



Results-based financing has potential but is not a silver bullet – *Theory-based evaluations and research can improve the evidence base for decision making*

By Ida Lindkvist and Per Øyvind Bastøe, the Evaluation Department, Norad

1. INTRODUCTION

Results-based financing (RBF) is hailed as a promising policy tool in everything from promoting maternal and child health to reducing deforestation and emissions from climate gasses. A focus on results, often combined with the idea that no costs will be born unless results materialize, is alluring. Emerging results from some sectors seems to offer promising results. Norway has taken a lead role in employing RBF in three sectors: health, climate change and energy and is currently advocating furthering its use in the education sector. To learn more about Norwegian experiences with employing RBF schemes, the Evaluation Department in the Norwegian Agency for Development Cooperation (Norad) has commissioned two desk studies of the experiences with RBF in Norwegian development assistance. The main finding from these studies is that the basis for decision making has largely been political and that results documentation, with some exceptions, remains limited. While the health initiative to some extent is an exception, we argue that all initiatives can benefit from more strategically embedding theory-based evaluations to ensure a better evidence base for decision making.

This discussion paper draws on two studies commissioned by the Evaluation Department, as well as other research on RBF. In the paper, we attempt to link findings from this research to the question of how to proceed with RBF in Norwegian Development Cooperation to ensure a better evidence base for decision making.

The two studies: "Experiences with results-based payments in Norwegian development aid" (study 1) and "Basis for decisions to use results-based payments in Norwegian development aid" (study 2) are written by Johan Helland and Ottar Mæstad and are available as separate publications at: <http://www.norad.no/en/front/evaluation/>

2. RESULTS-BASED FINANCING IN NORWEGIAN DEVELOPMENT ASSISTANCE

Projects involve RBF, or “payments by results” as Mæstad and Helland prefer to call it, if they include two central aspects: *1) the payment is based on achieved results, and 2) the relationship between payment and results is pre-defined.* The major difference between the different types of RBF is whether the schemes target the behaviour of states, organisations and/or individuals. Since 2007, Norway has advocated the use of RBF at all three levels, through major initiatives in health, climate and forestry, as well as in clean energy since 2011.¹

Under the health initiative, RBF targeting organisations and individuals have been used as a means to improve maternal and child health. Furthermore, RBF-targeting states have been employed to increase vaccination coverage, with the additional objective to improve health outcomes. The majority of RBF funding under the health initiative has been channelled through the Health Results and Innovation Trust Fund (HRITF), for which Norway is one of two donors, and Gavi, the Vaccine Alliance, for which Norway is also one of the alliance’s original donors. The fund is located in the World Bank, and Norway has committed NOK 2.1 billion to the fund’s activities during the period 2007–2022. Norway has also entered into bilateral agreements with countries interested in testing the effects of RBF schemes for improved maternal and child health. Norway has supported Gavi since the alliance’s start-up in 2001 with about USD 1.11 billion in direct funding. Apart from direct funding, Norway also contributes to the alliance through other funding streams. It should be noted, however, that the alliance’s work covers more than RBF to states.

The Norwegian International Climate and Forest Initiative (NICFI) employs the REDD+ (Reducing Emissions from Deforestation and Forest Degradation)

DEFINITION OF RESULTS-BASED FINANCING:

- 1) the payment is based on achieved results, and*
- 2) the relationship between payment and results is pre-defined.*

RBF schemes can target states, organisations and/or individuals. Payments targeting states are often called results-based aid, or cash on delivery (COD aid).

Payments to individuals supplying services are sometimes called pay for performance (P4P), while payments to individuals demanding services are often called conditional cash transfers.

mechanism to channel RBF to states as a means to reduce emissions from deforestation and forest degradation. Norway has paid about NOK 5.45 billion to Brazil and NOK 0.97 billion to Guyana through this REDD+ framework. Norway has also planned to employ RBF to promote clean energy through the initiative Energy+ initiative; however, Helland and Mæstad do not provide any figures for disbursements, possibly because the RBF component has not yet been implemented. For this reason, our focus in this discussion paper is mainly on the health and climate and forestry initiatives.

