

Evaluating Development Impact Bonds

A Study for DFID

Conducted by
Roger Drew and Paul Clist

February 2015

Commissioned and funded by



Department
for International
Development

This report and associated publications were commissioned and produced using funding from DFID. However, the views expressed are those of the independent academic authors and do not necessarily reflect the views of DFID

Executive Summary

- S1. Development Impact Bonds (DIBs) are a new and potentially exciting aid instrument. They are recognised by DFID as one of three different types of Payment by Results (PBR), alongside results-based aid (RBA) and results-based financing (RBF). These instruments differ primarily in terms of who receives payment from the outcome or ultimate funder, for example, DFID. In results-based aid, the payment is made to a country's government. In results-based financing the payment is made to a service provider. In a Development Impact Bond, the funder, e.g. DFID, makes the payment to an investor who would have pre-financed the provision of services through the activities of a service provider supported, in most cases, by an intermediary.
- S2. This study seeks to present concepts for understanding DIBs. These are based on economic theory. The study then considers issues relating to evaluation of individual DIBs and how data from those evaluations can be synthesised effectively into an overall evidence base related to DIBs.
- S3. The study recommends that DIBs are understood using a model based on the actions of six agents or groups. This is referred to as a 'six-agent model'. These agents include the four agents that would be active in other forms of payment by results, such as RBA or RBF. They are:
- An outcome payer, that is, the funder if the expected outcome is achieved. In many cases, this would be DFID.
 - A country's government, in the case of RBA or the service provider in the case of RBF
 - The target population, that is, the people benefiting from the services provided
 - A validating agency, that is, the agency that verifies that the results reported are correct.
- S4. DIBs involve two additional agents. These are specific to DIBs and are not involved in other forms of payment by results, such as RBA or RBF. They are the investor and, in most cases, an intermediary. Theoretically, it may be possible for DIBs to be implemented without an intermediary. However, all DIBs that have been designed to date have included an intermediary.
- S5. The addition of these two agents, the investor and the intermediary, changes the role of the other four agents so that their role in a DIB may differ from their role in other forms of payments by results. For example, in DIBs it is the investor who provides initial funding, while in RBA this is provided by the recipient government and in RBF by the service provider. Despite these differences and the differences in context in which the different forms of payments by results operate, these three forms of payment by results are similar in many ways. This means that much of the economic analysis of RBA and RBF is also relevant to DIBs.
- S6. Social Impact Bonds (SIBs) are closely related to DIBs. More than 40 such programmes now exist across several countries. However, no DIBs have yet started although several are at an advanced stage of design. Given that SIBs and DIBs are new and innovative, they need robust and rigorous evaluations. This must include evaluations of both process and impact.

- S7. The Centre for SIBs in the Cabinet Office in the UK proposes four excellent types of questions for evaluations of SIBs. These cover impact, process and economic evaluation and can also be applied to DIBs. There is emerging experience of evaluations of SIBs internationally. One randomised controlled trial is underway in New York. Approaches that are similar to experiments but without randomisation (quasi-experimental) have been used to assess outcomes for payment in the US (Rikers Island) and the UK (Peterborough). There have been a number of lesson learning evaluations and reviews of individual SIBs. There have also been evaluations of the economic benefits of SIBs and the development of a market for SIBs. There have been a number of initiatives to synthesise learning across multiple SIBs. These have largely been consultative exercises in which opinions are gathered from relevant stakeholders. In the UK, the Big Lottery Fund has set up an evaluation across its overall SIBs portfolio in addition to evaluations of individual SIBs. Concerning synthesising learning from DIBs, a number of respondents advocate the establishment of a DIBs Community of Practice to synthesise learning from evaluations into a body of relevant evidence.
- S8. This study presents an overall framework that could be used for evaluations of DIBs across DFID and beyond (see Figure 5). Key features of this framework are that it provides a framework that can be used to evaluate both processes and impact of DIBs; it can be mapped to the important criteria of relevance, effectiveness, efficiency, impact and sustainability which are recommended for use by OECD DAC in evaluations of development programmes; it allows mapping, analysis and identification of any gaps where a number of questions are proposed for a SIBs or DIBs evaluation; and it has been used to propose some general questions that could be asked in evaluations of DIBs (see Table 2)
- S9. This study also considers the types of methods that could be used for evaluations including situations when experimental methods, such as cluster-based randomised controlled trials might be possible and when they might not be suitable. It also considers different types of quasi-experimental approaches. It argues that the framework presented in Figure 5 provides a sound basis for a theory-based approach to evaluation. Such an approach could include data from many evaluation methods.
- S10. Finally, the study concludes by considering approaches that could be used to synthesise data from individual evaluations into an overall evidence base for DIBs. It seems unlikely that individual DIBs will be sufficiently similar in terms of context, sector, size etc. to allow them to be treated as individual experiments from which data could be aggregated. However, evaluation and lesson learning should be taken into account right from the start of designing a DIB. By having a common evaluation framework for DIBs, it should be possible to use this to synthesise evidence from individual DIBs evaluations. Doing this is likely to require some ongoing analytic capacity which could be provided through the proposed DIBs Community of Practice and/or an agent conducting this type of real-time synthesis on DFID's behalf. It seems likely that some form of active and intentional synthesis process will be needed to ensure that this happens. The model being used by the Big Lottery Fund in the UK of commissioning an evaluation across its portfolio of SIBs projects is one that DFID might consider emulating, perhaps by commissioning an evaluation of DFID's overall experience of DIBs. In addition, there is likely to be merit in periodic, retrospective reviews of evidence generated through DIBs evaluations. Conducting such synthesis exercises is likely to require similar techniques to those of a systematic review but the question is likely to be broader and there will be need to consider a wider range of evidence over and above that generated from randomised controlled trials.

Contents

Executive Summary	2
Contents	4
Glossary	5
Introduction.....	6
Overview of the Attributes of DIBs, Theory Supporting their Use in Development and Current Initiatives	7
Six Agents	9
Investor	9
Intermediary	9
Service Provider.....	9
Investor	9
Intermediary	11
Service Provider.....	11
Target Population.....	12
Validating Agency	12
Outcome Payer	13
Two paradigms: small and large DIBs	13
Cross Cutting Issues.....	14
Measure	14
Coordination	14
Risk Preferences.....	14
Interest Rates	15
Time Preferences	15
Innovation/Autonomy	15
Contractual Complexity	15
Strengths and Weaknesses of Existing Evidence and Evaluation Approaches and Methods	16
Experience of Developing SIBs	16
Experience of Developing DIBs	17
Relevant evaluation literature.....	18
Expectations of Evaluations of DIBs	19
Experiences of Evaluation.....	21
Experimental Approaches	21
Quasi-Experimental Approaches	21
Non-experimental Approaches: Individual Evaluations	22
Synthesis across Evaluations	24
Appropriate Evaluation Questions, Approaches and Methods for DIBs	28
Appropriate approaches and principles for evaluating DIBs.....	28
Appropriate questions for evaluations of DIBs.....	29
Appropriate methods to answer questions in evaluations of DIBs.....	34
A Strategy for Cumulating Knowledge across DIBs Initiatives	36
Evaluations should be designed around a common evaluation framework.....	36
Capacity is needed to conduct real-time synthesis.....	36
Periodic, retrospective synthesis exercises are likely to be very valuable	37
Lessons Learned and Key Recommendations	37
Annex 1: Terms of Reference	40
Annex 2: Bibliography	44
Annex 3: People Consulted.....	50
Annex 4: Questions Specified in the Terms of Reference for the Evaluation of the Commissioning Better Outcomes Fund.....	51
Annex 5: Mapping Proposed PBR Evaluation Questions to those Proposed for DIBs	55
Annex 6: Example Design for the Impact Evaluation of a DIBs Programme.....	56

Glossary

ABLE	Adolescent Behavioral Learning Experience
CGD	Center for Global Development
DAC	Development Assistance Committee
DFID	Department for International Development
DIB	Development Impact Bond
DWP	Department for Work and Pensions
FCA	Financial Conduct Authority
MOH	Ministry of Health
NGO	Non-Governmental Organisation
OECD	Organisation for Economic Cooperation and Development
PBR	Payment by Results
RBA	Results-Based Aid
RBF	Results-Based Financing
RCT	Randomised Controlled Trial
SIB	Social Impact Bond
SIBM	Social Impact Business Model
SIDA	Swedish International Development Agency
UK	United Kingdom
US	United States

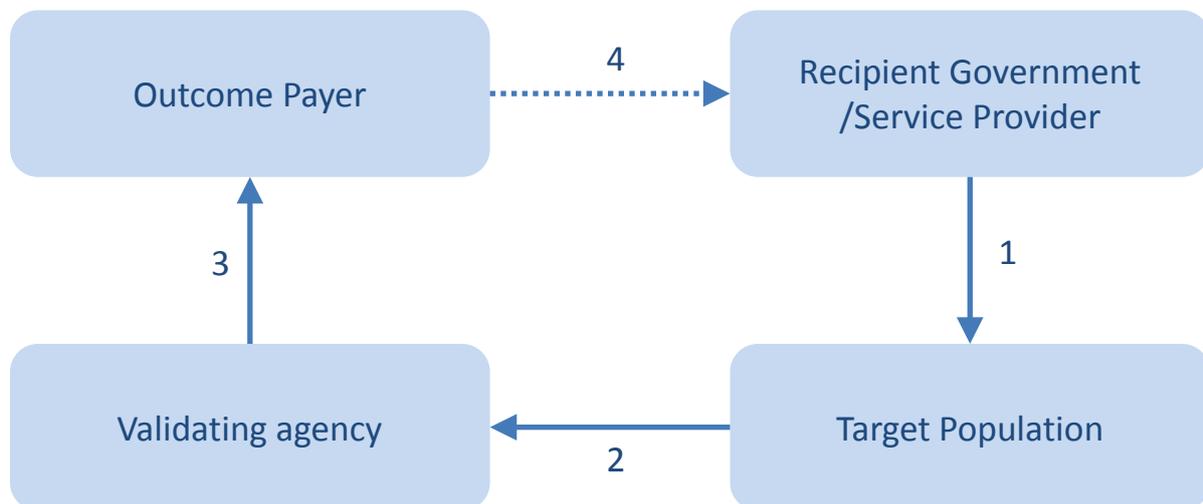
Introduction

1. There is a great deal of interest, within DFID and beyond, in Development Impact Bonds (DIBs) as a new aid modality for international development (see, for example, CGD and Social Finance, 2013a/b). DFID has recognised that DIBs, as a new aid modality, pose particular challenges for evaluation.
2. As a result, DFID commissioned this study, the main purpose of which is to assist DFID, other donors and policy makers to use the most appropriate evaluation approaches and methods for DIBs to enable wider learning. Terms of reference for the study are presented in Annex 1 (p40). This report is the main output of the study and presents its findings, lessons learned and key recommendations.
3. The study's method was described in more detail in an inception report (Drew and Clist, 2014). Essentially, the study has been conducted by a two person team who have reviewed relevant grey and academic literature and consulted a number of key informants. Details of documents reviewed are presented in Annex 2 (p44). Details of people consulted are presented in Annex 3 (p50).
4. This report is structured as follows. There is first a section which explores the attributes of DIBs and looks at the theory supporting their use. The section that follows explores the strengths and weaknesses of the current evidence base for DIBs and related instruments, such as SIBs. There are then sections on issues relating to evaluations of individual DIBs and how data might be synthesised from these to develop an overall evidence base for DIBs. Finally, there is a section which pulls together lessons learned and recommendations from this exercise.

Overview of the Attributes of DIBs, Theory Supporting their Use in Development and Current Initiatives

- DIBs are currently understood as one of three types of payment by results (DFID, 2014c), along with Results Based Aid (RBA) and Results Based Financing (RBF). DFID's preferred classification relates to who receives payment in each contract, contingent upon measured results. For RBA, aid is disbursed to the recipient government according to an agreed measure, and so the 'delivery risk' is borne by the recipient government. For RBF, aid is disbursed to a service provider according to an agreed measure, and so that risk is borne by the service provider (such as an NGO). Both of these cases can be summarised as interactions between four parties as displayed in Figure 1. The outcome payer, for example DFID, agrees to pay the recipient government or service provider on the basis of measured changes in the target population based on reports from the validating agency. The money that is disbursed is contingent upon measured success (hence the dotted line in Figure 1) may be a part or the whole of the overall contractual arrangement. Both of these types are being extensively piloted by DFID and other donors.

Figure 1: Four-agent model of RBA and RBF¹ (numbers denote order of processes)



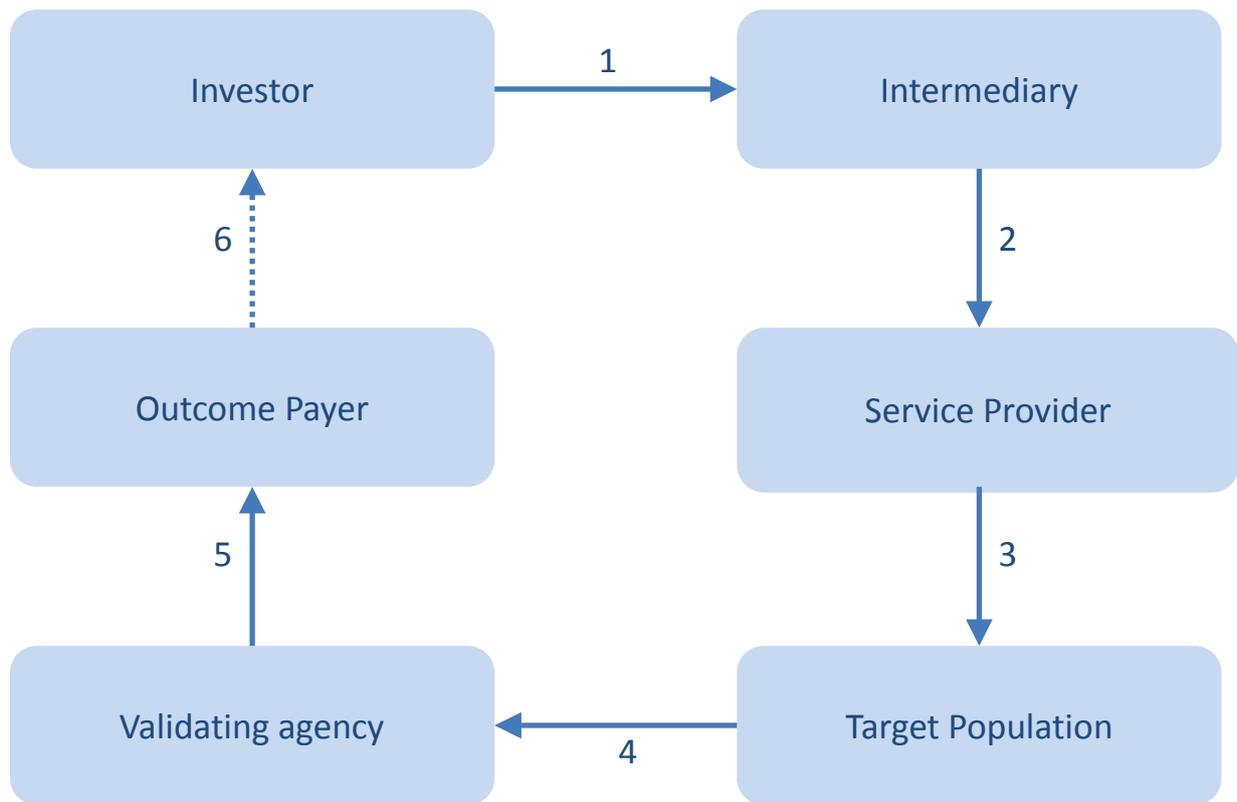
- As shown in Figure 2, the relationship is more complex in DIBs (or SIBs²), with two additional agents. With DIBs, an investor provides up front financing to an intermediary³. The intermediary uses this money to pay service providers to affect some target population. The outcome payer will then disburse funds to the investor on the basis of measured success as reported by the validating agency. As this payment is conditional it is represented by a dotted line.

¹ For similar diagrams see Hughes and Scherer (2014) and Burand (2013).

² For the purpose of this report, SIBs and DIBs are differentiated in the way proposed by Social Finance and Center for Global Development. In a SIB, the outcome funder is national or local government. In DIBs, the outcome funder is another donor or donors. As SIBs have a longer track record, they have a broader evidence base.

³ Although, as discussed below, an intermediary is not always strictly necessary in a DIB.

Figure 2: Six-agent model of DIBs (*numbers denote order of processes*)



7. The possible range of different contractual arrangements is vast, and so simplification is necessary. Table 1 provides a brief categorisation of three example projects using the six-agent model. The first is the most famous SIB: the Peterborough Prison Social Impact Bond, included because it has happened, even though it is not a DIB. The other two are DIBs for which there is a relatively high amount of publically available information⁴.

⁴ As both the Uganda Sleeping Sickness and Rajasthan DIBs are in design stage, many details are unknown and/or undecided.

Table 1: Examples of SIBs and DIBs using the six-agent model

Name	Peterborough Prison Social Impact Bond	Uganda Sleeping Sickness	Educational outcomes in Rajasthan, India
	Six Agents		
Investor	Mainly trusts and foundations	Health and agriculture focused trusts and foundations, high net worth individuals, Africa-focused impact investment funds	UBS Optimus foundation
Intermediary	Social Impact Partnership Ltd. (set up by Social Finance)	New body including universities	Instiglio
Service Provider	One Service (St Giles, Sava, Mind, Ormiston, John Laing, YMCA)	Unknown	Educate Girls
Target Population	Male Prisoners in Peterborough Prison	Initially Ugandan Cattle in high risk areas, ultimately the human population	Children in poorly performing schools in Rajasthan
Validating Agency	QinetQ and universities	Unknown	Unknown
Outcome Payer	Ministry of Justice and the Big Lottery Fund	Donor agencies e.g. DFID	Children's Investment Fund Foundation (CIFF)
	Other characteristics		
Control group	National male prisoners (live comparison)	Pre-intervention (Historical)	Historical
Underlying Aim	Reduction in reoffending	To reduce the risk for humans from Rhodesian sleeping sickness	Improve school learning outcomes for both genders in Rajasthan
Incentivised Measure	Reduction in reconviction	Spray and trypanocide (years 1-3) and spraying (years 4-8)	Unknown
Payment Trigger	Statistically significant improvement (10% in any of three cohorts, 7.5% overall)	65% treatment rate in high risk areas and staged target reduction in human infective parasite prevalence	Unknown
Notes	IRR return between -100% and 13% IRR	£1.5 million inception project,	\$238,000, return of 7-13%

- With the six-agent model, we can consider in turn the position of each agent before coming back to consider two paradigms and a number of cross-cutting issues.

Investor

- For the Peterborough Prison SIB, any investor considering the opportunity on purely financial terms would have been sorely disappointed. The return ranged from losing all of the initial capital, to making 13%. The idea was new, and consisted of an illiquid asset (i.e. it cannot be sold easily) that tied up money for several years, with the exact payment date unknown. In this situation the incentive for the initial investors was at least in part the social return, where investors valued the outcome of the intervention for its own sake. Blended return is the term used to describe these twin motivations of financial and social returns, and is a key part of impact investing.
- The DIBs market is not yet established, and there is recognition that foundations and trusts will need to play a larger part until the model is proven. In the Peterborough case, excitement in the model itself was added to the blended return motivation (Disley et al., 2011 pp.19-20). Rockefeller have expressed a willingness to 'pump prime' (Hughes and Scherer, 2014), and play a leading role in establishing the market by taking more risks than others may be willing to bear. While the return is always likely to be blended, any large DIBs will need to be viable as a business proposition, and the emphasis will become less on the social return and more on the financial return. The size of the expected DIB, the risks involved and the expected return will all need to be considered when estimating whether there will be sufficient demand for any given DIB. Even in future, it is quite likely that for any SIBs there will be a range of investor profiles, with different weightings on the social and financial returns, but any large DIB will need to attract investors whose interest is more geared to financial returns.

