GREEN CLIMATE FUND | Independent Evaluation Unit

TRUSTED **EVIDENCE**.
INFORMED **POLICIES**.
HIGH **IMPACT**.

Challenges in real-world impact evaluations: Some learning on costs and timeliness

Jyotsna Puri, Francis Rathinam

IEU Working Paper No. 03, 2019

Challenges in real-world impact evaluations: Some learning on costs and timeliness

Jyotsna Puri, Francis Rathinam
April 2019

175, Art center-daero
Yeonsu-gu, Incheon 22004
Republic of Korea
Tel.: (+82) 032-458-6428
Email: ieu@gcfund.org
https://ieu.greenclimate.fund

Citation
The suggested citation for this paper is:
Puri, J., & Rathinam, F. (2019). *Challenges in real-world impact evaluations: Some learning on costs and timeliness*. IEU Working Paper No. 03, Green Climate Fund. Songdo, South Korea

A FREE PUBLICATION

About the IEU

The IEU was established by the GCF Board as an independent unit, to provide objective assessments of the results of the Fund, including its funded activities, its effectiveness, and its efficiency. The IEU fulfils this mandate through four main activities:

Evaluation: Undertakes independent evaluations at different levels to inform GCF's strategic result areas and ensure its accountability.

Learning and communication: Ensures high-quality evidence and recommendations from independent evaluations are synthesized and incorporated into GCF's functioning and processes.

Advisory and capacity support: Advises the GCF Board and its stakeholders of lessons learned from evaluations and high-quality evaluative evidence, and provides guidance and capacity support to implementing entities of the GCF and their evaluation offices.

Engagement: Engages with independent evaluation offices of accredited entities and other GCF stakeholders.

About the IEU's Working Paper series

The IEU's Working Paper series is part of a larger effort to provide open access to the IEU's work and to contribute to a global discussion on climate change. The series disseminates the findings of work in progress and encourages an exchange of ideas about climate change. Their intention is to promote discussion. The findings, interpretations and conclusions are entirely those of the authors. They do not necessarily reflect the views of the IEU, the GCF or its affiliated organizations or the governments associated with it.

About this Working Paper

This working paper analyses real-world impact evaluations in development sectors in low- and middle-income countries. Using the example of grants-for impact evaluations given by the International Initiative for Impact Evaluation (3ie), it explores key drivers of costs, and causes for delays. The analyses provide insights into managing impact evaluations for real-world programmes where the programme team and the impact evaluation team are typically different, start with different objectives and have different timelines. It concludes with emerging lessons and directions for researchers, implementers and donors that are keen to build learning and impact evaluations into their programmes.

About the authors

Dr. Jyotsna Puri is the Head of the Independent Evaluation Unit at the Green Climate Fund. She also teaches "Impact evaluation in practice" at the School of International and Public Affairs, Columbia University, New York where she is Adjunct Associate Professor.

Francis Rathinam is the Senior Evaluation Specialist at the International Initiative for Impact Evaluation (3ie).

CONTENTS

TABLES

FIGURES

ACKNOWLEDGEMENTS

The authors would like to thank many seminar attendees at impact evaluation events including the annual event hosted by Global Development Network, the students at Columbia University where Dr. Puri teaches, Emmanuel Jimenez, Marie Gaarder, Mark Oldebeuving, Annette Brown, Kanika Jha and Avanika Khosla. All errors are ours.

ABSTRACT

The international development community has witnessed a surge in theory-based evaluations that use counterfactuals, also called impact evaluations, to establish the attributable impact of development interventions. Amid this growing interest in using impact evaluations, challenges exist in using these methods for real-world programmes and investments.

Real-world impact evaluations are studies that evaluate and measure the causal change of government, non-government or multilateral assistance programmes that are planned and rolled out in real-world settings with (all the) associated political, data, implementation and resource constraints. This paper examines key questions that are typically associated with conducting impact evaluations in real-world settings, including costs, complicated requirements for implementation and the ambition of policy influence.

To illustrate constraints and challenges, we draw on impact evaluations funded through the Open Window funding modality of the International Initiative for Impact Evaluation (3ie). This is one of few funding modalities that awarded study teams for undertaking policy-relevant, real-world impact evaluations, unrestricted by funding size request or by subject area. All real-world impact evaluations examined in this study use counterfactuals where programmes were either allocated randomly or where quasi-experimental designs were used to mimic this random allocation.

This study finds that:

- Real-world impact evaluations experience many obstacles. The costs of these evaluations vary according to the question being asked. There is no one formula.
- Close monitoring of projects and the implementation of impact evaluations can help the projects and impact evaluations.
- Good formative research can help to reduce risks in exposing both critical bottlenecks and unexpected challenges in hypothesized theories of change.
- Building flexibility into grant windows can help funders manage their own expectations as well as ensure that teams are protected from reputational concerns.
- Building in a strong financial due-diligence process can help manage cost overruns.
- Early communication and planning between programme teams and impact evaluation teams is a *sine qua non*.

ABBREVIATIONS

3ie	International Initiative for Impact Evaluation
CCT	Conditional Cash Transfer
EOI	Expressions of interest
IE	Impact evaluation
IEU	Independent Evaluation Unit
LEAP	Livelihood Empowerment Against Poverty
LoV	Letter of variation
NGO	Non-governmental organization
OW	Open Window
PI	Principal Investigator
RCT	Randomized Control Trial
USD	US Dollar

I. INTRODUCTION

The international development community has witnessed a surge in counterfactual-based impact evaluations of development programmes, which seek to establish the attributable impact of development interventions. A recent study by the International Initiative for Impact Evaluation (3ie) notes that of the 4,205 impact evaluations published from 1981–2015, some two-thirds happened in the past five years (Cameron et al., 2016; Sabet and Brown, 2018). Amid the growing interest seen among policymakers and donors in commissioning and using impact evaluations, considerable challenges exist in implementing their conclusions. Often-cited challenges include the cost of conducting rigorous evaluations and the delays in completing studies that might potentially miss the window of opportunity for influencing policy.