It is important to learn from existing experiences, especially since RBF appears to be in vogue. In the white paper Education for Development (MFA, 2014), the Norwegian Ministry of Foreign Affairs also advocates the use of RBF in the education sector.

¹ With the exception of Gavi, the sources of funding figures from the Norwegian initiatives are the two studies conducted by Helland and Mæstad. The figures for Gavi are from Gavi’s website: <http://www.gavi.org/funding/donor-profiles/norway/>. Figures on the website were accessed on the 14th of May 2015.

3. CURRENT STATE OF THEORIES OF CHANGE AND RESULTS DOCUMENTATION

The main conclusion from the two studies looking at the experiences with and the basis for decision making is that the basis has largely been political and that results documentation, with some exceptions, remains limited. To illustrate this finding analytically and to highlight strengths and weaknesses with existing results documentation, we have created Table 1 (page 4). In this table, we classify the different schemes as with respect to target groups for payments (individuals/organisations/states). We classify the state of theories of change at start-up, including the theoretical foundation today (1). When it comes to theoretical foundation today, we only consider the literature regarding the RBF component and not sector literature in general.

The current state of results documentation is assessed with respect to a) level of change measured, b) whether adverse effects are documented, and c) whether evaluations/research address attribution (2).

If (1) and (2) are present, evaluations can potentially explain whether interventions work as intended (3). In addition we have assessed whether evaluations and research can explain how and why changes occur (4). Even if evaluations can document whether programmes work, this will only offer limited information to policymakers, unless they can also explain why.

In the remainder of this paper, we will go into detail about lessons learned regarding the current state of results documentation and programme theories of change. The limits to both, perhaps with the exception of the health initiative, may imply that policymakers have limited evidence to base future decisions on. To improve the future evidence base, we offer some general lessons based on the findings from the two studies commissioned by the Evaluation Department.

3.1 Theory of change/theoretical basis

A good programme theory, a theory of change, is essential for RBF schemes. A theory of change can both ensure a well-thought-through programme design and evaluations that explain how and why changes occur. Theories of change can be very simple, especially

THEORY OF CHANGE

A theory of change explains what one wants to achieve, what needs to change and why, how this relates to the current context, and how changes are meant to come about (Vogel 2012).

Good theories of change should be explicit about underlying assumptions and link to existing theoretical and empirical literature. In cases where effects are well known, and unlikely to be context specific, theories of change can be very simple. This is unlikely to be the case for RBF where theories of change are important both as a programmatic tool, but also to ensure that we have good evidence regarding decisions to continue or discontinue projects.

so when programme effectiveness is well known and unlikely to be context specific. This is not the case for RBF, as we argue that theories of change should be both context and project specific.

As is demonstrated in Table 1, all programmes had weak theories of change at start-up. This does not imply however that these initiatives arose in a theoretical vacuum. RBF that targets individuals has a strong foundation in several disciplines. The theoretical basis for using RBF to states is more limited. In developing these theories, practitioners should be careful about generalizing theories developed to explain the effects of RBF on individual behaviour to states.

3.1.1 More work needs to be done on theories of change on the use of RBF to states

In study 2, Mæstad and Helland writes, “*The core idea of any results based approach is that incentives affect behaviour.*” This idea has a strong intuitive appeal, and for payments to individuals, it has a strong theoretical foundation. According to agency literature,² an employer (a principal) can use financial incentives to influence an employee’s behaviour when the employer and the employee have diverging interests (preferences).

² For an overview of agency theory, see (Eisenhardt, 1989).