11. There are two conceptual difficulties with the blended return of investors. First, while the blended return means investors have two separate incentives to invest (social and financial returns), these are positively correlated. If investors weight the two returns equally, they will feel a large disutility if an intervention fails, which would compound the disutility of their lost capital. One of the attractive features of DIBs is that governments can avoid the political difficulty of funding projects that fail. However, for investors to fund DIBs they receive no such guarantee.
12. Second, experimental and behavioural economic research has shown that people evaluate market and non-monetary transactions according to different criteria (Gneezy and Rustihini, 2000). This research relates to individuals, and so it is unclear how this research applies to different bodies. However, the research should cause us to consider the idea that blended returns will gradually become more balanced to the market proposition side. Essentially, charity and business propositions are evaluated according to different criteria, and interact in unexpected ways. To give a brief example, imagine a customer in a local charity shop. They may be willing to buy an item in the shop for £5 and donate £100 to the charity in the same year. However, they would generally not be willing to buy the same item for £105 and give no donation, despite the identical end result. Each transaction is assessed according to its own criteria, and while the £5 purchase may be a more generous price than another shop, it will be anchored to other prices. This implies that while blended returns mean that DIBs will not need to compete solely on the financial fundamentals, they will need to be more or less in line with other investments in order to be viable.
13. In the early days of SIBs and DIBs, one of the offered reasons for their use was that it would mean an increase in finance as new money was attracted to the sector (see discussion by Green, 2013). Recently this has become downplayed (Barder and Perakis, 2014), although it is unclear whether this aim is still in place over the longer term.
14. The investors are, however, expected to bring private sector knowledge. As CGD and Social Finance (2013b, p.26) put it: *“the private sector can offer expertise in measuring performance data and establishing feedback loops, as businesses’ survival often depends on their ability to collect (and quickly respond to) real-time data from customers.”* While the majority of investors do not get involved in the running and monitoring of projects, some do. This is also true at the contractual stage, where the majority of investors seem to trust that arrangements are sensible, delegating responsibility to more activist investors.
15. There is some diversity of opinion regarding how much the investor is able to drive innovation and learning. Often the investors attracted are money managers not business managers, and do not wish to take on extra responsibilities in micro-managing projects they invest in. Nando’s, who have been exploring becoming an investor in a DIB tackling malaria in Mozambique, are unusual in that they do see their innovation as important: *“We tried to understand what do we, as Nando’s, have to offer to the malaria work. One thing that we saw was lacking was creativity. We’re known for being a little bit cheeky, somewhat irreverent, but we realized that if we brought our creativity to bear we could make a way bigger difference than just donating cash.”* Nando’s spokesman, quoted in Saldinger (2013)
16. In some cases, the social return will potentially be correlated with commercial returns over a long time-horizon. However, in the vast majority of cases it is simply not credible for a business to consider investing in DIBs because social returns will translate into future financial returns. A better comparison in some cases may see DIBs as an alternative to other Corporate Social Responsibility activities.

Intermediary

17. Given the new market, there are not an abundance of intermediaries, with only Social Finance, Instiglio and Dalberg Capital known to be actively working in this area. In theory, an intermediary is not a necessary constituent of a DIB, but, in practice, all DIBs developed to date have involved an intermediary. DFID's current experience implies that this is likely to be the case for the foreseeable future. Investors have also indicated they are happy to delegate much of the active monitoring (Deloitte, undated). While the rhetoric of DIBs relates to bringing in private sector expertise, the section above makes clear that activist (or active) investors are the exception. In this case, it will often fall to the intermediary to manage and monitor the implementing agencies in a way that is more efficient than under traditional aid models. In Peterborough, this took the form of data-intensive monitoring of service provision. Related information was then provided to investors on a frequent basis, but it seems that it was the intermediary who actively monitored the performance of the service providers, with relatively little monitoring of the intermediary by the investors.
18. It is important to emphasise that this monitoring role of the intermediary is distinct from the role of the validating agency. The monitoring conducted by the intermediary has a strong focus on service improvement and delivery of results. The validating agency's role is to verify the accuracy of the reported results in order to justify payments (see paragraph 25). DFID emphasises the importance of keeping the roles of intermediary and validating agency separate to avoid conflicts of interest arising.
19. If the intermediary is the main source of innovation and monitoring, one may ask what value the investor adds to the agreement. First, while the number of active investors is expected to be small, they may be important if the intervention is failing. Second, where the outcome funder is a government, it is less able to provide hands-off funding than investors. While investors are able to write off individual investments that were unsuccessful, the same cannot be said for donors, whose mistakes are presented before the public accounts committee and where any failed projects loom large.
20. A further task of the intermediary, sometimes shared by the outcome funder, is to identify the SIB opportunity, provide the ground work to prove it is an investable opportunity and get it to market. This is discussed in much more detail later (see points 29 to 31).

Service Provider

21. For the service provider itself, there is not necessarily any change between a traditional aid project and a DIB. However, the different contractual details could define their role in very different ways. For example, one possible advantage of DIBs over traditional aid projects is expected to be greater contract flexibility to use inputs in different ways. This would occur if contracts with investors and intermediaries focus more on achieving results than on how inputs are used. Regardless of the contractual agreement, the service provider is likely to have greater reputation risk, attention and focus than in other types of aid because of the unambiguous way in which success will be measured. This in itself may provide an extra incentive to perform. Some of the effects may be unsustainable, if they are due to the Hawthorne effect, where agents change how they work because they are being studied.
22. One fundamental characteristic of the intermediary in a DIB appears to be the appetite for data-intensive monitoring of interventions. This may mean the service provider is better able to innovate and manage its own actions, because of the superior feedback system. A potential cost of the intermediary-provider relationship is that there may be

higher transaction costs and more costly managerial oversight. The balance of these two will be important, and quite different in different circumstances.

Target Population

23. In contractual terms, the target population will tend to be the silent partner in a DIB, be they Bolivian farmers or Rajasthani school children. The intermediary will probably be the first agent to consider whether a given target population is suitable for a DIB. For them, the incentive is to produce an investable opportunity – one that is of sufficient benefit to the outcome funder that they are willing to pay handsomely for it in the case of success, and one that is predictable enough that investors would be willing to take the risk. For the investor, the calculation will be devoted to whether the expected returns are sufficient to make an investable opportunity. For the benefit of the DIB model to fully be realised, the outcome funder will need to play a limited role: the intermediary will be rewarded for identifying and managing successful projects and it would not represent good value for money if the outcome funder felt the need to replicate much of this work. As such the type, scale, location and nature of the intervention should predominately be decided upon by the intermediary as much as possible, with the outcome funder judging each case after it has been made. For the outcome payer, the most important thing to consider should be the measure. A good measure will in all circumstances be strongly correlated with the underlying aims of the intervention. If the measure is not strong enough, it is possible that the outcome funder would fund illusory gains (see Clist and Verschoor, 2014).
24. To enable the outcome funder to be more hands-off, it should be satisfied that the target population will be heard, that there are no significant negative consequences of the agreement and that positive effects are likely and achievable. Much of the 'due diligence' will be conducted ahead of time but it should be remembered that part of the rationale for DIBs is that there is flexibility in delivery mechanisms – as such the intervention is unlikely to be amenable to extensive evaluation ahead of delivery. This places a larger burden on evaluation, in order to learn the lessons of the delivery and bring them into the public domain. This will tend to be the primary way in which the intended beneficiaries are included in the process.

Validating Agency

25. A validating agency is normally needed, as in other PBR modalities, to ensure that the reported results are accurate, given that the measure acquires some legal significance. Some of this may be a public good (in that new, useful data is created) but in practice the bulk of the cost is in verifying rather than creating new data.
26. One troubling irony of DIBs is that while they emphasis data-intensive monitoring, there will sometimes be no true impact evaluation. Some actors have argued that no impact evaluation is needed, as the validating agency certifies impact. There are several reasons why this is not satisfactory in some circumstances. First, as Clist and Verschoor (2014) show, a measure may lose its ability to provide information regarding the intended aim once it is incentivised. Second, the validation will not control for counterfactuals other than the one scenario that is enshrined in the contract. In cases where the measure is sophisticated this is justifiable, but in many cases this will not be the case. This limits the ability to learn, and may mask important unintended effects

Outcome Payer

27. The attractive logic of DIBs for the outcome payer is that they only pay if and when there is measured success. In addition, it sidesteps possible problems associated with Results Based Aid or Results Based Financing, as funding is provided to the implementing agency upfront. This means the capital is valued more highly by the implementer (as it is not time, volatility and risk discounted⁵) and provides the working capital to implement the project. In practice, it seems that the donor will sometimes pay for piloting and scoping (at least in the early stages of the market). These investments could lead nowhere and so DIBs do not completely remove the risk of funding unsuccessful projects. Another important detail for the outcome payer is that, in theory, they will be paying for additional results. Only a subset of possible designs guarantees this, as it is often difficult or expensive to recover an appropriate counterfactual. In other types of PBR, the ability to withhold is so far unproven over a longer time period. In DIBs this is unlikely to be a problem, as the outcome funder is not inherently interested in the investors' welfare in the same way that they are concerned about NGOs or governments in low-income countries.
28. While much of the rhetoric surrounding Payment by Results discusses risk transfer, it is not a rationale for DIBs. Donors such as DFID are closer in characteristics to insurance companies than insurance customers, given the number of diversified projects they are involved in. Transferring delivery risk to investors is only attractive if it leads to higher efficiency, not for its own sake. Part of the attraction in delegating project management is that it provides a more acceptable way for outcome funders to benefit from learning by experimentation. A worry with traditional aid projects is that unsuccessful interventions lead to negative press and uncomfortable hearings from the public accounts committee. By only disbursing funds when there is measured success, a donor is then able to defend its own spending. However, variable payments are not without their own costs. The money that will potentially be disbursed can't be used in the mean time, and so there is an opportunity cost of tying up funds for a period of time. The practical difficulties of variable payments also seem large.

Two paradigms: small and large DIBs

29. Two DIBs which have started the process of coming to market provide a useful illustration of the range of activities that are included under the DIB heading. In the Ugandan sleeping sickness DIB, DFID has agreed to provide up to £1.5 million to pilot the intervention. This process is designed to reduce the risk to the investor by investing upfront in a 'proof of concept' exercise. In this case DFID have hired Social Finance to undertake this work who also contributed to identifying the original problem. The Rajasthan DIB is quite different in scale and scope. Instiglio have approached an NGO that already works in the area, and the entire DIB is valued at \$238,000. The latter DIB's scale, risk and innovation are several orders of magnitude smaller. The investor is only asked to take on the risk that out-of-sample performance will not be drastically different from the existing evidence, whereas in the Ugandan sleeping sickness DIB the investor is asked to back a completely new intervention.
30. These two DIBs are answers to very different questions. The Rajasthan DIB is essentially allowing the scale up of a successful intervention. Here, the intermediary identifies high-performing actors and enables them to increase their activities by providing long-term capital. The investor here would provide little innovation or monitoring, but their discernment at the stage of investment choice is important. If

⁵ Time discounted means that any given amount of money is worth less to the recipient if it is to be transferred in the future than it would be if it were transferred today.

investors are able to distinguish between 'good' and 'bad' projects, they enable the best to be scaled up. This provides an important incentive for the intermediary to identify feasible projects, for which they are rewarded. The intermediary would typically also have some financial interest in the success of the project, and so has an incentive to monitor the intervention for the life of the project.

31. In both cases one of the benefits of a DIB which is not often highlighted is the market discipline that is provided as DIB goes to market. The intermediary must convince the investor, which provides an incentive to either search for an investable opportunity or conduct the necessary groundwork to create one. DIBs that failed to come to market are then potentially indicators of success, as they indicate that investors are providing a useful function of selecting which projects to finance.

Cross Cutting Issues

Measure

32. The major lesson from Clist and Verschoor (2014) and Clist and Dercon (2014) is that in payment by results contracts the measure itself is supremely important. It is not enough for the measure to be correlated with the true underlying goal, it must remain so once it is incentivised. In the Rajasthan DIB, the possible measures of school attendance and achievement are not irrelevant design details but determining factors in the true effect of the DIB. If school attendance is chosen, there is no guarantee of any increased learning, and the measure may be of limited value. As such, a good measure is a prerequisite for any PBR contract.

Coordination

33. One of the rationales for DIBs is it that they lead to better coordination between different parties. There is nothing inherent in the structure of a DIB that guarantees better coordination, though it does seem to have been a part of the SIBs rationale (Social Finance, 2012). In cases where several providers will be able to draw from the same pot (as in the Peterborough Prison SIB), the added coordination is both expected and beneficial. In cases of one implementer, the likelihood of increased coordination is less clear. More generally, the empirical and theoretical evidence surrounding coordination is quite thin, and so the hypothesis that DIBs will lead to higher coordination is worth monitoring in future evaluations.

Risk Preferences

34. While much of the rhetoric surrounding PBR relates to risk transfer, it is not a logical basis for increasing PBR (see paragraph 28). This is generally accepted by proponents of DIBs, who instead argue that DIBs provide a better allocation of different risk types amongst the different parties. For example, an intermediary, such as Social Finance, may be more able to manage performance risk under a DIB because they have room to be adaptive and not to be constrained by a rigid donor contract. The donor, e.g. DFID, may be better able to manage political risk. The basic argument regarding the attitude to risk is still used, for example, CGD and Social Finance (2013b) argue that the '*public sector struggles with complex problems, and is too risk and time averse.*' However, it does not seem that DFID are excessively risk averse in theory or in practice. In theory they have a large portfolio of diversified projects. In practice, they are willing to try innovative financing mechanisms: which is not consistent with a view of donors that alleges excessive risk aversion, or more risk aversion than private sector companies.

Interest Rates

35. One of the successes in bringing in financially motivated investors in SIBs was the New York state SIB, where Meryll Lynch received orders from 40 high net worth individuals for an average of \$350,000 (Reuters, 2013). In this case, Rockefeller put up \$1.32m to guarantee 10% of the investors capital and returns were up to 12.5% over 5.5 years. One contextual factor to consider is that interest rates across the world are at historically low levels. Over the time horizon envisaged (CGD and Social Finance, 2013b, state that the ideal time frame is 3-10 years) UK Gilt yields for 2, 5 and 10 year government bonds at the time of writing are 0.80%, 1.92% and 2.50% respectively. This means that impact investing may be popular partly because it is an opportune time. Alternative investment staples are yielding abnormally low returns. It will be interesting to track the extent to which the size of the impact investing market is affected by these, much safer, bond yields. Burand (2013) gives a sceptical view of the long-term attractiveness of SIBs and DIBs, arguing it offers '*...investors the least attractive features of both debt [low, capped returns] and equity [unlimited downside].*'

Time Preferences

36. Just as the risk attitude of each party is important in considering whether a successful agreement is possible, each party's attitude towards time is relevant. Investors will, for a given investment, seek a larger return the longer the time horizon. This follows logically from the opportunity cost of capital, and compound interest rates, and means the final payment is worth less as it is time-discounted. It is also less valuable to the outcome payer, assuming they do not have to tie-up the money for the duration of the contract. In the current setting this is not ensured as possible liabilities will need to be covered in existing accounting arrangements. However, the ability to contract over longer time horizons is likely to be an important element of the DIBs model. Evan Jones, involved in the Peterborough Prison SIB, stated that 'Normally, by the time we've realised what is working or isn't working on a contract, the funding has run out. This time we have space to say, "Let's do more of this and less of this."' (Guardian, 2013). However, this need for a longer time horizon is likely to be a major challenge for many DIBs given the contexts in which they are likely to operate.

Innovation/Autonomy

37. Innovation and autonomy are used interchangeably as reasons for DIBs, but they do not always go hand in hand. For example, in the Rajasthan DIB autonomy is to be expected: an established NGO will be given funds to expand. While it will be monitored by an intermediary, the scope for innovation is limited. A tried and tested method is being scaled up to new villages, but the size of the budget and time horizon means that true innovation is unlikely. By contrast, the Ugandan sleeping sickness DIB is a completely new intervention with relatively little evidence that it will work. The intervention in this case is almost innovative by definition as without the DIB financing model such an intervention is unlikely to have happened.

Contractual Complexity

38. One of the unavoidable costs of a DIB is that they are contractually complex, with Disley et al (2011, p.10) stating that they took 18 months to agree for the Peterborough SIB. In theory the initial set-up costs may reduce over time, as donors become more adept at contracting. However, Azemati et al (2013) report that with SIBs these costs have not fallen to the degree that they expected, and recommend \$20 million as a rule-of-thumb minimum size for a SIB to justify the expense. It is expected that each party knows all contractual agreements, and the different forms that these contracts take will

be very important: whether the investors can replace the intermediary, whether the intermediary can replace the implementing agents and so on.

The Donor's Role

39. Clist & Verschoor (2014) discuss various real world donor limitations that may impinge upon the theoretical functioning of various PbR contracts, including the time horizon and ability to disburse variable payments. Specific to DIBs are the role of the donor, and the ability to play a 'hands off' role. Specifically, if a donor feels it has a contractual obligation to specify the identity and functioning of the other five agents in the model, the benefits of DIBs (innovation, flexibility, longer time horizons, increasing results focus) will be foregone. It is a possible danger of DIBs that the contractual relationships between various agents will augment existing layers of bureaucracy rather than replace them. The donor should see their primary role in a DIB as the outcome funder, deciding which DIBs represent value for money. If the donor plays a large role in specifying contractual relationships or the nature of the intervention, it would be better placed using a more traditional form of aid, or another payment by results modality (where there a good measure exists). The intermediary will typically take the lead in deciding upon the role and identity of the six agents, and has a useful incentive to do so in a way that is attractive for each agent.

Strengths and Weaknesses of Existing Evidence and Evaluation Approaches and Methods

40. This section briefly reviews experience to date in developing and using Development Impact Bonds and related instruments, such as Social Impact Bonds. It then looks at issues related to evaluating the use of these instruments focusing on identified and documented evaluation experiences. It then considers strengths and weaknesses of existing evidence regarding DIBs and approaches and methods used to evaluate DIBs.

Experience of Developing SIBs

41. Although both SIBs and DIBs are relatively new and innovative, experience of using SIBs has been expanding extremely rapidly⁶, particularly in the United Kingdom, where 15 SIBs and a range of investors have been documented (Jupp and Hoare, 2014). A Cabinet Office database (undated, b) contains details of 40 SIBs⁷. Of these, 16 were in the UK, nine in the US, three each in Australia and Israel, two in Uganda and one each in Canada, Colombia, India, Mozambique, Pakistan, South Africa and Swaziland. Descriptions of individual SIBs are available from a number of countries including Australia (Australian Government, 2012), Israel (Social Finance Israel, undated) and New Zealand (MOH, 2014).
42. Griffiths and Meinicke's (2014) introduction to Social Impact Bonds and early intervention contains descriptions of a number of examples of SIBs from a variety of settings. In addition to the very well-known examples in the criminal justice systems in the UK and the US, individual reports describe experience of using SIBs in various parts of the social sector in the UK (Bridges Venture, 2013); tackling rough sleeping in London (Budget and Performance Committee, 2013); and teen pregnancy in Washington (Perakis, 2014).

