# *Current barriers to and facilitators of undertaking rigorous impact evaluations of Social Impact Bonds (SIBs) – a stakeholder-informed systematic review and thematic synthesis*




Thesis, MSc. in Evidence-Based Social Intervention and Policy Evaluation
Department of Social Policy & Intervention, University of Oxford (2016/2017)

Candidate Number: 1011230

Word count: 10,000

*(Excluding Abstract, Figures, Tables, Reference List and Appendices)*

This work has been reviewed and approved by the Departmental Research Ethics Committee (DREC) at the Department of Social Policy and Intervention, University of Oxford

# Table of Contents

ABSTRACT....4
CHAPTER 1 – BACKGROUND....6
1.1 Introduction....6
1.2 Payment-by-results (PbR)....7
1.3 Social Impact Bonds (SIBs)....7
1.3.1 Growth of the SIB model....9
1.3.2 'Rigorous' evaluation in SIBs....11
1.4 Rationale for Review....14
1.4.1 State of the evidence....14
1.4.2 What this thesis will add....15
1.4.3 Objectives and Research Question....15
CHAPTER 2 – METHODS....17
2.1 Introduction....17
2.2 Stakeholder involvement....17
2.3 Inclusion and exclusion criteria....19
2.4 Search strategy....20
2.4.1 Standard Search Strategy....20
2.4.2 Database search....21
2.4.3 Grey literature....23
2.4.4 Hand searching....25
2.4.5 Snow balling....26
2.4.6 Contacting experts....26
2.5 Data collection and analysis....26
2.5.1 Study selection....26
2.5.2 Data extraction....26
2.5.3 Quality appraisal....26
2.5.4 Analysis....27
2.5.5 Assessing the robustness of the synthesis....28
CHAPTER 3 – RESULTS....29
3.1 Search Results....29
3.2 Characteristics of included studies....30
3.3 Quality Appraisal....38
3.4 Excluded Studies....40
3.5 What factors affect rigorous evaluation in SIBs?....40
3.5.1 Descriptive themes....40
3.5.2 Sensitivity analysis....46

3.5.3 Contribution of individual studies ..... 48
3.5.4 Analytical themes ..... 50
CHAPTER 4 – DISCUSSION ..... 53
4.1 Perspectives of panel members ..... 53
4.2 Reflexivity ..... 54
4.3 Strengths and weaknesses of this study ..... 54
4.4 Strengths and weaknesses of this study in relation to other studies ..... 55
4.5 Implications for practice ..... 56
4.6 Implications for future research ..... 57
CHAPTER 5 – CONCLUSION ..... 58
REFERENCES ..... 59
Included studies ..... 59
Excluded studies ..... 61
Additional references ..... 69
Appendix 1: Expert Advisory Panel – Participant Information Sheet ..... 75
Appendix 2: Experts contacted and sample email ..... 78
Appendix 3: Data Extraction Form ..... 80
Appendix 4: Excluded studies and reason for exclusion ..... 83
Appendix 5: Summary of CERQual Assessment Ratings (per finding) ..... 87
Appendix 6: PRISMA Checklist ..... 91

## ABSTRACT

**Background:** Social Impact Bonds (SIBs) (an investor backed form of payment-by-results), have attracted considerable political attention and capital since 2010. In the May 2010 General Election, the opposition Conservative Party included SIBs in their manifesto "Big Society, Not Big Government". They promoted the use of SIBs as a way to achieve value for money and innovation, announcing that they would establish a social investment bank (Big Society Capital) to invest in innovative products such as SIBs (Conservative Party 2010). Since then, 89 SIBs have been launched in 20 countries, with a total investment of over 322 million USD (Social Finance 2017). Despite the political popularity, evaluation practices have differed significantly. In an initial scoping study, the reviewer identified that few SIBs have been subject to detailed evaluation. Many SIBs have not monitored performance against a robust counterfactual and there is a lack of consistency in the processes through which outcome payments are decided.

**Objectives:** This review aimed to identify the factors that act as barriers to or facilitators of conducting or being involved in rigorous impact evaluations of the efficacy of individual (SIB-funded) programmes. Evaluation of the efficacy of SIB commissioning versus other types of commissioning was outside the scope of this review.

**Methods:** To identify empirical literature that met the inclusion criteria, the reviewer searched 13 electronic databases, 91 grey literature sources and hand searched 33 academic articles. 92 experts were contacted, to identify additional articles. To enhance the usefulness of this research and minimise biases, this review was informed by an Expert Advisory Panel (10 experts). Prior to establishing the Panel, the project was accepted for ethical approval (CUREC_1A). Included studies needed to be primary research (any research design) or systematic reviews that focus on the design and/or implementation of SIBs and investigate barriers to or facilitators of undertaking or being involved in rigorous impact evaluations of SIBs (as identified by practitioners). Included studies were subject to quality appraisal. To generate new knowledge, a thematic synthesis of qualitative research was conducted. Barriers and facilitators were categorised into descriptive and analytical themes, which were generated inductively.

**Results:** 5,567 studies were retrieved for title and abstract review. 5,448 studies were excluded; 119 were retrieved for full text review. Of these studies, 25 were included and 94 were excluded. Included studies reported a range of factors which act as barriers to and/or facilitators of undertaking or being involved in rigorous impact evaluations of SIBs. The most frequently reported barriers included: 1) attribution challenges (e.g small sample sizes, no theory of change, overlap of services provided etc); 2) challenges in identifying appropriate outcome metrics; 3) lack of access to required data; 4) lack of financial resources; and 5) differing stakeholder incentives, views and needs. The most frequently reported facilitators included: 1) access to expertise in evaluation, design and data management; 2) access to required data; 3) availability of comparable counterfactual/control group; 4) early and ongoing stakeholder consultation to refine and test outcome metrics; and 5) regular monitoring and performance management.

# CHAPTER 1 – BACKGROUND

## 1.1 Introduction

SIBs have emerged at the intersection of three broad trends, related to the changing nature of public sector commissioning. Firstly, since the 1990's, the 'New Public Management' (NPM) approach has led to the growth of quasi-marketised models of service delivery, where public service provision is largely outsourced in order to increase efficiency and value for money (Warner 2013). Secondly, the voluntary sector's involvement in delivering public services has rapidly expanded and there has been a trend toward the professionalisation of the voluntary sector (Charity Commission 2012; Davison 2013). Thirdly, post economic crisis (2008) policies have largely focused on reduced public spending, associated with deficit and debt reduction (Taylor-Gooby & Stoker 2011). As SIBs have emerged at the intersection of these trends, they have been presented as a tool for: 1) transforming public service delivery; 2) supporting professional social sector organisations; and 3) reducing budgetary pressures.

To understand the impact of SIBs, it is important to rigorously evaluate their effectiveness. There are two broad levels of consideration when evaluating SIBs: 1) individual scheme level (i.e efficacy of individual SIB-funded programmes); and 2) delivery mechanism level (i.e efficacy of SIB commissioning model versus other types of commissioning). This paper is focused on the first consideration (individual scheme level evaluations), noting the importance of adopting consistent and reliable assessments of whether social outcomes have been achieved and thus whether payments to investors should be triggered. Evidently, focusing on the accuracy and reliability of measures used to decide whether private investors earn a return is important for transparency and public accountability.

***Table 1: Evaluation levels of consideration***

| Level | Consideration |
|---|---|
| Individual scheme level | Efficacy of individual programmes or initiatives that have been funded using a SIB structure |
| Delivery mechanism level | Efficacy of SIB commissioning model versus other types of commissioning |

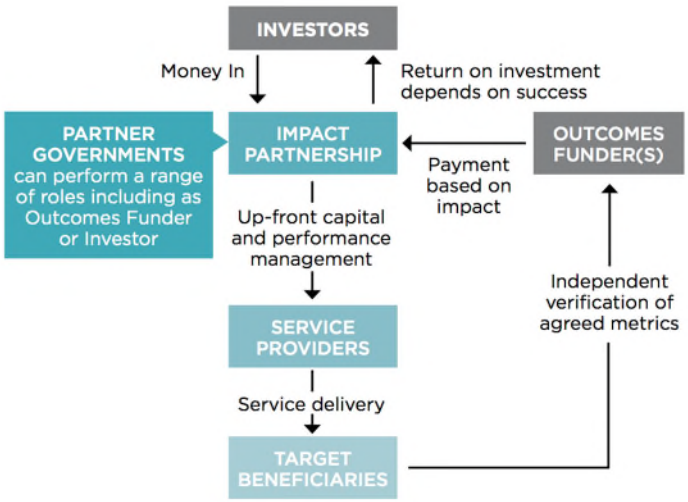
### 1.2 Payment-by-results (PbR)

Increasingly, many governments are using elements of PbR when commissioning services. The term PbR is most often used to refer to outcomes-based payment schemes where all or part of the payment depends on the service provider achieving outcomes specified by the commissioner (NAO 2015). While PbR is not a new mechanism, its usage has increased in recent years. The growth is often related to central government investment and part of wider reforms to public services (discussed earlier) (Lagarde et al., 2013). There are two key features related to the use of PbR in public service reform. Firstly, the focus is shifted away from inputs, outputs and processes, towards quantifiable social outcomes; these are linked to outcome payments (Edmiston & Nicholls 2017). Secondly, some risk is transferred as providers bear the risk of delivering services without reward if they fail to achieve outcomes. PbR is thus intended to increase the "*efficiency, effectiveness and equitability of public service*s" (Fox & Albertson 2011: 356–360). However, PbR is a technically complex form of contracting with many costs and risks. If executed poorly or used inappropriately, there may be negative impacts on value for money and/or quality of public services. For example, providers could focus efforts only on incentivised tasks or on clients most likely to achieve outcomes (NAO 2015).

### 1.3 Social Impact Bonds (SIBs)

The term "SIB" is highly contested in the literature and in practice. The initial applications of the SIB approach (e.g Peterborough Prison – UK) produced some ideas about what SIBs 'should' look like. However, SIBs have diverged significantly in practice since then. Broadly, SIBs are "the latest manifestation of the PbR policy priority" (Edmiston & Nicholls 2017: 2). Specifically, this paper defines a SIB as a form of PbR where the commissioner (often Government) pays for the outcomes achieved, rather than services delivered. The capital required to fund the intervention is obtained from external investors who earn a financial return if the intervention is successful (see Figure 1) (Ronicle et al., 2014).

***Figure 1: Typical SIB Structure***




*Source: Social Finance (2016)*

There are various features which distinguish SIBs from traditional PbR schemes; the main feature is the provision of upfront capital (Table 2).

***Table 2: Similarities and differences – SIBs & PbR***

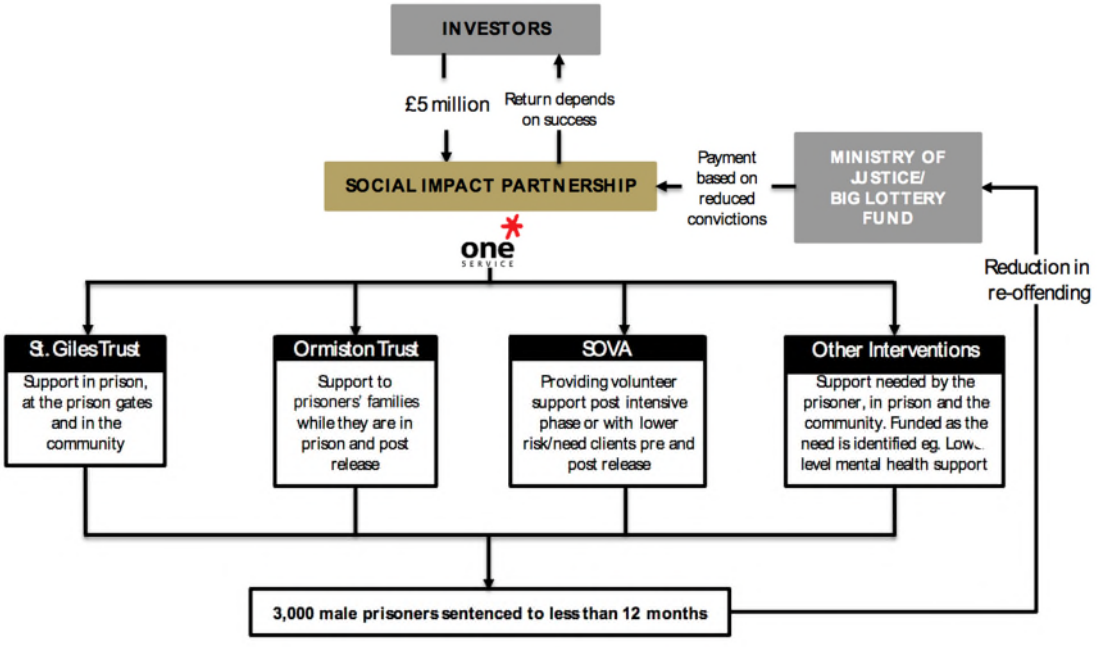
| *Feature* | *Social Impact Bonds* | *Payment-by-results* | *Traditional government debt* |
|---|---|---|---|
| Upfront capital is provided to cover service provider costs | ✓ | × | ✓ |
| Allows private investment into public services | ✓ | ✓ | × |
| Public sector only pays for success (delivery of outputs or outcomes, rather than inputs or processes) | ✓ | ✓ | × |
| Uses private sector rigour and performance management to enhance results | ✓ | ? | × |
| Risk of unsuccessful intervention is transferred away from public sector | ✓ | ✓ | × |
| Risk of unsuccessful intervention is transferred (partly) away from service provider/s | ✓ | × | × |
| Fosters cooperation between multiple service providers | ✓ | ? | × |
| Payments from social sector are proportional to improvement in social outcomes | ✓ | ? | × |
| Typically focused on delivery of human services (e.g compared to traditional infrastructure projects) | ✓ | ? | ? |
| Service interventions are preventative in nature | ✓ | ? | ? |
| Targeted towards specific populations with complex needs | ✓ | ? | ? |

SIB proponents hold that the introduction of private capital leads to: 1) innovation in service delivery; 2) improved social outcomes (e.g. improved rates of recidivism); 3) future cost savings; and 4) additionality (i.e. new flow of finance) (Edmiston & Nicholls 2017). Opponents note that SIBs may: encourage the use of an overly simplistic model of cause and effect; create incentives for service providers to shape service provision around contract terms rather than client needs; reduce government oversight and accountability; and/or lead to prioritisation of efficiency savings rather than enhanced services (Boyle 2011; Hudson et al., 2010; Joy & Shields 2013; McHugh et al., 2013).

