stern et al. (2012: 49) state that: ‘Some were “complex” because of their own characteristics whilst others were embedded in complex and sometimes dangerous country environments.’ One of the DFID programmes reviewed was the Civil Society Governance Fund (CSGF) which was later renamed Tilitonse, and is the subject of much of this paper. Stern et al. (2012: 59) note the following seven programme attributes that make it complex:

- An overlap with other interventions with similar aims
- Multiple and diverse activities and projects
- Customised non-standard activities/interventions in diverse contexts
- A programme working ‘indirectly’ through ‘agents’ and often at different levels and stages
- The likely impacts of programmes are long-term
- It is working in areas of limited understanding/experience
- The intended impacts are difficult to measure and partly intangible

Each of these attributes has an implication for the evaluation design. For example, one of the characteristics of a Fund is that there is a diversity of projects within a common intervention structure. This is something that implies measurement for at least two levels: the first with the intermediary agent, which often has greater capacity; and the second with the supported CSO projects, which are usually fairly small organisations with limited exposure to M&E. These two levels can cause tensions, with the former sometimes passing down donor pressures to demonstrate results on relatively inexperienced organisations with limited resources and limited capacity. The burden can sometimes be counterproductive, even with the best intention of providing a common evaluation framework by which to capture results.

The implications of these programme attributes goes beyond simply the baseline data collection. They do however at least show why a single baseline survey (as might be appropriate for a service delivery project in health, education or agriculture) may not be so straightforward when applied to empowerment and accountability work, particularly where it seeks to influence policy.
Baselines and situational analysis

There are different perspectives about what constitutes a baseline. There are many CSOs that view it broadly as ‘a recording of the situation at the start of the project’ (perhaps no different from a needs assessment), whereas for M&E purposes it more usually refers to specific evidence (often survey data) collected at the start and which can be repeatedly measured to demonstrate progress, results and impact. While the term ‘baseline situational analysis’ is used by some organisations, it is perhaps more helpful to make a clear distinction between the terms ‘baseline’ and ‘situational analysis’.

For instance, a situational analysis is usually a process internal to an organisation, and it is usually expected to cover things outside the control of the project (the broader context), while a baseline study tends to research only those factors that are expected to change as a consequence of the project. As a CARE (2008) guide points out: while both a baseline and situational analysis are pre-implementation activities, a situational analysis is usually an effort to gain understanding, whilst a baseline focuses on gathering information or data. More often than not, a situational analysis will tend to be used to inform the development of the programme, whereas a baseline will occur after the programme design has been determined.

Lessons from undertaking baselines

The focus here is not so much on the practical lessons from undertaking baseline surveys, even though this can be very challenging, but rather on the lessons about why a baseline may not be appropriate or possible in some situations. The lessons emerge mainly from recent work as part of the Independent Impact Evaluation Agent (IIEA) for Tilitonse in Malawi – a multi-million dollar fund for civil society. The Terms of Reference (ToR) for the IIEA assume that a baseline will be undertaken, and it simply stated that one of the deliverables was, ‘to design and commission national baseline, mid-point and end point public opinion (and other) surveys on issues the project aims to impact’. For several reasons, this proved to be less straightforward in practice. The following section identifies some key assumptions implicit in this requirement to undertake a baseline survey, and explains why they turned out to be somewhat misplaced.

Assumption 1: The baseline is ‘year zero’

Firstly, in a review of the first wave of new projects supported under the Tilitonse Fund it became evident that many were continuations of previous projects by the same CSOs. This was partly due to disbursement pressures on the managing agent to commence grant-making during the inception period, so to avoid an extended break in donor support to civil society. It is certainly not unusual for the design of new forms of donor support to civil society to become lengthy processes, especially when they involve several different donors (each with their own institutional arrangements and headquarter pressures). Indeed, the process from the original conception through to the approval of budgets, plus the tendering processes required to procure the managing agent, can sometimes leave a gap of a year or more between one funding mechanism and another. The pressure to disburse at start-up was therefore mainly about minimising disruption in support. For the managing agent and donors alike, it was less risky to support known organisations and continue work that had been funded under previous arrangements.