⁶ See Liebman, 2011; Mulgan et al., 2011; Social Finance, 2011b; Keohane, 2013; Social Finance, 2013a; Drexler and Noble, 2013; Khan, 2013; Liebman and Sellman, 2013; Barclay and Symons, 2013; Deloitte, 2014; Hughes and Scherer, 2014; Jupp and Hoare, 2014; Griffiths and Meinicke, 2014; Lewin, 2014; Wilson, 2014; Littlefield et al., 2014; Harvard Kennedy School, 2014; Butler et al., undated; and Cabinet Office, undated, a.

⁷ Some of the programmes included appear to be DIBs

43. Consequently, there has been huge interest in using SIBs, or similar instruments, in a wide range of countries and sectors, e.g. in early childhood development (Atinc and Gustafsson-Wright, 2013) and the environment (Nicola, 2013). Social Finance (Palandjian and Hughes, 2013 and 2014) have explored how the SIB market and field might develop in a particular country, such as the United States. They considered three possible scenarios (see Box 1).

Box 1: Three possible scenarios for development of the SIB market in the United States

From Palandjian and Hughes, 2013

Scenario 1 - Boom-Bubble-Bust – in which overblown expectations and ‘deep pockets of misunderstanding’ of the ‘latest craze’ results in the SIB market’s evolution being derailed.

Scenario 2 - SIBs are the wave of the future – and always will be – in this scenario, the SIB market ‘limps along’ – always appearing promising but never quite fulfilling its potential. A small number of SIBs would continue but with high transaction costs and high reliance on philanthropic support.

Scenario 3 - A successful market for social outcomes – the SIB concept would be proved over the next few years laying the foundation for a broader-scaled market appealing to blended-value investors.

44. However, there are many critical, sceptical voices regarding SIBs. There are those who question whether the savings or efficiency gains claimed by advocates of SIBs are occurring, arguing that the funds being used for SIBs should be used to better finance public institutions to address problems directly (NSPCC, undated; Godoy, 2014). Critics argue that potential investors, such as banks, rather than benefiting from these bonds, ‘at the expense of tax payers’ should pay more taxes because, if they did that, there would not be need for such private financing (Hughes and Helbitz, 2014). Experts can be produced on both sides of the debate (Guardian, undated) but it is difficult to judge between the different opinions because of extremely limited empirical evidence. A key question is whether SIBs are the best way to achieve a particular intended result or are there simpler, more efficient and more effective ways to finance programmes that realise the social benefits and produce expected savings (Demel, 2013)?

Experience of Developing DIBs

45. There has been huge interest in applying and adapting Social Impact Bonds to international development (Barder and Perakis, 2012; Keohane, 2013; CGD and Social Finance, 2013c; Burand, 2013; Belinsky, 2014b; OECD, 2014; Strickland, 2014; CGD and Social Finance, 2014; DFID, 2014a; Social Finance, undated; and SIBM, undated). The concept underlying Development Impact Bonds has been outlined by a working group led by the Center for Global Development and Social Finance (2013a).
46. Nevertheless, as with SIBs, there have been some who have been critical of DIBs (e.g. Oxfam, 2013) arguing that DIBs are being introduced based on limited experience of SIBs and, as a result, the evidence base is extremely thin and funds from foundations that might have been available as grants risk being diverted to DIBs. They argue that a better approach would be to sort out tax avoidance and evasion, using the funds generated to provide better public services.
47. Despite these concerns, there are a number of DIBs and SIBs under development in a range of different low- or middle-income countries. Examples are presented in Box 2. In addition, the working group on DIBs mapped known SIBs and DIBs globally (CGD and Social Finance, 2013b).

Box 2: Examples of SIBs and DIBs in low- and middle-income countries

Social Finance (2012) conducted a scoping exercise for DFID on using DIBs in family planning.

In Colombia, a SIB is being developed focused on addressing teenage pregnancy (Instiglio, undated, a).

In India, a DIB is being developed to seek to close the gender gap in education in Rajasthan (Instiglio, 2014 and undated, b; Mair, 2014; Crabtree, 2013). This example involves both Instiglio and an education NGO, Educate Girls.

A budget announcement in Mauritius (anonymous, 2014) included news that SIBs would be introduced by the Ministry of Social Integration and Empowerment. It was expected that SIBs would begin to be piloted in 2014 among 'ex-convicts and drug addicts'.

In Mozambique, a 'pioneering new coalition' is planning to provide sustainable, cost-effective funding for malaria control through DIBs (Han, 2014; Belinsky, 2014a). However, although there has been potential interest in this approach from investors, it is reported that progress has been slow for a number of reasons, including difficulties and delays in building the required partnerships.

In Pakistan, DIBs are being considered as an innovative financing mechanism for education in Punjab (CGD and Social Finance, 2013; Wheeler and Egerton-Warburton, 2012).

In Rwanda, DFID is planning a DIB in the education sector (DFID, undated, a). A scoping report has been produced which explores two broad options (Robinson, 2013) and suggests one which dovetails plans onto the expansion plans of a UK-based educational charity. The report also concludes that rewarding a basket of metrics would be preferable to just rewarding one. A number of possible absolute, outcome level metrics were proposed. Concerns raised by the scoping report included the high level of risk and likely low levels of return on investment.

In Uganda, there are reports of a number of DIBs under development, including one on education and another on sleeping sickness.

There are a number of other possible DIBs under development including in the criminal justice system in Capetown, South Africa and prevention of HIV and TB in Swaziland.

Relevant evaluation literature

48. Before considering specific experience of evaluating SIBs and DIBs, it may be helpful to consider other literature that may be relevant to evaluations of DIBs.
49. Some authors have emphasised the importance of evaluation in relation to Social Impact Bonds (e.g. Henderson, 2011). However, several manuals related to SIBs do not cover issues of evaluation (e.g. Social Finance, 2011b and 2013; Barclay and Symons, 2013). Liebman and Sellman's guide (2013) on SIBs for state and local governments does cover the issue of evaluation and emphasises the need to develop an evaluation methodology and for evaluation support during implementation. It suggests places to look for programmes with successful evaluations and provides a list of evaluation strategies. These include experimental approaches, such as a randomised controlled trial, quasi-experimental approaches, including regression discontinuity design and difference in differences, and non-experimental approaches, such as a comparison to an historic baseline. The guide states that '*if there is no credible approach to establishing a counterfactual outcome for a particular intervention, then a SIB may not be feasible.*' It also concludes that in most cases a comparison group will not be feasible.
50. There are some very practical manuals available related to assessing social impact (e.g. Russell, 2013). This covers a wide range of issues including cost-benefit analysis, social return on investment, the use of control groups and process evaluation. However, this manual is not specifically focused on SIBs. Similarly, there are authors (e.g. Haynes et

al., 2012) who argue for greater use of randomised controlled trials in relation to developing public policy, in general, without referring to SIBs specifically. Nevertheless, these documents make it clear that SIBs and DIBs will be operating in an environment in which there is increasing demand for evidence from rigorous evaluations including those that use experimental or quasi-experimental approaches and those that rigorously assess value for money.

Expectations of Evaluations of DIBs

51. Based on discussions with DFID staff, it is clear that there are certain expectations as to the kind of evidence evaluations of DIBs might be expected to generate and the types of questions they might be expected to answer. These include:

- a. Evidence on the extent to which markets for DIBs have been created and/or have developed.
- b. Data on the roles of different actors and the relationships between them.
- c. Data on the added value of using a DIB. For example, it appears that much of the original motivation for using SIBs was to generate cost savings for government. Can evaluations provide evidence on the extent of those cost savings?
- d. Evidence of the extent to which DIBs promote innovation.
- e. A focus on how DIBs function and operate and produce outcomes, i.e. the processes involved.

52. There is some excellent material on the Centre for SIBs website (2013) on their expectations of evaluations. These are articulated as four broad questions or question types (see Box 3)

Box 3: Proposed questions to be answered in evaluations of SIBs

From Centre for SIBs, 2013

What difference did the services make? This question is considered to constitute impact evaluation.

How were the services delivered? This question is considered to constitute process evaluation.

Did the benefits of the social impact bond justify the costs? This question is considered to constitute economic evaluation.

What was the effect of using a payment by results, and more specifically the social impact bond, model? This question appears to be focused on trying to differentiate between the benefits/effects of a service and the specific benefits of using a SIB as an instrument.

53. The summary report of the working group on DIBs (CGD and Social Finance, 2013a) is very clear that early DIBs should be evaluated rigorously and independently. They express the expectations that evaluations should:

- a. Include information on intervention costs and pricing of outcomes and results
- b. Assess whether and how the structure helped to lead to improved outcomes

- c. Include details of any positive or negative externalities
 - d. Be used by DIB actors to improve the future design of results-based contracts
54. The working group's full report (CGD and Social Finance, 2013b) also focuses on:
- a. Whether and how the structure changed incentives and led to greater transparency around the impact of donor funding
 - b. Whether and how the structure led to greater innovation
 - c. Whether and how it resulted in greater efficiency in terms of services, stakeholder relationships and value for money
55. It presents three options for evaluations (pp74-76). These are basically experimental, quasi-experimental and non-experimental approaches.
- a. The form of experimental approach proposed appears to be a cluster-based, randomised control trial. It gives some examples but does not really explain that some programmes are delivered in ways which make individual randomisation impossible and make cluster approaches necessary.
 - b. The quasi-experimental approach proposed is described as a 'live comparison group' which essentially appears to be a matched, non-random comparison. Reference is made to propensity score matching. But, there does not seem to be adequate recognition that the method would still probably require multiple intervention and comparison groups. Other quasi-experimental approaches, such as a discontinuity design are not considered.
 - c. The final (non-experimental) approach is referred to as a historical baseline. This may be possible where the baseline is static over time or there is a predictable trend. However, DFID's experience of trying to use this approach in evaluations of other forms of PBR has been problematic.
56. Several important issues emerged in discussing this report with DFID. The first relates to the importance of formative evaluations given how new and innovative DIBs are. There will be many opportunities to use formative evaluations to shape the design and set-up of DIBs. Building evaluation into DIBs from the earliest stage of design will enable better delivery and stronger evidence. Evidence generated from synthesising experience across DIBs is likely to be extremely useful at the formative and design stage of new DIBs.
57. The second relates to the role of the evaluator in a DIB. How, for example, do they fit into the six agent model described above? Essentially, they don't because they are not involved in the delivery of the DIB⁸. It would, for example, be possible to deliver DIBs without including evaluation. A key practical issue is whether the roles of evaluator could be combined with the role of data verifier. In principle, these are separate roles. For example, the evaluators might need to consider the role or cost of the data verifier when evaluating the DIB. This might be complicated or impossible if the evaluator and the data verifier were one and the same. On the other hand, there are practical and pragmatic reasons related to procurement and ongoing management where it might be easier if the roles were combined. On balance, given the early stage of experience of

⁸ The opposite could be argued on the basis that evaluations, particularly those that are formative, should influence design and delivery. However, to preserve relative simplicity, this paper has retained a six agent model rather than expanding it to seven.

evaluating DIBs, it is probably wise to keep these roles separate until more experience has been gained. It might be possible for one supplier to provide both evaluator and data verifier provided a firewall between the two could be ensured, e.g. through different organisations in a common consortium.

58. In theory, it is possible that an evaluation could be commissioned by any of the agents involved in the DIB acting separately or together. In practice, it seems likely that most evaluations would be commissioned by the outcome funder but it is possible that active investors may wish to commission evaluations either alone or with the outcome funder. It will be useful to monitor what actually happens in practice as more DIBs are established and evaluated.

Experiences of Evaluation

Experimental Approaches

59. According to Griffiths and Meinicke (2014), the Social Impact Bond in New York, focused on increasing employment among formerly incarcerated individuals and improving public safety, has been designed as a randomised controlled trial. Launched in 2013, this SIB claimed to be the largest at the time of its launch and the first to use a randomised controlled trial to determine outcome payments. Success criteria are expected to be differences between members of the treatment group and members of the control group in terms of starting a transition job; employment; earnings and number of days incarcerated.

Quasi-Experimental Approaches

60. Perhaps the most documented SIB relates to the one focused on reducing reoffending rates in Peterborough (Social Finance, 2011a/c and 2013b; Economist, 2013; Nicholls and Tomkinson, 2013; Big Lottery Fund et al., 2014; Cabinet Office, undated, c). Assessment of success in the Peterborough SIB was conducted by QinetiQ and the Universities of Greenwich and Leicester (Cave et al., 2012) and compared prisoners from Peterborough with a group of prisoners from other prisons who were said to be matched through propensity scores. Outcome measures included number of reconvictions in the 12 months after release from prison and the number of court appearances. There were three cohorts of 1,000 people and these were combined to form a fourth cohort. Payments were to be made if any cohort showed a 10% improvement or if the fourth cohort improved by at least 7.5%.
61. In addition, the Peterborough SIB was evaluated (Rand, 2011; Disley et al., 2011) through a methodology focused on collection of interview data. The evaluation generated some useful lessons learned. For example, it concluded that the development of a methodologically robust outcome measure, which had the confidence of all stakeholders, was a time-consuming and analytically complex process. It recommended that those involved in future SIBs and payment-by-results arrangements might wish to take into account the time and skills needed to develop robust outcome measures.
62. A follow-up report (Disley and Rubin, 2014) was produced based on 39 interviews although none of these were with service users. At the time of writing that report, the final results regarding reconviction rates were not available but the evaluators reported that the frequency of reconviction in Peterborough had declined by 11% over the period of the pilot while, over the same period, the national reoffending rate rose 10%. Care was urged, however, in interpreting these results as the method used differed from the method used to calculate payments due under the SIB. More recently, QinetiQ and the

Universities of Greenwich and Leicester published their results which showed an 8.4% reduction in reconvictions when compared to the comparison group. As this was below the 10% required, no payment was triggered (Birkwood, 2014).

63. Although the Peterborough experience is cited by quite a few sources as evidence for the success of SIBs, caution is needed in placing too much emphasis on one example, particularly when there are concerns and issues over data quality and there has been relatively little qualitative analysis of causal mechanisms which may limit the extent to which these findings could be considered applicable to other conditions and contexts.

Non-experimental Approaches: Individual Evaluations

64. An evaluation of Social Impact Bonds as a New Reentry Financing Mechanism in Maryland (Department of Legislative Services (2013) concluded that claims that a SIB could generate sufficient cost savings to self-finance were over-stated because the reduction in reoffending rates would not be sufficient to close a prison or a prison wing and this would be necessary to generate significant savings. However, this report lacks a clear description of method and it appears that this was essentially a financial modelling exercise which might be considered a formative evaluation. The conclusions of the evaluation were significant as the Department of Legislative Services recommended that the Department of Public Safety and Correctional Services continued to directly finance reentry programmes rather than introducing a SIB-financed programme.
65. A descriptive study of the Allia Future for Children Bond in the UK was conducted through interviews and document review (Rotheroe et al., 2013). It documented considerable problems with the market and these resulted in a decision not to go ahead with issuing the bond. While there was interest from investors and intermediaries, there were real barriers to effective distribution and capital raising. These included having to apply for the bond through financial advisers who were ill-equipped to promote social investment products and the costs of seeking financial advice for complex products were considered too high for small investments. The report did make recommendations for how the situation could be improved including through specific support from the Financial Conduct Authority including allowing financial advisers to rely on due diligence work conducted by experts in social investment with the approval of the FCA and allowing small investors who do not have a financial adviser to conduct some form of self assessment. A key conclusion of the report was that there should not be a focus only on individual products but on building the social investment market as a whole.
66. A lesson learning report concerning the New York City Social Impact Bond (Rudd et al., 2013) described the first operational SIB in the United States and the program financed by it, the Adolescent Behavioral Learning Experience (ABLE). Essentially, ABLE aims to equip adolescents ages 16 to 18 incarcerated in the New York City jail system with the social and emotional skills to help them make better life choices when they leave jail, yielding financial savings to city government by reducing readmissions to Rikers Island. The report attempted to present important lessons from the programme's early stages (see Box 4):

Box 4: Lessons learned from New York City SIB

From Rudd et al., 2013

A committed government partner is especially important to a SIB arrangement

Investors will have more confidence in a SIB if small differences in programme performance do not cause sharp distinctions between gaining a return on investment and experiencing a loss. They will also have more confidence in SIBs that incorporate early performance indicators, and in those whose transaction costs are lower.

Holding repayment until clear evidence that savings are in fact being achieved increases the government partner's confidence in the effectiveness of the programme.

Spreading out repayment to align with the value produced by the programme increases the government partner's ability to pay.

67. However, this report lacked any description of method and appeared to be more of a project description than an independent evaluation. There are a number of other relevant descriptions of this programme (e.g. New York, 2012; MDRC, undated). Drawing on their experience as intermediary for this programme, Butler et al. (undated) advocate for an approach to SIBs which moves beyond cost savings and focuses on other benefits, such as spurring innovation, knowledge building and rigorous evaluation. They argue that existing social programmes have been poorly-evaluated and that evidence for them is, at best, mixed and often absent. They also argue that SIBs have a built in evaluation element to them because payment is triggered by levels of performance in programme participants over and above those in a comparison group. However, they expressed concern that, based on conversations with potential SIB stakeholders in the US, support for high quality evaluations was not a priority and evaluations were often underfunded in efforts to ensure sufficient capital for the programme. The authors advocate strongly for the importance of impact studies and serious evaluations.
68. KPMG conducted an evaluation of the New South Wales Social Benefit Bonds in 2014 using a combination of interviewing 21 stakeholders, reviewing documents, survey data and analysis of transaction costs. This evaluation concluded that:
- a. Social benefit bonds are viable in New South Wales on the basis that two bonds had been successfully developed.
 - b. The development of the bonds had produced positive outcomes in terms of positive gains for both NGOs and government. There had been increased attention on and understanding of programme outcomes and how to measure them.
 - c. Capacity building is vital for the development of future bonds – within government, NGOs and financial intermediaries. Areas needing capacity building included development of bonds, data capacity, contracting capacity and market development. Specific suggestions included a social benefit bond unit within central government.

69. This evaluation also highlighted a number of areas in which there had been learning (see Box 5).

Box 5: Lessons learned from New South Wales Social Benefit Bond

From KPMG, 2014

Capacity – areas in which NGOs and government need skills include measurement, contracting and financial modelling.

Collaboration – key factors in effective collaboration included the role played by the New South Wales Treasury’s project manager; the goodwill associated with the trial; the successful involvement of a financial intermediary; and the degree of agency engagement.

Contracting – took several months for each bond. The extent to which the contract was seen as complex or not depended on participants’ familiarity with large transactions. Development of a plain English manual was considered helpful.

Prevention and early intervention – out-of-home care and recidivism were seen as appropriate areas for SIBs as they are expensive for government and outcomes can be specified in simple binary terms. However, there were challenges in measuring and capturing savings.

Measurement of outcomes – there were challenges in terms of different views of how savings should be measured and calculated and in determining outcome measures. However, using a SIB resulted in these matters being addressed.