TABLE 1: THE STATE OF RESULTS DOCUMENTATION

	TYPE OF RBF SCHEME	1. THEORY OF CHANGE AT START-UP/ THEORETICAL BASIS	2. RESULTS DOCUMENTATION: ASPECT OF DEVELOPMENT EFFECTIVENESS MEASURED			3. DOES IT WORK?	4. EXPLAINS HOW AND WHY
			a. Change measured (level)	b. Adverse effects	c. Attribution		
Health	Individuals/ organisations	Weak/strong theoretical basis	Focus on changes in service delivery/ utilization with some exceptions	Yes, to some extent	Yes	Possibly, evidence is mixed. Too early to tell.	Limited
	States	Weak/ weak theoretical basis	Vaccine uptake measured	Yes, to some extent	Limited	Possibly, mixed evidence	Limited
Climate and forestry initiative	States*	Weak/weak theoretical basis	Reduction in deforestation/ emission reduction	Little documentation	No	Do not know	Limited
Energy	Individuals/ organisations	Little at programme level/ relevant literature exists	n/a	n/a	n/a	n/a	n/a
	States		n/a	n/a	n/a	n/a	n/a

Note: The table is based on the two desk studies conducted by Helland and Mæstad (Norad, 2015a, Norad, 2015b).

* In addition, REDD also includes payments for environmental services (PES), which can be classified as RBF to individuals and organisations. However since PES is a voluntary, negotiated agreement, we do not include this here.

In other words, if the employee prefers to do other things than work, then the employer can impose a fine or a reward or fire the employee if she or he does not do what the employer wants. Or to put simply, employers can give someone money to induce them to do something they otherwise would not have done. Clearly, individuals have different types of motivation, and monetary incentives are unlikely to be the sole source of work motivation. That said, financial incentives have been demonstrated to be important in the workplace (Prendergast, 1999), especially in areas where it is easy to link payments to the outcomes of interest of the employer.

While the theoretical foundation for offering incentives to change individual behaviour is strong, we know less about how offering incentives to a government or state may work. Since both financing types are grouped together, it may be tempting to extrapolate theories of change from RBF schemes targeting individuals to schemes targeting states.

There is a danger however to transfer this simple idea from individuals to states or sometimes even to organisations. Paying a government for a reduction in deforestation need in fact not involve a single monetary transfer of resources to the individual that makes important decisions regarding deforestation. In fact, in summarising the experiences, Mæstad and Helland cannot find evidence that any changes in outcomes/outputs of interest are linked to the financial incentive as such. A lack of evidence does not necessarily imply a lack of effect, and even though offering financial incentives to governments is unlikely to work in the exact same way as for RBF to individuals, financial incentives can still play a role.

Increasing the resources available to policymakers could work as an incentive (to make them do something they otherwise would not have done), as additional money could increase a politician's political capital. However, in the case of the Amazon fund,³ the decision to spend

³ The Amazon Fund raise funds for "investments in efforts to prevent, monitor and combat deforestation, as well as to promote the preservation and sustainable use of forests in the Amazon Biome" (<http://www.amazonfund.gov.br/>). The Fund is managed by the Brazilian Development Bank.

the funds is to a large degree outside the control of the policymaker in charge of emission reductions or deforestation.

Payments to governments could also have a positive impact on, for example, deforestation through other channels. For an overview of different channels through which RBF to states (or results-based aid) can work, see Perakis and Savedoff (2015) as cited by Mæstad and Helland. Perakis and Savedoff argue that the attention to results is probably the most important. If attention is most important, then any policymaker considering employing RBF should also consider other modes to garner political attention.

It is worth mentioning that the attempts to use aid to influence policy choices fall into a long tradition in Western aid, known as policy conditionality, which to some degree can be compared with results-based financing. Research on its effectiveness has generally been disappointing (e.g. see (Riddell, 2007: chapter 14).

3.1.2 Theories of change need to be project and context specific

A lesson from evaluation and research funded through the health initiative is that RBF will not necessarily work in a different sector and context. This underscores the importance of context and project-specific theories of change for existing and future RBF projects.