In this study, we analyse real-world impact evaluations in development sectors in low- and middle-income countries. We define real-world impact evaluations as studies that evaluate and measure the causal change of government or NGO or multilateral assistance programmes that are planned and rolled out in real-world settings with all the associated political, data implementation and resource constraints (see for example Bamberger et al., 2007). They are different from researcher-driven "field experiments" or "laboratory experiments" because, in field/laboratory experiments, researchers have control over the design *and* implementation of both the programme *and* the impact evaluation. In contrast, real-world impact evaluations typically have separate teams from development agencies that are responsible for implementing programmes (the "programme teams"), and separate researcher teams that collaborate with programme teams to design and implement impact evaluations.

Real-world impact evaluations are therefore those that measure the causal and attributable effects of development programmes designed for and implemented in developing countries. They are led by government and/or NGO implementation teams with the aim of affecting change in their natural habitats and bringing benefits to the people of developing countries. All real-world impact evaluations examined in this study use counterfactuals (or comparison groups) where programmes were either allocated randomly or where quasi-experimental designs were used to mimic this random allocation. One of the key challenges in implementing real-world impact evaluations, where researchers have little control over implementations and other contextual factors, is the high likelihood of time delays and cost overruns.

Although we have seen a rise in the number of real-world impact evaluations over time (see, for example, Cameron, Mishra and Brown (2016), Sabet and Brown (2018) and Eduardo et al. (2019)), analyses of what is required for real-world impact evaluations, and managing the challenges thereof, have been mostly sectoral (see for example, Jimenez and Puri (2017); Barooah, Kaushish and Puri (2017), Naeve et al. (2017), and Puri (2015)) and have not focused on the challenges in implementing real-world impact evaluations. This paper fills that gap.

To illustrate key learnings, we draw on impact evaluations funded through the Open Window (OW) funding modality of 3ie. This is one of the few funding modalities we know of that awarded study teams for undertaking policy-relevant, real-world impact evaluations, unrestricted by funding size request or by subject area. Over the four OWs (i.e. OW1, OW2, OW3 and OW4) from 2009 to 2012, 3ie funded 89 impact evaluation grants worth approximately USD 38 million. This grant window has since been discontinued.

Section II of this paper briefly describes the nature and process of grant-making under the OW grant programme of 3ie.

Section III explores the cost drivers and the instances of time and cost overrun, while section IV examines some insights from

managing real-world grants, and specific reasons for delays. Section V discusses emerging lessons and directions for policymakers and donors.

II. Attributes of the 3ie Open Window grant programme

Between 2009 and 2012, 3ie supported four open grant windows. A grant "window" at 3ie is a grant programme that financially supports impact evaluations to measure the attributable causal change of development programmes. For OW grants announced by 3ie, the organization received 1,347 initial applications/expressions of interest (EOI) and issued 89 grants with a committed amount of USD 38 million (Table 1). Just over one third (37 per cent) of these grants supported the impact evaluations of interventions in *low*-income countries and close to three quarters (70 per cent) were for evaluations in South Asia and Sub-Saharan Africa. A little over a quarter (28 per cent) of these grants went to grant-holding institutions located in developing countries. Over time, these windows have become increasingly competitive in the sense that the number of applicants has increased. The acceptance rate for 3ie fell from 23 per cent of applications in the first OW (OW1) to just 7 per cent in OW4.

Table 1 Number of grant applications received and funded, 3ie Open Window (2009–2013)

	OW1 (early 2009)	OW2 (late 2009)	OW3 (2010)	OW4 (2012)	Total
Expressions of interest (EOIs)	n.a.	n.a.	365	634	999
Full proposal received	78*	270*	204	279	831
Selected for panel	37	95	60	66	258
Awarded a grant	17	30	22	20	89
Total budget (USD million)	4.27	15.41	9.44	8.81	37.93

Note: OW1 and OW2 had a one-stage process and did not require an initial expression of interest; OW3 and OW4 required an initial EOI.

III. Costs of Impact evaluations

Amid the increasing use of impact evaluations in international development, a key question that is often asked pertains to their cost. Many implementing agencies question the value for money of multi-year, real-world impact evaluations. The cost of doing impact evaluations depends on a number of factors unique to the programme being evaluated. These include its study design (everything held constant, randomized controlled trials have smaller sample sizes than those that use quasi-experimental methods for identification), the ambition of the programme (especially the desired effect size), other outcomes of interest and the local context (which affects, for example, intra-cluster correlation), and so forth.

In this section, we analyse the cost of 3ie-funded real-world impact evaluations across different geographical contexts, sectors and evaluation methods. We examine the different cost components of real-world impact evaluations, such as average reported data collection costs, personnel costs, travel costs, equipment costs, office expenses, sub-grants and indirect costs. We also explore and examine the average time overruns and cost overruns in OW grants, and analyse differences between approved budgets and actual budgets based on the audited final financial statements and letters of variations to assess key drivers of costs and reasons for

cost and time overruns in the 3ie grants. In this section, we only include OW impact evaluation grants that had "closed" by December 2015 (see Box 1). A total of 42 grants had closed by December 2015. About two-thirds of the closed grants were in developing countries in Africa and South Asia, and the remaining were in the middle- and lower-middle income countries in Latin America and East Asia.