### *1.3.1 Growth of the SIB model*

The number of SIB contracts is continuing to grow at a considerable rate (Fraser et al., 2016). The first SIB was launched in Peterborough (Eastern England) in 2010 (Figure 2). It aimed to reduce reoffending by short-sentenced prisoners released from Her Majesty's Prison Peterborough.

***Figure 2: Structure of Peterborough SIB***




*Source: Nicholls and Tomkinson (2013)*

Substantial attention and capital has been directed towards the development of SIBs since then (Ronicle et al., 2014). 89 SIBs have been launched in 20 countries, with a total investment of over 322 million USD (Social Finance 2017).

***Figure 3: SIB Development over time***

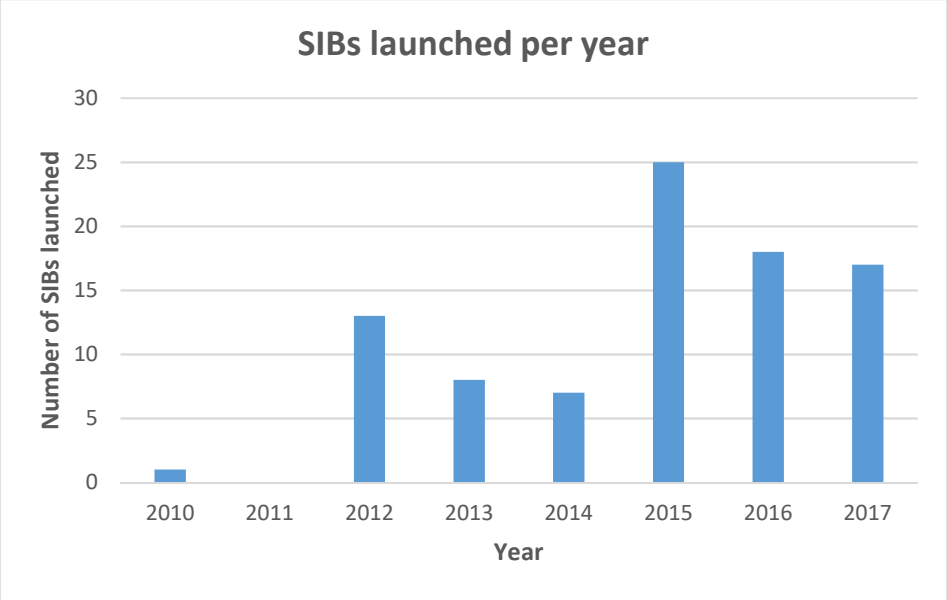
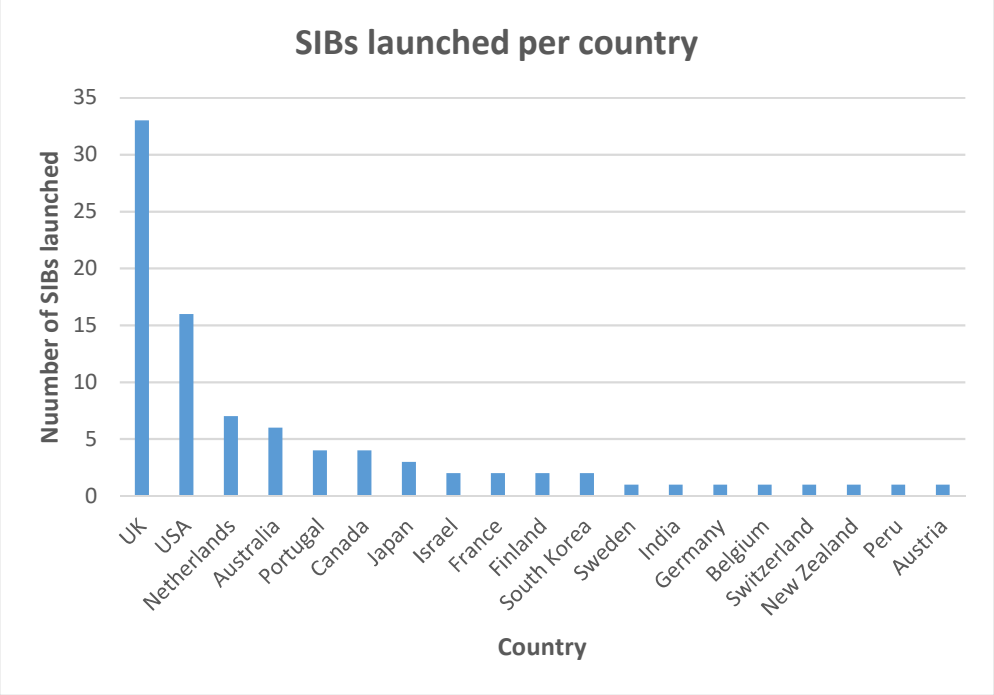
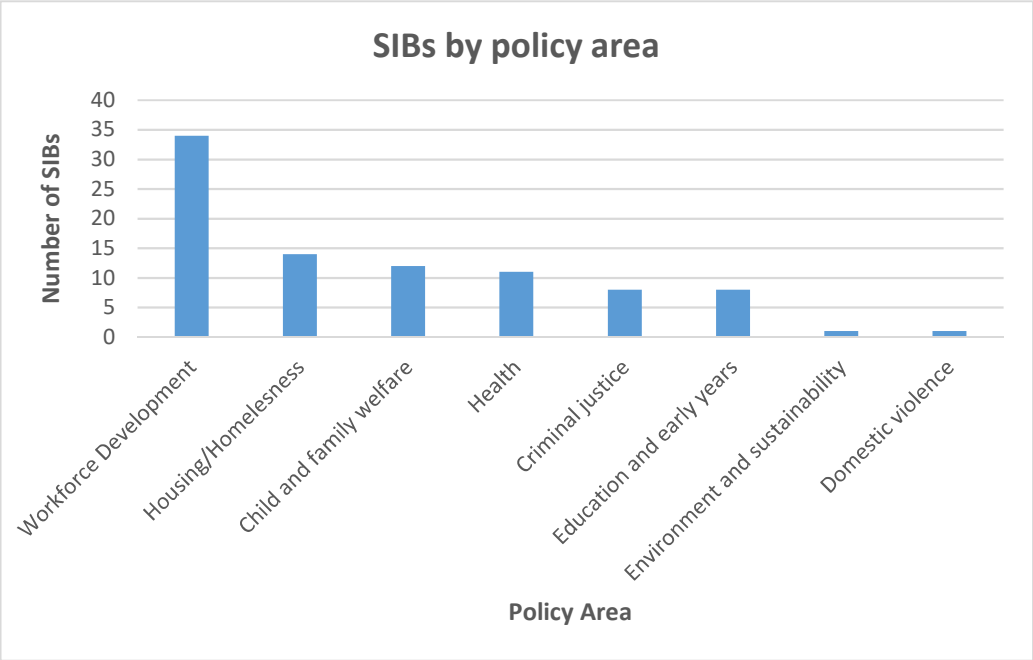



***Figure 4: SIBs by jurisdiction***



SIBs have been used to address various policy issues, including homelessness criminal justice, mental health, workforce development, health, child welfare, education, early years and environmental sustainability (Fraser et al., 2016).

***Figure 5: SIBs by Policy Area***

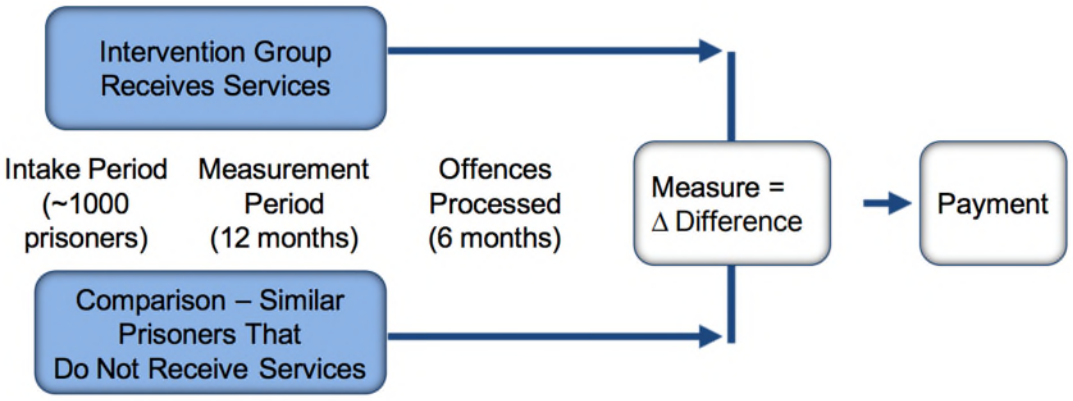



Despite increased interest, the evidence base regarding the efficacy of SIBs is limited (Edmiston & Nicholls 2017). Some preliminary, independent evaluations have suggested that SIBs are beginning to produce intended outcomes (e.g Gustafsson-Wright et al., 2015). However, cautionary narratives have highlighted the difficulties associated with this funding mechanism (discussed earlier) (Fraser et al., 2016).

### *1.3.2 'Rigorous' evaluation in SIBs*

Adopting a systematic approach for evaluating social interventions is important for overcoming sources of bias and understanding the 'true' impact of an intervention (Higgins et al., 2011). This is important in the context of SIBs as outcome payments are triggered by the achievement of social outcomes. Payments reward the positive difference that a SIB-funded intervention makes for a particular cohort. For example, in the Peterborough SIB, payments were triggered by comparing the outcomes of the treatment group and a similar group of prisoners that did not receive the intervention (Figure 6).

***Figure 6: Example of outcomes measurement and payment system (Peterborough SIB)***



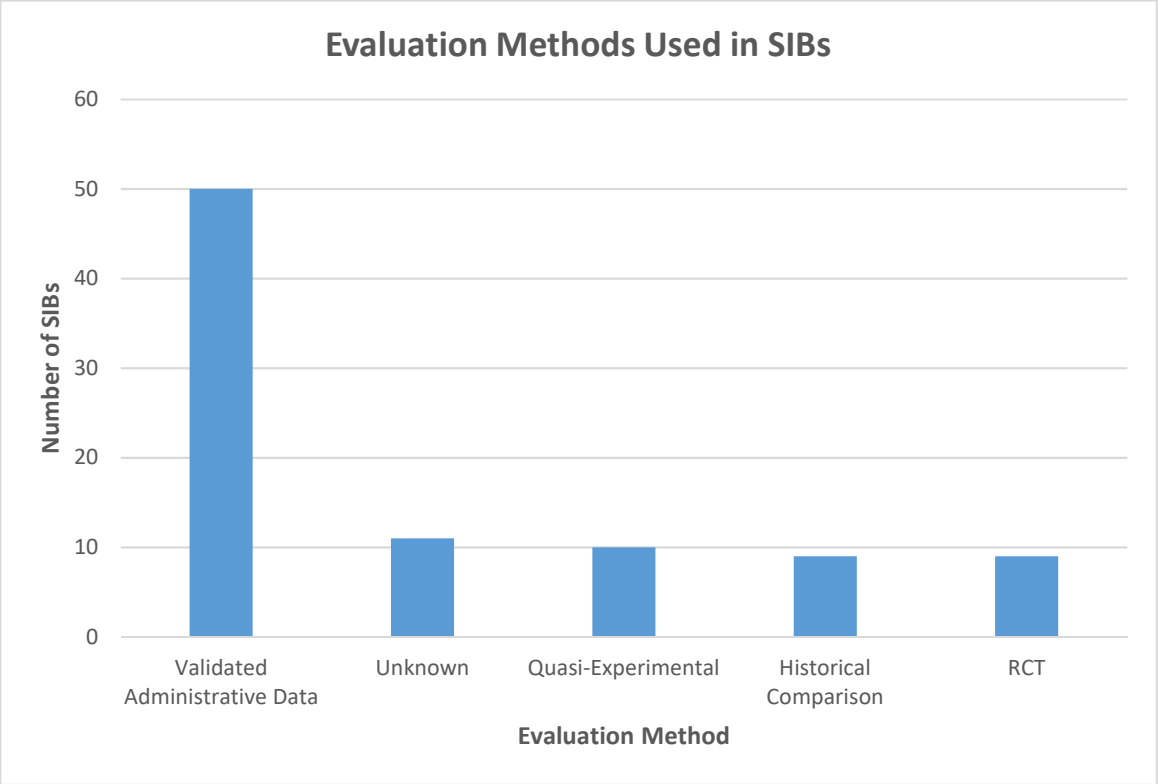

*Source: Nicholls and Tomkinson (2013)*

Without rigorous evaluation, outcomes could be over or under-valued. Payments could be triggered for outcomes that would have occurred without the intervention, thus undermining value for money. Examples of potential bias which can undermine attribution include publication bias, performance bias, selection bias, detection bias, reporting bias and attrition (Shadish et al., 2002). The method used to evaluate a SIB matters; "different methods might reveal different results" (Cabinet Office 2017). Weaker study designs (lower internal validity) are more likely to report positive outcomes and less likely to detect harm (Weisburd et al., 2001; Welsh et al., 2011). Thus, a critical aspect of rigorously evaluating research evidence, is the consideration of bias (Shadish et al., 1991).