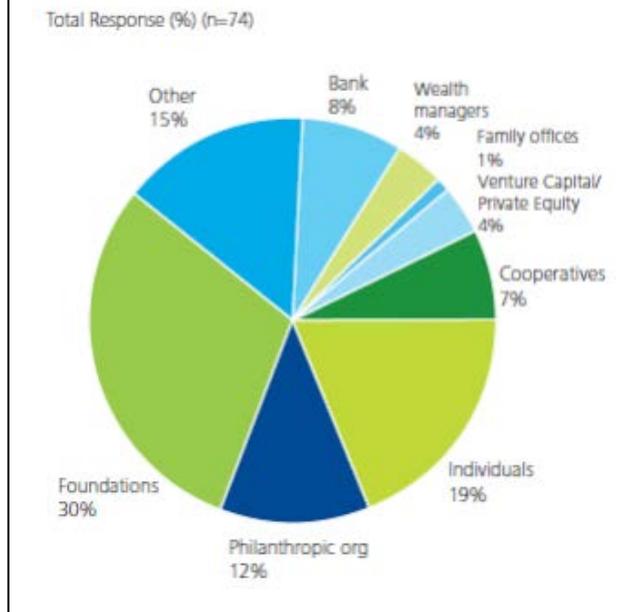
Innovation – the use of SIBs was considered to have been an exercise in innovation in a number of areas including financing, contracting and measurement. However, there appeared to be a contradiction between service innovation and developing a bond with a sound evidence base.

Social investment – the use of SIBs has generated conversation, engaged investors, raised awareness and created intrigue around financing solutions to social challenges aside from traditional financial sources. The bonds attracted investors but there were challenges in matching investor expectation with the product.

Synthesis across Evaluations

70. A number of opinion gathering exercises have been conducted in different countries in which relevant stakeholders were interviewed concerning their views on SIBs. For example, Hughes and Scherer (2014) reviewed experiences of SIBs in the United States by talking to representatives of Foundations. A range of issues and concerns were documented including, for example, concerns about the quality of data in the social sector, particularly for measuring outcomes (p36). Deloitte (2014) produced a report on government and the impact economy in the US based on interviews with more than 50 practitioners across the government, non-profit, philanthropic and investing sectors. Deloitte (undated) also conducted a similar exercise in Canada based on interviews with 80 investors. This exercise was largely formative as it was focused on discussing the possibility of introducing SIBs in Canada. Figure 3 shows the types of investors interviewed in that exercise. Box 6 briefly summarises some of the consultation’s findings which highlight investor perceptions on SIBs, at least in Canada.

Figure 3: Types of investors consulted in Canada
(from Deloitte, undated)



Box 6: Findings of consultation with investors in Canada

From Deloitte, undated

Given the right deal, investors are willing and able to fund a SIB. A vast majority of respondents would be interested in considering SIB investments.

Investors are willing to consider a broad range of issue areas and jurisdictions. Housing/homelessness, youth (notably unemployment and skills training), aboriginal focused issues, health and education surfaced as issue areas of potential.

The investor market has identified challenges to the successful implementation of any SIB deal in Canada, the most significant of which are collaborating with government, risk/perception of risk, liquidity and stakeholder capacity.

Investors would prefer to co-invest as part of a consortium in order to share capital commitments, due diligence, governance and learning as well as to allow for risk reduction.

Capital de-risking, such as guarantees, is preferred by investors to minimise exposure to execution risk.

Investors would prefer to conduct a SIB deal through an intermediary.

It is believed that a SIB market will develop in Canada if governments implement measures to remove barriers, and explore and test impact instrument models.

71. Azemati et al. (2013) identify a number of lessons learned from experience of developing and implementing SIBs in the United States (see Box 7). However, no method is presented. This appears to be an opinion piece based on experiences of those involved in the sector. As a result, it could be considered to lack independence.

Box 7: Lessons learned from developing and implementing SIBs in the United States

From Azemati et al., 2013

SIBs are spreading faster than expected – in the US and abroad

Different model variations emerging

Most important criterion for SIB is impact

Initial projects – there has been more innovation and learning and less replication than expected

It has proved difficult to find interventions that truly pay for themselves

Finding large enough samples may be difficult

Building government capacity requires dedicated staffing and expertise

Governments are taking several different approaches to identifying an intervention

Provider capacity is a significant challenge

Governments have several options for selecting intermediaries and service providers

New structures are necessary to enable government to commit to future payments

72. The working group on DIBS (Center for Global Development and Social Finance, 2013b) concluded that, in order to ensure that learning is shared, it would be helpful to establish a DIB Community of Practice including potential donors, investors, intermediaries and developing country governments to share learning from early DIB pilots and to advise of the development and application of the model in future. It was recommended that this group should also learn lessons from related areas including SIBs in developed countries and other forms of PBR contracts because of common challenges including defining appropriate outcome metrics; the need for multi-year donor funding commitments; and addressing public sector agencies' need to be accountable for programmes when they are not defining the way in which outcomes should be achieved.
73. In the UK, the Big Lottery Fund has commissioned Ecorys and ATQ to evaluate their Commissioning Better Outcomes portfolio which seeks to support the development of SIBs and other financial mechanisms among public bodies (see Big Lottery, 2014 and Ecorys, 2014). This evaluation does not focus on the impact of individual SIBs as this will be evaluated separately. It focuses on evaluating the development and impact of SIBs – both those funded by the programme and those funded by others. Methods include consultations and surveys with a range of stakeholders, collation of management information, and a number of more in depth case studies (or 'deep dives'). The terms of reference state that the evaluators are expected to assume that SIBs do work and specify a wide range of questions that the evaluators are expected to answer in both process and impact evaluations (see Annex 4, p51).
74. To conclude this section, Table 2 considers strengths and weaknesses in a number of areas. First, it assesses general approaches being taken towards evaluations of SIBs and DIBs before considering the strengths and weaknesses of the current evidence base for both SIBs and DIBs. It then considers specific methods being used to evaluate individual SIBs and DIBs and to synthesise learning across evaluations to contribute to the overall evidence base for SIBs and DIBs.

Table 2: Strengths and weaknesses of existing evidence and evaluation approaches and methods related to SIBs and DIBs: A summary

	Strengths	Weaknesses
Overall approaches to evaluation	<p>There is a strong commitment to evaluation of SIBs/DIBs from many stakeholders.</p> <p>Some forms of SIB require robust measures of outcome for payment to occur – as a result, SIBs/DIBs may contribute to a greater focus on robust data collection, including through evaluation.</p> <p>Clear expectations of evaluations of SIBs and DIBs are beginning to emerge.</p>	<p>Several manuals on SIBs do not cover issues of evaluation in much detail if at all.</p> <p>Expectations of what evaluations of SIBs and DIBs might produce are widely divergent. It is extremely unlikely that evaluations could meet all these expectations without extremely high levels of resourcing.</p> <p>The need for impact evaluations is not recognised by all SIB/DIB stakeholders and advocates. There is a documented tendency to scale back on evaluations where resources are tight.</p> <p>The value added of a full impact evaluation (over and above the results themselves) is downplayed by some, even in designs where payment is not automatically the same as truly desirable performance.</p>
Evidence base	<p>Some evidence is emerging from experience of SIBs. This is likely to be very relevant to DIBs.</p> <p>Some lessons are beginning to emerge regarding DIBs design.</p> <p>Some efforts have been made to synthesise learning from across experience of SIBs. There are plans to do similar work on DIBs.</p>	<p>Much of the evidence about SIBs to date is about process and based on collated opinion.</p> <p>Most of the evidence is through a positive lens, i.e. from advocates of impact bonds. Negative evidence is more limited although there is some. Much of the criticism of impact bonds is un-evidenced opinion.</p> <p>There is little information on development of markets for SIBs although there is some.</p> <p>There is as yet no evidence base specific to DIBs as there has, as yet, been no implementation experience.</p>
Methods for individual evaluations	<p>There has been some work on categorizing available methods for evaluation of SIBs and DIBs.</p> <p>At least one randomised controlled trial of a SIB is underway (in New York).</p> <p>Several SIBs (e.g. Peterborough, Rikers Island) use matched comparison groups as a basis for making payments.</p>	<p>Categorisation of available evaluation methods is incomplete. Other methods could include individual randomised controlled trials, a range of quasi-experimental approaches and the use of theory-based approaches.</p> <p>There has been relatively little use of experimental or quasi-experimental approaches and more reliance on interviews as the main means of primary data collection.</p> <p>Many of the reports of evaluations and reviews lack clear descriptions of methods which would mean that they would be likely to be downgraded in terms of any assessment of the quality/strength of evidence presented. They might be excluded from some forms of data synthesis, e.g. systematic reviews.</p>
Methods for synthesising learning	<p>Efforts to synthesise learning from experience of SBs has begun.</p> <p>The value of having capacity available to conduct real-time synthesis of evidence from evaluations has been recognised for both SIBs (e.g. the evaluation of the Commissioning Better Outcomes portfolio in the UK) and for DIBs (e.g. the call to establish a Community of Practice).</p> <p>There are a number of reports of SIB experience and evaluations which could be interpreted as a willingness to share information openly and transparently.</p>	<p>SIBs and DIBs are being developed and implemented in very different ways and in different contexts. The applicability of evidence and lessons learned from a SIB/DIB in one context to SIBs/DIBs in other contexts is unclear. Simplistic methods intended to aggregate learning across SIBs/DIBs are unlikely to work.</p> <p>There is reluctance among some stakeholders to talk openly about their experiences and plans. Reasons are complex but include concerns about biasing procurement processes and giving away valuable intellectual property. This reluctance or 'secrecy' is a major potential barrier to synthesising evaluation evidence into a body of evidence.</p>

Appropriate Evaluation Questions, Approaches and Methods for DIBs

75. This section briefly considers what might be appropriate approaches and principles for evaluating DIBs. It then considers appropriate questions for such evaluations before concluding with consideration of the methods that might be used to answer those questions.

Appropriate approaches and principles for evaluating DIBs

76. Overall, it is extremely important that robust evaluation is a key component of any DIB design for the foreseeable future. Reasons for this include the innovative nature of DIBs as an aid instrument and the relative paucity of the evidence base related to both SIBs and DIBs. It seems essential that evaluations should not only focus on processes, as proposed by some advocates of DIBs, but they should also examine the impact of such bonds.

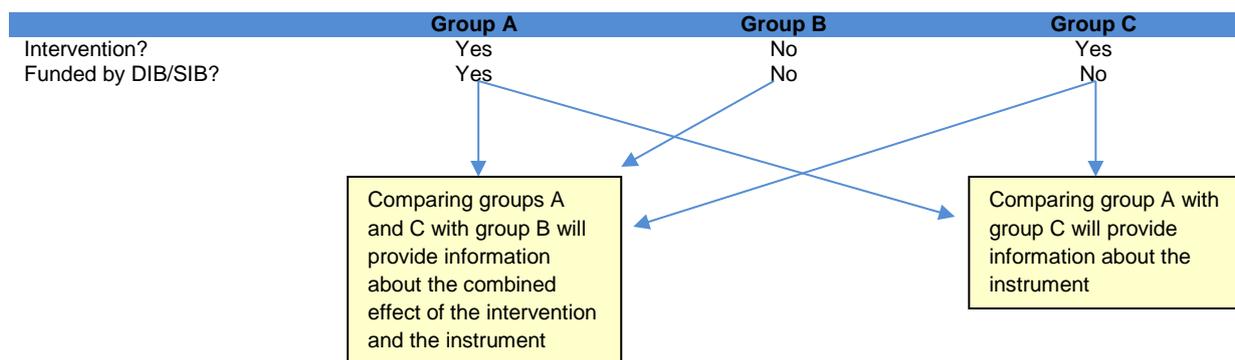
77. One distinctive feature of some SIBs has been that robust assessment of achievement of outcomes, including contrasting results with those achieved by a comparison group, has been part of their design and implementation. If such approaches are included in DIBs, measurement of such outcomes, which would be necessary for payment purposes, would be built into the programme's design. However, this would not be a sufficient assessment of impact for evaluation purposes. It would also be important to assess impact on non-payment outcomes and measures. However, evaluations need to go beyond assessing whether or not DIBs achieve impact, they need to also examine how this occurs, i.e. the processes involved.

78. A key issue, which perhaps has not yet received enough attention in writings on evaluations of DIBs and SIBs is the distinction between evaluating the intervention and the instrument. For example, if a group of prisoners (A), who receive an intervention focused on reducing reoffending financed through a SIB, are compared with a group of prisoners (B) who receive no intervention, the findings provide information about the combined effect of the intervention and the instrument. If, however, a further comparison group were added (C) – a group of prisoners who received a similar intervention but financed through a different instrument⁹, e.g. direct funding to an NGO, it would be possible to use those findings to distinguish between the effects of the intervention and the SIB/DIB instrument specifically. This type of evidence is likely to be needed to respond to the concern of some critics of DIBs and SIBs who believe that equivalent or better results could be achieved through other funding means. These issues are illustrated in Figure 4.

⁹ This other instrument could be a conventional or 'regular' aid instrument, for example, based on financing inputs or it could be an alternative form of payment by results, e.g. results-based aid or results-based financing. Specific issues that could be considered in a DIB/RBA or RBF comparison might include:

- How different incentive structures between DIBs and other forms of PBR influence process and outcomes.
- How having more agents in DIBs and more complex relationships influences efficiency, process and outcomes.
- The extent to which RBA and RBF are more applicable in more 'risky' sectors because investors may be more risk averse than donors.

Figure 4: Ways in which evaluation findings might be used to distinguish between effects of interventions supported by DIBs/SIBs and the effect of the instruments themselves



79. In addition, there is need to generate evaluation evidence concerning the creation of DIBs and SIBs markets as a whole. While evaluations of individual SIBs and DIBs may contribute to this, synthesising evidence from such evaluations will always present a partial picture as there are a number of issues which would not be picked up by such evaluations, e.g. the SIB or DIB that was not launched because of insufficient investor interest. Cross-portfolio evaluations, e.g. of DIBs or SIBs could be extremely useful in generating such data.

Appropriate questions for evaluations of DIBs

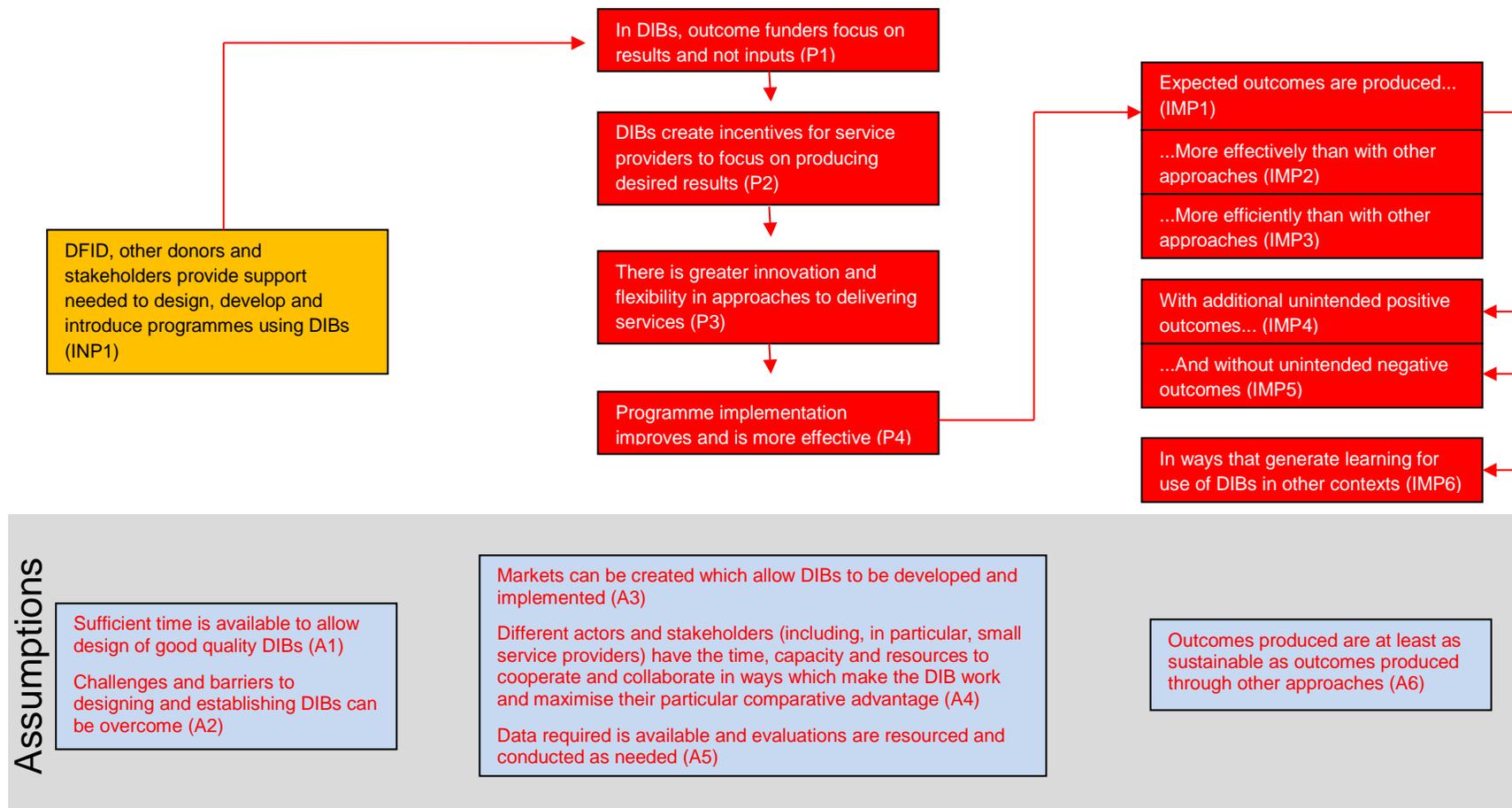
80. Commonly, questions are developed for an individual evaluation through a process of stakeholder consultation. While this is a useful approach in terms of ensuring that important stakeholder questions are raised and answered, there are a number of risks with this approach, including proliferation of questions and the risk that questions may be ad hoc and unsystematic. While this may create some difficulties for an individual evaluation, it becomes hugely problematic if a series of evaluations are expected to contribute to an overall body of evidence on a particular topic.
81. To address this challenge, it may be useful to have an organising framework for evaluating Development Impact Bonds within DFID and beyond. A proposed draft of such a framework is presented in Figure 5. This is loosely organised around a theory of change which shows how inputs into DIBs (INP1) lead to a number of processes (P1-P4¹⁰) which are expected to produce (a) particular impact(s) (IMP1-6). A number of assumptions have been identified at each stage of the framework (A1-A6)¹¹.

¹⁰ P4 is intended to correlate broadly to the expected outputs of the programme financed through the DIB.

¹¹ In general, elements which might be expected to be addressed by an individual DIB programme are included within the elements and linkages of the framework with other broader issues included as assumptions. It is recognised that in some DIBs, there may be particular activities/processes to address assumptions, such as developing markets or promoting relationships among stakeholders, which might mean these issues would be better shown as elements within the DIB programme rather than as an assumption.

