BALANCING EVIDENCE AND RISK

CHILDREN'S
INVESTMENT FUND
FOUNDATION
CIFF

CONTENTS

Introduction

Impact Bonds[1] aim to measurably improve the lives of the people most in need by driving resources towards better, more effective programmes.

At their core, Impact Bonds are public-private partnerships which fund effective service provision through outcomes-based contracts. Impact Bonds enable federal, state and local governments and donors to partner with high-performing service providers by using private investment to expand effective programmes.

In their involvement of investors, Impact Bonds differ from other forms of results-based financing. Solving the pre-financing needs of service providers through private investment enables governments and donors to focus a greater portion of their payments on outcomes. This in turn creates flexibility to move away from input or activity based monitoring. The contractual scope for service adaptation to varied and changing local needs, through rapid feedback loops, is a key strength of the Impact Bond model and, in turn, should drive the best possible social outcomes.

Central to each Impact Bond's effective operation is the confidence that both outcomes funders (governments or donors) and investors have around the validity of metrics on which payments will be made.

As use of Impact Bonds has grown over the last couple of years[2] much has been written about the basis on which such contracts are evaluated. Debate, particularly in markets like the US, is increasingly polarised among those that maintain that only randomised control trials (RCTs)[3] will do, and those that advocate less intensive approaches in order to accelerate the market.

Polarisation of any debate is always concerning as it, almost inevitably, ends up stifling more nuanced discussion. This paper outlines a

---

1 Also known as 'Pay for Success' or 'Social Benefit' Bonds in the US and Australia respectively.

2 More than 60 worldwide to date with at least as many again under development.

3 RCTs randomly assign members of the potential target population to either receive the intervention or not (thereby creating a control group). By observing how the two groups differ after the intervention, the effectiveness of the intervention can be assessed.

pragmatic way forward, taking as its starting point consideration of the contractual risks that different evaluation approaches would represent to outcomes funders and investors in different contexts.

The paper describes three levels of levels of evaluation that may take place within an Impact Bond (Figure 2), but focuses on considerations around 'Contractual Outcomes Metrics' - the metrics against which Impact Bond payments are or aren't made that carry with them both financial and reputational risk for outcomes funders and investors.

The objectives of outcome funders in commissioning an Impact Bond – whether to scale an evidence-based programme or to drive innovation in programme design and delivery – are a key determinant of what an appropriate evaluation approach would look like. The less contractual control investors and service providers have over what services are delivered and how, the more appropriate it will be for Contractual Outcomes Metrics to relate to the quality of service delivery rather than ultimate end user outcomes.

It is to these Contractual Outcomes Metrics that we apply the term 'Counterfactual Risk' – the risk to both outcomes funders and investors generated by not having an accurate assessment of what would have happened in the absence of an intervention. If we do not know what would have happened in the absence of a service then outcomes might be over- or under-valued.

We conclude that key determinants of Counterfactual Risk include:

- The stability of target population outcomes to external events and over time;

- The strength of the evidence base for the intervention (if specified) for the target population in a relevant context; and

- The length and scale of the Impact Bond contract.

We recommend that evaluations of Contractual Outcome Metrics:

- Focus on determining whether contractual payments should be made and reducing Counterfactual Risk rather than seeking to build the 'global evidence base' for interventions;

- Do not restrict, through their design, the effective delivery and adaptation of Impact Bond funded services to meet the varied and changing needs of the target population; and

• Are proportional (in effort and cost) to both the size of the Counterfactual Risk to each party and the scale of the Impact Bond contract.

Evaluation for the purposes of determining payment in Impact Bonds is high stakes - this paper outlines a framework for thinking about which evaluation approaches might be relevant and proportionate in different contexts.

We acknowledge that practical considerations – including target population size, regulatory constraints and ethical considerations – will also need to be taken into account.