Box 1 Glossary

Approved budget: An approved budget is what is cited in the initial grant agreement between 3ie and the applicant. Applicants that are successful, or grantees, submit a budget along with detailed narratives and a rationale for the proposed budget under each cost heading, such as personnel, consultancy, travel, survey cost, office expenses, equipment, indirect costs (overheads), and sub-grants. They also provide an analysis of how much money will be allocated for evaluation design and preparation, data collection, data analysis, report preparation, policy influence, monitoring, and capacity-building among key stakeholders. The proposed budget is reviewed and approved by the 3ie finance team in consultation with the evaluation team, providing insights into whether the proposed budget is required for the technical proposal that has been submitted.

Letter of Variation: A letter of variation (LoV) is an amendment made to the executed grant agreement. A formal request for an LoV is submitted by the grantee, which is reviewed and approved by 3ie before a grant amendment is processed. Letters of variation are issued by 3ie for the following reasons:

- Change in the evaluation design
- Budget revision
- A no-cost extension in case of delays
- Change in key personnel, especially lead principal investigator (PI)
- Changes in the deliverable and disbursement tranche payment schedule

Actual budget: This is the final audited financial statement (for grants over USD 500,000), or the final financial utilization statement submitted by the grantees at the end of the grant period, reviewed and approved by the 3ie finance team.

Cost overrun: Cost overrun is the difference between the proposed budget and the actual budget. A grant is classified as a "no overrun" grant if the difference between the original and the actual budget is within ±5 per cent of the approved budget. If the actual budget is more than +5 per cent above the approved budget, it is defined as a cost overrun, and if it is more than 5 per cent below the approved budget, it is defined as underspent.

Closed grant: This is defined as a grant for which all the agreed deliverables – including the final report and data that meet the standards of 3ie for impact evaluations – are received, and the final audited financial utilization statement from the grantee has been received and approved by 3ie.

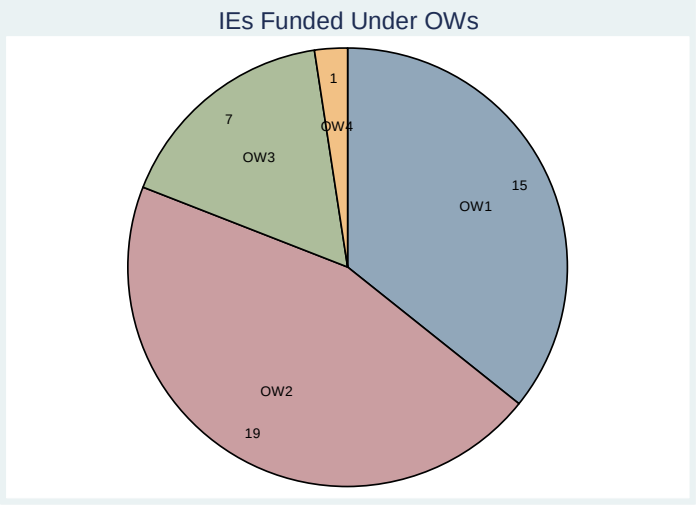

Figure 1 Closed impact evaluation grants supported by 3ie Open Window (2009–2015)

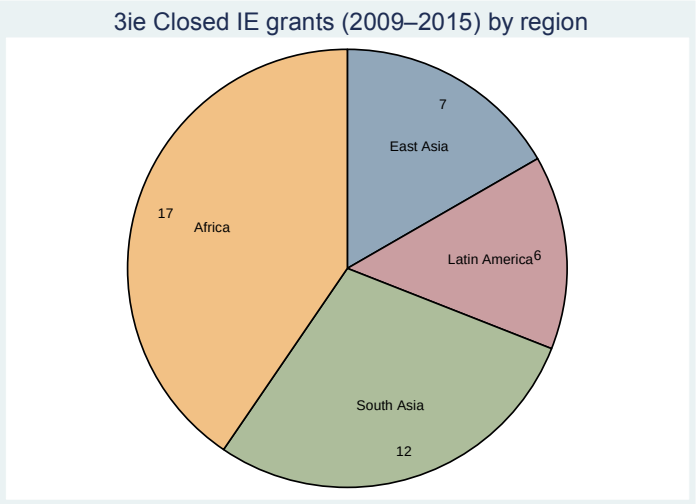

Figure 2 Closed impact evaluation grants supported by 3ie Open Window, regional distribution

A. Cost drivers

On average, OW-funded closed impact evaluation grants had a budget of USD 344,021, while a typical (median) 3ie-funded impact evaluation (IE) costs around USD 334,000[1]. Average costs varied by design, duration, sector and by the region in which the programme was implemented. Costs of IEs increased with study duration (see Table 2).

A comparison between the proposed budgets for the OW grants of 3ie with the actual budgets for closed grants (2009–2015) reveals some interesting findings. First, the average proposed budget for OW grants was USD 421,000, whereas the average closed impact evaluation grant from all windows had an average budget of USD 344,021. This is because more than half of the grants are still not closed, and these are the grants that are likely to have higher costs. However, the difference is the same when you compare the proposed costs of the studies that have closed, and also compare proposed budgets with the actual costs of impact evaluations. Part of this can be explained by the fact that the management process is a good oversight process that helps evaluation teams to rationalize their budgets as they go along. Part of this is also explained by the fact that applicants routinely overstate their budgets and include risks that they foresee while erring on the side of caution. The actual budgets of closed impact evaluations tend to be somewhat smaller than the proposed budget, as the proposed budgets are reviewed and each budget component benchmarked against the past context and sector-specific averages from 3ie funded grants.[2] Furthermore, the average cost of closed grants is also skewed by a handful of smaller grants (see Table 2 for more details).