Rigorous evaluation is also crucial as interventions may cause harm, even when properly implemented and based on sound program theory (Chalmers 2005; Petrosino et al., 2003; McCord 2003; Lilienfield 2007). More specifically, there is evidence of perverse effects in PbR contracts, such as providers neglecting non-incentivised activities (NAO 2015). The risk of harm is always present to some extent when intervening to address a social problem; thus, it may be unrealistic to follow the Hippocratic Oath of "first do no harm."—as in the medical sciences. A more appropriate stance is that harm should be avoided where possible; steps should be taken to identify harm early and to stop programs if they cause too much harm. To achieve this when delivering SIBs, it is important to use robust evaluation methods; regularly monitor outcomes and adhere to appropriate guidelines of best reporting practice (Schulz et al., 2010; Moher et al., 2010).

While the existing literature emphasises the importance of being able to measure attribution in a SIB, evaluation practices have differed significantly. The use of a strong and valid comparator has varied. The most common method for determining outcome payments has been validation of service provider or government administrative data (Figure 7) (Gustafsson-Wright et al., 2015). This means that client data is collected as verification of a particular outcome or outcomes (e.g number of days spent in foster care); outcome payments can be triggered using the client data (without a counterfactual).

***Figure 7: Evaluation Methods Used in SIBs***




Evaluation methods have differed significantly across jurisdictions, particularly USA and the UK. In USA, most SIBs have been evaluated using RCTs, potentially because robust evaluation is typically required for payment validation. In the UK, beneficiary outcomes are often measured without a counterfactual. For example, the IAAM Adoption Bond uses no comparator and assumes that none of the children in the cohort would have found a permanent home without the intervention. Therefore, the achievement of *any* positive outcome is understood to be an improvement (Cabinet Office 2017). Additionally, many UK

SIBs have used a ‘rate card’; this is a list of outcomes that the government is seeking to achieve and the price it is willing to pay per outcome (Figure 8).

***Figure 8: Rate Card Example***

| OUTCOME | PAYMENT PER INDIVIDUAL |
|---|---|
| Improved behavior at school (Measured by a letter from a teacher) | £800 |
| Stop persistent truancy (absent for over 10% of school days per year) | £1,300 |
| Achievement of First National Qualifications Framework (NQF) Level 2 qualification | £2,200 |
| Achievement of First NQF Level 1 qualification | £700 |
| Entry into first employment including a training element | £2,600 |
| Entry into sustained employment | £1,000 |
| Completion of first NQL Level 3 training/ vocational qualifications | £3,300 |
| Successful completion of an ESOL course | £1,200 |
| Entry into education at NQF level 4 | £2,000 |

*Source: Gustafsson-Wright et al (2015)*

Service providers may submit claims for outcome payments, using various forms of administrative data, either on a monthly or quarterly basis. Thus, robust evaluation is not necessarily required for payment validation. In light of the differing evaluation practices, this study will provide integral insights into the barriers and facilitators experienced by practitioners, when evaluating the efficacy of individual programmes funded by SIBs.

## 1.4 Rationale for Review

### *1.4.1 State of the evidence*

There is little systematic evidence focused on the evaluation practices and methods used to evaluate the efficacy of individual programmes that have been funded using a SIB structure. Consequently, the full body of evidence has not yet been synthesised. Two systematic reviews focused on SIBs have been published to date. One is limited to the context of health and social care; it does not focus on barriers to and facilitators of rigorous evaluation (Tan et al., 2015). The most recent review (Fraser et al., 2016) aims to explore different narratives that have emerged in the literature; it also does not explore barriers to or facilitators of rigorous evaluation. Some secondary research is relevant to the topic of evaluation practices. Lagarde et al. (2013), investigated challenges related to design,

implementation and evaluation. However, the report was non-exhaustive and focused on the broader pay-for-performance landscape.

Most existing reports consider the SIB market within a specific country or policy area, resulting in a disjointed overview of the landscape. There are few publicly available evaluation reports. There is a dearth of quantitative data; existing evaluations primarily rely on qualitative data (Ronicle & Stanworth 2015; Disley et al., 2011; Tan et al., 2015; DCLG 2014; Disley et al., 2015; Disley and Rubin 2014). Study sample sizes are typically limited (Vennema 2016; Gustafsson-Wright & Gardiner 2016). Unpublished material is often authored by stakeholders with an interest in the success of SIBs (e.g KPMG 2014). Many academics studying SIBs have strong views on the role of private investment in the area of health and social care (e.g McHugh et al., 2013). Regardless, there are some balanced perspectives on the use of contracting out for public services (e.g Warner 2013).

### *1.4.2 What this thesis will add*

This thesis will shed light on why SIB evaluation practices remain "fragmented, diverse and often idiosyncratic", despite growing interest in outcomes measurement (Nicholls 2009; Harlock 2013). It will systematically review existing primary evidence to understand the current state of research and to identify new barriers to and facilitators of rigorous evaluation of SIBs. Existing research shows considerable variability in the evaluation practices and methods used to verify whether a SIB-funded intervention has improved outcomes for beneficiaries and thus whether investor repayments should be triggered. Regardless, few authors have attempted to analyse why this may be the case and the implications. This thesis will overcome an existing gap in the literature, thus contributing to an enhanced understanding of how to improve evaluation practices in this field.

### *1.4.3 Objectives and Research Question*

Using the SPIDER tool (Cooke et al., 2012), the proposed research question asks: *"What are practitioners' perceptions about the main barriers to and facilitators of undertaking rigorous impact evaluations of SIBs?"* There is limited independent evidence on the effectiveness of SIBs (Ronicle et al., 2014; Edmiston & Nicholls 2017). Thus, the overarching objective is to systematically assess the literature, to identify the barriers to and

facilitators of rigorously evaluating SIBs, including factors perceived by different groups of practitioners. Table 3 provides further details about the research question.

***Table 3: SPIDER Research Question***

| SPIDER Framework Component | Research Question Components |
|---|---|
| ***Sample*** – group of participants that will be studied | 'Practitioners' – defined broadly as those who either have credible domain expertise (including academic body of knowledge) or those with experience in the development, operation or evaluation of SIBs. |
| ***Phenomenon of interest*** – similar to the intervention component of PICO. Aimed at understanding how and why evaluation practices differ | Undertaking or being involved with rigorous impact evaluation/s of SIBs. 'Rigorous' evaluation offers strong proof that the intervention has led to the desired difference in outcomes. |
| ***Design*** – techniques used by researchers to gather data | Interviews, focus groups, observations, round tables, surveys |
| ***Evaluation*** – similar to the outcomes component of PICO | Barriers to or facilitators of rigorous evaluation (perceptions or attitudes regarding barriers or facilitators) |
| ***Research type*** – type/s of research searched for | Any primary research design or systematic review |

In a SIB transaction, *all* parties need to agree to the evaluation method and metrics. While independent researchers often conduct evaluations, various stakeholders are involved in activities related to evaluation. Thus, the research question defined 'practitioners' broadly to include stakeholders who are either: 1) directly involved with the development, operation or evaluation of SIBs; or 2) have credible domain expertise (including an academic body of knowledge). The phenomenon of interest was undertaking or being involved in rigorous impact evaluations of SIBs, as the goal was to identify how more rigorous evaluation approaches might be employed when evaluating the efficacy of SIB-funded programmes. This review adopted the World Bank's definition of evaluation, which notes that 'impact evaluation' seeks to determine whether the intervention has the desired effects on beneficiaries and whether those effects are attributable to the intervention (Adam 2006). There may be various levels of metrics, data collection and evaluation within a single SIB. For pragmatic reasons, this paper is focused on impact evaluations of the efficacy of individual SIB-funded programmes. 'Rigorous' evaluation offers strong proof that the intervention has led to the desired difference in outcomes. The outcome measures included practitioners' perceptions or attitudes regarding the barriers to or facilitators of rigorous evaluation. All primary research designs were included, as qualitative inquiry provides key insights into practitioner perspectives, while mixed methods can help to conceptualise findings (Oliver et al., 2014).

# CHAPTER 2 – METHODS

## 2.1 Introduction

This study is a systematic review of empirical literature related to barriers to or facilitators of undertaking or being involved in rigorous evaluations of SIBs. The study was conducted and reported following the PRISMA guidelines (Appendix 6). A participatory approach was used, as the review was informed by an Expert Advisory Panel. This review is the first systematic, exhaustive attempt to synthesise existing research on the factors that support or hinder rigorous impact evaluations of SIBs.

## 2.2 Stakeholder involvement

Participatory approaches have become established elements of many research programmes, as several commentators have noted that professionals are no longer the "sole holders of expertise" (Chalmers 1995; Oliver 1997). Practitioners working on systematic reviews may bring their own subjective viewpoints about the issues and questions being addressed. The existence of and potential impact of different perspectives is exemplified by comparing the results of two systematic reviews, which both focused on preventing obesity (Doak et al., 2006; Summerbell et al., 2005). While the reviewers identified the same studies, they presented very different findings in terms of what they judged to be effective. Gough et al (2012), holds that the differing findings primarily arose from differences in the conduct of the reviews, relating to: 1) different ways of perceiving the challenge of obesity and potential solutions; or 2) differing viewpoints regarding how to judge an intervention as 'effective'. In light of the potentially subjective elements of systematic reviews, researchers are now beginning to draw on various perspectives, through participatory research. Stakeholder involvement may be perceived as an important indicator of quality (Brouwers et al., 2010).

An Expert Advisory Panel was established to inform the key steps of the review. Prior to establishing the Expert Advisory Panel, the project was reviewed and accepted for ethical approval (CUREC_1A) by the internal Ethics Committee of the Department of Social Policy and Intervention, University of Oxford. Prospective panel members were explicitly defined, using a purposive sampling approach. The list of prospective members consisted of a broad range of stakeholders who could inform on barriers and facilitators, as well as other aspects of the systematic review. The identified individuals had either: 1) been directly involved in

the funding, administration and/or implementation of SIBs; 2) published academic articles related to SIBs; or 3) been part of the wider policy debate (e.g policy makers who have worked on SIB-related issues). An invitation email was sent to each prospective member, with an information sheet (Appendix 1), introducing the study and researcher. Nine prospective members accepted the invitation; one prospective member nominated another individual (with equal seniority) from their organisation. See Table 4 for a list of members.

***Table 4: Advisory panel members***

| Name | Organisation | Position |
|---|---|---|
| Roger Bullen | Triodos Bank | Lead – Commissioner Advisory Services |
| Dr Clare FitzGerald | Government Outcomes Lab | Research and Policy Officer |
| Dr Nicholas Hicks | Cobic | Founder and Chief Executive |
| Marcus Hulme | Big Society Capital | Social Impact Director |
| Andrew Levitt | Bridges Ventures | Partner – Social Sector Funds |
| James Magowan | Government Inclusive Economy Unit, Office for Civil Society | Senior Policy Advisor (focused on impact bond development) |
| Dr Alex Nicholls | Said Business School | Professor of Social Entrepreneurship |
| Louise Savell | Social Finance | Director (Co-founder of Social Finance) |
| David Royce | Independent Consultant | Consultant – Strategies for Social Investment |
| Stefanie Tan | London School of Hygiene and Tropical Medicine | Research Fellow – Policy Innovation Research Unit (PIRU) |

By providing ongoing feedback and guidance, the panel enhanced the conduct, rigour and relevance of the review. While the ultimate responsibility remained with the principal researcher, panel members provided feedback relating to key systematic review steps (see Table 5). For pragmatic reasons, the panel members provided their feedback in three distinct phases. Additionally, they answered further questions via email, phone or in person as required.

***Table 5: summary of panel input stages***

| Stage | Requirement | Summary |
|---|---|---|
| **Stage 1** | Provide input on the research question and proposed methods, in person, by phone or via email | Panel members reviewed the following: research question and objective; background to review; inclusion and exclusion criteria; databases search strategy; grey literature search strategy; data collection and analysis methods |
| **Stage 2** | Provide comments on studies retrieved, in person, by phone or email | Panel members advised on the following: 1) whether any additional studies should be included; and 2) whether any of the (tentatively) included studies should be excluded |
| **Stage 3** | Provide input on the interpretation of findings, either in person, by phone or email | Panel members provided feedback on the following: search results; characteristics of included studies; critical appraisal; excluded studies; author's interpretations of findings from narrative synthesis |

## 2.3 Inclusion and exclusion criteria

An explicit inclusion and exclusion criteria was established and applied (Table 6), to reduce bias and to distinguish the systematic review from a narrative review (Littell et al., 2008). The criteria were initially piloted on a sample of 15 papers (including some thought to be eligible; not eligible; and doubtful). The inclusion criteria were structured in a hierarchical manner, assessed in order of importance. The first 'no' response was considered the primary reason for exclusion.