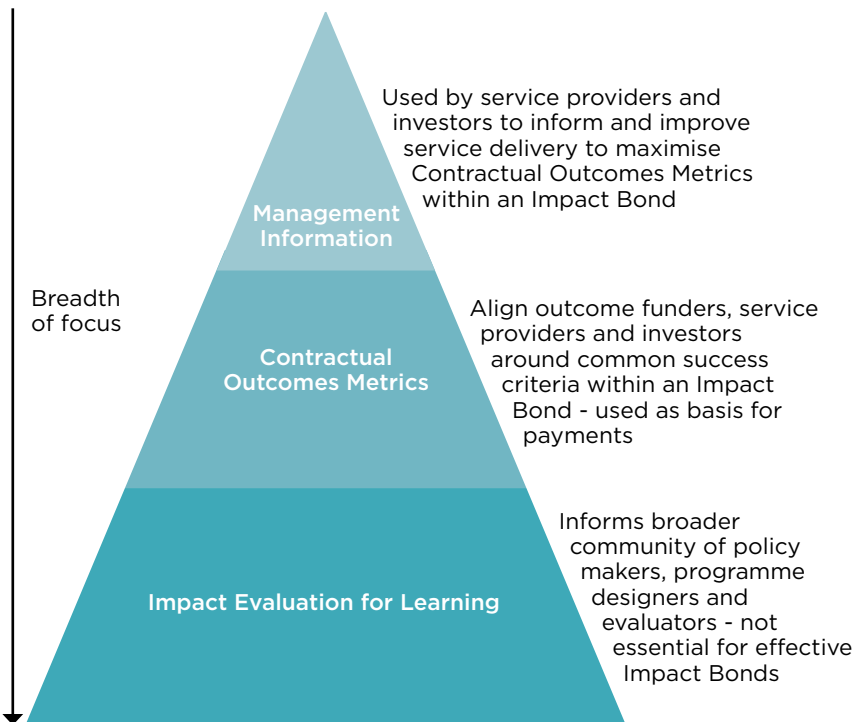## FIGURE 1: TYPICAL IMPACT BOND STRUCTURE
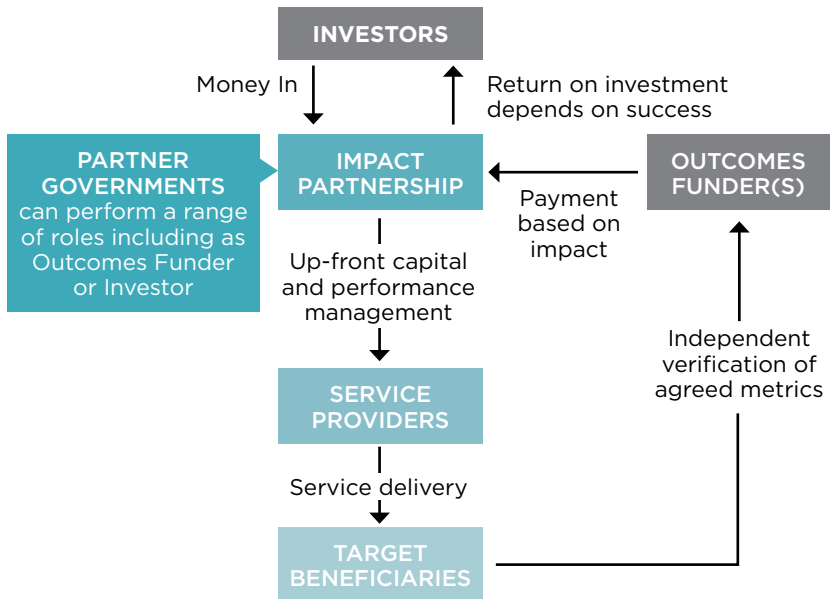
1 What do we mean by Impact Bond evaluation?

Some of the debate around appropriate Impact Bond evaluation approaches has been clouded by a lack of clarity around the purpose of such evaluations. Before progressing to consider contractual risk as it relates to outcome measurement in more detail it is useful to step back to consider the levels of data collection that might be desired.

The core of an Impact Bond contract is the Contractual Outcomes Metrics. These are the metrics against which Impact Bond payments are or aren't made and carry with them financial and reputational risk for both outcomes funders and investors. Only these metrics relate to the risk associated with payment triggers for payor (outcomes funders) and payee (investors and, sometimes, service providers).

FIGURE 2: LEVELS OF DATA COLLECTION WITHIN IMPACT BONDS

Breadth of focus

Management Information — Used by service providers and investors to inform and improve service delivery to maximise Contractual Outcomes Metrics within an Impact Bond

Contractual Outcomes Metrics — Align outcome funders, service providers and investors around common success criteria within an Impact Bond - used as basis for payments

Impact Evaluation for Learning — Informs broader community of policy makers, programme designers and evaluators - not essential for effective Impact Bonds

Management Information is collected and analysed at the discretion of investors and service providers with the intention of informing decisions around how the programme should be delivered and adapted to best meet varied and changing local needs. Such data are a tool for maximising the chances that Contractual Outcomes Metrics will be achieved. The contractual scope for such adaptation through rapid feedback loops, is a key strength of the Impact Bond model and, in turn, should drive the best possible social outcomes.[4]

Impact Evaluation for learning seeks to extract broader lessons around longer-term programme design, to inform public policy decisions and to contribute to a global evidence base. Such Impact Evaluation is desirable, but not necessary for effective Impact Bond implementation.[5]

Contractual Outcomes Metrics are the focus of the remainder of this paper.

---

[4] In some instances investors may contractually require government to make available relevant data to facilitate adaptive performance management. This is used as a mechanism for managing their risk.