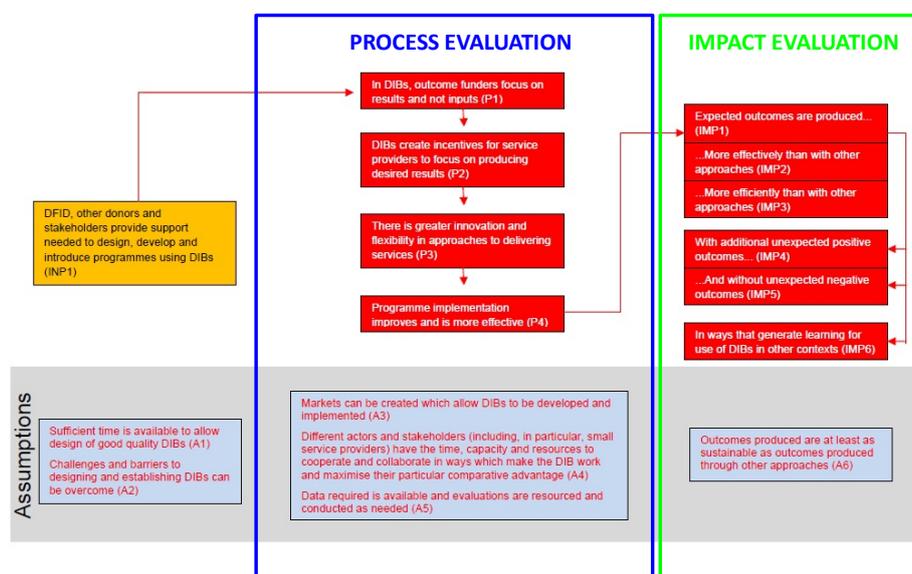
Figure 5: Possible framework for synthesising evidence from evaluations of DIBs to constitute a body of knowledge

Colour of boxes and linking arrows indicate current strength of evidence where green is very strong or strong; yellow is medium; amber is limited and red is no evidence (see DFID, 2014b, Table 2, p20). Boxes within the framework are considered to represent inputs (INP), processes (P) or impacts (IMP). Assumptions (A) for different stages of the proposed framework are shown in blue boxes at the bottom of the diagram. The strength of evidence that these assumptions hold true is denoted with the colour of text using the same system outlined above.



82. The framework also contains an assessment of the current level of evidence for particular elements, linkages and assumptions of the framework using a colour coding system based on DFID's approach to assessing the strength of evidence (DFID, 2014b). This assessment relates to DIBs specifically and concludes that there is currently no empirical evidence for (or against) most of the elements, linkages and assumptions. There is limited evidence that stakeholders are supporting the design and development of DIBs although no DIBs are currently operational. If evidence from SIBs was also included in this diagram, there would be some evidence for some of the elements, linkages and assumptions. Given the huge level of interest in both SIBs and DIBs currently, it is likely that level of evidence for elements, linkages and assumptions within this framework will develop and evolve extremely rapidly.
83. Adopting a framework of this nature could have a number of significant advantages for DFID and others in evaluating DIBs in the foreseeable future. First, it allows clear distinction between process and impact evaluations (see Figure 6). This is likely to be important given the issue documented in relation to SIBs (see paragraph 67, p23), i.e. that although the importance of rigorous impact evaluation is recognised by some stakeholders, there is a tendency for resources for these to be reduced under general budgetary pressures. This is an issue that DFID is likely to be pressed on, as it has with evaluations of Payment by Results initiatives, in general, with advocates of DIBs arguing that impact evaluations are not needed or are not a priority. DFID is likely to want to resist these pressures arguing that it is important to understand both the processes within a DIB that explain how it works and also the impact that programmes supported through DIBs have, particularly in comparison to more conventional aid instruments and other forms of payment by results. This framework could provide DFID with a useful tool for making these arguments.

Figure 6: Proposed framework shows how both process and impact evaluations are needed for DIBs



84. Second, although the framework is not explicitly designed around the OECD DAC (1991) evaluation criteria¹², it has been designed with these in mind and it is possible to show how each of these maps to the framework (Table 3).

¹² Of relevance, effectiveness, efficiency, impact and sustainability

Table 3: Mapping elements of the proposed evaluation framework (see Figure 5) to the OECD DAC evaluation criteria (Note: the labels INP, P, IMP and A relate to inputs, processes, impacts and assumptions shown in Figure 5)

Framework element	OECD DAC criteria				
	Relevance	Effectiveness	Efficiency	Impact	Sustainability
INP1: DFID, other donors and stakeholders provide support needed to design, develop and introduce programmes using DIBs	✓				
A1: Sufficient time is available to allow design of good quality DIBs					
A2: Challenges and barriers to designing and establishing DIBs can be overcome					
P1: In DIBS, outcome funders focus on results and not inputs	✓				
P2: DIBS create incentives for service providers to focus on producing results	✓				
P3: There is greater innovation ¹³ and flexibility in approaches to delivering services					
P4: Programme implementation improves and is more effective		✓			
A3: Markets can be created which allow DIBs to be developed and implemented					
A4: Different actors and stakeholders (including, in particular, small service providers) have the time, capacity and resources to cooperate and collaborate in ways which make the DIB work and maximise their particular comparative advantage ¹⁴					
A5: Data required is available and evaluations are resourced and conducted as needed					
IMP1: Expected outcomes are produced.				✓	
IMP2: More effectively than with other approaches		✓			
IMP3: More efficiently than with other approaches			✓		
IMP4: With additional unexpected positive outcomes				✓	
IMP5: And without unexpected negative outcomes				✓	
IMP6: In ways that generate learning for use of DIBs in other contexts					
A6: Outcomes produced are at least as sustainable as outcomes produced through other approaches					✓

85. Third, adopting this framework could potentially be useful for analysing questions and findings of relevant evaluations. For example, Annex 4 (p51) uses the framework to analyse questions developed for the evaluation of the Big Lottery Fund’s Commissioning Better Outcomes portfolio. While caution is needed in doing this, because that programme is quite specific to a UK context, use of this framework could be useful in analysing lessons learned from that evaluation that might be relevant to DIBs more broadly. In addition, it also potentially identifies gap areas that might not be covered by a particular evaluation. Annex 5 uses the framework to map the extent to which evaluation questions proposed for PBR, in general (DFID, 2014d) are included in this proposed framework for DIBs.

86. Fourth, the framework could be useful for developing generic questions for evaluations of DIBs. This could potentially be used by DFID, and perhaps others, to seek to ensure that individual DIBs evaluations contribute knowledge to a body of evidence regarding DIBs. These generic questions could be modified and/or supplemented for particular DIBs’ evaluations to generate context-specific material. Ideas for such generic questions are presented in Table 4.

87. Finally, the framework is relevant whatever stage evaluation is conducted. Given the new and rapidly developing DIBs field, it is likely that many initiatives to design DIBs may be guided by formative evaluations. These may be structured around the proposed framework as effectively as later summative evaluations.

¹³ Innovation is considered by some to be a hidden benefit of DIBs. Other hidden benefits might include spillover effects, for example. Within this framework, they would be captured under IMP4.

¹⁴ This assumption also relates to potential hidden costs of DIBs, such as transaction costs for DFID and beyond, e.g. related to contracts.

Table 4: Possible generic questions for DIBs evaluations derived from proposed evaluation framework (see Figure 5) (Note: the labels INP, P, IMP and A relate to inputs, processes, impacts and assumptions shown in Figure 5)

Framework element	Question(s)
INP1: DFID, other donors and stakeholders provide support needed to design, develop and introduce programmes using DIBs	To what extent have DFID, other donors and other stakeholders provided support needed to design, develop and introduce programmes using DIBs? What worked well? What worked less well?
A1: Sufficient time is available to allow design of good quality DIBs	Were any time constraints faced in designing the DIB? If yes, what were these and how were they overcome? Did they adversely affect DIBs design?
A2: Challenges and barriers to designing and establishing DIBs can be overcome	What challenges were faced in setting up the DIB? How were these overcome?
P1: In DIBs, outcome funders focus on results and not inputs	To what extent did outcome funders shift their focus to results away from inputs? If outcome funders continued to be focused on inputs and processes, why was this? If not, what were the incentives and how did they influence outcome funders?
P2: DIBs create incentives for service providers to focus on producing results	To what extent did the DIB create incentives for service providers to focus on results? What were the incentives and how did they influence service providers?
P3: There is greater innovation ¹⁵ and flexibility in approaches to delivering services	To what extent did the focus on outcomes create more flexibility, autonomy and innovation within the interventions?
P4: Programme implementation improves and is more effective	To what extent were programmes implemented through DIBs effective? What were the programme's outputs? To what extent were these produced as planned?
A3: Markets can be created which allow DIBs to be developed and implemented	To what extent has a market for DIBs been created? What type of investors are supporting DIBs? How is this changing over time?
A4: Different actors and stakeholders (including, in particular, small service providers) have the time, capacity and resources to cooperate and collaborate in ways which make the DIB work and maximise their particular comparative advantage ¹⁶	Who were the main stakeholders in the DIB? What role did they play? To what extent did this role make maximum use of their comparative advantage?
A5: Data required is available and evaluations are resourced and conducted as needed	To what extent was data required available? What was the effect of the DIB on data availability and quality?
IMP1: Expected outcomes are produced.	Did the DIB deliver the proposed outcomes and benefit to beneficiaries?
IMP2: More effectively than with other approaches	To what extent was the DIB more effective in delivering the outcome than more conventional aid instruments?
IMP3: More efficiently than with other approaches	To what extent was the DIB more efficient in delivering the outcome than more conventional aid instruments? Did it offer better value for money than other approaches?
IMP4: With additional unexpected positive outcomes	Did the DIB produce any unintended positive outcomes?
IMP5: And without unexpected negative outcomes	Did the DIB produce any unintended negative outcomes?
IMP6: In ways that generate learning for use of DIBs in other contexts	What lessons were learned that could be relevant for DIBs in other contexts?
A6: Outcomes produced are at least as sustainable as outcomes produced through other approaches	To what extent are the outcomes produced by the DIB likely to be sustainable? How does that compare with the sustainability of outcomes produced by other aid instruments?

¹⁵ Innovation is considered by some to be a hidden benefit of DIBs. Other hidden benefits might include spillover effects, for example. Within this framework, they would be captured under IMP4.

¹⁶ This assumption also relates to potential hidden costs of DIBs, such as transaction costs for DFID and beyond, e.g. related to contracts.

Appropriate methods to answer questions in evaluations of DIBs

88. A range of different methods could be used to collect data to answer these questions. These can be grouped into three main categories as documented in the working group's report on DIBs (see paragraph 55, p20) – experimental, quasi-experimental and non-experimental/other. These are briefly discussed here.
89. The order in which they are discussed should not be taken as endorsing the perhaps implied hierarchy of evaluation methods (Deaton, 2010), i.e. experimental methods at the top (the “gold-standard”) followed by quasi-experimental and non-experimental methods. While this may have some truth in terms of assessing impact of individual interventions, other questions may be better answered using other methods or approaches. For example, useful responses to questions about why and how things worked may be gained from non-experimental methods.
90. In addition, it is important to stress that methods from different categories can be used in combination. For example, experimental evaluations will also benefit from strong theory-based qualitative analysis if they are to correctly identify and understand causal mechanisms.
91. Experimental approaches are likely to be particularly valuable for answering some questions, particularly those which rely on comparing what happened when a DIB was used with what might have happened had some other approach been used (e.g. questions IMP2 and IMP3) or no intervention had taken place. Experimental approaches require statistically comparable groups to be allocated to different interventions and/or no intervention. Randomisation is usually used to generate statistically comparable groups. Individuals may be allocated randomly to a particular intervention or no intervention and this is the approach that is commonly used for biomedical research, e.g. into a new medicine. However, most social interventions are applied to groups or clusters and it is likely that this would be the case for most, if not all, DIBs. It is the groups or clusters that are randomly allocated to a particular intervention or no intervention. Whether or not an individual receives a particular intervention or no intervention is determined by which group or cluster that person belongs to. A key issue concerning cluster-based randomised controlled trials is that statistical power is, in most cases, influenced more by the number of clusters than by the number of individuals within the cluster.
92. Experimental methods are likely to be possible only when certain conditions can be met. There would need to be a reasonable number of groups or clusters who could potentially benefit from the DIB. These groups/clusters could be geographical areas, such as districts or wards, or particular institutions, such as hospitals, schools or prisons. In addition, for experimental methods, the decision about which units receive a particular intervention or no intervention needs to be made on a random basis. So, there are a number of situations when experimental methods are not likely to be possible. These include:
 - a. When no comparison group is possible – e.g. when a large DIB operates at a national level.
 - b. When the number of intervention/comparison clusters is small – e.g. when one intervention cluster is compared with one comparison cluster. It may be possible to use quasi-experimental approaches in such a situation but experimental methods may lack statistical power.

- c. Where pre-existing decisions have been taken as to which clusters receive a particular intervention or no intervention (although quasi-experimental methods may be possible).
- 93. It seems likely that there will be relatively few situations where DIBs can be evaluated using experimental methods. There is also a risk that trying to evaluate DIBs in this way may delay implementation unreasonably and/or may not be acceptable to some stakeholders, e.g. investors. Nevertheless, it would be ideal if some DIBs could be evaluated in this way as such evaluations are likely to add greatly to the knowledge base concerning DIBs. Such evaluations would be able not only to determine whether an intervention achieved outcomes to a greater effect than no intervention but they could also determine whether DIBs allowed interventions to produce greater levels of outcome than supporting the same interventions through other more conventional aid instruments.
- 94. The relative rarity of finding contexts in which high control can be exercised does mean that where it can be, it should be done to separate the effect of the DIB from the intervention. Experimental and non-experimental methods are able to separate the intervention from the financing of the intervention given planned (or fortunate) variation. In the majority of cases an evaluation of a DIB will not be able to get at the added-value of a DIB over and above the intervention itself. In rare cases where the possibility exists to design a DIB such that the same intervention is delivered in various ways (e.g. one through a traditional financing mechanism and another through a DIB), and to have a control where no intervention takes place, these should be grasped.
- 95. There are a wide variety of quasi-experimental methods. These can be used when comparison groups are available but it is not possible to allocate these to particular interventions/no intervention using random allocation. Such approaches can be used when:
 - a. Small numbers of intervention and comparison clusters are available. This involves different techniques to try to match the intervention and comparison clusters, including statistical techniques, such as propensity score matching.
 - b. Clusters are allocated to receive or not receive an intervention based on some numerical criterion. Clusters around the cut-off can then be compared using regression discontinuity design.
- 96. It may be possible to use experimental or quasi-experimental methods for a national programme if this is phased in in some way, e.g. with some districts beginning implementation before others. If choice of district can be made randomly, experimental methods may be possible with subsequent implementation waves acting as controls initially. If choice of district is non-random, quasi-experimental methods may be possible.
- 97. A range of non-experimental methods are also available for evaluations. The working group (see paragraph 55) refer to the use of historical baselines for assessing the impact of a DIB. While this might be possible in situations where the baseline is static and stable, it is unlikely that many such situations will be identified. In addition, such an approach would not give any information about what might have happened had the same intervention be supported with a different aid instrument. It would also not clearly rule out that any observed change could have been caused by something other than the intervention. Overall, the use of historical baselines as a basis for impact evaluation of DIBs is likely to be highly problematic.

98. However, this does not mean that non-experimental methods do not have any place in evaluations of DIBs.
99. Currently, there is a great deal of interest in using theory-based methods of evaluation. These require proposing a theory of change as to how an intervention produces expected changes and then collecting data to support or refute that theory. The proposed framework in Figure 5 would be highly suited for theory-based evaluations of DIBs.

A Strategy for Cumulating Knowledge across DIBs Initiatives

100. DFID is keen to ensure that when evaluations of DIBs are conducted, they should not only be relevant to that particular intervention but should also contribute to the overall body of evidence relating to DIBs. Three complementary approaches are proposed to seek to ensure that this happens.

Evaluations should be designed around a common evaluation framework

101. One option would be for DIBs to be established and evaluated as a series of experiments. However, given that DIBs are likely to differ greatly from each other in terms of many variables (including context, sector, size, implementing partners etc.), this is likely to be largely impractical, at this stage of DIBs design and development. For this reason, this option is discounted.
102. However, this does not mean that evaluations of DIBs could not be developed based on a common evaluation framework, e.g. as proposed in Figure 5. Doing this would allow individual evaluations to generate evidence and lessons which could be used to populate the framework. Of course, such evidence is likely to be context-specific but as more evaluations are conducted, the body of evidence would expand, allowing generalisations to be proposed and tested.

Capacity is needed to conduct real-time synthesis

103. However, just adopting a common framework for evaluation of DIBs is unlikely to result in evidence and lessons being synthesised into a body of knowledge without the availability of some capacity to ensure actively and intentionally that this happens. This will be particularly the case if evaluations of DIBs are commissioned on an individual basis.
104. Options for this include:
 - a. Establishing a consultative body or group to discuss and share learning among interested stakeholders. This could be along the lines of a DIBs community of practice as proposed by Center for Global Development (see paragraph 72, p26). However, there is a risk that credibility of such synthesis may be reduced if it is produced by staunch advocates of DIBs
 - b. Appointing an agent or organisation to be responsible for collecting and synthesising evidence generated from evaluations of DIBs. Such an agent/organisation could support and facilitate the community of practice outlined above. It is likely that such an agent or organisation is needed to ensure that the consultative body or group operates effectively and that lessons learned are synthesised into a body of evidence.

- c. Conducting cross-portfolio evaluations in addition to evaluations of individual DIBs. This is the approach being taken by the Big Lottery Fund in relation to SIBs funded through Commissioning Better Outcomes. A similar approach could be followed by DFID, either specifically focused on DIBs, or focused on more innovative forms of Payment by Results, more generally. A major added value of such a cross-portfolio evaluation would be that they could consider the market as a whole. The culmination of individual evaluations will not consider important cross-cutting issues such as DIBs that never got funded, the effect of bond yields on investor interest or the number of intermediaries. Such an evaluation would be needed to discover the true added-value of DIBs.

105. A major barrier to the real-time synthesis of lessons learned is secrecy concerning experiences of developing and implementing DIBs. While some caution in sharing information is perhaps inevitable because of concerns regarding fair procurement and/or protecting intellectual property, solutions to this issue need to be found if evidence regarding DIBs is to be synthesised effectively and promptly.

Periodic, retrospective synthesis exercises are likely to be very valuable

106. In addition to any real-time exercise that might be established, it is also likely to be useful to conduct retrospective synthesis exercises periodically. While approaches used for systematic reviews could be useful, there is a risk that these could be too narrowly focused on very specific questions and too reliant on data from randomised controlled trials to be particularly useful in relation to DIBs at this stage. It may be helpful to conduct an annual evidence review which includes a relatively systematic approach to identifying and assessing evidence relevant to DIBs but does not only consider evidence from randomised controlled trials.

Lessons Learned and Key Recommendations

107. DIBs and SIBs differ from other forms of payment by results (RBA and RBF) in that there are additional actors, e.g. investors and their intermediaries. It may be helpful for DFID to adopt the six-agent model proposed in this report for considering and evaluating DIBs. This provides a simplifying framework to assess each actor and their role.

108. Given the new and innovative nature of both SIBs and DIBs, rigorous and robust evaluation is of crucial importance. Although this is widely recognised in the field, the importance of evaluations is sometimes overlooked in manuals and/or squeezed out when finances are pressed. In addition, some advocates of DIBs argue for a greater focus on process evaluation and do not appear to see the need for additional impact evaluation over and above what is needed for authorising payment. DFID could play a useful role in advocating for and financing robust impact evaluations of DIBs as it has sought to do with other, innovative forms of payment by results.