***Table 6: Inclusion and Exclusion Criteria***

| Inclusion Criteria | Exclusion Criteria |
|---|---|
| Study is either primary research (any study design) or a systematic review | Study is not primary research or systematic review |
| Study is about design and/or implementation of SIBs | Study focuses on subject other than design and/or implementation of a SIB |
| Study considers barriers to and/or facilitators of undertaking or being involved in rigorous evaluation of SIBs | Study does not consider barriers to and/or facilitators of undertaking or being involved in rigorous evaluation of SIBs |
| Study includes sample of practitioners who either have credible domain expertise (including an academic body of knowledge) or experience in the development, operation or evaluation of SIBs | Study does not include relevant sample of practitioners |
| Study includes perceptions of practitioners regarding barriers or facilitators to rigorous evaluation | Study does not include perceptions of practitioners regarding barriers or facilitators to rigorous evaluation |
| Study is English-language document | Non-English language document |
| Study is retrievable | Not retrievable |

To be included, studies had to be primary research (any study design – including but not limited to qualitative evaluations) or systematic reviews that discussed barriers to or facilitators of conducting, or being involved in rigorous impact evaluations of SIBs (as identified by practitioners). 'Impact evaluation' is defined as an attempt to quantify or demonstrate the impact, performance or outcomes of the services provided in a SIB-funded

intervention. 'Barriers' and 'facilitators' are defined as factors that hinder or support impact evaluation; they did *not* need to be the primary outcome of interest. This paper considers the defining feature of a SIB to be the provision of upfront capital from external investors who earn a return if the intervention succeeds. Broader literature relating to other types of outcomes based commissioning (which do not involve upfront working capital) were excluded. While SIBs are often defined as a subset of PbR, the inclusion of *all* PbR literature would have widened the scope of the review significantly. As SIBs are a relatively new funding mechanism and there is little published literature available, the inclusion criteria did *not* specify that articles need to be peer-reviewed and published in scientific literature. For practical reasons (time and cost), only English-language documents were included. No date restriction was applied; although the first SIB was introduced in 2010, similar mechanisms may have been used before then.

## 2.4 Search strategy

### *2.4.1 Standard Search Strategy*

As the evidence base for SIBs is still emerging, the standard search string (Table 7) was designed to retrieve a large number of studies related to the evaluation of SIBs. For example, the search did *not* include each component of the research question; study design filters and research type were omitted as they would limit the results (Higgins & Green 2011). The string was organised around the key terms of: SIBs; evaluation; and barriers or facilitators. To maximise search sensitivity, synonyms were tested in formative searches and included.

***Table 7 – standard strategy***

| Search # | Topic | Searched for |
|---|---|---|
| S1 | SIBs | AB,TI(payment by results OR "social impact bonds" OR (social AND impact AND bonds) OR pay for performance bonds OR outcomes based commissioning OR social invest*" OR "impact invest*")) |
| S2 | Evaluation | AB,TI(evaluat* OR assess* OR impact OR measurement OR measuring OR experience* OR learning OR performance OR "EBP" OR "evidence based practice" OR "evidence-based practice" OR outcome) |
| S3 | Barriers or facilitators | ALL(barrier* OR facilitat* OR challenge* OR impediment OR benefit* OR success* OR constrain* OR difficult* OR enhanc* OR influen* OR interfer* OR motivat* OR obstruct* OR problem* OR promot* OR restrain* OR restrict* OR disincentive* OR incentive* OR factor* OR capacity) |
| S4 | Combined search | S1 + S2 + S3 |

*2.4.2 Database search*

Sensitivity maximising searches were conducted in each of the databases outlined in Table 8. Searches were tailored to each database, to accommodate for independent indexing systems (Higgins & Green 2011). SIBs are operating in various policy areas and the evidence base is emerging. To retrieve a broad range of published *and* unpublished literature, it was deemed important to search a diverse array of databases, including databases with peer-reviewed articles, and databases containing dissertations and conference proceedings. 13 databases were found to be relevant. Beyond this number, preliminary searches demonstrated that there diminishing returns for each additional database searched.

***Table 8: Search strategy***

| | Database | Date of final search | No. of results |
|---|---|---|---|
| **1.** | **ABI/INFORM Global** | 18 Jul 17 | 602 |
| | **Search 1:**<br>(AB,TI(payment by results) OR AB,TI("social impact bonds") OR AB,TI(social AND impact AND bonds) OR AB,TI(pay for performance bonds) OR AB,TI(outcomes based commissioning) OR AB,TI("social invest*") OR AB,TI("impact invest*")) AND<br><br>**Search 2:**<br>ALL((barrier* OR facilitat* OR challenge* OR benefit* OR success* OR constrain* OR difficult* OR enhanc* OR influen* OR interfer* OR motivat* OR obstruct* OR problem* OR promot* OR restrain* OR restrict* OR disincentive* OR factor* OR capacity) NEAR/6 (evaluate OR evaluation OR evaluating OR assess* OR impact OR measurement OR measuring OR experience* OR learning OR performance OR "ebb" OR "evidence based practice" OR "evidence-based practice" OR outcome))<br><br>**Search 3:**<br>1 + 2 | | |
| **2.** | **Applied Social Sciences Index & Abstracts (ASSIA) Abstracts** | 18 Jul 17 | 231 |
| | Standard search strategy | | |
| **3.** | **COS Conference Papers Index** | 18 Jul 17 | 23 |
| | (AB,TI(payment by results) OR AB,TI("social impact bonds") OR AB,TI(social AND impact AND bonds) OR AB,TI(pay for performance bonds) OR AB,TI(outcomes based commissioning) OR AB,TI("social invest*") OR AB,TI("impact invest*")) | | |
| **4.** | **International Bibliography of the Social Sciences (IBSS)** | 18 Jul 17 | 676 |
| | Standard search strategy | | |
| **5.** | **MEDLINE** | 18 Jul 17 | 553 |
| | **Search 1:**<br>(AB,TI(payment by results) OR AB,TI("social impact bonds") OR AB,TI(social AND impact AND bonds) OR AB,TI(pay for performance bonds) OR AB,TI(outcomes based commissioning) OR AB,TI("social invest*") OR AB,TI("impact invest*")) AND<br><br>**Search 2:**<br>ALL((barrier* OR facilitat* OR challenge* OR benefit* OR success* OR constrain* OR difficult* OR enhanc* OR influen* OR interfer* OR motivat* OR obstruct* OR problem* OR promot* OR | | |

| | | | |
|---|---|---|---|
| | restrain* OR restrict* OR disincentive* OR factor* OR capacity) NEAR/6 (evaluate OR evaluation OR evaluating OR assess* OR impact OR measurement OR measuring OR experience* OR learning OR performance OR "ebb" OR "evidence based practice" OR "evidence-based practice" OR outcome))<br><br>**Search 3:**<br>1 + 2 | | |
| **6.** | **PAIS Index** | 18 Jul 17 | 254 |
| | Standard search strategy | | |
| **7.** | **Policy File Index** | 18 Jul 17 | 173 |
| | Standard search strategy | | |
| **8.** | **Proquest Dissertations and Theses: UK & Ireland** | 18 Jul 17 | 331 |
| | Standard search strategy | | |
| **9.** | **Proquest Dissertations and Theses: Global** | 18 Jul 17 | 773 |
| | **Search 1:**<br>(AB,TI(payment by results) OR AB,TI("social impact bonds") OR AB,TI(social AND impact AND bonds) OR AB,TI(pay for performance bonds) OR AB,TI(outcomes based commissioning) OR AB,TI("social invest*") OR AB,TI("impact invest*"))<br><br>**Search 2:**<br>ALL((barrier* OR facilitat* OR challenge* OR benefit* OR success* OR constrain* OR difficult* OR enhanc* OR influen* OR interfer* OR motivat* OR obstruct* OR problem* OR promot* OR restrain* OR restrict* OR disincentive* OR factor* OR capacity) NEAR/6 (evaluate OR evaluation OR evaluating OR assess* OR impact OR measurement OR measuring OR experience* OR learning OR performance OR "ebb" OR "evidence based practice" OR "evidence-based practice" OR outcome))<br><br>**Search 3:**<br>1 + 2 | | |
| **10.** | **Social Services Abstracts** | 18 Jul 17 | 126 |
| | Standard search strategy | | |
| **11.** | **SCOPUS** | 18 Jul 17 | 1,919 |
| | **Search 1:**<br>TITLE-ABS (payment by results OR “social impact bonds” OR (social AND impact AND bonds) OR pay for performance bonds OR outcomes based commissioning OR social invest*" OR “impact invest*”))<br><br>**Search 2:**<br>TITLE-ABS (evaluat* OR assess* OR impact OR measurement OR measuring OR experience* OR learning OR performance OR “EBP” OR “evidence based practice” OR “evidence-based practice” OR outcome)<br><br>**Search 3:**<br>TITLE-ABS-KEY (barrier* OR facilitat* OR challenge* OR benefit* OR success* OR constrain* OR difficult* OR enhanc* OR influen* OR interfer* OR motivat* OR obstruct* OR problem* OR promot* OR restrain* OR restrict* OR disincentive* )<br><br>**Search 4:**<br>1 + 2 + 3 | | |
| **12.** | **Web of Science** | 18 Jul 17 | 179 |
| | Standard search strategy | | |
| **13.** | **Worldwide Political Science** | 18 Jul 17 | 175 |
| | Standard search strategy | | |
| | **Total database results:** 6,073 | | |

*2.4.3 Grey literature*

A grey literature search was an important component of this systematic review due to the nascence of the literature (Table 9) (Adams et al., 2016; Mahood Q et al., 2014). Various stakeholders (e.g governments, intermediaries and non-profit organisations) have released different forms of commentary, which are rarely published in electronic databases. At the outset, reference lists of existing literature reviews were hand searched, to identify sources to search in more detail. Further, the reviewer cross-checked identified grey literature articles with three (unpublished) reference lists compiled by experts from the Government Outcomes (GO) Lab (Blavatnik School of Government, University of Oxford).

***Table 9: Grey literature Sources***

| | Source | No. of results |
|---|---|---|
| 1. | Google Scholar | 207 |
| 2. | Urban Institute | 33 |
| 3. | Social Finance | 26 |
| 4. | Government Performance Lab (Harvard) | 25 |
| 5. | Saïd Business School | 19 |
| 6. | Office of Social Impact Investment - NSW | 18 |
| 7. | Gov.uk | 16 |
| 8. | Global Impact Investing Network | 10 |
| 9. | Big Lottery Fund | 9 |
| 10. | Center for American Progress | 8 |
| 11. | Instiglio | 7 |
| 12. | Impact Investing Australia | 6 |
| 13. | National Council for Voluntary Organisations (NCVO) | 6 |
| 14. | RAND Europe | 6 |
| 15. | Stanford Social Innovation Review | 5 |
| 16. | Boston Consulting Group | 4 |
| 17. | Big Society Capital | 4 |
| 18. | Bridges Ventures | 4 |
| 19. | Carleton University | 4 |
| 20. | Deloitte | 4 |
| 21. | Rockefeller Foundation | 4 |
| 22. | Harvard University – Business School | 3 |
| 23. | New Philanthropy Capital | 3 |
| 24. | Canadian Centre for Policy Alternatives | 2 |
| 25. | Centre for Social Impact (CSI), NSW | 2 |
| 26. | Duke University | 2 |
| 27. | Global Social Impact Investment Steering Group | 2 |
| 28. | Government Outcomes Lab | 2 |
| 29. | Government of the Netherlands | 2 |
| 30. | Mission Investors | 2 |
| 31. | New Philanthropy Capital | 2 |
| 32. | OPM Group | 2 |
| 33. | PwC | 2 |

| | | |
|---|---|---|
| 34. | The Young Foundation | 2 |
| 35. | Third Sector Capital Partners | 2 |
| 36. | Venture Philanthropy Partners | 2 |
| 37. | World Economic Forum | 2 |
| 38. | 2020 Public Services Trust | 1 |
| 39. | Acumen | 1 |
| 40. | Barclays Wealth | 1 |
| 41. | Bill & Melinda Gates Foundation | 1 |
| 42. | Brookings Institute | 1 |
| 43. | Center for Global Development | 1 |
| 44. | Center for Law and Social Policy (CLASP) | 1 |
| 45. | Central European University | 1 |
| 46. | Cicero Consultancy | 1 |
| 47. | City of London | 1 |
| 48. | Corporation for National & Community Service | 1 |
| 49. | Criminal Justice Inspectorates | 1 |
| 50. | Early Intervention Foundation | 1 |
| 52. | European Services Strategy Unit | 1 |
| 53. | Evaluating Impact Investing (Jackson & Associates) | 1 |
| 54. | Federal Reserve Bank of Atlanta | 1 |
| 55. | General Assembly of Maryland | 1 |
| 56. | Hewlett Foundation | 1 |
| 57. | Impact Assets | 1 |
| 58. | Investing for Good | 1 |
| 59. | Istanbul Policy Center | 1 |
| 60. | Kauffman Foundation | 1 |
| 61. | KPMG | 1 |
| 62. | London.gov.uk | 1 |
| 63. | Manpower Demonstration Research Corporation (MDRC) | 1 |
| 64. | MaRS | 1 |
| 65. | McKinsey & Company | 1 |
| 66. | Milken Institute | 1 |
| 67. | Monitor Institute | 1 |
| 68. | National Australia Bank (NAB) | 1 |
| 69. | National Audit Office (NAO) | 1 |
| 70. | NHS Networks | 1 |
| 71. | Nonprofit Finance Fund | 1 |
| 72. | OECD | 1 |
| 73. | Omidyar Network | 1 |
| 74. | Parliament of NSW | 1 |
| 75. | Philanthropy Impact | 1 |
| 76. | PIRU | 1 |
| 77. | Policy and Politics (Journal Blog) | 1 |
| 78. | Princeton University | 1 |
| 79. | REES Centre, University of Oxford | 1 |
| 80. | Shelter SA | 1 |
| 81. | Social Impact Investment Taskforce | 1 |
| 82 | Social Market Foundation | 1 |
| 83. | Social Venture Technology Group | 1 |
| 84. | The Australian Sociological Association | 1 |
| 85. | The Centre for Social Justice | 1 |
| 86. | The Institute for Government | 1 |
| 87. | Tides Foundation | 1 |
| 88. | University of Leeds | 1 |

| | | |
|---|---|---|
| 89. | Vera Institute of Justice | 1 |
| 90. | World Bank | 1 |
| 91. | Parliament.uk | 1 |
| | **Total number of results:** | 513 |

### *2.4.4 Hand searching*

The search strategy also included hand searching (Table 10), noting that a combination of electronic searching *and* hand searching is required to identify all relevant reports (Hopewell 2007). Some reports are not included in electronic bibliographic databases, while others (even if included) may not properly indexed (Dickersin 1994).