[5] There is a broader debate in the literature around the extent to which Impact Evaluations offer transferable learning: "Lessons travel poorly...most social interventions have millions of design possibilities and outcomes depend on complex combinations between them." https://www.project-syndicate.org/commentary/evidence-based-policy-problems-by-ricardo-hausmann-2016-02#DJusjUGqB2oBv6iO.99

CASE STUDY

EDUCATION IMPACT BOND IN RAJASTHAN, INDIA

CONTRACTUAL OUTCOMES METRIC EVALUATION APPROACH –
EXPERIMENTAL DESIGN (RCT)

Contractual Outcome Metrics: Number of out of school girls enrolled and increase in learning gains

Investor: UBS Optimus Foundation

Outcome funder: The Children's Investment Fund Foundation

Implementer: Educate Girls

Evaluator: IDInsight

More info: http://instiglio.org/educategirlsdib/

This Development Impact Bond (DIB) was primarily set up as a 'learning DIB', to demonstrate that the structure is feasible and practical, and to generate lessons for future DIBs. As the implementer, Educate Girls is working to improve education for 18,000 children in Rajasthan, India.

The counterfactual risk for this Impact Bond is medium: learning outcomes and enrolment are unlikely to be greatly affected by external events and there is one independent RCT that shows positive impact of Educate Girls' intervention in a very similar context. The working group agreed to use an RCT design to evaluate outcomes as both the investor and outcome payer wanted to reduce counterfactual risk as much as possible given some of the other risks (e.g. dependence on government approvals) inherent in entering into a novel financial agreement. Payments are made for every additional girl enrolled in school and every additional gain in learning levels achieved over and above the control group. This was weighed against the relatively high cost of the evaluation itself and the cost to the implementer (and indeed all parties in the DIB) of coordination to ensure that the RCT was implemented robustly.

2 Choosing metrics based on objectives

PROGRAMME SCALE-UP VS. SERVICE INNOVATION

While some commentators regard Impact Bonds primarily as a mechanism for scaling-up evidence-based programmes with a high level of 'fidelity',[6] many more have focused on their potential to drive innovation in programme design and delivery through ongoing service adaptation and improvement.[7,8] In the US, much attention is currently focused on the scale-up of evidence-based programmes, but such Impact Bonds may also incorporate innovative elements – like novel referral pipelines or institutional relationships – while maintaining fidelity to a core intervention approach.

The objectives of outcome funders in commissioning an Impact Bond are a key determinant of what an appropriate evaluation approach would look like.[9] The less contractual control investors and service providers have over what services are delivered and how, the more appropriate it will be for Contractual Outcomes Metrics to relate to the quality of service delivery rather than ultimate end user outcomes.

Take, for example, Impact Bonds that primarily seek to roll-out an existing evidence-based intervention - for instance, to improve access to and usage of bed nets to prevent the transmission of malaria. Extensive evidence already exists that correct usage of bed nets significantly reduces the chances of contracting malaria – bed nets work – it would therefore be unnecessary to evaluate this premise again. The key Contractual Outcome Metrics in this case could instead focus around bed net coverage (how many households have

---

6   "Fidelity may be defined as the extent to which delivery of an intervention adheres to the protocol or program model originally developed." Mowbray et al. (2003), American Journal of Evaluation.

7   http://www.huffingtonpost.com/john-roman-phd/governments-big-problem-a_b_8223184.html

8   http://www.brookings.edu/research/reports/2015/07/social-impact-bonds-potential-limitations

9   Other important considerations include the power (sample size, baseline probability of outcomes being achieved, etc.) and practicality (ability to collect data, cost of data collection, etc.) of an evaluation approach.

a functional bed net) and usage (how many households are regularly and appropriately using their bed net).[10] Where appropriate, use of such metrics may improve the cost-effectiveness of Impact Bonds by enabling earlier payments from outcomes funders and requiring both less effort and cost to be expended in measuring Contractual Outcomes Metrics.

However, if the intervention approach is tightly defined, but it is not possible to accurately measure the quality of service implementation, it may be necessary to consider using a broader outcome metric (e.g. changes in malaria rates). In such circumstances, the contractual risk would increase as a number of factors, beyond simply distribution of bed nets, will affect whether this outcome is achieved. This may mean that investors and service providers want to broaden the range of interventions to manage their risk.

Contrast this with an Impact Bond that aims to encourage service innovation to deliver an outcome that has been historically difficult to achieve and is sensitive to exogenous shocks – for instance, supporting the long-term unemployed into work. The intentionally undefined nature of the intervention approach would necessitate an evaluative focus on the ultimate impact on the target group. In this case access to and maintenance of employment.

TARGETED OUTCOME IMPROVEMENT VS. BUILDING A GLOBAL EVIDENCE BASE

The purpose of Contractual Outcomes Metrics is to determine whether payments should be made in relation to a particular Impact Bond contract. Impact Bonds aim to measurably improve the lives of the people most in need by driving resources towards better, more effective programmes.