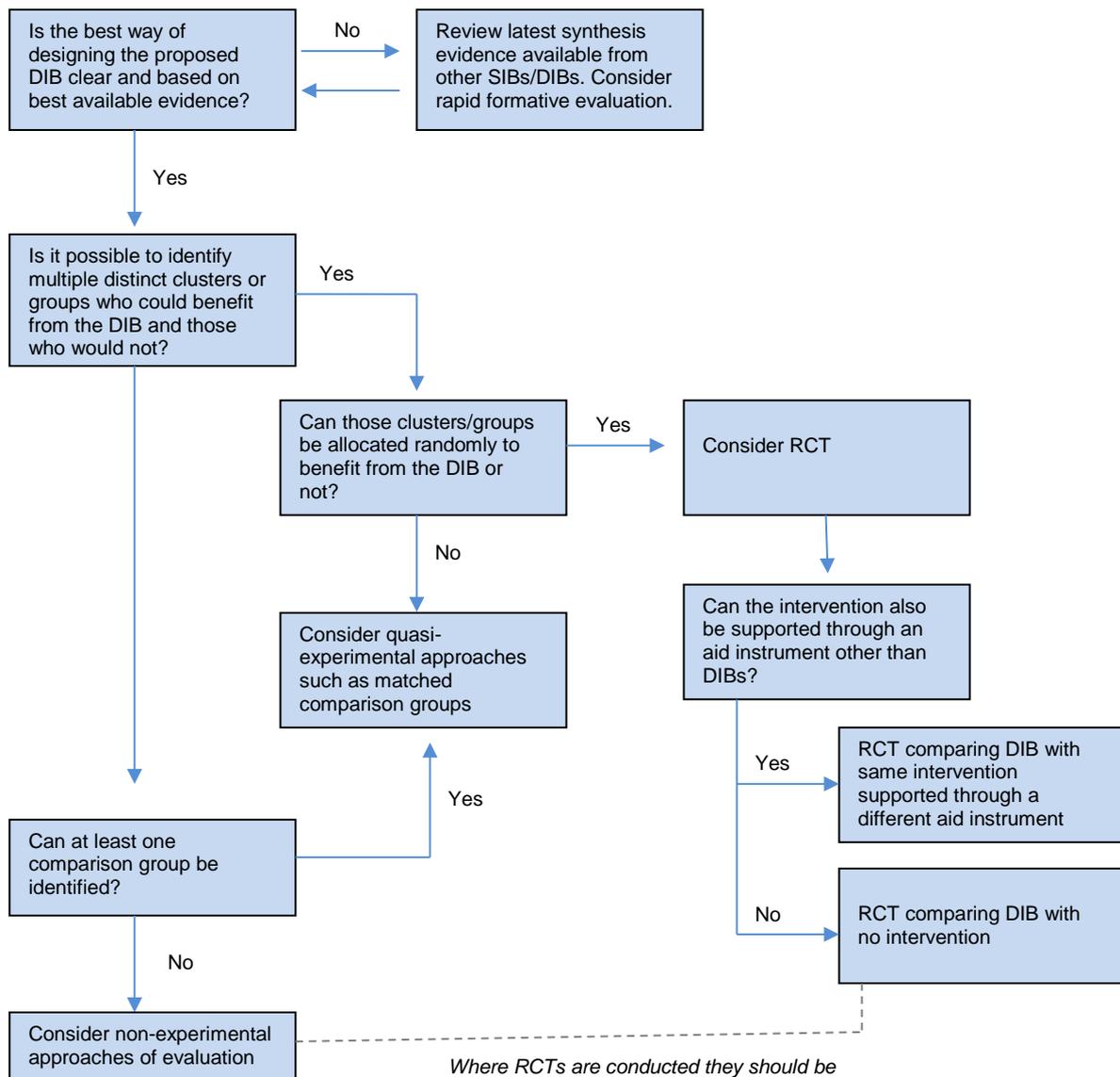
109. Given that DFID would like evaluations of individual DIBs to contribute data to an overall DIBs evidence base, it could be helpful to adopt a generic evaluation framework for DIBs within DFID and perhaps beyond. A first draft of such a framework is presented in Figure 5.

110. Generic evaluation questions have been derived from the draft evaluation framework (see Table 4) and these could be provided as a preliminary basis for individual evaluations of DIBs. Proposed generic terms of reference for an evaluation of DIBs are

presented in Annex 6 (p55). Both the terms of reference and the evaluation questions could be modified according to the specific contexts of individual DIBs.

111. A range of individual methods are suitable for evaluation of DIBs including experimental, quasi-experimental and non-experimental approaches. A possible flow chart for selecting among these methods is presented in Figure 7. This illustrates the importance of considering the need for a formative evaluation incorporated into DIB design. As it will normally be difficult to separate the intervention from the financial instrument, it is important that in cases where it is possible to design DIBs in such a way that they can be separated, it is done.

Figure 7: Flow chart for selecting methods to evaluate DIBs



112. Three distinct approaches are presented for synthesising learning from individual evaluations into an overall DIBs evidence base. They include using a common framework for evaluations of DIBs across DFID; contracting an actor or actors to conduct real-time synthesis of relevant DIBs-related learning and conducting periodic retrospective syntheses of learning from evaluations. It is suggested that DFID consider a combination of these three. It is likely that real-time synthesis of relevant DIBs-related learning could be essential given the speed at which DIBs are developing. Similarly, a cross-portfolio evaluation is likely to be particularly valuable to understand issues and experiences which affect the DIBs market as a whole and that might be overlooked by individual DIBs evaluations.

Annex 1: Terms of Reference¹⁷

Introduction

DFID is seeking a contractor – preferably a small team of researchers strong in economic theory and evaluation – to provide analytical support in identifying the most appropriate evaluation questions, approaches, methods and data for cumulating a body of evidence for Development Impact Bonds (DIBs). DIBs are a new aid modality for DFID and pose unique challenges for evaluation. The main purpose of the study is to assist DFID, other donors and policy makers to use the most appropriate evaluation approaches and methods for DIBs to enable wider learning. DFID is committed to learning from its expanding portfolio of payment by results (PbR) programmes, of which DIBs are one of the most innovative modalities.

The output from this exercise should clearly identify challenges in evaluating programmes that use DIBs; suggest cross-cutting evaluation questions that need to be answered to understand DIBs; appropriate experimental, quasi-experimental and non-experimental approaches to evaluating DIBs; and evidence and data that needs to be collected across evaluations to build a body of knowledge for DIBs.

Audience

The audience for the report are global policy makers, researchers, donors and programme teams working with PbR and DIBs programmes and evaluations. Study findings and recommendations need to be presented in a way that is accessible for both technical and non-technical audiences.

Objectives and scope

The contractor is expected to deliver a short report (maximum of 30 pages, excluding annexes) that should identify and synthesise evidence for the use of robust experimental, quasi-experimental and non-experimental approaches to evaluating the impact of development impact bonds. The final recommendations should provide a balanced perspective on the merits of different approaches, and suggest how best to measure the specific effect of a DIBs initiative.

As DIBs are a new modality, the scope of work for this study is to identify how best to evaluate their attributes and effects using suitable evaluation approaches and methods. DIBs are intended to improve development cooperation through gains in efficiency and by encouraging innovation. To understand the DIBs approaches (level of incentive, stakeholders involved, etc.), development objectives, partners, and contexts in which development cooperation improves through DIBs it is important that evaluations are designed to collect evidence in a way that enables cumulating a body of evidence. A key concern will be to ensure that each individual evaluation captures evidence of the effectiveness of DIBs in a way that can be synthesised and compared across programmes.

As DIBs are still largely untested in development, the evidence needed and selection of evaluation approaches will need to be informed by academic research and evaluations of programmes with similar attributes. Programmes with similar attributes to DIBs (e.g. levels of actors and complexities) will likely be found in other initiatives, such as social investment, financial markets, and M4P programmes, which may be in any development sector.

The objectives of the study are to:

¹⁷ These terms of reference were updated from the version used to procure services. The main changes involved including actual dates for deliverables rather than relative dates, i.e. number of weeks after contract signing.

- a) Recommend cross-cutting evaluation focus areas/questions that need to be answered to understand the value of DIBs as an aid modality (e.g. When is it appropriate to use a DIB instead of other PBR tools?);
- b) Recommend approaches and methods for evaluating DIBs programmes to answer cross-cutting evaluation questions;
- c) Recommend data to be collected across DIBs programmes, the external validity of this evidence, and the most appropriate ways of aggregating and synthesising;
- d) Recommend steps that can be taken (e.g. coordination mechanisms, etc.) to encourage evaluations designed for cumulating evidence.

This study should include the following:

- e) A systematic analysis of the theory, attributes and expected benefits of DIBs, including:
 - a. An analysis of the existing theory (e.g. economic and social) and evidence used to inform the design of DIBs initiatives;
 - b. Intended value addition of DIBs in areas like efficiency and innovation;
 - c. Expectation about which outcomes, impacts and beneficiaries are most suitable for benefitting from DIBs;
 - d. Theories of change for achieving intended outcomes and impacts, including underlying assumptions;
 - e. Levels and types of communities involved in achieving intended outcomes (e.g. donors, government, investors, implementers, etc.);
 - f. Approach to verification and remuneration;
 - g. Scale (e.g. regional, national, or specific communities) and length of programmes;
 - h. Contextual factors (e.g. conflict areas, climatic conditions, etc.)
 - i. Risks and challenges associated with DIBs to be tested;
- f) The designs, approaches, and methods currently used to evaluate DIBs programmes;
- g) Assessment of the strength of data and evidence available and collected and its external validity and suitability for synthesis;
- h) An assessment of strengths and weaknesses of different evaluation approaches and methods for generating evidence that is suitable for cumulating a body of knowledge on the effectiveness of DIBs;
- i) Lessons learned for DIBs evaluations that will affect the selection and use of proposed evaluation approaches and methods. These should be presented in a way that is accessible for non-technical audiences engaged in managing evaluations of their own initiatives.

Deliverables and outputs

The following deliverables and outputs are expected as part of the project:

- An inception report/analytical framework for the report **to be completed /submitted by 11th July 2014.** This must include -
 - A clear definition for what is meant by development impact bond, evaluation and impact evaluation (IE);
 - Inclusion criteria for academic papers, evaluations and programmes (DIBs and other) to be analysed and sampling strategy for the assessment of existing evidence;
 - Clear methodology for analysing DIBs, DIB programme attributes, evaluation approaches and methods, and data sampled;
 - Work-plan and timeline for completing the study;
 - Dissemination/communication plan for the final report;
- Draft Final report – **To be completed/submitted by 29th August 2014**

- Final report, taking on board suggestions and revisions to the draft final report – **To be completed/submitted by 26st September 2014**
- Presentation of the report to DFID and/or external audiences and participation in any pre-agreed dissemination/communication events (**dates to be agreed with supplier and Evaluation Department**).

Methods

The analysis and conclusions contained in the report should be based on the following:

- Desk review of academic papers, evaluation and programme documents;
- Desk review of relevant literature on development impact bonds (DIBs) and similar funding modalities (social investments, etc.);
- Desk review of DFID's PBR strategy documents and guidance;
- Sampling and assessment of the strength and validity of data and wider evidence collected from initiatives similar to DIBs;
- Interviews and fact checking with relevant staff from DFID, Centre for Global Development, the World Bank Health Results Innovation Trust Fund and Global Partnership for Output-Based Aid, and other organisations who may engage in DIBs evaluations;
- Example evaluation design for evaluating an DIB programme in a way that answers cross-cutting questions and provides evidence suitable for synthesis.

Skills Required

Key professional requirements for the contractor are:

- Knowledge of international development, specifically in the areas of development finance, social investing, private sector development, investment climate, and governance;
- Knowledge of the theoretical grounding for development impact bonds (e.g. actor-agent models, risk transfer, etc.)
- Significant experience with quantitative and qualitative data and analysis;
- Significant experience working with evaluation approaches and methods;
- Significant experience with research and synthesis;
- Publication record in relevant topics areas, specifically development impact bonds, social investing, aid effectiveness and/or development finance;
- Access to a network of evaluation specialists and policy stakeholders;

Timing of the study: Expected start/end dates are 23 June 2014 to 26 September 2014

The study should take a maximum of 30 days from the start of the contract.

The study team is expected to manage their inputs but an indicative project timetable is given below:

- Up to 1 day FTE for inception report;
- Up to 25 days FTE of desk-based research, interviews and meetings with relevant individuals, analysis of findings, checking results and formulating final recommendations;
- 3 days FTE of writing the report (including interim discussions with DFID);
- 1 day FTE of dissemination and communication activities.

Structure of the report

The report should comprise of the following sections:

- Executive Summary
- Introduction

- Overview of the attributes of DIBs, theory supporting their use in development and current initiatives
- Strengths and weaknesses of existing evidence and evaluation approaches and methods
- Appropriate evaluations questions, approaches and methods for DIBs
- A strategy for cumulating knowledge across DIBs initiatives
- Lessons learned and key recommendations
- Annex: Example design for the impact evaluation of a DIBs programme

Contractual issues

The project is contracted by DFID and is accountable to DFID. The contractor will report to Jonas Heirman (j-heirman@dfid.gov.uk) and Carol Travers (c-travers@dfid.gov.uk). The report should credit DFID for its contribution to the project. DFID will provide a logo for use in the report.

Annex 2: Bibliography

- Anonymous (2014) *Building a Better Mauritius: Creating the Next Wave of Prosperity*
- Atinc, T. and Gustafsson-Wright, E. (2013) *Early Childhood Development: The Promise, the Problem and the Path Forward* see <http://www.brookings.edu/research/articles/2013/11/25-early-childhood-development-atinc-gustafsson> Accessed 15.07.14
- Australian Government (2012) *Senate Economics References Committee Report – Investing for Good: The Development of a Capital Market for the Not-for-Profit Sector in Australia: Government Response*
- Azemati, H., Belinsky, M., Gillette, R., Liebman, J., Sellman, A. and Wyse, A. (2013) *Social Impact Bonds: Lessons Learned so Far* Community Development Investment Review
- Barclay, L. and Symons, T. (2013) *Introduction to Social Impact Bonds Workshop*
- Barder, O. and Perakis, R. (2012) *What if you could Invest in Development?*
- Barder, O. and Perakis, R. (2014) *Alchemy vs Chemistry: Does Innovative Finance Really Bring in New Money?* See: <http://international.cgdev.org/blog/alchemy-vs-chemistry-does-innovative-finance-really-bring-new-money> Accessed 28.08.14
- Belinsky, M. (2014a) *Details of Dalberg's Mozambique Malaria Development Impact Bond* See <http://michaelbelinsky.com/2014/04/01/details-of-dalbergs-mozambique-malaria-development-impact-bond/> Accessed 15.07.14
- Belinsky, M. (2014b) *Development Impact Bonds: Success Depends on a Supportive Network* See <http://www.theguardian.com/global-development-professionals-network/2014/jan/02/development-impact-bonds-success-network> Accessed 15.07.14
- Big Lottery Fund (2014) *Commissioning Better Outcomes and the Social Outcomes Fund* See <http://www.biglotteryfund.org.uk/sioutcomesfunds> Accessed 15.07.14
- Big Lottery Fund, Local Government Association and Social Finance (2014) *Case Study: The Peterborough Social Impact Bond*
- Birkwood, S. (2014) *Peterborough Prison Social Impact Bond Pilot Fails to Hit Target to Trigger Repayments* Third Sector – See <http://www.thirdsector.co.uk/peterborough-prison-social-impact-bond-pilot-fails-hit-target-trigger-repayments/finance/article/1307031> Accessed 22.08.14
- Bridges Ventures (2013) *Bridges Social Impact Bond Fund* See <http://bridgesventures.com/social-sector-funds/social-impact-bond-fund/> Accessed 22.08.14
- Budget and Performance Committee (2013) *Transcript of Agenda Item 5: GLA Social Impact Bond to Tackle Rough Sleeping*
- Burand, D. (2013) *Globalizing Social Finance: How Social Impact Bonds and Social Impact Performance Guarantees can Scale Development* NYU Journal of Law and Business, Vol 9:447
- Burde, D. and Linden, L.L. (2012). *The Effects of Village-based Schools: Evidence from an RCT in Afghanistan* See <http://www.povertyactionlab.org/evaluation/effect-village-based-schools-evidence-rct-afghanistan> Accessed 15.07.14
- Burde, D. and Linden, L.L. (2013) *Bringing Education to Afghan Girls: A Randomized Controlled Trial of Village-Based Schools* American Economic Journal: Applied Economics, 2013, 5(3):27-40
- Butler, D., Bloom, D. and Rudd, T. (undated) *Using Social Impact Bonds to Spur Innovation, Knowledge Building and Accountability* Community Development Investment Review

Cabinet Office (undated, a) *Guidance on the Template Contract for Social Impact Bonds and Payment by Results*

Cabinet Office (undated, b) *SIBs Database* (Excel)

Cabinet Office (undated, c) *Ministry of Justice: Offenders Released from Peterborough Prison*

Cave, S., Williams, T., Joliffe, D. and Hedderman, C. (2012) *Peterborough Social Impact Bond: An Independent Assessment: Development of the PDM Methodology* Ministry of Justice Research Series 8/12

Centre for Social Impact Bonds (2013a) *New South Wales Government: Children in Out-of-home Care* See http://data.gov.uk/sib_knowledge_box/new-south-wales-government-children-out-home-care Accessed 15.07.14

Centre for Social Impact Bonds (2013b) *Evaluation* See http://data.gov.uk/sib_knowledge_box/evaluation Accessed 15.07.14

Centre for Social Impact Bonds (undated) *About Us* See <http://blogs.cabinetoffice.gov.uk/socialimpactbonds/about-sib/> Accessed 15.07.14

CGD and Social Finance (2013a) *Investing in Social Outcomes: Development Impact Bonds: Summary Report of the Development Impact Working Group*

CGD and Social Finance (2013b) *Investing in Social Outcomes: Development Impact Bonds: The Report of the Development Impact Working Group*

CGD and Social Finance (2013c) *Making Development an Investment Opportunity*

CGD and Social Finance (2014) *Development Impact Bonds Briefing Note*

City of New York (2012) *Mayor Bloomberg, Deputy Mayor Gibbs and Corrections Commissioner Schiri Announce Nation's First Social Impact Bond Program*

Clist, P and Dercon, S. (2014) *12 Principles for Payment By Results (PbR) in International Development* See http://r4d.dfid.gov.uk/pdf/outputs/Misc_Infocomm/clist-dercon-PbR.pdf Accessed 28.08.14

Clist, P. and Verschoor, A. (2014) *The Conceptual Basis of Payment by Results in International Development* See <http://r4d.dfid.gov.uk/Output/195131/> Accessed 28.08.14

Crabtree, J. (2013) *A Rajasthan Project Wants to Bring a New Model to Charity* See <http://www.ft.com/cms/s/0/c5f074c2-0fcd-11e3-99e0-00144feabdc0.html#axzz37WliSaUT> Accessed 15.07.14

Deaton, A (2010) *Instruments, Randomization, and Learning about Development*, Journal of Economic Literature 48, pp. 424–455

Deloitte (2014) *Government and the Impact Economy*

Deloitte (undated) *Social Impact Bonds in Canada: Investor Insights*

Demel, A. (2013) *Second Thoughts on Social Impact Bonds* NYU Journal of Law and Business, Vol 9:503