***Table 10: Hand searching***

| | **Source** | **No. of results** |
|---|---|---|
| 1 | Journal of Sustainable Finance and Investment | 8 |
| 2 | Community Development | 5 |
| 3 | Journal of Social Entrepreneurship | 3 |
| 4 | Research in International Business and Finance | 2 |
| 5 | Journal of Positive Management | 2 |
| 6 | Entrepreneurship Research Journal | 2 |
| 7 | Accounting, Organizations and Society | 1 |
| 8 | Administrative Sciences | 1 |
| 9 | Business & Society | 1 |
| 10 | Canadian Journal of Nonprofit {&} Social Economy Research | 1 |
| 11 | Company and Securities Law Journal | 1 |
| 12 | Corporate Ownership and Control | 1 |
| 13 | Cultural Studies | 1 |
| 14 | European Scientific Journal | 1 |
| 15 | Financial Analysts Journal | 1 |
| 16 | Florida Tax Review | 1 |
| 17 | Fordham Law Review | 1 |
| 18 | Global Economics and Management Revie | 1 |
| 19 | International Journal of Public Administration | 1 |
| 20 | International Journal of Public Policy | 1 |
| 21 | International Society for Third Sector Research | 1 |
| 22 | International Studies Review | 1 |
| 23 | Journal of Business Ethics | 1 |
| 24 | Journal of Comprehensive Health | 1 |
| 25 | Journal of Contemporary European Research | 1 |
| 26 | Journal of Law & Business | 1 |
| 27 | Probation Journal | 1 |
| 28 | Procedia - Social and Behavioral Sciences | 1 |
| 29 | Research on Social Work Practice | 1 |
| 30 | Universal Journal of Accounting and Finance | 1 |
| 31 | University of Pennsylvania Journal of Business Law | 1 |
| 32 | University of Pittsburgh Law Review | 1 |
| 33 | VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations | 1 |
| | **Total:** | 49 |

### *2.4.5 Snow balling*

To identify additional studies, reference lists for relevant studies and reviews were also searched. While this was a useful review component, it is worth noting that this approach is “far from objective” as reference lists may selectively cite sources (Higgins et al., 2011).

### *2.4.6 Contacting experts*

As this is a relatively new field of research, all identified authors of relevant studies were contacted via email and asked whether they could provide unpublished studies or general assistance identifying additional studies. 92 authors were contacted in total. At the time of submission, 39 authors had responded (Appendix 2).

## 2.5 Data collection and analysis

### *2.5.1 Study selection*

After identifying relevant studies through electronic databases, grey literature, contacting experts and snowballing, studies were merged (using “Covidence” software) and duplicates were removed. The reviewer screened all articles for relevance on title and abstract. Any obviously irrelevant studies were removed. The reviewer undertook a full-text review of the remaining studies. For pragmatic reasons, articles were not double screened, although panel members provided feedback on included and excluded studies, to ensure rigour.

### *2.5.2 Data extraction*

Studies were screened and stored using the “Covidence” software. A data extraction form (Appendix 3) was created, using guidance from the EPPI Centre, Cochrane Collaboration and National Institute for Health and Care Excellence (NICE). Data were extracted on factors such as study aims, study design, sample characteristics, theoretical background and findings (Higgins et al., 2011).

### *2.5.3 Quality appraisal*

The aim of a thematic synthesis (the method used for data analysis) is not to determine the size of an effect. Despite this, critical appraisal was considered integral for

assessing confidence in the findings and conducting sensitivity testing. Thus, all included studies were subject to appropriate critical appraisal. There are various limitations to prescriptively using checklists or scales (Higgins & Green 2011). Regardless, using an appraisal tool helped the reviewer to treat each study in the same way (Gough et al., 2012). While primary studies of any research design could be included, the included studies all used qualitative methods to identify barriers to or facilitators of rigorous evaluation. A small number of reports employed mixed methods; however, the quantitative data were not relevant to identifying factors operating as barriers or facilitators. Thus, the reviewer used critical appraisal tools developed for qualitative research studies (e.g CASP 2017; Tong, Sainsbury, & Craig 2007).

### *2.5.4 Analysis*

This review focused on the factors (barriers or facilitators) affecting the conduct of rigorous impact evaluations of SIBs, as identified by practitioners. Included studies could be qualitative (e.g interviews, focus groups, case studies), cross sectional or mixed methods. Despite this, all included studies used a qualitative methodology. Analysis was conducted using the thematic synthesis approach outlined by Thomas and Harden (2008). This method is commonly used to bring together findings of research from many different types. It was deemed to be the most appropriate method, as the included studies consisted of many diverse forms of grey literature. The studies each had different aims; none of the included studies explicitly aimed to identify barriers to or facilitators of rigorous evaluation. Consequently, there were few pre-specified themes. The reviewer was required to generate themes inductively, using their interpretation to move beyond individual studies and generate new knowledge (Gough et al., 2012).

The analysis involved the three key stages outlined by Thomas and Harden (2008); 1) coding text; 2) developing descriptive themes; and 3) generating analytical themes. Firstly, each included study was read in detail; line-by-line (inductive) coding was conducted on the findings. A key challenge in this step is deciding what constitutes the ‘findings’ of included studies. Even if a selective approach is taken (e.g only including quotations from participants), the ‘findings’ may still reflect the authors’ original perspectives and priorities. Thus, the reviewer treated the entire paper or report as ‘data’. Each sentence and paragraph was carefully read, to identify underpinning themes. Relevant text was highlighted and

associated with a particular code. Secondly, codes were compared and contrasted across different studies to develop descriptive themes. The aim of this step was to identify and articulate relationships between themes, and to associate conceptually similar themes with each other. The final step involved generating analytical themes which progressed beyond the content of the included studies, thus offering new conceptualisations of barriers and facilitators.

### *2.5.5 Assessing the robustness of the synthesis*

After completing each of the above steps, the reviewer checked the robustness of the synthesis. The reviewer created a grid which depicts the relative contribution of each study to the synthesis, as suggested by Rees et al (2009). This helped to identify: whether important parts of the synthesis depend upon few studies; whether this matters; and whether there are associations between specific populations and/or quality judgements of studies and groups of themes. If an area of the synthesis is based on more studies than others, it does not necessarily mean it is more trustworthy. This approach was used to raise key questions about the robustness of the findings.

Additionally, the reviewer used the Confidence in the Evidence from Reviews of Qualitative research (GRADE-CERQual) approach for qualitative syntheses (Guyatt et al., 2011), to assess how much confidence to place in each individual finding (Lewin et al., 2015). This involved an assessment of the following four criteria, for each finding: 1) methodological limitations; 2) coherence; 3) relevance; and 4) adequacy of data (Table 11) (Lewin et al., 2015). Based on this, the reviewer then assigned one of the following categories for each finding: ‘high’; ‘moderate’; ‘low’; or ‘very low’.

***Table 11: GRADE-CERQual assessment components***

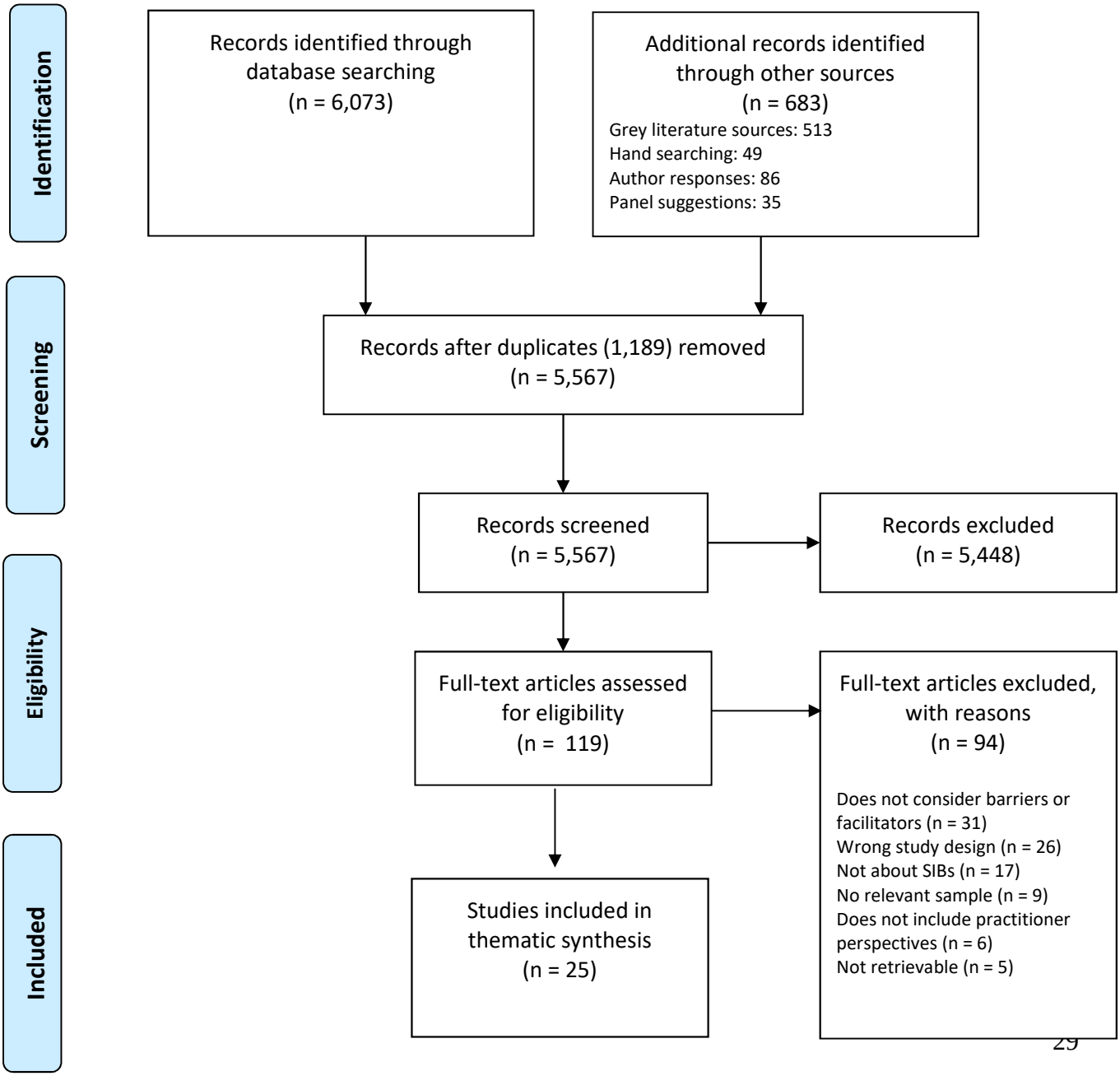
| Component | Definition |
|---|---|
| Methodological limitations | The extent to which there are problems in the design or conduct of the primary studies that contributed evidence to the review finding. |
| Relevance | The extent to which the body of evidence from the primary studies supporting a review finding is applicable to the context (perspective or population, phenomenon of interest, setting) specified in the review question. |
| Coherence | The extent to which the review finding is well grounded in data from the contributing primary studies and provides a convincing explanation for the patterns found in these data. |
| Adequacy of data | An overall determination of the degree of richness and quantity of data supporting a review finding. |

## CHAPTER 3 – RESULTS

### 3.1 Search Results

The searches of electronic databases, grey literature sources and hand searches were conducted between 1st July and 20th July 2017. The electronic database searches yielded 6,073 citations. The reviewer collected an additional 683 citations from other sources. After removing duplicates, 5,567articles were retrieved for title and abstract review. 5,448 articles were excluded on title and abstract; 119 were retrieved for full text review. Of these, 25 were included and 94 were excluded (Figure 9). 19 of the 25 included studies were retrieved from the grey literature search, thus highlighting the importance of searching beyond electronic databases.