Attempts to use such metrics to inform a 'global body of knowledge' around what works may confuse or dilute the incentives within a given

---

10 It should be noted that bed net coverage alone may not be an appropriate metric in this context due to, potentially significant, local differences in knowledge about proper bed net usage between one community and another. It may sometimes be necessary for evaluators to validate proxy metrics for evidence-based interventions before finalising Impact Bond contracts to ensure that the theory of change holds up in a new context and no additional metrics are required.

Impact Bond contract increasing risk for both outcomes funders and investors.[11]

Such broader lessons are best learned through Impact Evaluations for Learning, which have no contractual bearing, and should be expressly designed not to compromise effective implementation and adaptation of Impact Bond funded services. To contractually restrict Impact Bond service delivery in order to facilitate an evaluation is to miss out on much of the value of outcomes-based contracts, and to unnecessarily increase the contractual risk borne by outcome funders and investors.

---

[11] Social Finance recommends identifying the smallest number of Contractual Outcomes Metrics that create the right service provider incentives. For instance, an Impact Bond targeting a reduction in reoffending rates may have only a single Contractual Outcomes Metric, against which payments are made, to ensure that the focus of service providers and investors is clear. To build the evidence base around what works, the policy community may nevertheless want to evaluate what influenced reductions in reoffending within an Impact Bond (e.g. access to housing and benefits, family reconnection, substance misuse counselling, etc.). To trigger contractual payments against such metrics, however, would presume prior knowledge about the detailed needs of the target population in terms of the intervention mix they require that is unlikely to be available. Such metrics would also risk distracting service providers from their core objective to reduce reoffending and instead place the focus on the way in which services are delivered negating much of the value of an outcomes-based approach.

CASE STUDY

REOFFENDING IMPACT BOND IN PETERBOROUGH, UK

CONTRACTUAL OUTCOMES METRIC EVALUATION APPROACH – LIVE, BUT NON-EXPERIMENTAL COUNTERFACTUAL

Contractual Outcome Metrics: % reduction in the reconviction rates of short-sentenced male prisoners leaving HMP Peterborough

Investors: Barrow Cadbury Foundation, Esmée Fairbairn Foundation, Friends Provident Foundation, Panahpur Charitable Trust, the Tudor Trust, the Rockefeller Foundation, Lankelly Chase and others

Outcome funders: Ministry of Justice and Big Lottery Fund UK

Implementer: The One* Service (a 6 provider consortium)

Evaluators: Qinetiq and the University of Leicester, and NIESR

More info: https://data.gov.uk/sib_knowledge_box/ministry-justice-offenders-released-peterborough-prison

The Peterborough Social Impact Bond was set up in 2010 and is an outcomes based model. Social investors provided £5m to fund interventions to reduce reoffending among three cohorts of 1000 short-sentenced male prisoners leaving Peterborough prison.* Prisoner engagement with the SIB-funded programmes was voluntary.

The Counterfactual Risk for this Impact Bond was judged to be high – reconviction rates can be affected by a number of factors both local and national, from changes in social policy, to changes in policing levels, probation policy and eligibility for benefits. As a result, a live counterfactual was judged to be necessary, but it was considered neither ethical nor practical to randomly assign offenders from Peterborough Prison to intervention and control groups in the way that a fully experimental measurement approach would require.

Instead, outcomes are measured across cohorts of up to 1,000 individuals. The attribution mechanism is a matched control group, against whom the number of reconviction events is compared over a 12 month period following release. The control group is selected from the Police National Computer using Propensity Score Matching. For every one individual worked with in Peterborough, up to ten statistically similar individuals are selected for the matched control group. Individuals can access the Peterborough intervention for 12 months post-release.

---

* Due to the introduction of a national rehabilitation programme, the Peterborough Social Impact Bond only worked with 2,000 offenders.

3 Introducing the concept of Counterfactual Risk

When establishing an Impact Bond, both the investor and outcome funder take on a certain amount of risk which is typically reflected in the pricing of outcomes and expected investor returns. Investors take the risk that the interventions they fund will deliver improvements in Contractual Outcomes Metrics. Outcome funders take the risk that Contractual Outcomes Metrics would not have been achieved in the absence of investor-funded interventions.

The definition and evaluation approach for Contractual Outcomes Metrics is of central importance in managing this risk for both parties.

For example, a back-to-work scheme may have an impressive number of clients finding work but in reality this could be due to a surge in economic growth. In this case an outcome payer may feel that they are overpaying for outcomes.

Conversely, a bad harvest may lead to a sharp increase in malnutrition and an increase in neonatal mortality despite the efforts of a programme to train community midwives in safe childbirth practices. In this case an investor may feel that they are being underpaid for outcomes.

This is the 'Counterfactual Risk' – the risk generated by not having an accurate assessment of what would have happened in the absence of an intervention. If we do not know what would have happened in the absence of a service then outcomes might be over- or under-valued. The table overleaf outlines the different factors affecting Counterfactual Risk.