For this first set of projects, a baseline could not truly be achieved – if viewed strictly as the evidence collected before an intervention starts. But then, this also points to a more general point about baselines never truly being at ‘year zero’ and a lesson that perhaps much more can be made of existing knowledge and data. Even where new projects are starting, they will be preceded by past interventions that have shaped (for good or bad) the context and likelihood of success. In terms of citizen engagement for example, the work in any particular area will be unlikely to require a completely new form of mobilisation by the CSO, but it will build on a (perhaps) lengthy history of citizen-state engagements. This may include a series of both successful and failed past projects.

Capturing the key pre-baseline factors that have led to the current intervention – whether through primary or secondary data – can therefore be important in better understanding the eventual impact. And importantly, this can be essential in assessing why one intervention may be more effective in one location rather than another, and ultimately whether the lessons are generalisable.

Assumption 2: The design is sufficiently detailed from the start

In the case of Tilitonse and several other E&A funds, the design continues to evolve after the initial grant-making rounds have begun. So, while there was a theory of change prior to the managing agent coming on board, this was further refined during inception, and will continue to do so during the grant-making process. Importantly, and even at the official start of implementation, the theory of change provided only a broad framework about how civil society-state relations are viewed in Malawi, and how the Fund (through its grantees) may enable change. Indeed, while three prime
indicators at the outcome and impact levels were developed during the inception period, these were essentially ‘holding places’ that would subsequently be specified as more became known about the projects being supported by the Fund. The three core indicators were:

- The quality of engagement between civil society and government
- Improvements in the performance of service delivery (access, resource allocations, etc)
- Equitability, transparency and inclusiveness of institutions (changes in the political economy of different interest groups)

At the outset it was not possible to know what or where the interventions were likely to occur: grants could be working in any number of sectors (health, education, water, agriculture), could be of different types (capacity support and mobilisation, budget tracking, policy advocacy, etc), at different levels (grassroots, local government, national, etc), and with different stakeholders (civil servants, politicians, private firms, middle classes, NGO staff, etc). For example, a water project seeking to change citizen-state relations could improve the quality of engagement (say between Water User Associations, the Water Board and local government), service delivery (in terms of continuity of supply through water kiosks), and institutional change (where the middle classes, private firms and the poor align interests to put pressure on the complaint system to address infrequent supply). While a baseline around such a project is relatively straightforward to design in terms of the specific variables for a survey questionnaire, at the start of the E&A Fund, the specifics about each project (its sector, location, target groups, etc) were completely unknown. Once the grant-making rounds began to take place and selected projects became known, then the above mentioned three indicators would be further specified to the particularities of each project.

More specifically the original ToR required that a general public opinion survey should be undertaken. Apart from not being able to determine the specific variables about each project in advance, it was also problematic to devise a sampling strategy; the interventions could be located in any number of districts (and locations within those districts), and so a general public opinion survey (even with national generalisability like the Afrobarometer) would be unlikely to have sufficient concentration of data-points to capture the effects of one or even a number of grant-funded interventions. Put simply, the concentration of the grants was likely to be so diffuse that the aggregate impact would not be discernible from a national-level survey. A rolling baseline at the start of each grant-making round might be feasible, but putting a baseline in place at that start of the Fund itself (as required by donor procedures), was unlikely to prove useful in measuring impact.

**Assumption 3: Demand-led designs can be treated the same as blueprint approaches**

As discussed in the previous section, the very notion of a baseline assumes that the design of the intervention can be largely specified from the start. In the case of Tilitonse, this was more than simply that some aspects could not be fully determined before the grants had been approved. Instead, the Fund itself was deliberately designed to be flexible and respond to demand – with lessons incorporated into subsequent rounds. Indeed, a grant-making fund is a classic example of a demand-led approach: where the detailed design of the fund, its theory of change, priorities and protocols will respond to the demand of its applicants. In particular, the more open the calls for applicants are, the more likely that the range of funded projects will vary between each round.