***Figure 9: PRISMA diagram***

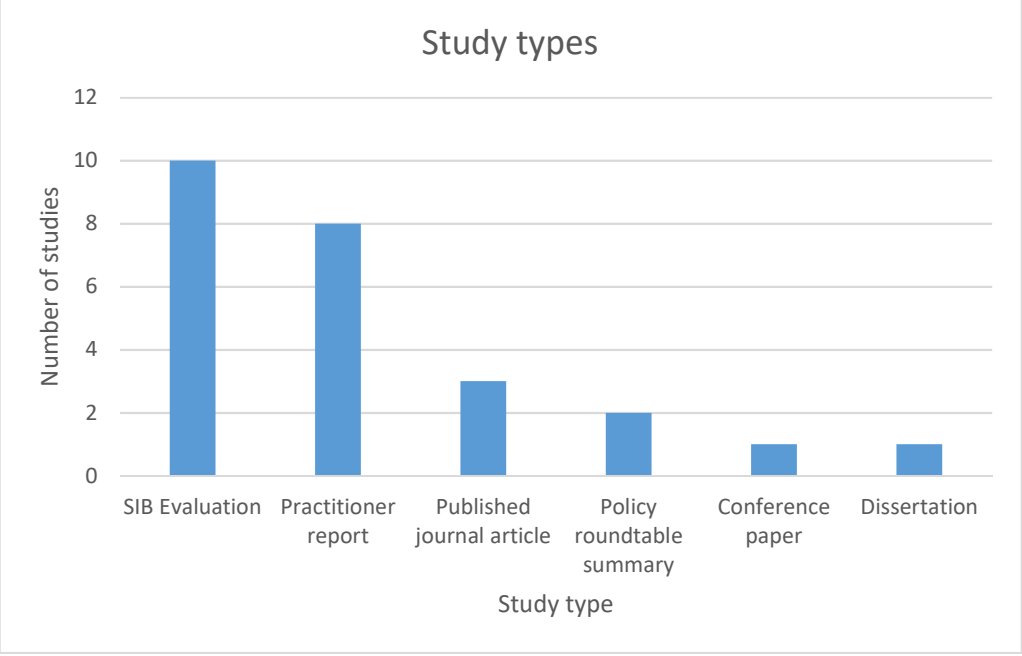


### 3.2 Characteristics of included studies

25 studies met the inclusion criteria. Each of the studies primarily used qualitative methods to collect data (interviews, round tables or case studies). Some reports included limited quantitative data (DCLG 2015; KPMG 2014; Murray et al., 2012; Roberts & Cameron 2014; Ronicle et al., 2014). However, the quantitative data were not relevant to identifying barriers to or facilitators of rigorous evaluation and were excluded from the analysis. One of the studies included a systematic review in addition to qualitative interviews; the review component was not relevant for the research question (Tan et al., 2015).

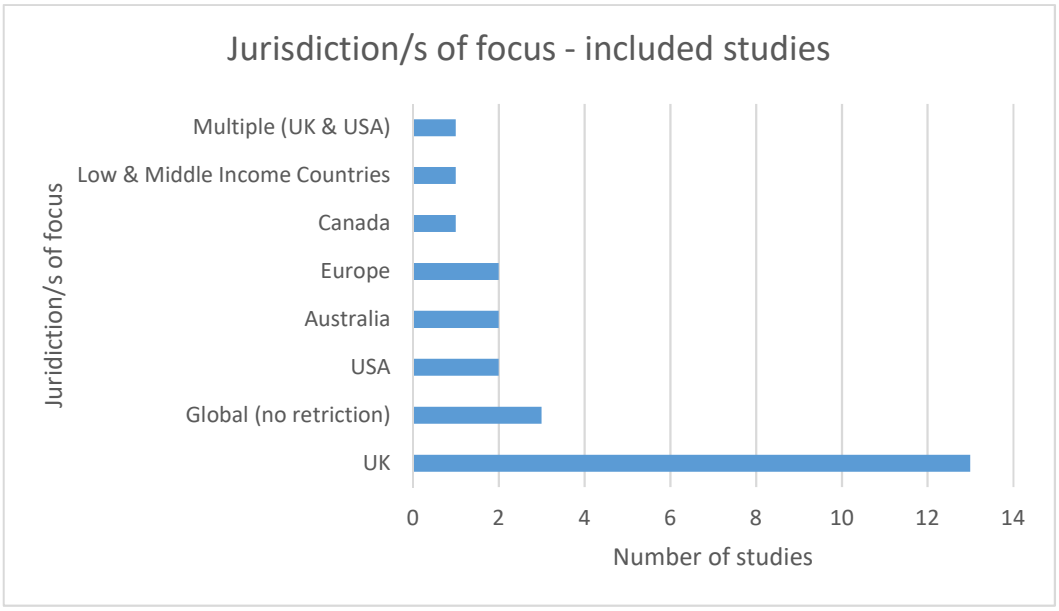
Of the included studies, one was a dissertation (Vennema 2016), one was a conference paper (Cooper et al., 2013), two were policy roundtable summaries (CRESSI 2016; Rotheroe 2014), three were published journal articles (Edmiston & Nicholls 2017; Reeder et al., 2015; Warner 2013); eight were practitioner reports (Galley et al., 2014; Gustafsson-Wright et al., 2015; Gustafsson-Wright and Gardiner 2016; Hughes & Scherer 2014; James 2016; Reeder et al., Rudd et al., 2013; Tomkinson 2012) and ten were evaluation reports (DCLG 2014; DCLG 2015; Disley et al., 2011; Disley and Rubin 2014; KPMG 2014; Murray et al., 2012; Roberts and Cameron 2014; Ronicle et al., 2014; Ronicle & Stanworth 2015; Tan et al., 2015).

***Figure 10: Types of studies***



In terms of context, thirteen studies focused solely on SIBs implemented in the UK (Cooper et al., 2013; CRESSI 2016; DCLG 2014; DCLG 2015; Disley et al., 2011; Disley & Rubin 2014; Edmiston & Nicholls 2017; James 2016; Murray et al., 2012; Roberts & Cameron 2014; Ronicle & Stanworth 2015; Rotheroe 2014; Tan et al., 2015); two focused solely on USA (Rudd et al., 2013; Hughes & Scherer 2014); two focused entirely on Australia (KPMG 2014; Tomkinson 2012); two focused on Europe (Reeder et al., 2014; Reeder et al., 2015); one focused on Canada (Galley et al., 2014); one focused on low and middle-income countries (Gustafsson-Wright & Gardiner 2016); one focused on both USA and UK (Warner 2013); and three had a global focus (Gustafsson-Wright et al., 2015; Ronicle et al., 2014; Vennema 2016).

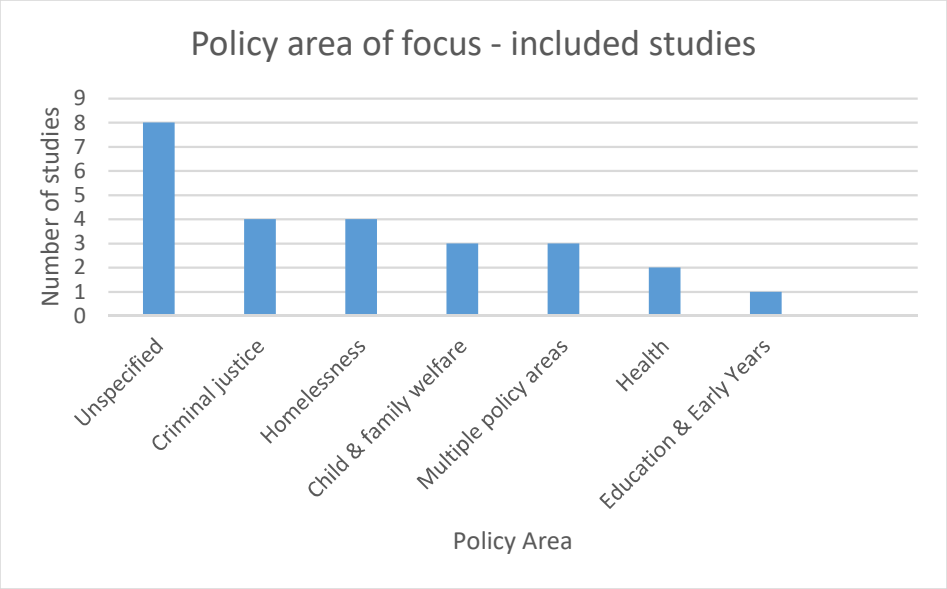
***Figure 11: Jurisdictions of focus***




With regards to policy area, four studies focused specifically on homelessness (Cooper et al., 2013; DCLG 2014; DCLG 2015; James 2016); four focused on recidivism (Disley et al., 2011; Disley & Rubin 2014; Murray et al., 2012; Rudd et al., 2013); three focused on out-of-home-care (KPMG 2014; Roberts & Cameron 2014; Tomkinson 2012); two focused on health (Ronicle & Stanworth 2015; Tan et al., 2015); one focused on recidivism and early childhood education (Warner 2013); two focused on homelessness, out-of-home-care and youth not in employment, education or training (NEET) (CRESSI 2016; Edmiston & Nicholls 2017); one focused on early childhood development (Gustafsson-

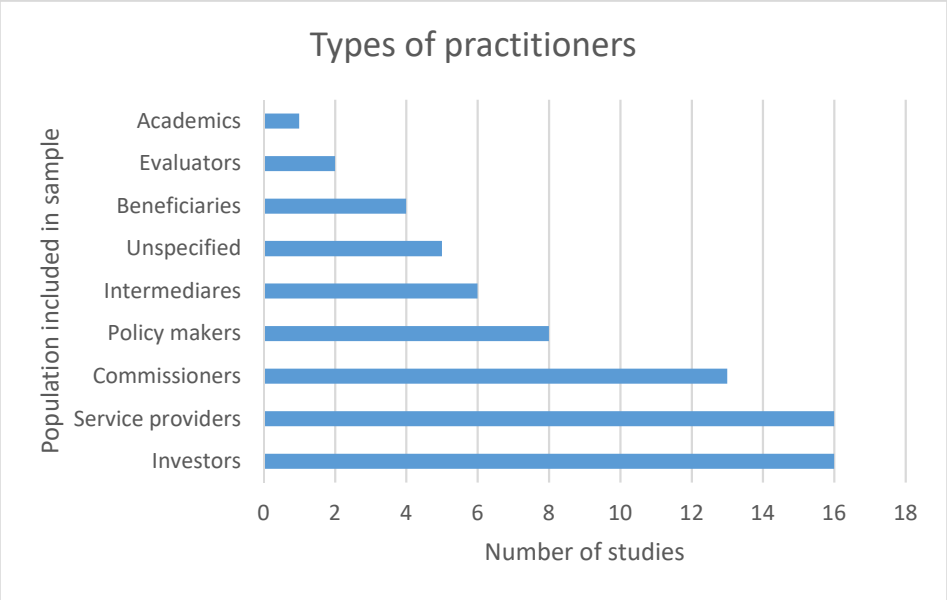
Wright & Gardiner 2016); and eight did not specify a policy area of focus (Galley et al., 2014; Gustafsson-Wright et al., 2015; Hughes & Scherer 2014; Reeder et al., 2014; Reeder et al., 2015; Ronicle et al., 2014; Rotheroe 2014; Vennema 2016).

***Figure 12: policy area of focus***




Study samples included a mix of stakeholders involved with SIBs, such as investors, intermediaries, service providers and outcome funders. Figure 13 demonstrates that some stakeholders (particularly evaluators and academics) were significantly under-represented in the samples.

***Figure 13: Sample populations***



It is also important to consider who the studies were written by. Nine studies were written by academics or people with academic affiliations (Cooper et al., 2013; CRESSI 2016; Edmiston & Nicholls 2017; Reeder et al., 2014; Reeder et al., 2015; Tan et al., 2015; Tomkinson 2012; Vennema 2016; Warner 2013); nine were written by independent evaluators commissioned by government to conduct evaluations (consulting firms in 7 studies; think tanks in 2 studies) (DCLG 2014; DCLG 2015; Disley et al., 2011; Disley and Rubin., 2014; KPMG 2014; Murray et al., 2012; Roberts & Cameron 2014; Ronicle et al., 2014; Ronicle & Stanworth 2015); six were written by non-profit or charitable organisations (Gustafsson-Wright et al., 2015; Gustafsson-Wright & Gardiner 2016; Hughes & Scherer 2014; James 2016; Rotheroe 2014; Rudd et al., 2013); and one was conducted by an independent think tank (Galley et al., 2014).