TABLE 1: FACTORS AFFECTING COUNTERFACTUAL RISK

Counterfactual Risk factor	High Risk	Low risk
Availability of historical data on outcome	Little or no baseline data on outcome and low understanding of variability across populations and over time (e.g. prevalence of child slavery).	Good understanding of the outcome in the relevant target group and how this outcome has fluctuated over time (e.g. rates of HIV infection across sub-Saharan Africa).
Dependence of outcome on external events	Outcome likely to be highly affected by external events such as recession, extreme weather or policy change (e.g. employment or malnutrition).	Outcome closely linked to service provision (e.g. literacy programme).
Strength of evidence base for target group in relevant context[12] (if intervention approach is contractually defined)	No previous impact evaluation, link between interventions targeted at the outcome and population of interest unquantified in the context (e.g. privatisation of schooling).	Quantified link between interventions targeted at the outcome and population of interest well established in context (e.g. nutritional supplements to infants).
Scale of service provision or social issue	Small scale (e.g. city or county level) or low incidence event (e.g. teen pregnancy or neonatal mortality), outcomes data subject to significant noise.	Large scale (e.g. national or regional level), outcomes detectable at a population level.
Duration of Impact Bond	A long time period (e.g. 5 years) over which context is likely to change.	Short time period (e.g. 1-2 years) over which context is likely to remain stable.

12 There is evidence that context is as important for the strength of evidence. Pritchett and Sandefur (2013) show that, in the education literature, the impact of a programme is better predicted by estimates from non-experimental studies in the same context than by estimates from RCTs in different contexts. http://www.cgdev. org/publication/context-matters-size-why-external-validity-claims-and-development-practice-dont-mix

These factors can be considered together using the following diagram. Counterfactual Risk will be amplified for bigger and longer contracts.

FIGURE 3: ASSESSING COUNTERFACTUAL RISK

---
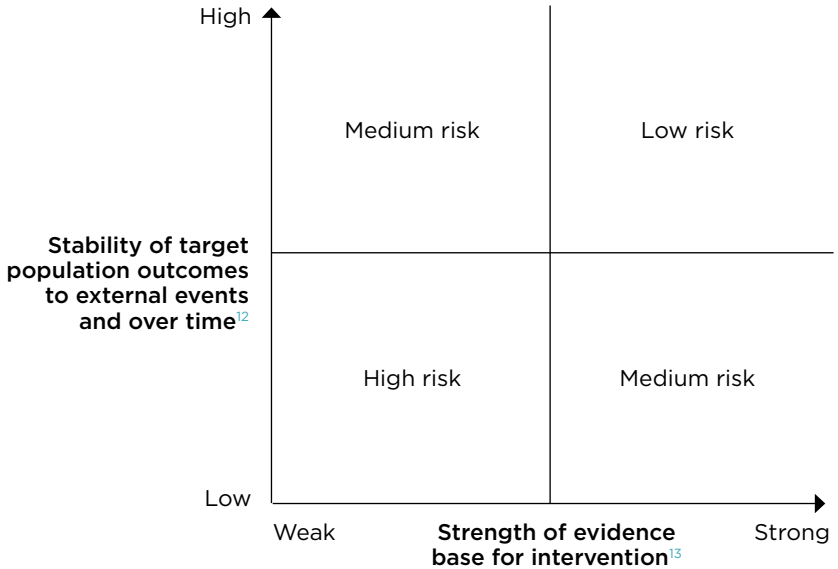
13 Proximity of this evidence base to the target context and population is a key consideration here. It is unlikely there will ever be a perfect match, but the closer the fit between the historic evidence and planned intervention context / population the stronger the evidence base can be considered to be for the purposes of Counterfactual Risk assessment.

CASE STUDY

MULTI-SYSTEMIC THERAPY IMPACT BOND IN ESSEX COUNTY, UK

CONTRACTUAL OUTCOMES METRIC EVALUATION APPROACH – CONSTRUCTED COUNTERFACTUAL

Contractual Outcome Metrics: average number of care days saved relative to a historical comparison group

Investors: Big Society Capital, Bridges Ventures and others

Outcome funder: Essex County Council

Implementer: Action for Children

Evaluator: OPM

More info: https://data.gov.uk/sib_knowledge_box/essex-county-council-children-risk-going-care

This Impact Bond raised £3.1m from investors to enable Essex County Council to provide Multi-Systemic Therapy (MST) to 380 young people at risk of entering care and their families over a 5.5 year service delivery period. MST is an evidence-based programme that delivers family therapy in the home through highly qualified therapists.

As analysis of historical case file data revealed that the likelihood of at risk adolescents entering care is relatively unaffected by external events, and MST has a strong evidence base with similar populations, the Counterfactual Risk was judged to be medium. A live counterfactual was judged to be both ethically complicated and unnecessary given the historical stability of outcomes for the target group. The comparison review figure is the proportion of individuals in the comparison group entering care over a 30-month period, which is then applied to the average number of days spent in care during the life of a child admitted into care, generating the number of care days against which the intervention population is compared. The cohort is tracked for outcomes for a comparable 2.5 years (giving a total programme length of 8 years). The comparison review figure was established prior to the signing of contracts, based on historical case file data.