The role of 3ie budget and financial due diligence is evident when we compare proposals funded by 3ie with those that are not funded during the OW call for proposals. On average, 3ie-funded proposals have lower budgets than those that were not funded (see Table 3).

[2] The cost of evaluation is calculated from 3ie grant budget figures alone. These evaluations are partly funded from other sources as well. About 40 percentage of the studies did not receive any non-3ie support at all. While the median non-3ie contribution—either from the study teams themselves or from alternative sources—is USD 45946, the average is skewed by a couple of studies that reported to have received over a million USD from other sources. The non-3ie contributions reported by the grantees were not verified by 3ie.

[2] In a separate analysis of the grant modalities of 3ie, we found that in the initial OWs, staff time for each award was approximately five days at the time of the award, with approximately two to four days for subsequent tranche payments. However, this has subsequently changed as the appraisal time for feasibility, relevance and engagement in each grant has increased (also see Box 1)

Figure 3 3ie OW-supported, closed IE grants (2009–2015) by study duration

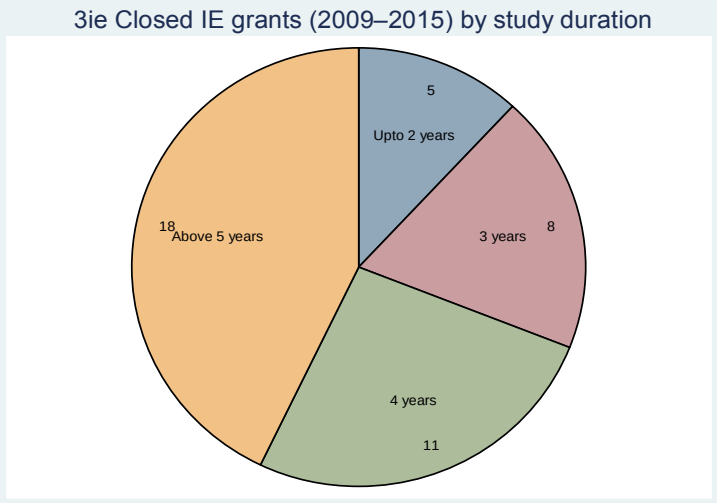

Table 2 Average actual budgets for 3ie OW-supported closed IE grants (2009–2015) by method and duration, unit: USD'000

	DURATION OF THE GRANT				
	<2 years	**2-3 years**	**3-4 years**	**>4 years**	**Average**
Randomized Control Trial (RCT)	116.2 (n = 3)	170.0 (n = 5)	404.8 (n = 9)	468.4 (n = 16)	373.8 (n = 33)
Quasi-experimental	133.2 (n=2)	145.0 (n = 3)	638.4 (n = 2)	672.0 (n = 2)	234.7 (n = 9)
Average	123.0 (n = 5)	160.5 (n = 8)	447.2 (n = 11)	423.8 (n = 18)	344.0 (n = 42)

Table 3 Average budgets for rejected proposals, OWs of 3ie (2009–2013), unit: USD'000

	<1 YEAR	1–2 YEARS	2–3 YEARS	>3 YEARS
Mixed methods + RCT	256.5	438.6	587.3	984.4
Mixed methods + Quasi	230.0	499.0	699.0	845.5
RCT	315.6	340.8	582.7	620.7
Quasi-experimental	269.5	393.7	580.6	636.5

Disaggregating grants by design and length of study (see Table 2) shows that the *ex-post* studies that have quasi-experimental designs cost the least.[3] Although there are only nine studies that use quasi-experimental designs, the size of the budget is skewed by two studies.[4] The first is the study of alternative models of early childhood education, which focuses on children in the favelas of Brazil. The study authors conducted extensive psychological tests. The second study is a comparison of vocational study versus general schooling in China. The study examined five different interventions and

[4] This is expected, as some of these designs do not collect baseline data, or indeed any survey data, and some do not undertake qualitative work

[4] The figures in Table 2 and Table 4 - cost of IEs across designs and across regions - should be read as indicative rather than representative given the small sample size in some of the cells.

included extensive data collection for all five treatment arms, and the sample sizes were very large. The study also used tests to measure learning outcomes, which are usually very expensive.

A little more than half of the completed grants used RCT designs, which cost around USD 373,823. The average actual budget for RCTs is statistically different from the average cost of quasi-experiment studies. It should be noted that this distribution is skewed because of some of the short duration studies. If smaller grants that cost less than USD 100,000 are excluded from the sample, then the average cost for the IEs is USD 419,336. Table 3 shows that multi-year studies with multiple rounds of data collection using mixed methods are expensive. Importantly, the period of study is dependent on the types of outcomes that agencies and researchers aim to measure.

Table 4 Average cost of closed 3ie IE grants by research method and region, unit: USD'000

	RESEARCH METHOD		
	RCT	**Quasi-experiment**	**Total**
East Asia	304.9 (n=4)	315.6 (n=3)	309.4 (n=7)
Latin America	332.8 (n=4)	253.0 (n=2)	306.2 (n=6)
South Asia	482.3 (n=11)	159.2 (n=1)	455.4 (n=12)
Africa	320.0 (n=14)	167.0 (n=3)	293.0 (n=17)
Total	373.8 (n=33)	234.7 (n=9)	344.0 (n=42)

Disaggregating the data by region shows that it is cheaper to carry out evaluations in Africa and Latin America, and more expensive in South Asia (see Table 4). One of the key cost drivers in overall cost is the cost of surveys. The average survey cost for a typical 3ie IE grant is USD 133,773, though the median expense is USD 90,957 (see Table 5 below). The average increases to USD 176,235 for multi-year, multi-round survey studies. There is only a slight difference between survey cost for RCT and (self-identified) mixed method studies.