***Figure 14: authors of studies***

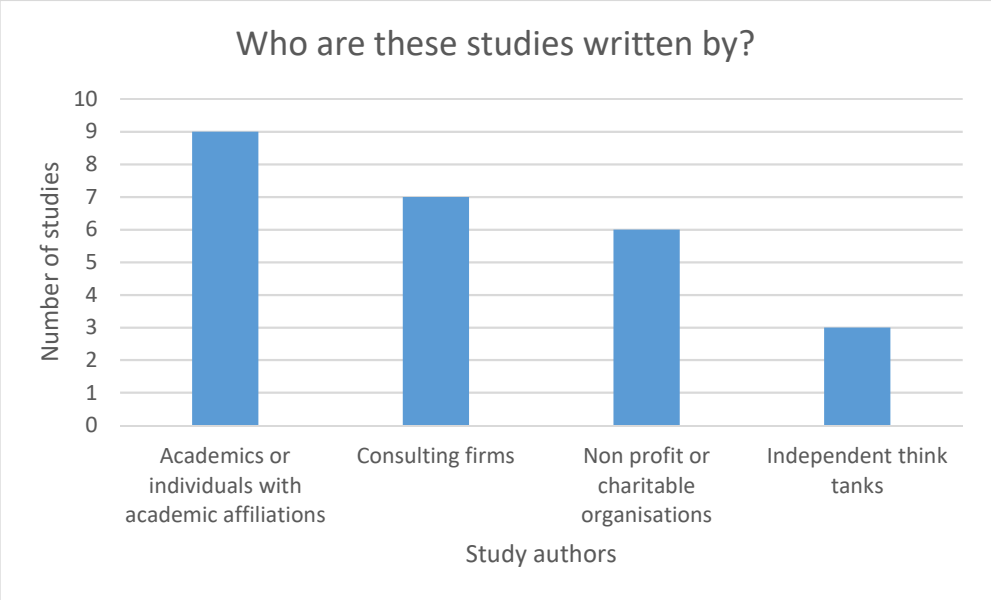


**Table 12: Characteristics of included studies**

| Citation | Study type | Country | Policy Area | Design | Data collection | Sample | Barriers – themes discussed | Facilitators – themes discussed |
|---|---|---|---|---|---|---|---|---|
| Cooper et al (2013) | Conference paper | UK | Homelessness | Qualitative | Interview data | 3 practitioners (an investor and staff from Greater London Authority) | Data & analytics; measurement practices | Data & analytics; contact & collaboration |
| CRESSI (2016) | Policy roundtable summary | UK | Youth NEET; OOHC | Qualitative | Roundtable discussion | 32 practitioners (commissioners, academics, policy makers investors, service providers) | Measurement practices | Measurement practices |
| DCLG (2014) | Evaluation report | UK | Homelessness | Qualitative | Interviews | 73 stakeholders (commissioners, investors, intermediaries, providers, London boroughs, beneficiaries) | Resources; data and analytics; contact & collaboration; measurement practices | Resources; data & analytics; contact & collaboration; measurement practices |
| DCLG (2015) | Evaluation report | UK | Homelessness | Qualitative | Interviews | 49 stakeholders (commissioners, investors, providers, beneficiaries) | Measurement practices | Resources; measurement practices |
| Disley et al (2011) | Evaluation report | UK | Recidivism | Qualitative | Interviews | 22 practitioners involved with Peterborough SIB (organisations/ roles not specified) | Resources; contact and collaboration; measurement practices | Resources; data and analytics; measurement practices |
| Disley and Rubin (2014) | Evaluation report | UK | Recidivism | Qualitative | Interviews | 39 practitioners involved with Peterborough SIB (organisations/roles not specified) | Data and analytics; measurement practices | Resources; data and analytics; contact and collaboration; measurement practices |
| Edmiston and Nicholls (2017) | Journal article | UK | Homelessness; OOHC; youth unemployment | Qualitative | Interviews; practitioner seminar; roundtable | Unspecified number of practitioners (policy makers, commissioners, | Resources; contact and collaboration; | N/A (not discussed) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | providers, intermediaries, investors and service users). | measurement practices | |
| Galley (2014) | Practitioner report | Canada | Unspecified | Qualitative | In-depth interviews | 22 practitioners (NFP service providers, funders, and policy makers who have direct experience with SIBs). | Data and analytics; contact and collaboration; measurement practices | Resources; data and analytics; contact and collaboration |
| Gustafsson-Wright et al (2015) | Practitioner report | UK; USA; Australia; Continental Europe | Unspecified | Qualitative | Interviews (structured and unstructured) | Unspecified number of practitioners (investors, intermediaries, service providers, outcome funders) from each of the 38 SIBs included in the study) | Resources; data and analytics; contact and collaboration; measurement practices | Resources; data and analytics; contact and collaboration; measurement practices |
| Gustafsson-Wright and Gardiner (2016) | Practitioner report | Low & Middle Income Countries | Early Childhood Development | Qualitative | Interviews | 10 practitioners (service providers and investors) | Resources; data and analytics; measurement practices | Resources; data and analytics; contact and collaboration; measurement practices |
| Hughes and Scherer (2014) | Practitioner report | USA | Unspecified | Qualitative | Interviews | Unspecified number of practitioners (thought leaders at 30 different foundations and organisations involved in impact investment) | Resources; data and analytics; contact and collaboration; measurement practices | Resources; data and analytics; contact and collaboration; measurement practices |
| James (2016) | Practitioner report | UK | Homelessness | Qualitative | Case study | 4 organisations, including Depaul - organisation with a SIB contract worth more than £1m | Measurement practices | Measurement practices |
| KPMG (2014) | Evaluation report | Australia | OOHC | Qualitative | Interviews; survey; review of literature; transaction costs data; survey | 21 practitioners involved in SIB trial (policy makers, commissioners, service providers, investors, intermediaries) | Resources; data and analytics; contact and collaboration; evaluation practices | Resources; data and analytics; contact and collaboration; |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | | measurement practices |
| Murray et al (2012) | Evaluation report | UK | Recidivism | Qualitative | Semi-structured interviews | Unspecified number of stakeholders (Ministry of Justice, service providers, partner agencies, beneficiaries) | Data and analytics; measurement practices | Data and analytics; contact and collaboration |
| Reeder et al (2014) | Practitioner report | Europe | Unspecified | Qualitative | Interviews | Unspecified number of practitioners from 15 investment firms | Resources; data and analytics; measurement practices | Resources; measurement practices |
| Reeder et al (2015) | Journal article | Europe | Unspecified | Qualitative | Interviews | Unspecified number of practitioners from 16 investment firms | Data and analytics; contact and collaboration; measurement practices | Measurement practices |
| Roberts and Cameron (2014) | Evaluation report | UK | OOHC | Qualitative | Interviews; consultations/workshops | Unspecified number of MST stakeholders directly involved with Essex SIB | Resources; data and analytics; measurement practices | Resources; data and analytics; contact and collaboration; measurement practices |
| Ronicle et al (2014) | Evaluation report | Unspecified | Unspecified | Qualitative and quantitative | Literature review; in-depth consultations; surveys | Unspecified number of practitioners (commissioners, investors, policy makers and service providers) | Resources; data and analytics; contact and collaboration; measurement practices | Resources; data and analytics; measurement practices |
| Ronicle and Stanworth (2015) | Evaluation report | UK | Health | Qualitative | Consultations/interviews with stakeholders | Key stakeholders (unspecified number) involved in the SIB (service providers, investors, commissioners, intermediaries) | Resources; contact and collaboration; measurement practices | Data and analytics |
| Rotheroe (2014) | Policy Roundtable Summary | UK | Unspecified | Qualitative | Roundtable discussion | Unspecified number of practitioners (sector experts and charities involved in SIBs) | Resources; measurement practices | Resources; data and analytics; contact and |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | | collaboration; measurement practices |
| Rudd et al (2013) | Practitioner report | USA | Recidivism | Qualitative | Interviews | Stakeholders (unspecified number) in the New York City SIB (investors, service providers, commissioners, evaluator) | Measurement practices | Resources; data and analytics; contact and collaboration; measurement practices |
| Tan et al (2015) | Evaluation report | UK | Health | Qualitative | Interviews | Key informants (unspecified number) involved in UK SIB development undertaken between May and November 2014 | Resources; data and analytics; contact and collaboration; measurement practices | Resources; data and analytics; contact and collaboration; measurement practices |
| Tomkinson (2012) | Practitioner report | Australia | OOHC | Qualitative | Stakeholder perspectives collected during Social Finance Forum | All organisations involved in NSW SIB trial; broader government & investment representatives; Social Finance UK representatives. | Resources; measurement practices | Resources; data and analytics; contact and collaboration; measurement practices |
| Vennema (2016) | Dissertation | Unspecified | Unspecified | Qualitative | Open interviews | 6 interviewees including 2 service providers, an evaluator, a commissioner, an intermediary and an investor. | Resources; contact and collaboration; measurement practices | Resources; data and analytics; contact and collaboration; measurement practices |
| Warner (2013) | Journal article | USA and UK | Recidivism; early childhood education | Qualitative | Interviews | Unspecified number of practitioners (investors, policy makers, service providers) | Resources; data and analytics; contact and collaboration; measurement practices | Resources |

### 3.3 Quality Appraisal

Included studies were appraised for quality, noting that systematic bias can result in inaccurate or untrustworthy findings (Mhasker et al., 2009). See Table 13 for a summary of appraisal results. Applying the CASP checklist (also informed by COREQ), 3 studies (12%) were rated ‘high quality’; 13 studies (52%) were rated ‘medium quality’); and 9 studies (12%) were rated ‘low quality’.

**Table 13: Quality Appraisal using CASP tool ("methodological limitations" section of CERQual Assessment)**

| Quality criteria | Cooper et al (2013) | CRESSI (2016) | DCLG (2014) | DCLG (2015) | Disley et al (2011) | Disley and Rubin (2014) | Edmiston and Nicholls (2017) | Galley et al (2014) | Gustafsson-Wright et al (2015) | Gustafsson-Wright et al (2016) | Hughes and Scherer (2014) | James (2016) | KPMG (2014) | Murray et al 2012 | Reeder et al (2014) | Reeder et al (2015) | Roberts & Cameron (2014) | Ronicle et al (2014) | Ronicle & Stanworth (2015) | Rotheroe (2014) | Rudd et al (2013) | Tan et al (2015) | Tomkinson (2012) | Vennema (2016) | Warner (2013) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Was there a clear statement of the research goals? | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Is a qualitative methodology appropriate? | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ? | ✓ | ? | ✓ | ✓ | ✓ | ✓ |
| Was the research design appropriate to address the aims of the research? | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ? | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ? | ✓ | ? | ✓ | ✓ | ✓ | ✓ |
| Was the recruitment strategy appropriate to the research aims? | ✗ | ? | ? | ? | ✓ | ? | ✓ | ? | ? | ? | ? | ? | ✓ | ✓ | ✓ | ✓ | ? | ? | ? | ? | ? | ✓ | ✗ | ? | ✗ |
| Was the data collected in a way that addressed the research issue? | ? | ? | ? | ? | ✓ | ✓ | ? | ? | ? | ? | ? | ? | ✓ | ✓ | ? | ? | ? | ? | ? | ? | ? | ? | ✗ | ? | ✗ |
| Has the relationship between researcher & participants been considered? | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| Have ethical issues been taken into consideration? | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ? | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ? | ? | ✓ | ✗ | ✗ | ✗ |
| Was the data analysis sufficiently rigorous? | ✗ | ✗ | ✗ | ✓ | ✗ | ✓ | ? | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ | ? | ? | ? | ? | ? | ? | ? | ? | ✓ | ✗ | ✗ | ✗ |
| Is there a clear statement of findings? | ✗ | ✓ | ✓ | ✓ | ✓ | ✓ | ? | ? | ✓ | ✓ | ✓ | ? | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ? | ? | ✓ | ✓ | ✓ | ✓ | ✓ |
| How valuable is the research? | ? | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ? | ✓ | ✓ | ? | ? | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ? | ✓ | ✓ |
| **Quality Score (out of 10)** | 3 | 5 | 5 | 8 | 5 | 7 | 5 | 3 | 5 | 4 | 4 | 3 | 8 | 7 | 6 | 6 | 5 | 5 | 1 | 4 | 3 | 8 | 4 | 5 | 5 |

**Key:** 8-10 (High quality: Minor methodological concerns); 5-7 (Medium quality: moderate methodological concerns); 1-4 (Low quality: major methodological concerns)

### 3.4 Excluded Studies

At full-text review, 94 studies were excluded (Appendix 4). Articles were primarily excluded for not investigating barriers to or facilitators of rigorous evaluation (n = 31). Other exclusion reasons included; not primary research (i.e they did not include primary data collection) or a systematic review (n = 26); not about SIBs (n = 17); not retrievable (n = 5); lacks relevant sample (n = 9); and does not include perceptions of practitioners (n = 6).

### 3.5 What factors affect rigorous evaluation in SIBs?

Included studies reported a range of factors that act as barriers to and/or facilitators of undertaking or being involved in rigorous impact evaluations of SIBs. No studies ranked identified factors hierarchically or suggested a specific weighting. While barriers and facilitators were often related, they did not directly mirror each other. Table 14 outlines the top cited barriers and facilitators.

***Table 14: Top cited barriers and facilitators (n = number of studies cited in)***

| Top 5 facilitators to undertaking or being involved in rigorous impact evaluations of SIBs | Top 5 barriers to undertaking or being involved in rigorous impact evaluations of SIBs |
|---|---|
| Access to expertise in evaluation design and data management (n = 12) | Attribution challenges (e.g small sample size, overlapping service provision, no theory of change etc) (n = 13) |
| Access to required data (n = 11) | Challenges in identifying accepted outcome metrics (n = 13) |
| Availability of comparable counterfactual/control group (n = 10) | Lack of access to required data (n = 12) |
| Early and ongoing collaboration with stakeholders (n = 10) | Lack of financial resources for evaluation (n = 9) |
| Regular monitoring and performance management (n = 9) | Differing stakeholder incentives, views and needs (n = 8) |

#### *3.5.1 Descriptive themes*

To better interpret the factors reported by practitioners, the barriers and facilitators were categorised into descriptive themes. Across the 25 included studies, 4 overarching descriptive themes were identified, describing a total of 75 barriers to and facilitators of undertaking or being involved in rigorous impact evaluations. Figure 15 (barriers) and Figure 16 (facilitators) show the top reported descriptive themes; there are slight differences regarding the prevalence of themes.