4 Mitigating Counterfactual Risk

Counterfactual risk can be managed through appropriate evaluation of Contractual Outcomes Metrics. Often the most important design component is the methodology. Table 2 outlines how more rigorous research design can reduce Counterfactual Risk.[14,15]

There will often be a trade-off between the extent to which Contractual Outcomes Metric evaluation can reduce Counterfactual Risk and the cost. Investing more resources can also give more accurate impact estimates, and thus reduce Counterfactual Risk, by increasing sample sizes, increasing the frequency of data collection and using third parties to collect data rather than relying on audited Management Information.

If Counterfactual Risk is high, for example an untested employment service, then it makes sense to invest in an approach to outcome measurement that will a) give the outcomes funder confidence that they are not paying for outcomes that would have been achieved anyway; and b) give the investor confidence that they are not being underpaid for outcomes that were more difficult to achieve than expected due to external events. However, if Counterfactual Risk is low, for example roll-out of a literacy programme that has been well tested with the target group, then the additional cost of more robust approach to outcome measurement may not be worth the corresponding reduction in risk.

---

14 The costs provided cover a broad range and are relative due to the multitude of other factors affecting cost that may not be within the control of the parties to an Impact Bond. For example, whether outcomes can be assessed via existing data sets, the need for household surveys, or the need to collect biometric information.

15 A brief description of common evaluation methodologies can be found in Appendix 1.

TABLE 2: CONTRACTUAL OUTCOMES METRIC MEASUREMENT METHODOLOGIES

Outcome measurement methodology	**Experimental design that controls for both observed and unobserved variables (e.g. randomised controlled trial, regression discontinuity design)**
Impact on Counterfactual Risk	Large. Well specified counterfactual that controls for all variables
Relevant level of Counterfactual Risk	High
Typical costs	£££
Other considerations	Prevents population level interventions as some groups must, by definition, be denied services to form a control Complex design so more difficult to ensure that the research is executed as planned

Outcome measurement methodology	**Live but non-experimental counterfactual (e.g. purposively selected control geographies)**
Impact on Counterfactual Risk	Medium. Impact of external events (e.g. economic shocks) and measurable variables (e.g. gender and income) accounted for.
Relevant level of Counterfactual Risk	Medium
Typical costs	££–£££
Other considerations	Prevents population level interventions as some groups must, by definition, be denied services to form a control. Can be difficult to maintain a pure control.

Outcome measurement methodology	'Constructed' counterfactual with no live control (e.g. comparison to an historical benchmark)
Impact on Counterfactual Risk	Medium. Only measurable variables accounted for and no calibration for real-time impact of external events
Relevant level of Counterfactual Risk	Low
Typical costs	£–££
Other considerations	Can be implemented at full scale

Can require good quality administrative data that may not be available |

Outcome measurement methodology	No counterfactual (e.g. pay per outcome achieved with no adjustment)
Impact on Counterfactual Risk	None
Relevant level of Counterfactual Risk	Very Low
Typical costs	£
Other considerations	Can be implemented at full scale (e.g. for nationwide programmes)

CASE STUDY

DEPARTMENT OF WORK & PENSIONS INNOVATION FUND IN THE UK

CONTRACTUAL OUTCOMES METRIC EVALUATION APPROACH – NO COUNTERFACTUAL*

Contractual Outcome Metrics: outcomes for gaining and sustaining future employment. Outcomes include re-engaging with education, such as addressing truancy and behavioural issues; gaining educational qualifications; and entering apprenticeships and employment.

Investors: Various across 10 Impact Bond contracts commissioned under the Fund

Outcome funder: Department of Work & Pensions, UK

Implementers: Various across 10 Impact Bond contracts commissioned under the Fund

Evaluator: National Centre for Social Research and Insite Research and Consulting

More info: https://data.gov.uk/sib_knowledge_box/department-work-and-pensions-innovation-fund

The Innovation Fund aimed to test new social investment and delivery models to support disadvantaged young people, and those at risk of disadvantage, aged 14 to 24 years. The Innovation Fund was commissioned over two rounds via an open competition. This resulted in ten Impact Bonds in total, testing a range of social investment and innovative delivery models.

The Department of Work and Pensions (DWP) specified a maximum amount they were willing to pay per outcome, which represented a proportion of the benefit savings associated with moving a disadvantaged young person into work. A list of payable outcomes was published in the specifications for each round. Bidders were invited to pick and mix from this list and work toward outcomes appropriate for their particular group of young people. Bidders also proposed discounts for the outcome payments. As no contractual counterfactual was defined for these contracts, there was no off-set of Counterfactual Risk for either DWP or Impact Bond investors.