This lack of specificity was not only due to the attributes of the programme being demand-led, but it is also something inherent in civil society advocacy and influencing work that means that interventions need to respond to the opportunities in the policymaking context. So, while ex post evaluations are not currently in vogue (and mostly for good reason), it is sometimes necessary to trace backwards in order to evaluate policy change. This is particularly so where the policy impact may not have been anticipated in advance (or indeed may be totally unintended or unexpected). A good example of this is by Solesbury (2003) who traces the influence of research on a particular policy shift (the evolution of DFID policy towards the sustainable livelihoods approach). This was conducted in 2003 and yet maps the causal influences going back for more than a decade. The CIPPEC (2011) guide for the International Initiative for Impact Evaluation (3ie) similarly provides a framework for a case study approach for mapping the effect of impact evaluations (rather than research) on policy. Although the guide does not necessarily advocate for an ex post approach to assessing policy impact, much of the material focuses on the inherent complexity of policy processes that requires a case study approach to: ‘identify factors that help or hinder rigorous impact evaluations from influencing policy and improving policy effectiveness’ (CIPPEC 2011: 3). As the guide then goes on to explain:
Factors and forces not only explain ex post how a piece of research may or may not influence a policy decision but also help in establishing policy influence objectives... Sometimes, these factors may be identified and assessed from the outset while at other times they may appear as driving forces in later stages (CIPPEC 2011: 16).

Assumption 4: Agency can be important, and should not be underestimated

Who determines the intervention and its ability to achieve effects and impacts? In most development programmes it is perhaps fair to say that while there is a degree of agency, the implementation will follow a certain path as a result of organisational structures and management. Projects in particular are usually managed by a single Project Manager under which there is a clear project management structure to ensure implementation. In contrast, partnership and network-based structures, as well as grant-making funds, have a greater degree of agency. In a grant fund for example, each CSO (or grantee) can operate somewhat independently from the fund management – and perhaps to a larger extent than a managing agent or donor would like to admit. Many CSO projects are extensions of existing work or the priorities of that particular organisation, with some of the more shrewd CSOs being adept at meeting proposal/ funding application requirements (and the language imposed) while also maintaining their own agendas for particular work.

Baselines are dependent on knowing enough about the design in advance, and that the design is a fair reflection of the intended intervention. Where CSO support is provided through a large fund, then the design and implementation of the intervention(s) is devolved to perhaps hundreds of separate organisations. Each organisation may comply with the managing agent’s requirements – sometimes superficially – whereas in reality they retain a great degree of autonomy to further design, develop and implement the intervention in a way that they see fit. Attempts at an overall baseline survey – or standardisation across individual baseline instruments and indicators – may be able to overcome some of this variability, but ultimately it may miss the very specific impacts of interventions designed by individual organisations (operating in specific contexts).

Concluding remarks

The paper is not an argument against convention, as baselines are often essential for monitoring and evaluation purposes – and they are not done often enough. It is also not an exploration of the many challenges involved in undertaking (or reconstructing) baseline data. Rather, it is a note of caution: the assumption that a baseline must be always done at the start, and that it provides the only basis against which to measure change, does not always hold true. This is particularly so in measuring interventions which are more adaptive and cannot be fully planned from the beginning. This includes the likes of civil society engagement initiatives and interventions that aim to achieve policy change. As the CIPPEC (2011: 44) guide states:

Yet, the effective measurement of policy influence... is another large challenge, which deserves special attention and further development. The literature has for long recognized the deep complexity of the task of assessing the impact and role of a certain piece of research on public policy because of the nature of research and its related activities and because – as already highlighted in this paper – policymaking is a very dynamic process with a multiplicity of actors and relationships.

Furthermore, it is problematic to devise a baseline for interventions where a great deal of agency is inherent in the design – and where individuals and other entities can shape and change the specific outputs and outcomes somewhat independently of the overarching design. Grant-making funds are an example of this; where despite grantees being required to fit within an overarching structure and results agenda (such as through the proposal requirements, approval and reporting processes), the individual grantees retain a great deal of agency to design and determine the ultimate results and impact. This contrasts with a classic implementation programme (such as for vaccines), where the roll-out of the design may fit a blueprint that can be largely determined from the start.