Department of Legislative Services (2013) *Evaluating Social Impact Bonds as a New Reentry Financing Mechanism: A Case Study on Reentry Programming in Maryland*

DFID (2014a) *UK Development Bonds will Combat Global Poverty*

- DFID (2014b) *Assessing the Strength of Evidence: How to Note*
- DFID (2014c) *DFID's Strategy for Payment by Results: Sharpening Incentives to Perform*, DFID Policy Paper
- DFID (2014d) *Designing and Delivering Payment by Results Programmes – Technical Guidance*
- DFID (undated, a) *Secondary Education Development Impact Bond in Rwanda*
- DFID (undated, b) *Open Data Strategy: April 2012 – March 2014*
- Disley, E., Rubin, J., Scraggs, E., Burrows, N. and Culley, D. (2011) *Lessons Learned from the Planning and Early Implementation of the Social Impact Bond at HMP Peterborough*
- Disley, E. and Rubin, J. (2014) *Phase 2 Report from the Payment by Results Social Impact Bond Pilot at HMP Peterborough*
- Drexler, M. and Noble, A. (2013) *Impact Investing: Putting Ideas into Practice* See http://www.ssireview.org/blog/entry/impact_investing_putting_ideas_into_practice Accessed 15.07.14
- Drew, R.S. and Clist, P. (2014) *Evaluating Development Impact Bonds: Inception Report*
- Economist (2013) *The Peterborough Principles*
- Ecorys (2014) *Ecorys Wins Ground Breaking Evaluation into Social Impact Bonds (SIBs)* See <http://www.uk.ecorys.com/news/ecorys-wins-ground-breaking-evaluation-social-impact-bonds-sibs> Accessed 15.07.14
- Godoy, E. (2014) *Critics Question Impact of 'Pay for Success' Bonds* Inter Press Service See <http://www.ipsnews.net/2013/10/critics-question-impact-of-pay-for-success-bonds/> Accessed 15.07.14
- Griffiths, A. and Meinicke, C. (2014) *Introduction to Social Impact Bonds and Early Intervention*
- Gneezy, U. and Rustichini, A. (2000) "Pay Enough or Don't Pay at All", *Quarterly Journal of Economics*, 115(3), pp.791-810
- Guardian (undated) *Social Impact Bonds: Is the Dream Over?*
- Guardian (2013) *We Weren't Achieving Potential before Social Impact Bond, says Charity*, See <http://www.theguardian.com/social-enterprise-network/2013/nov/06/good-deals-social-impact-bond> Accessed 28.08.14
- Han, L. (2014) *Malaria in Mozambique: Trialling Payment by Results* See <http://www.theguardian.com/global-development-professionals-network/2014/mar/31/malaria-control-payment-by-results> Accessed 15.07.14
- Harvard Kennedy School (2014) *Social Impact Bond Technical Assistance Lab: About Us* See <http://hks-siblab.org/about-us/> Accessed 15.07.14
- Haynes, L., Service, O., Goldacre, B. and Torgerson, D. (2012). *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials*
- Henderson, D. (2011) *Social Impact Bonds will Fail without Solving the Evaluation Problem First* See <http://www.fullcontactphilanthropy.com/2011/04/04/social-impact-bonds-will-fail-without-solving-the-evaluation-problem-first/> Accessed 15.07.14
- Hughes, J. and Helbitz, A. (2014) *The Social Impact Bond Wars: The Defense Responds and the Prosecution Rests* See <http://www.pbs.org/newshour/making-sense/social-impact-bond-wars-defense-responds-prosecution-rests/> Accessed 15.07.14

Hughes, J. and Scherer, J. (2014) *Foundations for Social Impact Bonds: How and Why Philanthropy is Catalyzing the Development of a New Market*

Instiglio (2014) *Development Impact Bond: India* See <http://www.instiglio.org/en/girls-education-india/> Accessed 15.07.14

Instiglio (undated, a) *Teen Pregnancy in Colombia* See *Our Projects* webpage <http://www.instiglio.org/en/the-innovation/projects/> Accessed 15.07.14

Instiglio (undated, b) *Gender Gap in Education in India* See *Our Projects* webpage <http://www.instiglio.org/en/the-innovation/projects/> Accessed 15.07.14

Jupp, B. and Hoare, H. (2014) *Introduction to Procurement Social Impact Bonds Workshop*

Keohane, G.L. (2013) *Can Social Impact Bonds Unlock Private Money for Public Goods? Innovation in Pay-for-Success and Social Finance: Policy Note*

Khan, Z. (2013) *Developing Social Impact Bonds: Where Next?* See <http://www.rockefellerfoundation.org/blog/developing-social-impact-bonds-where> Accessed 15.07.14

KPMG (2014) *Evaluation of the Joint Development of the NSW Social Benefit Bonds Trial*

Lewin, M. (2014) *Making Impact Investing a Priority* See http://www.ssireview.org/blog/entry/making_impact_investing_a_priority Accessed 15.07.14

Liebman, J.B. (2011) *Social Impact Bonds: A Promising New Financing Model to Accelerate Social Innovation and Improve Government Performance*

Liebman, J. and Sellman, A. (2013) *Social Impact Bonds: A Guide for State and Local Governments*

Littlefield, E., Strauss, M. and Kimball, A. (2014) *Creating a Future Impact Investing Strategy* See http://www.ssireview.org/blog/entry/creating_a_future_impact_investing_strategy Accessed 15.07.14

Mair, V. (2014) *UBS and CIFF Launch Development Impact Bond in Education* See http://www.civilsociety.co.uk/finance/news/content/17666/ubs_and_ciff_launch_development_impact_bond_in_education Accessed 15.07.14

MDRC (undated) *Social Impact Bond Project at Rikers Island* See http://www.mdrc.org/project/social-impact-bond-project-rikers-island#featured_content Accessed 15.07.14

Ministry of Health (2014) *Social Bonds – New Zealand Pilot* See <http://www.health.govt.nz/our-work/preventative-health-wellness/social-bonds-new-zealand-pilot> Accessed 15.07.14

Mulgan, G., Reeder, N., Aylott, M. and Bo'sher, L. (2011) *Social Impact Investment: The Challenge and Opportunity of Social Impact Bonds*

Nicholls, A. and Tomkinson, E. (2013) *The Peterborough Pilot Social Impact Bond*

Nicola, D.J. (2013) *Environmental Impact Bonds* Center for the Advancement of Social entrepreneurship (CASE): The CASE Initiative on Impact Investing (CASE i3)

NSPCC (undated) *Payment by Results: Opportunities and Challenges for Improving Outcomes for Children*

O'Brien, A.M. (2014) *Leveraging Expertise across Asset Classes for Institutional Impact* See http://www.ssireview.org/blog/entry/leveraging_expertise_across_asset_classes_for_institutional_impact Accessed 15.07.14

O'Brien, T. and Kanbur, R. (2013) *The Operational Dimensions of Results Based Financing*

- OECD (1991) *DAC Principles for Evaluation of Development Assistance* See <http://www.oecd.org/development/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm> Accessed 26.08.14
- OECD (2014) *Innovative Approaches to Support more Inclusive Growth* Background Note – Session 2; 2014 Global Forum on Development
- Oxfam (2013) *Development Impact Bonds and Impact Investing – Genuine Impact or Snake Oil?* See <http://oxfamblogs.org/fp2p/development-impact-bonds-and-impact-investing-genuine-impact-or-snake-oil/> Accessed 28.08.14
- Palandjian, T. and Hughes, J. (2013) *The Social Impact Bond Market: Three Scenarios for the Future* See http://www.ssireview.org/blog/entry/the_social_impact_bond_market_three_scenarios_for_the_future Accessed 15.07.14
- Perakis, R (2013) *New Money? Cash Savings?* A Response to Duncan Green about Development Impact Bonds See <http://www.cgdev.org/blog/new-money-cash-savings-response-duncan-green-about-development-impact-bonds> Accessed 28.08.14
- Perakis, R. (2014) *Washington Tries a Social Impact Bond to Reduce Teen Pregnancy* See <http://www.cgdev.org/blog/washington-tries-social-impact-bond-reduce-teen-pregnancy> Accessed 15.07.14
- Probono Australia (2012) *NSW Govt Announces Social Bond Pilot Selection* See <http://www.probonoaustralia.com.au/news/2012/03/nsw-govt-announces-social-bond-pilot-selection> Accessed 15.07.14
- Rand (2011) *Evaluation of the Social Impact Bond: Lessons from Planning and Early Implementation at HMP Peterborough*
- Reuters, (2013) *Merrill Taps Rich Investors for Ex-Convict Social-Impact Bond* See <http://www.reuters.com/article/2013/12/30/us-merrill-bond-prison-idUSBRE9BT0QG20131230> Accessed 28.08.14
- Robinson, M. (2013) *Rwanda Education Development Impact Bond: A Scoping Report for DFID Rwanda*
- Rotheroe, A., Lomax, P. and Joy, I. (2013) *The Future for Children Bond: Identifying the Lessons Learned from Allia's Bond Offer to Retail Investors*
- Rudd, T., Nicoletti, E., Misner, K. and Bonsu, J. (2013) *Financing Promising Evidence-Based Programs: Early Lessons from the New York City Social Impact Bond*
- Russell, S. (2013) *Journey to Impact: A Practitioner Perspective on Measuring Social Impact*
- Saldinger, A. (2013) *How a Restaurant Chain Pioneered a Social Impact Bond to Fight Malaria* Devex Impact, See: <https://www.devex.com/news/how-a-restaurant-chain-pioneered-a-social-impact-bond-to-fight-malaria-82212> Accessed 28.08.14
- Social Finance (2011a) *Social Impact Bonds: The One Service: One Year On: Reducing Reoffending among Short Sentenced Male Offenders from Peterborough Prison*
- Social Finance (2011b) *A Technical Guide to Developing a Social Impact Bond: Vulnerable Children and Young People*
- Social Finance (2011c) *Peterborough Social Impact Bond*
- Social Finance (2012) *Family Planning Development Impact Bond: Initial Scoping Report to DFID*

Social Finance (2013a) *Guide to Social Impact Bond Development: A Technical Guide to Developing Social Impact Bonds*

Social Finance (2013b) *Peterborough Interim Figures – 31 October 2013*

Social Finance (undated) *Development Impact Bonds* See <http://www.socialfinance.org.uk/work/developmentimpactbonds> Accessed 15.07.14

Social Finance Israel (undated) *Employment Opportunities for Arab Citizens of Israel* See <http://www.socialfinance.org.il/social-impact-bonds/21/employment-opportunities-for-arab-citizens-israel> Accessed 15.07.14

Social Impact Business Models (undated) *Instiglio – Bringing Social Impact Bonds to Developing Countries* See <http://sibm.co/2013/04/21/instiglio-bringing-social-impact-bonds-to-developing-countries/> Accessed 15.07.14

Strickland, B. (2014) *Survey of Trends in Private Sector Partnerships for International Development and Modalities for Engagement*

Wheeler, K. and Egerton-Warburton, C. (2012) *Scoping Paper*

Wilson, G.E.R. (2014) *Enhancing Financial Returns by Targeting Social Impact* See http://www.ssireview.org/blog/entry/enhancing_financial_returns_by_targeting_social_impact Accessed 15.07.14

Annex 3: People Consulted

Owen Barder, Center for Global Development
Mike Belinsky, Instiglio
Ellie Cockburn, DFID
Geraldine De Boisse, Instiglio
Emma Disley, Rand Europe
Toby Eccles, Social Finance
Sofia Ericsson, SIDA
Lesley Hamill, DFID
Jonas Heirman, DFID
Anna Hentinnen, DFID
Barbara Kong, Dalberg
Ross Masood, DFID
Peter Nicholas, Social Finance
Liesbet Peeters, Dalberg
Rita Perakis, Center for Global Development
David Rinnert, DFID
Bill Savedoff, Center for Global Development
Louise Savell, Social Finance
Ajay Sharma, DFID
Gail Warrander, DFID
Gemma Wilson-Clark, DFID
Jonathan Wong, DFID

Annex 4: Questions Specified in the Terms of Reference for the Evaluation of the Commissioning Better Outcomes Fund (Big Lottery, 2014)

Letters and numbers in brackets map onto the proposed DIBS evaluation framework (see Figure 5)

The Contractor will be expected to use a mixture of qualitative and quantitative methods to address **at least** the following points, and will also be invited to comment on the questions proposed and suggest further questions:

Process evaluation

Development and set up phase

- What “works” when setting up a SIB? (INP1)
- What are the challenges in setting up a SIB and how can they be overcome? (A2)
- Which of these barriers would apply again if the SIB is repeated, and what could help mitigate them? (A2)
- What learning is there with regards to metrics development, for example timing, learning, consultation/methodology, robustness, data systems and sharing? (INP1; A5; IMP6)
- How can we give delivery partners standard metrics which work and which they can present back to commissioners/investors to secure contracts? (A5)
- To what extent is risk transferred away from the commissioner? (A4)
- When does the cost of risk transfer offer value for money when compared to alternative contract mechanisms? (IMP3)
- How effective are the differing approaches to mitigating risk? (IMP2)
- What due diligence processes should be undertaken when developing SIBs? (INP1)
- Why did commissioners who could have engaged only do so partially, or not at all? (INP1)

Mature phase

- What variations on the SIB model were funded through CBO? (P3)
- What are the key aspects of each variant? (P3)
- What are the benefits and disbenefits of SIBs (and each variant) for:
 - Commissioners
 - (VCSE) delivery partner(s)
 - Beneficiaries and/or policy areas
 - Investors (including their balance of risk) (A4)
- What is the value for money of each variant? (IMP3)
- What role do intermediaries play? (A4)
- What is the importance and effectiveness of the intermediary role, and are there any risks associated? (A4)
- How have partnerships developed? (A4)
- Are stakeholders staying engaged? (A4)
- Are there any metrics which proved particularly effective and which could be replicated in the future? (A5)
- What learning is there with regards to the way interventions were priced? (IMP3; IMP6)
- How is data being used to make decisions regarding ongoing service delivery, and how have the metrics informed service delivery changes during the project? (P3; A5)
- Which communication and managerial processes in maintaining and delivering the SIB are most effective? (P4)
- Is the SIB approach better than ‘normal’ commissioning arrangements and/or other forms of funding interventions to address social issues? (IMP2)
- Could any of the SIBs funded by CBO be replicated elsewhere? (IMP6)

- Are any of the SIBs funded by CBO scalable? (IMP6)

Impact evaluation

- Who are the different interventions funded through CBO most and least effective for, and how/why? (P4)
- Did the focus on outcomes create more flexibility within the interventions? (P3)
- Did the interventions deliver the proposed outcomes and benefit to beneficiaries? (IMP1)
- Are there significant differences in effectiveness between the SIBs, and what are the reasons for these differences? (IMP2)
- Are SIBs more appropriate for certain sectors? (IMP6)
- Did the SIBs funded through CBO demonstrate a route to greater efficiency in service delivery through innovation? (IMP3)
- Did CBO encourage or facilitate better collaboration between different parts of the public sector, and how/why (not)? (A4)
- Were the estimated savings achieved, and how did commissioners perform in 'cashing' these savings? (IMP1)
- How effective was the support and engagement contract? (A4)
- How effective was the development funding which was used to purchase technical support? (P4; A4)
- Is there still a requirement for technical support and development funding to help create more SIBs? (INP1; A4)
- Did CBO act as a catalyst to grow the social investment market and create more SIBs? (A3)
- Did CBO help correct market failure? (A3)
- Did CBO bring about a change in the pool of investor types (for example, more high net-worth individuals)? (A3)

Table A1 shows how these questions might be organised around the proposed evaluation framework in Figure 5. Where questions could be presented in more than one part of the framework one has been selected. Caution should be exercised in analysing this example as the proposed framework is for DIBs specifically and the example analysed is for a SIBs portfolio in the UK. However, overall, there appears to be a good match between the proposed questions and the framework. Some possible gaps could be identified from this analysis:

- There is no specific question related to time taken for setting up DIBs (A1). However, it could be argued that this would be covered by the question on challenges. Should this assumption simply be included in the more general assumption A2?
- There are no specific questions about the changing role of outcome funders (P1) or the effects of DIBs in terms of creating incentives (P2).
- There are no specific questions about unintended outcomes (IMP4 and IMP5).
- There is no question specifically about sustainability (A6). It is recognised that this may be more of an issue in a developing country context than in the domestic UK context.

Table A1: Example of how evaluation questions in the terms of reference for the evaluation of Commissioning Better Outcomes might be organised according to the framework presented in Figure 5

Framework element	Question
INP1: Outcome funders and other stakeholders provide support needed to design, develop and introduce programmes using DIBs	<p>What “works” when setting up a SIB?</p> <p>What due diligence processes should be undertaken when developing SIBs?</p> <p>Why did commissioners who could have engaged only do so partially, or not at all?</p>
A1: Sufficient time is available to allow design of good quality DIBs	
A2: Challenges and barriers to designing and establishing DIBs can be overcome	<p>What are the challenges in setting up a SIB and how can they be overcome?</p> <p>Which of these barriers would apply again if the SIB is repeated, and what could help mitigate them?</p>
P1: In DIBS, outcome funders focus on results and not inputs	
P2: DIBS create incentives for service providers to focus on producing results	
P3: There is greater innovation and flexibility in approaches to delivering services	<p>What variations on the SIB model were funded through CBO?</p> <p>What are the key aspects of each variant?</p> <p>How is data being used to make decisions regarding ongoing service delivery, and how have the metrics informed service delivery changes during the project?</p> <p>Did the focus on outcomes create more flexibility within the interventions?</p>
P4: Programme implementation improves and is more effective	<p>Which communication and managerial processes in maintaining and delivering the SIB are most effective?</p> <p>Who are the different interventions funded through CBO most and least effective for, and how/why?</p>
A3: Markets can be created which allow DIBs to be developed and implemented	<p>Did CBO act as a catalyst to grow the social investment market and create more SIBs?</p> <p>Did CBO help correct market failure?</p> <p>Did CBO bring about a change in the pool of investor types (for example, more high net-worth individuals)?</p>
A4: Different actors and stakeholders (including, in particular, small service providers) have the time, capacity and resources to cooperate and collaborate in ways which make the DIB work and maximise their particular comparative advantage	<p>To what extent is risk transferred away from the commissioner?</p> <p>What are the benefits and disbenefits of SIBs (and each variant) for:</p> <ul style="list-style-type: none"> ○ Commissioners ○ (VCSE) delivery partner(s) ○ Beneficiaries and/or policy areas ○ Investors (including their balance of risk) <p>What role do intermediaries play?</p> <p>What is the importance and effectiveness of the intermediary role, and are there any risks associated?</p> <p>How have partnerships developed?</p> <p>Are stakeholders staying engaged?</p> <p>Did CBO encourage or facilitate better collaboration between different parts of the public sector, and how/why (not)?</p> <p>How effective was the support and engagement contract?</p> <p>How effective was the development funding which was used to purchase technical support?</p> <p>Is there still a requirement for technical support and development funding to help create more SIBs?</p>
A5: Data required is available and evaluations are resourced and conducted as needed	<p>What learning is there with regards to metrics development, for example timing, learning, consultation/methodology, robustness, data systems and sharing?</p> <p>How can we give delivery partners standard metrics which work and which they can present back to commissioners/investors to secure contracts?</p> <p>Are there any metrics which proved particularly effective and which could be replicated in the future?</p>
IMP1: Expected outcomes are produced.	<p>Did the interventions deliver the proposed outcomes and benefit to beneficiaries?</p> <p>Were the estimated savings achieved, and how did commissioners perform in ‘cashing’ these savings?</p>