In the Norwegian White paper Education for Development (MFA, 2014), RBF is specifically mentioned as one of the new tools for development assistance to education. Norway is mentioned as a pioneer in the field, and success stories from RBF for improved maternal and child health are highlighted, perhaps to demonstrate that RBF can offer positive results when more traditional approaches have failed. While this may certainly be true, the two studies that we have commissioned provide an important insight. Similar types of projects may end up with different outcomes if they are altered slightly or moved from one context to another, even when the simple idea behind these schemes is the same. While individuals do respond to incentives, one may argue that how they respond



PHOTO: KEN OPPRANN

Under Gavi, governments receive payments for additional vaccinations exceeding a baseline. This is an example of a RBF scheme targeting states.

and ultimately effectiveness will crucially depend upon a) what you are paying for, b) how the financial incentive is perceived by individuals and a group (preferences), c) the opportunity set of the individual and d) the relative size of funds. A change in any of these may mean that a RBF scheme that is effective in one context and sector may not work in a different sector or context.

Policymakers should not interpret this as RBF not being effective, only that success from some areas cannot be easily transferred to others, and as such, employing RBF implies that investments in good programme theories and theory-based evaluations are essential to ensure a good evidence base for decision making.

Another important lesson from the theoretical literature on results-based payments to individuals is that possible adverse effects,⁴ that is, negative non-intended effects, make it difficult to predict the effectiveness of a project (see e.g. (Oxman and Fretheim, 2009) for a summary). Adverse effects described in relation to health tend to fall into two categories: crowding out of tasks and crowding out of motivation.

4 Observe that adverse effects are not precisely the same as unforeseen effects, as adverse effects are negative effects that arise and which theory may or may not be able to predict. If adverse effects are not predicted by theory, they are unforeseen. Unforeseen effects may also be positive. Given the large theoretical literature on RBF to individuals, the type of adverse effects can in many cases be predicted, although the form it takes may not.

In addition, one may be concerned about corruption, that is, doctoring the numbers rather than the patients.⁵ The basic idea behind the first crowding out category is that since we are paying for something other than what we actually want to achieve, that is, improved health outcomes,⁶ health workers may focus their energy only on aspects of work that maximize rewards and ignore other aspects that may be important for health outcomes. Examples include focusing on patients that are easy to treat or improving quality aspects observable to patients, while ignoring others that are not. Motivational crowding out simply means that workers' main source of motivation changes from intrinsic (internal) to extrinsic (external)⁷ (Frey and Jegen, 2000, Benabou and Tirole, 2003). Even if motivational crowding out occurs, this may not necessarily mean that the scheme will be ineffective; the effect of the pecuniary incentive may be so strong that it counteracts any motivational crowding out. That said, it might matter a great deal if RBF projects are discontinued.

Adverse effects may also arise when it comes to RBF schemes targeting states, and new theories developed to explain how such schemes work should also look closer at potential adverse effects.

It is important to note that even though theory may predict the possibility of adverse effects, this does not mean that RBF may have a negative impact. For one, adverse effects may not be present. For another, the positive effects may counteract any negative effects that arise. Hence, the question of adverse effects is to some extent an empirical question. It is still important to include such effects in a theory of change to increase the chance that an evaluation will be able to pick up such effects.

5 If health workers alter the number of visits to achieve a bonus, without there being any actual improvement, this can be categorized as corruption under the standard definition: abuse of public power for private gain.

6 This is not a critique of the choice of what one pays for. Paying directly for improved maternal and child is probably not a good idea for obvious reasons.

7 A simple allegory to explain the difference is a girl playing a piano: enjoyment of playing is an example of intrinsic motivation, while playing to achieve a gold star is an example of extrinsic motivation.