*   DWP claims to have used a 'review of evidence and assessment of deadweight' to both determine maximum outcome payment values and to retrospectively assess value for money within individual Innovation Fund contracts. Such calculations are important, but as they have no bearing on which payments are or aren't triggered within a given contract they do not serve to offset counterfactual risk in the sense that it is described in this paper. On this basis we classify such contracts as having 'no counterfactual'.

5 A Question of Balance

To ensure that Impact Bonds offer cost-effectiveness as well as strong incentives for greater impact, it is important to keep evaluation costs in proportion to Impact Bond outcomes payments. The framework presented in this paper is intended to support both investors and outcome funders to assess the level of Counterfactual Risk inherent in a particular transaction. This should enable informed decisions around the most appropriate approach to evaluating Contractual Outcome Metrics.

The objectives of outcome funders in commissioning an Impact Bond should be the ultimate determinant of what Contractual Outcomes Metrics look like. However, this needs to be balanced against the Counterfactual Risk that investors are asked to bear. If Contractual Outcomes Metrics, and the related evaluation approach, set too high a threshold for payments to be made, investment may be hard to come by and expensive, undermining the value-for-money that should come from creating a contractual focus on outcomes and enabling an adaptive implementation approach. Outcomes funders will need to balance the effort and money spent on Impact Bond evaluation with the additional investment cost that they may be asked to bear if the Counterfactual Risk to investors is too great.

Contractual Outcomes Metrics should, ideally, be limited to the smallest number of payment triggers needed to incentivise the right service provider behaviour. Their measurement should balance the Counterfactual Risk between outcomes funders and investors. The focus should be on effective contractual measurement not broader evaluation to inform a 'global knowledge base'. To get maximum value out of Impact Bond contracts, care should be taken to ensure that the outcome evaluation methodology neither restricts nor interferes with the implementation of investor-funded services.

Evaluation for the purposes of determining payment in Impact Bonds is high stakes - this paper outlines a framework for thinking about which contractual measurement approaches might be relevant and proportionate in different contexts. We acknowledge, however, that practical considerations – including target population size, regulatory constraints and ethical considerations – will also need to be taken into account.

Appendix 1: Overview of Common Evaluation Methodologies

Methodology	Description
Randomised Control Trial (RCT)	An *impact evaluation* design in which *random assignment* has been used to allocate the *intervention* amongst members of the *eligible population*. Since there should be no correlation between *participant* characteristics and the *outcome*, any differences in *outcome* between the treatment and control can be fully attributed to the intervention, i.e. there is no *selection bias*. However, RCTs may be subject to several types of *bias* and so need follow strict protocols. Also called *Experimental Design*.*
Regression Discontinuity Design	An *impact evaluation* design in which the *treatment* and *comparison* groups are identified as being those just either side of some threshold value of a variable. This variable may be a score or observed characteristic (e.g. age or land holding) used by program staff in determining the *eligible population*, or it may be a variable found to distinguish *participants* from non-participants through data analysis. Regression Discontinuity Design is an example of a *quasi-experimental design*.[14]
Purposively Selected Controls	An *evaluation* design in which the treatment and comparison groups are not assigned at random but are selected individually. This introduces bias. As treatment groups may be selected based on political considerations unobservable factors cannot be controlled for. This is not an experimental or quasi-experimental design.

---

16 International Initiative for Impact Evaluation (3ie) 2012. Impact Evaluation Glossary. New Delhi: 3ie

Methodology	Description
Constructed counterfactual (e.g. historical control)	An *evaluation design* in which there is no comparison group tracked over the course of the evaluation. The impact on the treatment group is compared to outcomes for a group tracked in another data set (e.g. comparing employment outcomes to baseline levels of employment in the general population, or against a historical data set).
Individual tariffs	A contractual design in which payment is made for each outcome achieved (e.g. for every person getting a job or girl enrolling in school) without any reference to a comparison group.

Appendix 2: Other Useful Terms[17]

Attribution
The extent to which the observed change in outcome is the result of the intervention, having allowed for all other factors which may also affect the outcome(s) of interest.

Baseline survey and baseline data
A survey to collect data prior to the start of the intervention. Baseline data are necessary to conduct double difference analysis, and should be collected from both treatment and comparison groups.

Comparison Group
A group of individuals whose characteristics are similar to those of the treatment groups (or participants) but who do not receive the intervention. Under trial conditions in which the evaluator can ensure that no confounding factors affect the comparison group it is called a control group.

Confounding factors
Factors (variables) other than the programme which affect the outcome of interest.

Control Group
A special case of the comparison group, in which the evaluator can control the environment and so limit confounding factors.

Counterfactual
The state of the world in the absence of the intervention. For most impact evaluations the counterfactual is the value of the outcome for the treatment group in the absence of the intervention. However, studies should also pay attention to unintended outcomes, including effects on non-beneficiaries.

External Validity
The extent to which the results of the impact evaluation apply to another time or place.