Framework element	Question
IMP2: More effectively than with other approaches	<p>How effective are the differing approaches to mitigating risk?</p> <p>Is the SIB approach better than 'normal' commissioning arrangements and/or other forms of funding interventions to address social issues?</p> <p>Are there significant differences in effectiveness between the SIBs, and what are the reasons for these differences?</p>
IMP3: More efficiently than with other approaches	<p>When does the cost of risk transfer offer value for money when compared to alternative contract mechanisms?</p> <p>What is the value for money of each variant?</p> <p>Did the SIBs funded through CBO demonstrate a route to greater efficiency in service delivery through innovation?</p>
IMP4: With additional unexpected positive outcomes	
IMP5: And without unexpected negative outcomes	
IMP6: In ways that generate learning for use of DIBs in other contexts	<p>What learning is there with regards to the way interventions were priced?</p> <p>Could any of the SIBs funded by CBO be replicated elsewhere?</p> <p>Are any of the SIBs funded by CBO scalable?</p> <p>Are SIBs more appropriate for certain sectors?</p>
A6: Outcomes produced are at least as sustainable as outcomes produced through other approaches	

Annex 5: Mapping Proposed PBR Evaluation Questions to those Proposed for DIBs (see DFID, 2014d)

Question proposed for PBR	Comment on mapping to proposed DIB framework
Key Questions	
Under what sets of circumstances is an incentive approach appropriate?	Not covered – would be an important question for lesson learning and synthesis
What is the optimal size and nature of the incentive required?	Not covered – would be an important question for lesson learning and synthesis
Were the results DFID paid for greater than what would have occurred in the absence of PbR (i.e. did PbR have an impact on the outcome of interest, relative to what would have happened in the absence of PbR)	Key element of framework – IMP1 and IMP2
What is the actual nature of the intervention?	Quite a descriptive question which is not really covered in the proposed framework. There is a question about the nature of incentives.
Was the pilot better able to demonstrate 'what we're buying' better than other forms of aid? (i.e. the <u>additional</u> results achieved.)	Key element of framework – IMP2
Was the DFID pilot more cost effective than other forms of aid or interventions in the same sector (taking into account verification costs; any pricing of risk and incentive; and the value of increased certainty around results)?	Key element of framework – IMP3
Did the pilot have any impact (positive or negative) on accountability to citizens?	Implied as part of IMP1 if this is an intended outcome – and under IMP4 and 5 if unintended.
Did the pilot create an incentive for an increased results focus by the partner government?	Perhaps the issue of results focus is more relevant to other agents in a DIB – covered for outcome funder under P1 and for service provider under P2
Was the quantity and quality of DFID's policy dialogue with government affected?	Issues of relative roles and changes in them are covered in the framework under A4
Additional questions	
Attention to potential unintended effects.	Key element of framework – IMP4 and IMP5
What is the impact on equity?	Implied as part of IMP1 if this is an intended outcome – and under IMP4 and 5 if unintended.
To what extent, and under what circumstances, are PbR approaches compatible with other development principles and desired outcomes?	Not covered – question is quite complex and precise meaning may be unclear
To what extent do PbR approaches lead to improvements in quality?	If quality of service is considered an outcome - Implied as part of IMP1 if this is an intended outcome – and under IMP4 and 5 if unintended.
What are the effects of incentives on the broader systems?	Implied as part of IMP1 if this is an intended outcome – and under IMP4 and 5 if unintended.
What are the long-term effects of incentive approaches?	Key element of framework – A6
What impact has the intervention had on peace building and state building & on the development of sustainable and inclusive institutions?	Implied as part of IMP1 if this is an intended outcome – and under IMP4 and 5 if unintended.

General Comments

- While the proposed evaluation questions for PBR are potentially valuable, they appear to lack any organising framework. So, for example, a descriptive question about the nature of interventions is sandwiched between two high level questions about outcomes. Without some form of structure/organising framework, it is difficult to see how such questions could be used in practice for an individual evaluation. In particular, it is hard to see how they could be used as an organising framework for synthesising learning related to PBR.
- There are some very specific questions about particular impacts, e.g. on accountability to citizens and on peace/state building. It is unclear if these specific issues would apply in all contexts and are comprehensive. It is unclear why they are not grouped together.

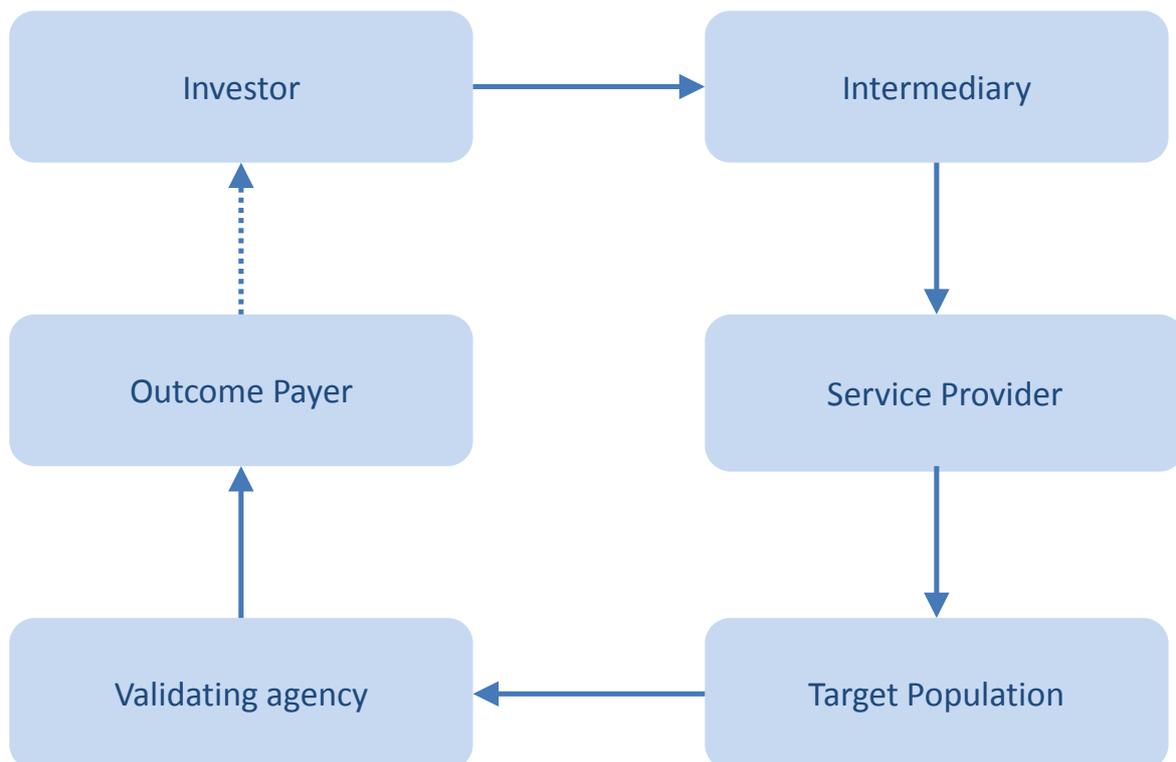
Annex 6: Example Design for the Impact Evaluation of a DIBs Programme

This annex seeks to provide a generic template for an impact evaluation of a DIBs programme. This is quite difficult because DIBs are likely to vary hugely in terms of context, sector and size. As a result, there are a lot of gaps which are shown in red. It might be possible to develop more detailed terms of reference for a specific DIB as an example.

Context

1. Specific for particular evaluation
2. Development Impact Bonds (DIBs) are a relatively new instrument, based on the experience of Social Impact Bonds (SIBs) in developed countries (CGD and Social Finance, 2013). DIBs are recognised by DFID as one of three main types of approach to payment by results (DFID, 2014). These three approaches differ from each other in terms of who DFID, as the outcome payer, pays when agreed results are achieved. In results-based aid (RBA), DFID would pay a country's government. In results-based financing (RBF), DFID would pay a service provider. In a Development Impact Bond, DFID would pay one or more investors who had pre-financed activities through a service provider, usually through the services of an intermediary (see Figure T1).

Figure T1: DFID's DIBs six agent model (see Drew and Clist, 2014)

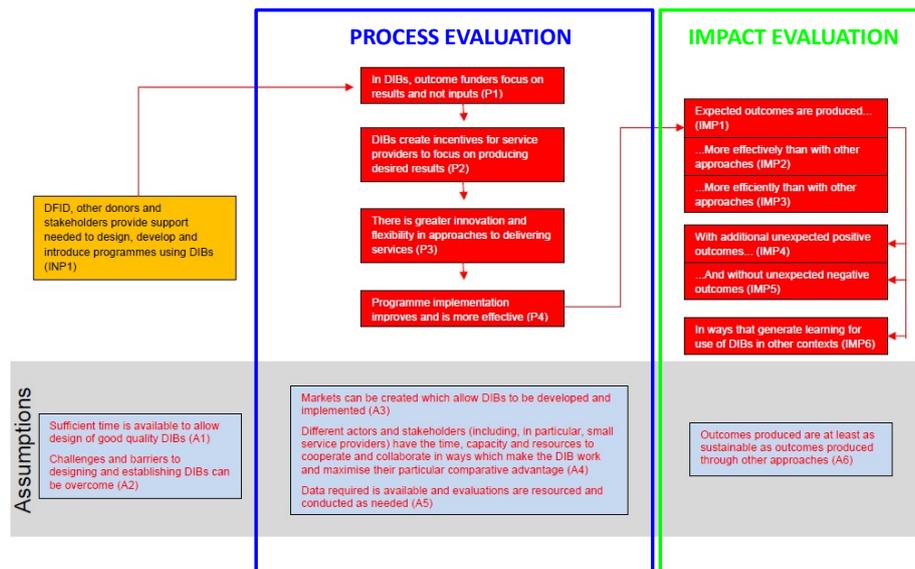


3. Context-specific information on why DIBs being introduced specifically here and plans to date.

Purpose

- This evaluation has a dual purpose. The first element is to determine the impact of the Development Impact Bond and the second is to understand more fully the process by which impact has occurred. DFID has developed a generic framework to guide its evaluations of DIBs (see Appendix 1) and this can be used to show the two key elements of process and impact evaluation (see Figure T2).

Figure T2: Proposed framework shows how both process and impact evaluations are needed for DIBs



- The evaluation is intended to have a strong focus on learning lessons both for improving this particular DIB and for contributing to the overall evidence base concerning DIBs. It is not expected that the evaluation will have a significant focus on accountability as that is the main focus of a separate data validation process.

Scope and Objectives

- The evaluation is expected to evaluate the impact of both the intervention supported by the DIB and the DIB itself as an aid instrument. Bidders are expected to show how they would do this and to demonstrate that they understand clearly the distinction between these two issues.
- The specific objectives of the evaluation:
 - Context specific**
- The evaluators are expected to conduct the evaluation using a logic-based evaluation framework. This should be based on the generic DIBS evaluation framework (see Appendix 1) but this may be modified for local context subject to agreement from DFID. Bidders are expected to demonstrate their understanding of the generic framework and to explain how they would modify it and use it in practice. It is expected that the final evaluation framework would be agreed between DFID and the contractor during inception.

9. Table T1 presents a number of key questions which have been developed by DFID for evaluations of DIBs based on understanding of the framework in Appendix 1. However, DFID expects that these might be modified for specific DIBS' evaluations. Modifications could include deletions, changes and additions. Bidders are expected to propose any modifications they would expect to make. Justifications for these modifications should be provided including how they will benefit the individual DIB being evaluated and how they will ensure that lessons learned can still be synthesised into the growing evidence base related to DIBs. DFID expects that the final set of questions for the evaluation would be agreed with the selected contractor during inception.

Table T1: Generic questions for DIBs evaluations derived from proposed evaluation framework (see Appendix 1)

Framework element	Question(s)
INP1: DFID, other donors and stakeholders provide support needed to design, develop and introduce programmes using DIBs	To what extent have DFID, other donors and other stakeholders provided support needed to design, develop and introduce programmes using DIBs? What worked well? What worked less well?
A1: Sufficient time is available to allow design of good quality DIBs	Were any time constraints faced in designing the DIB? If yes, what were these and how were they overcome? Did they adversely affect DIBs design?
A2: Challenges and barriers to designing and establishing DIBs can be overcome	What challenges were faced in setting up the DIB? How were these overcome?
P1: In DIBs, outcome funders focus on results and not inputs	To what extent did outcome funders shift their focus to results away from inputs? If outcome funders continued to be focused on inputs and processes, why was this?
P2: DIBs create incentives for service providers to focus on producing results	To what extent did the DIB create incentives for service providers to focus on results? What were the incentives and how did they influence service providers?
P3: There is greater innovation and flexibility in approaches to delivering services	To what extent did the focus on outcomes create more flexibility, autonomy and innovation within the interventions?
P4: Programme implementation improves and is more effective	To what extent were programmes implemented through DIBs effective? What were the programme's outputs? To what extent were these produced as planned?
A3: Markets can be created which allow DIBs to be developed and implemented	To what extent has a market for DIBs been created? What type of investors are supporting DIBs? How is this changing over time?
A4: Different actors and stakeholders (including, in particular, small service providers) have the time, capacity and resources to cooperate and collaborate in ways which make the DIB work and maximise their particular comparative advantage	Who were the main stakeholders in the DIB? What role did they play? To what extent did this role make maximum use of their comparative advantage?
A5: Data required is available and evaluations are resourced and conducted as needed	To what extent was data required available? What was the effect of the DIB on data availability and quality?
IMP1: Expected outcomes are produced.	Did the DIB deliver the proposed outcomes and benefit to beneficiaries?
IMP2: More effectively than with other approaches	To what extent was the DIB more effective in delivering the outcome than more conventional aid instruments?
IMP3: More efficiently than with other approaches	To what extent was the DIB more efficient in delivering the outcome than more conventional aid instruments? Did it offer better value for money than other approaches?
IMP4: With additional unexpected positive outcomes	Did the DIB produce any unexpected positive outcomes?
IMP5: And without unexpected negative outcomes	Did the DIB produce any unexpected negative outcomes?
IMP6: In ways that generate learning for use of DIBs in other contexts	What lessons were learned that could be relevant for DIBs in other contexts?
A6: Outcomes produced are at least as sustainable as outcomes produced through other approaches	To what extent are the outcomes produced by the DIB likely to be sustainable? How does that compare with the sustainability of outcomes produced by other aid instruments?

Evaluation Criteria

10. The proposed DFID evaluation framework for DIBs (Appendix 1) has been developed with the OECD DAC evaluation criteria (relevance, effectiveness, efficiency, impact, sustainability) in mind. More details of how this framework maps onto those criteria are contained in the study on evaluating DIBs (Drew and Clist, 2014). Bidders are expected to confirm their understanding of this but are not expected to use the OECD DAC criteria as an organising framework for the evaluation.

Methodology

11. Bidders are expected to propose methods for conducting this evaluation. In general, DFID welcomes proposals that include experimental or quasi-experimental methods, particularly for evaluating impact. **Context specific guidance as to whether these are considered feasible in this particular DIB.**
12. Overall, whatever specific methods are used, DFID expects that the evaluation will be theory-based using a framework derived from Appendix 1 as a basis for this. Bidders should explain how they understand theory-based evaluation in general and specifically how they would apply that understanding for this evaluation.

Data

13. In general, the evaluators will not be expected to collect primary data in relation to any payment metrics. It is expected that such primary data will be collected by service providers and verified by an independent data verification contractor. The evaluators are welcome to comment on the adequacy or otherwise of those processes but they are not expected to compensate for any deficiencies they might identify in those processes.
14. Bidders are expected to explain any other primary data collection they might expect to undertake, e.g. for relevant non-payment outcome measures with explanation as to how they would collect such data, particularly where those methods involve activities other than interviews with groups or individuals.
15. The evaluation will be expected to draw heavily on existing sources of data, including service delivery records and reports of the independent data verifier. It is expected that the evaluation will consider both quantitative and qualitative data. **Context specific information on data sources.**

Outputs

16. Specific outputs of the evaluation include:
 - i. An inception report.
 - ii. A final evaluation report focused on lessons learned for the specific DIB evaluated.
 - iii. A summary of lessons learned that would be considered relevant to other DIBs both inside and outside DFID. It is expected, in particular, that this summary would be structured around the elements, linkages and assumptions identified in Appendix 1. It is not expected that every evaluation would generate evidence on every part of this framework but it is expected that every evaluation would summarise what the evaluators believe the

individual evaluation has contributed to the overall evidence base regarding DIBs.

17. It is expected that all outputs of the evaluation will be made freely and publicly available without copyright/ownership restrictions. Bidders should explain briefly in their proposals how they would ensure this, including arrangements for storage and accessibility of any data generated through the learning process.

Workplan

18. Context specific material on time frame, deliverables and budget (if appropriate).

19. It is proposed that the evaluation would be conducted by a team of evaluators with at least the following skills and expertise.
 - i. Understanding of payments by results programme, in general (essential) and DIBs or SIBs, in particular (desirable)
 - ii. Experience of evaluating both interventions and instruments using theory-based approaches (essential)
 - iii. Experience of generating lessons learned from specific evaluations that can be used to contribute to an evidence base in a new and rapidly evolving area (essential)
 - iv. **Context specific skills**
 - v. Excellent written and verbal communications skills (in English) (essential).

Responsibilities

20. The primary recipient of the evaluation is DFID **context specific material for the evaluation** Other audience for the findings of the specific evaluation include:
 - i. Other outcome funders – **context specific material for the evaluation.**
 - ii. Investors - **context specific material for the evaluation**
 - iii. National and local government
 - iv. Intermediaries - **context specific material for the evaluation**
 - v. Service providers - **context specific material for the evaluation**
21. There are a number of audiences for the broader lessons learned from the evaluation. These include stakeholders involved in other DIBs, including those supported by DFID. **Could mention specific actors and/or any community of practice that might be established?**
22. Day to day management of the learning process will be conducted **by context specific material for the evaluation. Also context specific material on reference groups.**

Other matters

23. Bidders are expected to explain how they would ensure the independence of the evaluation explaining, in particular, how they would define, identify and deal with any conflicts of interests.
24. Bidders should also explain how they would ensure that the evaluation is ethical in nature. Bidders are advised to refer to DFID's ethics principles for research and evaluation (2011).

Key documents

- CGD and Social Finance (2013) *Investing in Social Outcomes: Development Impact Bonds: The Report of the Development Impact Working Group*
- DFID (2011) *DFID Ethics Principles for Research and Evaluation*
- DFID (2014) *DFID's Strategy for Payment by Results: Sharpening Incentives to Perform*
- Drew, R. and Clist, P. (2014) *Evaluating Development Impact Bonds: A Study for DFID*

Appendix 1: Generic framework for evaluations of DIBs in DFID

Colour of boxes and linking arrows indicate current strength of evidence where green is very strong or strong; yellow is medium; amber is limited and red is no evidence (see DFID, 2014b, Table 2, p20)

Assumptions for different stages of the proposed framework are shown in blue boxes at the bottom of the diagram. The strength of evidence that these assumptions hold true is denoted with the colour of text using the same system outlined above.