General recommendations for the development of programme theories of change are given below:

- > **RBF and RBF are not the same.** Be careful about extrapolating theories of change or evidence from RBF schemes. Theories of change need to be context and project specific. Theoretical development is probably required to understand RBF to states (results-based aid).
- > **RBF is likely to have an effect, but need not be effective.** Theories of change should include the possibility of adverse effects.

3.2 Limited available results documentation

The main finding from Table 1 is that it is difficult to answer the question of whether any of these initiatives work. This is partly due to the limited available results documentation.

3.2.1 Strong results documentation in the pipeline for RBF health projects to individuals/organisations

As can be seen from Table 1, the health initiative is likely to have the strongest level of results documentation when evaluations are ready. Although the specifics of these RBF health schemes vary from project to project, they all tend to tie financial incentives to service utilization and sometimes to quality indicators of maternal and new born-care services. Incentives typically target health workers, facilities and sometimes higher-level management, with the aim to improve maternal and child health. The reason that we expect results documentation to be stronger is that all these projects tend to have an evaluation component, typically an impact evaluation, and sometimes a process evaluation. The use of impact evaluations (see definition in the box to the right ⁸) ensures that the initiative measures changes convincingly and can attribute changes to the programme.

As for the answer to the question in Table 1 of whether these initiatives work, the short answer is that evidence

IMPACT EVALUATION

“Impact evaluations compare the outcomes of a program against a counterfactual that shows what would have happened to beneficiaries without the program. Unlike other forms of evaluation, they permit the attribution of observed changes in outcomes to the program being evaluated by following experimental and quasi-experimental designs.”

The Development Impact Evaluation Initiative (DIME)

is mixed and that it is too early to tell since most evaluations are still in the pipeline. Helland and Mæstad identify five studies looking at the effect of the Norwegian funded⁹ RBF schemes aimed at improved maternal and child health. Of the published studies, some of them indicate positive results, while others do not. Several studies indicate that RBF to health workers and patients can increase the number of women who deliver at health clinics and can possibly improve health outcomes for children (Basinga et al., 2011, Gertler and Vermeersch, 2013, Gertler and Giovagnoli, 2014). Other studies, such as the evaluation documenting the effects of the RBF scheme in the Congo, have found no impact on the indicators of interest (HRITF, 2014). This confirms the theoretical prediction that RBF schemes are likely to be project and context specific.

While some of the published evaluations look at health outcomes for children, the impact on maternal mortality is less certain, since neither the schemes nor the published evaluations look at maternal health outcomes. Documenting effects with respect to health outcomes may be extremely costly, particularly for maternal health. When incentives are not directly tied to the outcome of interest to the principal (i.e. health outcomes), the existence of adverse effects may crowd out the positive effects of the incentives. If adverse effects are not systematically documented and we do not evaluate projects with respect to the outcomes of interest (for example, health), then there is a real danger that we overestimate the positive effects of these programmes.

⁸ Source: http://siteresources.worldbank.org/INTDEVIMPEVAINI/Resources/DIME_project_document-rev.pdf

⁹ The HRITF is funded by Norway and the UK (DFID).

For the Norwegian initiative for maternal and child health, early experiences indicate that adverse effects need not dominate. Studies from Rwanda and Argentina indicate positive effects on child health outcomes (Gertler and Vermeersch, 2013, Gertler and Giovagnoli, 2014). Two of the five impact evaluations (Tanzania and the DCR) looked at task crowding out and found no indication that health workers reduced efforts to the provision of non-incentivized services (Binyaruka et al., 2014, HRITF, 2014). Furthermore, the evaluation from Tanzania indicates that informal user fees (“gifts” to health workers) were in fact reduced (Binyaruka et al., 2014). This study also looked for evidence of motivational crowding out and found no indication of this. On the other hand, qualitative research from Tanzania also found indications of less positive effects, such as coercive practices to increase demand (Chimhutu et al., 2014). That said, it is less clear whether such documentation is systematically collected across the projects, as some types of adverse effects appear to be documented in some projects and not others.