The other major cost driver in IEs are personnel costs for the high-quality research staff required for conducting the IEs. Personnel costs cover the time and effort put in by study teams in designing real-world evaluations and surveys, and their analyzing of data, writing of the report and the dissemination of findings. On average, 3ie grantees spent in the region of USD 109,991 on personnel costs in OWs that had closed by 2015. The actual average spent on travel was USD 19,822.

Table 5 Average budget for key budget components for 3ie IE grants (2009–2015), unit: USD'000

STATS	MEAN	MEDIAN	MAX
Budget	344 (235)	334	878
Survey cost	134 (125)	90.9	429
Personnel cost	110 (95)	87.7	421
Travel cost	21 (20)	16	90
Overheads	21.1 (55)	0	280

B. Cost Overruns

The 3ie requires grantees to assess the duration of the study and costs. For example, the study duration is determined with inputs from the implementing agencies based on the programme timeline, and the reasonable gestation time required to measure the impact of the programme. However, government-implemented real-world development programmes face immense implementation and budget challenges, that are likely to affect impact evaluation. Cost and time overruns in impact evaluations are therefore not uncommon. We define cost overruns as the difference between the budget that was originally approved, and the actual budget. A grant is classified as a "no overrun" grant if the difference between the original and the actual budget is within ±5 per cent of the approved budget.

Table 6 below shows that the average cost overrun is negative (i.e. on average, around USD 655 is unspent). We suspect that one of the reasons for small cost overruns is the budget review process that 3ie follows during the proposal approval stage. During this stage, a budget review takes place, allowing technical and context specialists from the grantee teams to work closely with the 3ie finance team, to arrive at reasonable estimates for personnel, surveys, travel and other costs. The 3ie finance team draws on the experience of 3ie in a variety of sectoral contexts and costs, and benchmarks proposed budgets against context and sector-specific averages. However, a further breakdown of terms of study duration, and actual budgets and research methods, shows substantial variation between the studies. The largest variations we find are in studies that proposed the lowest study duration and budget.

Table 6 Cost overruns in 3ie closed grants (2009–2015)

		COST OVERRUNS (USD)
Research methods	RCT	-4,505 (n=33)
	Quasi experiment	13,465 (n=9)
	Average	-655 (n=42)
Study duration	>2 years	22,801 (n=5)
	3 years	13,814 (n=8)
	4 years	-10,301 (n=11)
	<5 years	-20,296 (n=18)
Actual budget	Less than USD 300,000	18,360 (n=19)
	USD 300,000 – 500,000	-3,780.6 (n=13)
	Above USD 500,000	-32,718.8 (n=10)

IV. DELAYS IN REAL-WORLD IMPACT EVALUATIONS

In this section, we examine possible delays in conducting real-world impact evaluations and draw on the results of 3ie OW grants to discuss potential challenges and solutions. The analysis of the delays covers the first three OWs. We excluded the fourth OW as it started too recently. A delay in a 3ie grant is defined as having occurred if the final report or the draft final report reaches the programme office of 3ie after its due date, or if a deliverable associated with a tranche payment reaches 3ie after its due date (see Box 2). Using this definition, on average, as of July 2013, 28 per cent of the primary grants of 3ie were delayed for more than a year (see Table 7). In Table 8, we see the difference between the planned period for the evaluations and the actual time period.

Box 2 Definition of a delay

In order to identify what the relevant delay is, first, the relevant end date has to be identified, as this is sometimes different from the revised end date. If there are no future tranche payment dates, and the last tranche has been paid (hence, the grant is complete), the relevant end date is the last tranche payment date, minus 56 days accounting for the payment process within 3ie.

If there are no future tranche payment dates, but the last tranche has not been paid, then the relevant end date is set as July 1 (minus 56 days accounting for the payment process within 3ie).

If there are future tranche payment dates, then the revised end date is either the last tranche due date, or the revised end date (minus 56 days accounting for the payment process within 3ie), whichever is later.

It is important to note that the date that is derived from this is the only one necessary for calculating delays, and does not necessarily reflect the actual end date of the grant. This is especially relevant for grants that are awaiting the processing of the last tranche payment. The end date is set for 1 July, and the delay only captures the time up to that point.

Table 7 Distribution of grants by length of delay, 3ie Open Window grants, 2009–2012

	None	<6 months	6 months–1 year	>1 year
OW1 (n=17)	1	3	5	8
OW2 (n=30)	8	1	11	10
OW3 (n=22)	9	3	6	4
Total (n=69)	18	7	22	22

Note: OW4 is not included.

Table 8 Length of grants, 3ie Open Window grants, 2009–2012

	Mean planned length of grants	Mean actual length of grants
Delayed (51)	2.3 years	3.4 years
Non-delayed (18)	3.4 years	3.4 years
Total overall (69)	2.6 years	3.4 years

Note: Does not include OW4 grants.

A. When and why?

When and why do delays occur? The grantees are required to produce at least one or more technical deliverable – such as a pre-analysis plan, a baseline report, midline or qualitative report, final report, et cetera – for every year of the grant period, and each tranche payment is associated with a deliverable product. Once the deliverable package reaches 3ie, it enters a "review stage": each deliverable in the package is reviewed by the relevant 3ie office and the external project advisor. An incomplete or inadequate deliverable would be sent back to the grantee. In many cases, reminders are sent to grantees for a response within two weeks. Once responses are received and 3ie staff are assured of the quality of the deliverable, the tranche payment is processed.

Two kinds of delays occur: (1) a delay in the submission of deliverables; and (2) a delay in the review stage. Figure 4 displays (by OW and by tranche) the average number of days that a grant was delayed for these two reasons. It shows that, first, except for one tranche, most delays were due to the late submission of deliverables. This accounted for 70 per cent of the total number of delays observed. It is interesting to see that overall (in number and time) delays reduced in later OWs.