---

17 International Initiative for Impact Evaluation (3ie) 2012. Impact Evaluation Glossary. New Delhi: 3ie

Impact evaluation
A study of the attribution of changes in the outcome to the intervention. Impact evaluations have either an experimental or quasi-experimental design.

Quasi-Experimental Design
Impact evaluation designs used to determine impact in the absence of a control group from an experimental design. Many quasi-experimental methods, e.g. propensity score matching and regression discontinuity design, create a comparison group using statistical procedures.
The intention is to ensure that the characteristics of the treatment and comparison groups are identical in all respects, other than the intervention, as would be the case from an experimental design. Other, regression-based approaches, have an implicit counterfactual, controlling for selection bias and other confounding factors through statistical procedures.

Random assignment
An intervention design in which members of the eligible population are assigned at random to either the treatment group or the control group (i.e. random assignment). That is, whether someone is in the treatment or control group is solely a matter of chance, and not a function of any of their characteristics (either observed or unobserved).

Treatment group
The group of people, firms, facilities or whatever who receive the intervention. Also called participants.

Unobservables
Characteristics which cannot be observed or measured. The presence of unobservables can cause selection bias in quasi-experimental designs, if these unobservables are correlated with both participation in the programme and the outcome(s) of interest.

ABOUT THE AUTHORS

Louise Savell

Louise jointly leads Social Finance's international development team – designing, developing and delivering payment-by-results and social investment structures for a range of clients including DFID, USAID, Grand Challenges Canada, the Global Innovation Fund and the Shell Foundation, in both sub-Saharan Africa and South Asia. In the UK, Louise has led high impact projects on financial inclusion, homelessness, health and regulatory reform. Before joining Social Finance, Louise headed the Eastern European programmes of ARK where she worked with government and service providers to accelerate the reform of child welfare systems.

Lucy Heady

Lucy has over ten years of experience of measuring social impact in both the UK and internationally. In June 2016 Lucy moved to NESTA Impact Investments as their Impact Director. Prior to this, she worked as Evidence, Measurement and Evaluation Manager at the Children's Investment Fund Foundation (CIFF) where she designed and managed a variety of evaluations of public and private sector initiatives working with partners including the World Bank, Government of Kenya and Sesame Workshop. Lucy was also responsible for driving an evidence-based approach within the Education Team. Most recently she set up and managed the outcome evaluation for the Educate Girls Development Impact Bond. Before moving to CIFF, Lucy was the Senior Economist at Building Markets where she measured job creation and private sector growth in Afghanistan, Haiti and Liberia. Prior to this she was Head of Measurement at NPC, working with charities to measure their impact in a variety of sectors from crime prevention to violence against women.

ABOUT THE PARTNERS

Social Finance
Social Finance is a not for profit organisation that partners with governments and NGOs, social sector and financial community to find better ways of tackling social problems in the UK and beyond. Since its formation in 2007, Social Finance has mobilised over £100 million of investment and helped to design a series of programmes to improve outcomes for vulnerable individuals. Social Finance is the originator of the Social Impact Bond model.

In the UK, work includes support for 2,000 short sentence offenders released from Peterborough Prison, 380 children on the edge of care in Essex, 2,500 young people at risk of becoming NEET and 1,400 homeless youth and rough sleepers. Internationally, Social Finance is working with the Global Fund, World Bank, Grand Challenges Canada, the Inter-American Development Bank, USAID, DfID and others on Development Impact Bonds and other innovative outcomes-based financing mechanisms to address challenges in low and middle income countries.

www.socialfinance.org.uk

CIFF
The Children's Investment Fund Foundation is an independent philanthropic organisation, headquartered in London with offices in Nairobi and New Delhi. CIFF was set up in January 2004 by Chris Hohn and Jamie Cooper, to demonstrably improve the lives of children living in poverty in developing countries through strategies that have lasting impact. With a rigorous business-like approach to philanthropic funding, the foundation has been focused on clear returns for children from the outset. The emphasis on data to measure impact has been widely adopted by other development funders.

CIFF works with a wide range of partners seeking to transform the lives of poor and vulnerable children in developing countries. This involves supporting bold ideas to seemingly intractable challenges for children. Areas of work include children and mothers' health and nutrition,

children's education, deworming and welfare, and smart ways to slow down and stop climate change.

With an endowment of over $4 billion and a commitment to urgent transformational change for children, the foundation's independent board, professional staff, and partners are constantly seeking ways to increase impact at scale.

https://ciff.org/

Social Finance is a not-for-profit organisation that partners with government, the social sector and impact investors to find better ways of tackling social problems and improving the lives of people in need.

Social Finance Ltd
131–151 Great Titchfield Street
London, W1W 5BB

T +44 (0)20 7667 6370
info@socialfinance.org.uk
www.socialfinance.org.uk

Social Finance is authorised and regulated by the Financial Conduct Authority
FCA No: 497568

© Social Finance 2016