For evaluations to be able to pick up these effects, programmes and projects should be followed by both quantitative and qualitative research. Employees may engage in innovative strategies to maximize payments, and even a very good theory of change need not necessarily be able to pick up these in advance.

A closer look at the extent to which adverse effects are systematically measured may offer more information on the quality of results documentation.

3.2.2 Weaker results documentation for RBF schemes targeting states

In general, results documentation for RBF to states is weaker than for RBF schemes to individuals and organisations. One reason is the difficulty in attributing changes to a programme. It is unlikely that attribution can be established through impact evaluations for any of these initiatives. Instead, theory-based evaluations can discuss causality by systematically analysing evidence along the theory of change. See for example

White and Phillips (2012) for an overview of different social science methods that can be employed systematically to discuss causality and attribution. A precondition for theory-based evaluations that discuss causality is probably better-developed programme theories.

That said, the different initiatives also have different challenges with respect to results documentation. For Gavi, one difficulty concerns establishing the accuracy of changes in vaccination numbers. While there is an indication that support has increased vaccination rates, some have questioned whether vaccination numbers are inflated.

When it comes to the effects on deforestation and emission reductions, evaluators are reasonably sure that changes have occurred, although there is more uncertainty with respect to whether results can be attributed to the programmes.

Although we have fewer theoretical insights into the potential adverse effects for RBF schemes targeting states, Helland and Mæstad report results from an evaluation of Gavi’s vaccination programme that indicate that vaccination coverage did not drop for those vaccinations that were not covered under the programme. Without good programme theories that systematically address adverse effects, there is a risk however, that evaluations do not pick up such effects even if they exist.

In conclusion, it is difficult to answer whether any of these initiatives have worked as intended. Given the weak state of programme theories of change of RBF programmes to states, it is not surprising that evaluators may struggle to address the causal effects of these initiatives.

General lessons for improved evaluations of RBF schemes targeting individuals, organisations and states are summarised below:

> **Evaluations should ideally test the effect of RBF schemes on outcome or impact indicators and be able to explain why changes occur and whether they can be attributed to the programme.** It may be easier to address attribution for RBF schemes targeting individuals or organisations through impact evaluations; however, causality can also be discussed for results-based financing to states. Theories of change are likely to be particularly important in such discussions.

> **Adverse effects should always be addressed.** This is particularly important if RBF schemes are related to output, rather than outcome or impact indicators, or if addressing attribution is difficult.

3.3 Evaluations do not explain how and why RBF is supposed to work

Few of the evaluations available can explain whether it is the RBF component that has resulted in observed changes in the indicators of interest. This is unfortunate because it limits the value of the information for policymakers. An obvious alternative to conditional funding is unconditional funding.

3.3.1 RBF schemes are rarely tested against viable options

As stated above, theory-based evaluations are likely to be particularly important as a tool for policymakers when it comes to RBF. When it comes to RBF to individuals and organisations, evaluations can be utilized, not only to attribute effects to the programme but also to test different approaches against each other. With a few exceptions, this has rarely been the case. Normally, RBF is tested against the status quo and not against viable options. Clearly, the best alternatives may depend upon the specific context of the programme. That said, a good option is simply to test the effectiveness of RBF compared to increasing the resources available. This can also explain why changes occur. If the financial incentive is important, then tying funds to conditions is likely to be more effective than not.



PHOTO: ESPEN RØST
The Norwegian International Climate and Forest Initiative (NICFI) employs the REDD+ mechanism to channel RBF to states as a means to reduce emissions from deforestation and forest degradation.

An important question arising from the two studies is whether changes occurring after the introduction of RBF are due to the incentive in offer or simply the provision of additional resources. The difference may seem trivial; however, if effects are due to the additional resource provision rather than the incentive, then simply offering additional resources is likely to be more cost-effective, since monitoring is no longer as important. Incidentally, the Norwegian-financed study in Rwanda by Basinga et al. (2011) does test this, as they offer the control group a similar amount of resources and find that linking resources to results is more effective than simply offering more resources. However, it is by no means certain that this would be true for all types of RBF schemes.