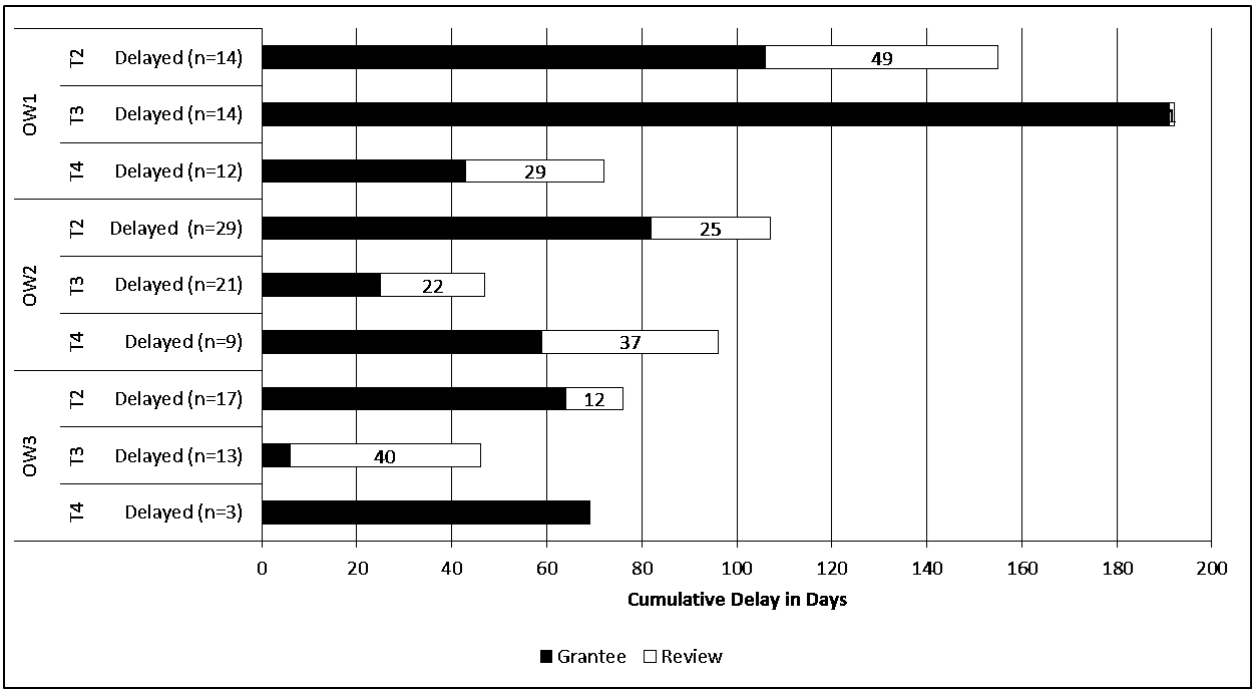

Figure 4 OW tranche delay by stage for delayed studies: a diagrammatic representation (number of days)

B. Why are there delays in submissions in real-world impact evaluations? A qualitative analysis of OW grants

In this section, we examine some causes for delays in real-world impact evaluation grants submitted to 3ie. In our analysis, we partition reasons into six groups and use examples from a sample of 3ie impact evaluation grants. These grants are listed in Appendix I. The six causes of delays are:

- Initial delays because of lack of engagement
- Process and political delays
- Low take-up and the risk of underpowered studies
- Implementation obstacles
- Delays related to surveys and data analyses
- Overall quality of studies

Initial delays because of lack of engagement

An important lesson for real-world impact evaluations is that there needs to be close engagement between the impact evaluation team (the research team or the grantee) and the project implementation team. Ensuring this linkage is established and continuous is a requirement in most real-world impact evaluation grant proposals. This is important so that impact evaluation designs can be incorporated into projects, and measurements can be developed. Early engagement also ensures that real-world impact evaluations are relevant to the programme. Despite this, most grantees started engaging in earnest with the implementing agency (the one implementing the policy/programme/project) only after having been notified of 3ie awarding the grant. This meant that there were changes in many grants, because the implementing agency did not understand the full consequences of undertaking an impact evaluation for the implementation of their project, or because the timing of implementation had changed in the interim.

In one grant, the country of implementation changed after grant signing, as a more responsive implementing agency was found. In another, data became available much later

than promised. In a third, the survey firm that was initially tipped to take on the survey was let go because of quality concerns and replaced by another. In total, eight studies were delayed because of changes in the plans and designs of the programme and lack of early engagement.

Process and political delays

Several studies have been delayed due to political reasons beyond the control of the research team. Two studies, in India and Sierra Leone, were delayed because elections were postponed. After the contract was signed, a grantee in Mexico realized that it needed to have authorization from the authorities to enter communities for a study on parenting. In a study in India that focused on reducing tuberculosis incidence through community counselling, implementation of the programme by the state government took much longer than anticipated. This meant that data collection for the mid-term and end line surveys was delayed. In a study in Peru, discussions with the Government on randomly assigning the programme took much longer than expected. A study of mid-day meals in India was delayed because the Government held back implementation. Two other grants were cancelled after the contract was signed: a study grant in Mauritania was discontinued because of political conflict, and a study in Ecuador was stopped because the newly installed Government did not allow the research study to continue.