It is not that baselines are impossible, but that in such cases process tracking and ex post assessments may be necessary to capture the full extent of the results and impacts where programmes are flexible, demand-led, and working on change areas that cannot be fully specified from the outset. Developing greater robustness around methodologies to evaluate the work of civil society – particularly E&A initiatives that seek to advocate and influence policy change – should therefore not be limited to simple baseline (plus end-line) survey traditions.
References


End notes

1 Common practice suggests that a ‘pre/ post design with a comparison’ (especially when randomised) provides amongst the most robust measure of impact – and conversely, ex post designs (with no comparison group) provide the least robustness. A baseline is necessary for the former in order to provide the pre-intervention data to assess the ‘before/after’ change, which when combined with a ‘with/without’ comparison can be used to measure the net impact. See Bamberger et al. (2006: 49).

2 For example, the Tilitonse programme in Malawi, the State Accountability and Voice Initiative (SAVI) in Nigeria, and the Strengthening Transparency Accountability and Responsiveness programme in Ghana (STAR-Ghana).

3 The paper also draws on discussions from an ILT Learning Lab conducted in 2012.

4 The term implies an apparent dichotomy between ‘demand-side’ and ‘supply-side’ governance, which in turn creates an artificial divide between support provided to civil society organisations (CSOs) and support to the national government. There is now a body of knowledge (such as from ODI and IDS) that suggests that the focus should instead be reoriented towards state-society relations – with a greater emphasis on the political economy and contextual dynamics aspects of this relationship, rather than supporting one side or the other.

5 Stern et al. (2012: 2) conceive the selection of the evaluation design as a process of aligning three aspects: the evaluation questions with the available repertoire of designs, plus the attributes of the programme.

6 The sample size for the 2012 (Round 5) of the Afrobarometer in Malawi was doubled to around 2,400 respondents, based on a nationally representative sample. This increase in the sample size was so that more could be done with the data in terms of disaggregated analysis. Even so, this will still only allow disaggregation by regions or groups of special interest – and nothing like the concentration of data-points necessary to assess the impact of Tilitonse-funded projects.
Learning about Measuring Advocacy and Policy Change: Are Baselines always Feasible and Desirable?

About the Impact and Learning Team (ILT)

What makes development research accessible, relevant or appropriate for people outside the research community? Does development research get its due in policymaking and practice? What would be value for money in research communication?

The Impact and Learning Team at IDS are interested in how communication of research brings about change - in particular, what happens when people and technology mediate between researchers and decision makers. We use the term ‘intermediary’ to describe people and technology acting in this way. We think they play a critical role in making knowledge accessible, relevant and responsive to demand.

The work we are doing in the Impact and Learning Team (ILT) is exploring and testing this assumption using action research. We support people to think about the difference they want to make as well as how they are going to go about it. We draw insights and approaches from IDS’s history of research, and the fields of marketing, strategic planning and evaluation, and capacity development.

This Practice Paper is an output from our work.

Recent papers in this set

IDS Practice Paper In Brief 5, Brighton: IDS

IDS Practice Paper In Brief 6, Brighton: IDS

Brighton: IDS

IDS Practice Paper In Brief 8, Brighton: IDS


Download for free at: www.ids.ac.uk/publications/ids-series-titles/practice-papers-in-brief

Credits

This Practice Paper In Brief was written by Chris Barnett. The opinions expressed are those of the authors and do not necessarily reflect the views of IDS or any of the other institutions involved.

Readers are encouraged to quote and reproduce material from issues of Practice Papers In Briefs in their own publications. In return, IDS requests due acknowledgement and quotes to be referenced as above.

Funding for the production of this issue was provided by the IDS Knowledge Services team through the Mobilising Knowledge for Development (MK4D) programme.

MK4D is funded by the UK Department for International Development.

© Institute of Development Studies, 2013