Conditional versus unconditional cash transfers is another good example. Providing conditional cash transfers to send kids to school is a RBF scheme targeted to families. The underlying idea is that not sending kids to school is a choice made by parents, and by offering them money to send kids to school, then the parents will change their decision (due to the incentive). If this is correct, then it is crucial to monitor whether kids are actually sent to school and to remove the cash transfer if kids do not show up. However, if the decision is simply a result of poverty, that is, if parents

cannot afford it, then an unconditional cash transfer will be equally effective. An unconditional cash transfer is however not RBF, and incidentally, it may be much cheaper to implement because it requires no monitoring.

Hence, if the real reason for low school attendance is poverty, then the project policymakers should scale up is unconditional cash transfers, as this may be more effective. A recent systematic review by (Baird et al., 2013) found that both conditional and unconditional cash transfer schemes increase school enrolment and that the difference between unconditional and conditional schemes is insignificant. However, when they divide conditional schemes by the intensity of the conditions, they find larger differences in favour of conditional cash transfers being more effective in increasing school enrolment.¹⁰

The final lesson is that, while RBF is in vogue, policymakers and practitioners should carefully consider whether simply increasing resources available is equally effective. Furthermore, when feasible, impact evaluations of RBF schemes can potentially allow for more rigorous testing of other alternatives as well.

- > **Test alternatives.** When possible, RBF should be tested against viable policy options. An alternative to most conditional transfers of funds are unconditional transfers.

¹⁰ They find, however, at best, only small effects on test scores.

4. CONCLUDING REMARKS

While the three initiatives have several aspects in common – for example, weak theories of change at the start-up phase and difficulties in explaining why initiatives work – results documentation appears to have been strongest for the health initiative. This initiative has routinely employed impact evaluations to test the effectiveness of projects, and even though it is too early to conclude with respect to effectiveness, one should ask whether it is at all possible to demand the same amount of scientific rigour from the other initiatives. We have argued that RBF to states and individuals differ fundamentally, and this matters for the choice of evaluation methodologies. For one, the nature of RBF schemes targeting individuals and

organisations implies that evaluation methods can credibly establish attribution and document effects through impact evaluations. This is not as easy for RBF that targets states, as there is no counterfactual. For another, RBF that targets individuals and organisations can rely on a strong theoretical literature; even though empirical studies are scarce, the same is probably not the case for the climate and forest initiative.

That said, improved theories of change, combined with theory-based evaluation, specifically addressing the issue of causality, could potentially also improve the evidence base for these initiatives. In addition, all initiatives should be able to explain why programmes work.

THEORY-BASED EVALUATIONS AND RESEARCH CAN IMPROVE THE BASIS FOR DECISION MAKING

Embed programme and project-specific theories of change explaining how and why changes are expected to occur, into the planning of projects.

RBF and RBF are not the same. Be careful about extrapolating theories of change or evidence from RBF schemes. Theories of change need to be context and project specific. Theoretical development is probably required to understand RBF to states (results-based aid).

RBF is likely to have an effect, but need not be effective. Theories of change should include the possibility of adverse effects.

Introduce theory-based evaluations that test whether change occurs and can be attributed to the programme. If possible, also test RBF schemes against viable policy options.

Evaluations should ideally test the effect of RBF schemes on outcome or impact indicators and be able to explain why changes occur and whether they can be attributed to the programme. It may be easier to address attribution for RBF schemes targeting individuals or organisations through impact evaluations; however, causality can also be discussed for results-based financing to states. Theories of change are likely to be particularly important in such discussions.

Adverse effects should always be addressed. This is particularly important if RBF schemes are related to output, rather than outcome or impact indicators, or if addressing attribution is difficult.