Low take-up and the risk of underpowered studies

In several studies, we witnessed a lower take-up than had originally been anticipated by the grantees. While in many cases such an occurrence can still provide measures of intention to treat, most studies aim to go beyond these measures to understand the effect on the treated. However, lower-than-expected uptake meant that sample sizes were too small for robust measures of impact. Frequently, this meant that the study team had to think of creative ways to increase uptake or alter the intervention altogether, all of which take time. Delays because of these reasons were seen, for example, in the evaluation of the Mann Deshi insurance programme being implemented in western India, a weather insurance programme (also in India), a bio-fortification programme in Uganda, a study examining the uptake of low-cost farm equipment, and a study examining the factors affecting the uptake of male circumcision.

Implementation obstacles

In several studies, grantees encountered implementation obstacles that were outside of their control. In one study examining the effectiveness of SMS messages on tuberculosis treatment adherence in Pakistan, messages did not work properly (they did not get through) and sticks prepared for urine testing did not work. In another study that examined the impact of mobile phones on people's savings in Sri Lanka, mobile phones set up by Etisalat with the help of a software company did not register the transfers made to treated households' bank accounts. Additionally, software problems were encountered in the sending of Sinhalese SMS messages. In another study in Ghana that looked at the relative effectiveness of cash transfers and health insurance, bank accounts did not register the increases in amounts. In the above cases, the study team witnessed low take-up as a consequence of a lack of credibility in the systems being set up. In another study in Andhra Pradesh, the study team ignored the fact that a microfinance programme had been implemented universally in the state.

Surveys and data analysis

Most impact evaluation studies collect primary data for their results. In many cases, study teams underestimate the amount of time required to plan, train for and roll out surveys, as well as the time needed to clean data, undertake double data entry and then analyse data. Frequently, studies underestimate how long these processes take. For example, a study of anaemia in Bihar, India took longer than anticipated because the baseline and roll-out surveys took longer than expected.

Another study examining the efficiency and satisfaction derived from different targeting mechanisms in Indonesia was delayed because data collection took longer than initially planned. A study of a safety net programme in Ethiopia had to be postponed by a year because the researchers did not insert their questions into the annual government-led survey in time. While a study of entrepreneurship in Chile became delayed because fieldwork took longer than expected. The same occurred during a study of the vocational training programme of China and during a study of tenant farmers in Bangladesh.

Quality of study reports

Despite several checks and balances, poorly written study reports often get submitted. These have very little detail: no balance tables, no theory of change, no description of the intervention, no investigation of the literature, no description of attrition, spillover effects, no sensitivity tests to different specifications, no discussion of results, and, last but not least, no actionable policy recommendations. In such cases, these studies are sent back and detailed revisions are requested. This was witnessed, for example, with a study on agricultural vouchers in Tanzania, a study of slum housing in Peru, and a study of health insurance in Ghana. There are other, less frequent causes of delays, such as the inability of the study team to retain essential technical staff. Theory-based impact evaluations often take place over long time periods, and this means that teams change, experiencing turnover in their technical and analysis teams that delays study reports.

V. Discussion and Conclusions

It is clear from these analyses that real-world impact evaluations experience an abundance of obstacles. In many cases, such obstacles are beyond the control of the evaluation team, the programme team and 3ie. There are others, however, that can be controlled. Here we discuss several changes that 3ie instituted in its programme monitoring and its technical appraisal of studies to mitigate risks.

Closer monitoring of projects and impact evaluations

Close monitoring of projects and the implementation of impact evaluations can help their efficacy. In response to challenges witnessed in its OW grants, 3ie creates tools that closely monitor and track projects and studies. Its programme management system is also used as a monitoring device. Tranche payments, deliverables and delays are also easily tracked through these tools. Automated reminders have also been built into the grant management system of 3ie.

Prior formative research

Good formative research can help to reduce risks in both critical bottlenecks and unexpected challenges in hypothesized theories of change. It can also help in the understanding of take-up and aid in the formation of realistic expectations for implementation and field realities. Formative research as a pre-requisite for theory-based real-world impact evaluations can help to manage impact evaluations, while also helping research teams to interact more with project implementation teams.[5]

Building-in flexibility

Building flexibility into grant windows can help funders manage their own expectations as well as ensure that teams are protected from reputational concerns. The 3ie builds this flexibility into its grant cycles by introducing letters of variation (LoVs). These allow for changes in timelines, turnover among teams and also changes in designs, pre-analysis plans and impact evaluation questions (see Table 9). To prevent teams from misusing these allowances, 3ie requires that LoVs are signed by the head of the grant-

[5] This is similar to the process undertaken in the Policy Window of 3ie.

holding institution.

Despite all these efforts, it is clear that delays are still likely. In subsequent grant cycles, 3ie has considered incorporating reputation built from previous studies into its assessment of grant applications. It has also considered incorporating penalties and incentives into its grant cycle and has looked into a dialogue on rating the quality of final reports, which will be publicly available. None of these have yet been formalized, but we would encourage similar agencies to consider these suggestions in their grant cycles and management.

Cost overruns are manageable

Building in a strong financial due diligence process can help in the management of cost overruns. On average, 3ie grants faced only a minor underspend. This was due to a rigorous budget review process where personnel, surveys, travel and other costs were benchmarked against past studies in a comparable context. However, there is some variation among these studies. Those with large budgets and very ambitious timelines tend to be off the mark.

Real-world theory-based impact evaluations are difficult and complex. They require effort, skill, time and money. To undertake and manage successful real-world impact evaluations, organizations that undertake and manage them need to innovate and adapt constantly. Organizations need not to focus only on technical quality but also pay attention to mixing methods while striving for relevance. The focus must also be given to building capacities amongst a wider cross-section of disciplines, and to incorporating a variety of methods and approaches.

Table 9 LoVs received, 3ie grants, 2009–2013

Window	Total number of LoVs	Mean number of LoVs per grant
OW1	20	1.2
OW2	39	1.3
OW3	17	0.8
TW1	8	0.8

Note: OW4 is not included.

BIBLIOGRAPHY

Bamberger, M., Rugh, J. and Mabry, L. (2012). Real World Evaluation: Working under Budget, Time, Data, and Political Constraints. *Sage Publications.*

Barooah B, Kaushish B and Puri J. (2017). Understanding financial risks for smallholder farmers in low and middle-income countries: what do we know and not know? *Scoping paper 9, 3ie*.

Cameron DB, Mishra A and Brown AN. (2016). The growth of impact evaluation for international development: how much have we learned? *Journal of Development Effectiveness*, 8:1, 1-21, DOI: 10.1080/19439342.2015.1034156

Edoardo M., Rathinam F., Nath M., Vigneri M., and Wood B. (2019) Designing successful impact evaluations. CEDIL Inception Paper 6

Jimenez E and Puri J. (2017). The wicked cases of education and climate change: the promise and challenge of theory-based impact evaluations. *In:* Van den Berg RD, Naidoo I and Tamondong SD (eds.) *Evaluation for Agenda 2030: Providing Evidence on Progress and Sustainability*.

Naeve K, Fischer-Mackey J, Bhatia R, Yegbemey RN and Puri J. (2017). Evaluating advocacy: an exploration of evidence and tools to understand what works and why. *Working paper 29, 3ie*.

Puri J. (2015). Embracing Complexity: Some experiences from Impact Evaluations of development programs. *eValuation Matters.* African Development Bank.

Puri J. (2017). Using Mixed Methods to Assess Trade-offs between Agricultural Decisions and Deforestation. *In*: *Evaluating climate change action in the context of sustainable development.*

Sabet SM and Brown AN. (2018). Is impact evaluation still on the rise? The new trends in 2010–2015. *Journal of Development Effectiveness*, 10:3, 291-304, DOI: 10.1080/19439342.2018.1483414

APPENDIX I: LIST OF GRANTS USED FOR ILLUSTRATING LESSONS IN SECTION IV.

OW1.20: Building a brighter future – a randomised experiment of slum-housing upgrading in Mexico.

OW2.031: Environmental and socio-economic impacts of the Payment for Ecosystem Services Programme of Mexico.

OW2.223: Alternative models of early childhood education: daily centre-based care versus parental training in Brazil.

In OW3.1236: Making networks work for policy: evidence from agricultural technology adoption in Malawi (there were delays in starting).

In OW4.1190: Land property rights and agricultural investments in the Philippines: experimental evidence on impacts and mechanisms. There were changes in the study team (the local PI became head of the National Economic and Development Authority, of the Philippines) and new PIs had to be selected after the grant was signed.

OW2.099: An impact evaluation of information disclosure on elected representatives' performance; and TW1.1042: The impact of conditional and unconditional transfers on livelihoods and conservation in Sierra Leone.

OW3.1079: Estimating the effects of a low-cost early stimulation and parenting education programme in Mexico.

OW3.1218: Fighting tuberculosis through community-based counselors in north Indian slums: a randomized evaluation of performance-based incentives.

TW1.1020: Impact evaluation of the non-contributory social pension programme 70 y mas.

OW3.1217: Female empowerment and occupational impacts of vocational training in the day-to-day life of the oases: evidence from a randomized evaluation in rural Mauritania.

OW1.57: Rigorous assessment of micro-credits in Ecuador.

OW2.161: The economics and psychology of long-term savings and pensions: a randomized experiment among low-income entrepreneurs in Maharashtra, India; OW2.103: Smallholder access to weather securities, demand and impact on consumption and production decisions; OW2.142: Sustainability of impact: dis-adoption, diffusion and social learning following a bio-fortification programme to reduce Vitamin A deficiency in Uganda; OW1.45: Monitoring and assessing the impacts of Kickstart's low-cost farm equipment on poverty reduction in Africa; OW2.165: Scaling-up male circumcision service provision.

OW1.48: X out TB: monitoring patient compliance with tuberculosis treatment regimens.

OW2.192: Enabling micro-savings through bank-linked mobile phones and mobile banking in Sri Lanka.

OW3.1075: Cash transfers, health insurance and health outcomes in Ghana's Livelihood Empowerment Against Poverty (LEAP).

OW3.1195: Effects of debt relief on the portfolios, consumption and welfare of the rural poor of Andhra Pradesh.

OW3.1205: Evaluating the impact of supplying double fortified salt through the public distribution system (PDS) on anaemia in Bihar, India.

OW3.1055: Improving targeting in conditional cash transfer programmes – a randomized evaluation of targeting methods in the CCT Programme of Indonesia.

TW1.1065: Assessing general equilibrium impacts of productive safety nets in Ethiopia.

OW2.200: Micro-entrepreneurship support programme in Chile – an impact evaluation.

OW2.208: Investment in vocational vs. general schooling – evaluating the expansion of vocational education in China and laying the foundation for further vocational education.

OW3.1222: Impact assessment of credit programme for tenant farmers in Bangladesh.

OW2.171: Enhancing food production and food security through improved inputs: an evaluation of the National Agricultural Input voucher scheme of Tanzania with a focus on gender impacts.

OW1.20: Building a brighter future – a randomized experiment of slum-housing upgrading in Mexico.

OW3.1075: Cash transfers, health insurance and health outcomes in the Livelihood Empowerment Against Poverty (LEAP) of Ghana

Independent Evaluation Unit
Green Climate Fund
175 Art center-daero, Yeonsu-gu,
Incheon 22004, Republic of Korea
Tel. (+82) 032-458-6428
ieu@gcfund.org
https://ieu.greenclimate.fund

GREEN CLIMATE FUND | Independent Evaluation Unit