Test alternatives. When possible, RBF should be tested against viable policy options. An alternative to most conditional transfers of funds are unconditional transfers.

The Evaluation Department, located in Norad, initiates evaluations of activities financed over the Norwegian aid budget. The Department is governed under a specific mandate and reports directly to the Ministry of Foreign Affairs. The evaluations are carried out by independent evaluators, and all evaluation reports are made public.

May 2015 **Norad**

No of copies: 200 Norwegian Agency for Development Cooperation
ISBN: 978-82-7548-806-8 www.norad.no
Cover photo: Ken Opprann post-eval@norad.no

REFERENCES

BAIRD, S., FERREIRA, F. H. & WOOLCOCK, M. 2013. Relative effectiveness of conditional and unconditional cash transfers for schooling outcomes in developing countries: A systematic review. *Campbell Systematic Reviews*, 9.

BASINGA, P., GERTLER, P. J., BINAGWAHO, A., SOUCAT, A. L., STURDY, J. & VERMEERSCH, C. M. 2011. Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: An impact evaluation. *The Lancet*, 377, 1421-1428.

BENABOU, R. & TIROLE, J. 2003. Intrinsic and extrinsic motivation. *Review of Economic Studies*, 70, 489-520.

BINYARUKA, P., PATOULLARD, E., GRECO, G., POWELL-JACKSON, T., MÆSTAD, O. & BORGHI, J. 2014. Effect on universal health coverage of a pay for performance programme in Tanzania: A controlled before and after study. Discussion paper.

CHIMHUTU, V., LINDKVIST, I. & LANGE, S. 2014. When incentives work too well: locally implemented pay for performance (P4P) and adverse sanctions towards home birth in Tanzania-a qualitative study. *BMC health services research*, 14, 23.

EISENHARDT, K. M. 1989. Agency theory - an assessment and review. *Academy of Management Review*, 14, 57-74.

FREY, B. S. & JEGEN, R. 2000. Motivation crowding theory: A survey of empirical evidence, University of Zurich, Institute for Empirical Research in Economics.

GERTLER, P. & VERMEERSCH, C. 2013. Using performance incentives to improve medical care productivity and health outcomes. National Bureau of Economic Research.

GERTLER, P. J. & GIOVAGNOLI, P. I. 2014. Rewarding provider performance to enable a healthy start to life: Evidence from Argentina's Plan Nacer. World Bank Policy Research Working Paper.

HRITF 2014. Performance base financing for health in Haut-Katanaga. Impact Evaluation Report. Executive summary. www.rbfhealth.org. Washington D.C.: World Bank.

MFA 2014. White paper: Report No. 25 to the Storting: Education for Development. Oslo.

NORAD 2015a. Study a: Experiences with results-based payments in Norwegian development aid. Oslo: The Evaluation Department.

NORAD 2015b. Study b: The basis for decisions to use results-based payments in Norwegian development aid. Oslo: The Evaluation Department.

OXMAN, A. & FRETHEIM, A. 2009. Can paying for results help to achieve the Millennium Development Goals? Overview of the effectiveness of results-based financing. *Journal of Evidence-Based Medicine*, 2, 70-83.

PERAKIS, R. & SAVEDOFF, W. 2015. Does results-based aid change anything? Pecuniary interests, attention, accountability and discretion in four case studies. Center for Global Development.

PRENDERGAST, C. 1999. The provision of incentives in firms. *Journal of Economic Literature*, 37, 7-63.

RIDDELL, R. C. 2007. Does foreign aid really work?, Oxford University Press.

VOGEL, I. 2012. Review of the use of 'theory of change' in international development. Review Report. London DFID.

WHITE, H. & PHILLIPS, D. 2012. Addressing attribution of cause and effect in small n impact evaluations: Towards an integrated framework. Working Paper 15. International Initiative for Impact Evaluation 3iE. London, UK.