***Figure 15: Top reported themes – factors operating as barriers***

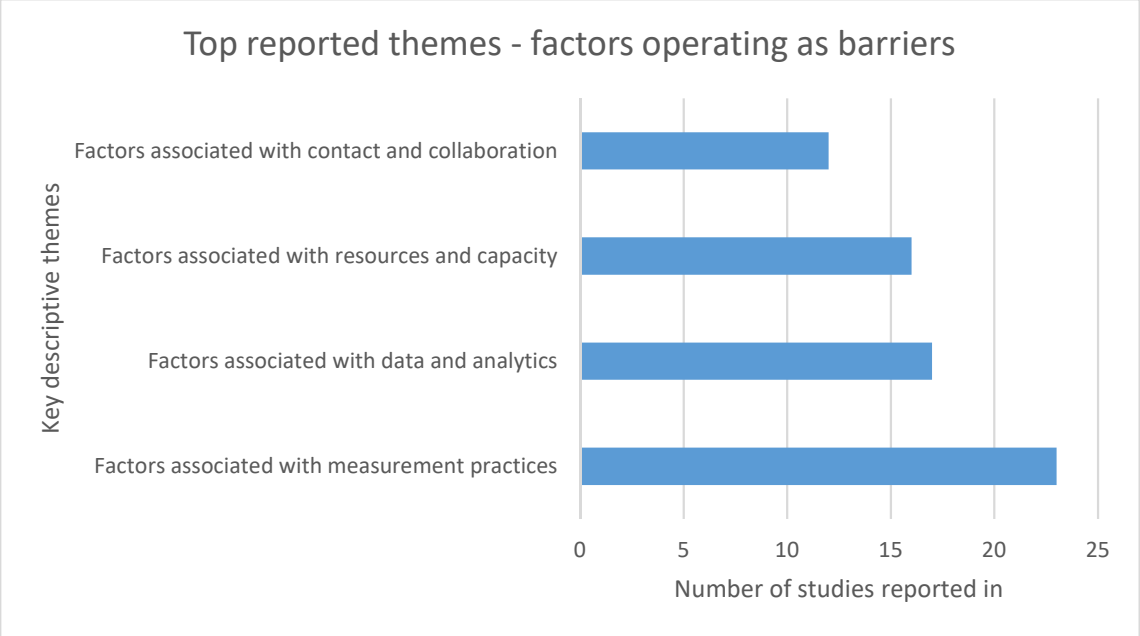



***Figure 16: Top reported themes – factors operating as facilitators***

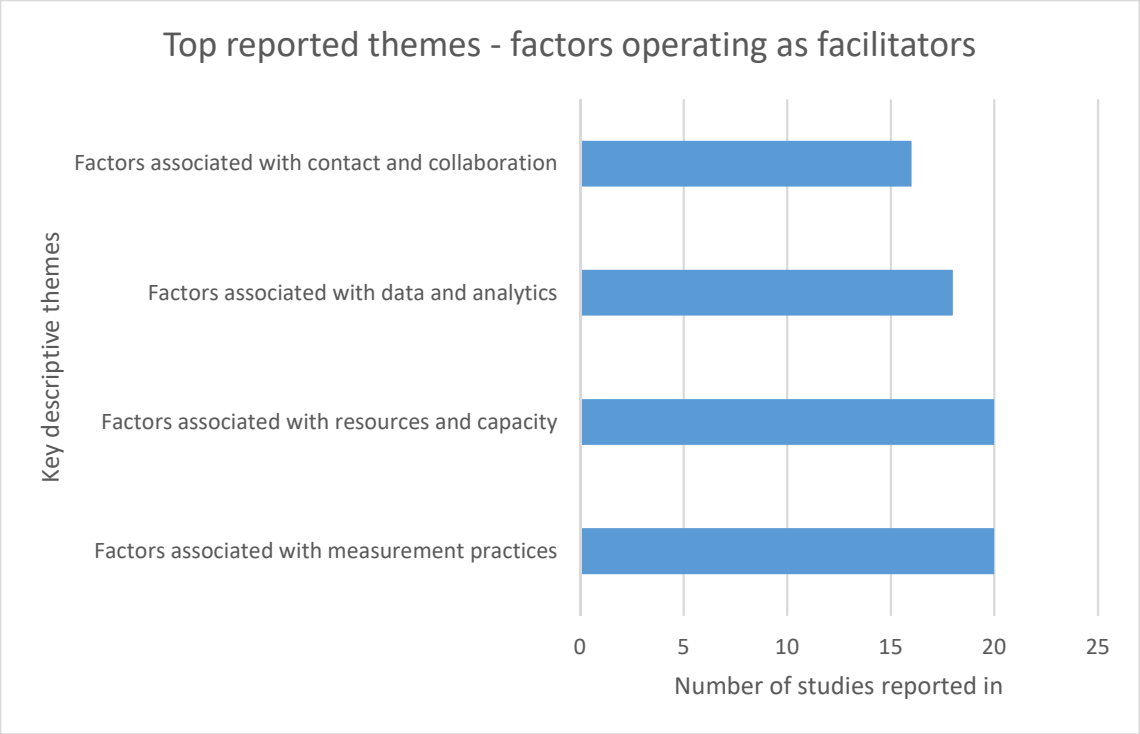



Table 15 summarises each of the descriptive themes and factors reported within them, including a CERQual confidence rating for each identified barrier and facilitator. See

Appendix 5 for further information about the CERQual confidence ratings. Below the table, each of the descriptive themes are described in detail, with supplementary information.

***Table 15: Barriers and Facilitators categorised into themes (n = number of studies cited in)***

| | Theme/finding | Number of studies cited in | Confidence in finding |
|---|---|---|---|
| | **Resources and capacity** | | |
| | ***Barriers*** | **16** | |
| 1 | Financial resources | 9 | High confidence |
| 2 | Administrative burden | 5 | High confidence |
| 3 | Time consuming/lack of time | 4 | Moderate confidence |
| 4 | Perceived compromise between evaluation and service delivery | 4 | High confidence |
| 5 | Staff or personnel resources | 3 | Moderate confidence |
| 6 | Lack of technical expertise and knowledge in measurement | 2 | High confidence |
| 7 | Resource intensive to hire external evaluators | 1 | Low confidence |
| | ***Facilitators*** | **20** | |
| 8 | Access to expertise in evaluation design and data management | 13 | High confidence |
| 9 | Use of independent evaluator | 7 | Low confidence |
| 10 | Adequate in house capacity | 6 | Moderate confidence |
| 11 | Capacity building and skills development | 6 | Moderate confidence |
| 12 | Additional staff responsible for monitoring and data collection | 6 | High confidence |
| 13 | Culture of monitoring and evaluation | 5 | Moderate confidence |
| 14 | Strong government support | 5 | Low confidence |
| 15 | Workable evaluation and data collection tools | 4 | Low confidence |
| 16 | Alliances with bodies who can fund evaluation | 4 | Low confidence |
| 17 | Investing in resources for data collection and analysis | 1 | Moderate confidence |
| | **Data and analytics** | | |
| | ***Barriers*** | **17** | |
| 18 | Data access | 12 | High confidence |
| 19 | Data quality | 5 | High confidence |
| 20 | Data systems issues | 5 | High confidence |
| 21 | Lack of independent performance data | 4 | Low confidence |
| 22 | Confidentiality and data protection issues | 2 | Moderate confidence |
| 23 | Enhanced data requirements | 2 | Moderate confidence |
| 24 | Lack of data collection and monitoring in social services | 2 | Moderate confidence |
| 25 | Data appropriateness | 1 | Moderate confidence |
| 26 | Data collection capacity | 1 | Very low confidence |
| 27 | Complexity of data monitoring and measurement | 1 | Moderate confidence |
| | ***Facilitators*** | **18** | |
| 28 | Access to required data | 11 | High confidence |
| 29 | Adequate systems for reporting, data collection and performance management | 7 | Low confidence |
| 30 | Data sharing/transparency | 6 | Low confidence |
| 31 | Benchmark data availability | 6 | Low confidence |
| 32 | Improved capacity for data capture and analysis | 5 | Low confidence |
| 33 | Strong data collection and analytics | 5 | Very low confidence |
| 34 | Culture of data collection and analysis | 4 | Low confidence |
| 35 | Clarity on data requirements | 2 | Low confidence |

| | | | |
|---|---|---|---|
| | **Contact and collaboration** | | |
| | ***Barriers*** | **12** | |
| 36 | Differing stakeholder incentives, views and needs | 8 | High confidence |
| 37 | Challenging to reach consensus between stakeholders | 6 | Low confidence |
| 38 | Perceived micro management by funders | 1 | Low confidence |
| 39 | Relationship tensions due to more detailed project management | 1 | Very low confidence |
| 40 | Cultural differences or tensions | 1 | Very low confidence |
| 41 | Power structures influence evaluation decisions | 1 | Very low confidence |
| 42 | Lack of understanding about role of investors and intermediaries | 1 | Very low confidence |
| | ***Facilitators*** | **16** | |
| 43 | Early and ongoing stakeholder consultation to refine and test outcome metrics | 10 | Low confidence |
| 44 | Regular monitoring and performance management | 9 | High confidence |
| 45 | Clarity on shared objectives and roles | 7 | Moderate confidence |
| 46 | High level of trust between parties involved | 3 | Very low confidence |
| 47 | Alignment of incentives between different stakeholder groups | 2 | Low confidence |
| 48 | Stakeholder agreement on success metrics | 2 | Very low confidence |
| 49 | Commitment to learn from each other | 2 | Moderate confidence |
| | **Measurement practices** | | |
| | ***Barriers*** | **23** | |
| 50 | Attribution challenges | 13 | High confidence |
| 51 | Challenges in identifying accepted outcome metrics | 13 | High confidence |
| 52 | Perverse incentives not removed or addressed | 8 | Moderate confidence |
| 53 | Trade-off between simplicity and robustness | 6 | High confidence |
| 54 | Difficult to standardise | 6 | High confidence |
| 55 | Ethics/practicalities of randomised approaches | 5 | Low confidence |
| 56 | Broad spectrum for 'acceptable' methodological practice | 5 | Moderate confidence |
| 57 | Vulnerability to policy changes | 5 | Low confidence |
| 58 | Concern about use/suitability of proxy measures | 5 | Low confidence |
| 59 | Few agreed metrics, tools, benchmarks | 3 | Low confidence |
| 60 | Ethics/practicalities of control group | 3 | Low confidence |
| 61 | Challenging to establish accurate baseline | 2 | Moderate confidence |
| 62 | Challenges of setting up and measuring multiple outcomes | 1 | Low confidence |
| | ***Facilitators*** | **20** | |
| 63 | Use of comparable counterfactual/control group | 10 | Low confidence |
| 64 | Safeguards against perverse incentives | 8 | High confidence |
| 65 | Easily quantifiable and measurable outcomes | 7 | Low confidence |
| 66 | Clear causal link between intervention and result | 7 | Moderate confidence |
| 67 | Clear theory of change | 6 | Very low confidence |
| 68 | Streamlined measurement methodology | 5 | Very low confidence |
| 69 | Use of measures that cannot be manipulated and are independent | 5 | High confidence |
| 70 | Focus on objective and hard outcomes | 4 | Very low confidence |
| 71 | Consideration of practical constraints | 4 | Very low confidence |
| 72 | Desire for rigorous evaluation of efficacy | 3 | Moderate confidence |
| 73 | Use of mixed methods | 3 | Very low confidence |
| 74 | Consideration of all system interactions | 2 | Moderate confidence |
| 75 | Measurement of processes | 2 | Very low confidence |

*Resources and capacity*

Some factors related to resources and capacity were frequently reported as hindering or supporting rigorous impact evaluations. The most frequently reported barrier was 'financial resources', with various studies (n = 9) emphasising that it is expensive to build

robust evaluation. Many studies noted that while evaluation methods such as RCT's are desirable in terms of high validity, they may make a SIB 'too expensive.' Any cost to the SIB (including evaluation) is covered by the commissioner; complex evaluation methods increase the cost of the outcome and may be perceived as impractical. Another frequently reported barrier (n = 5) was 'administrative burden' as significant resources are required for monitoring and evidencing outcomes. Studies noted that the requirements can be particularly onerous for service providers, and they often do not align to the realities of their work. Many studies (n = 4) noted that it is time consuming to build robust evaluation and there is often a perceived compromise between evaluation and service delivery. The most frequently reported facilitator (n = 12) was 'access to expertise in evaluation design and data management'. Studies emphasised the important role that evaluation experts can play in technical aspects such as specifying and developing outcome measures, as well as data management. Another frequently reported facilitator was 'capacity building and skills development (n = 7), with studies emphasising the importance of capacity building in government and NGO human service sectors. Skills could be developed in measurement, contracting, data collection and financial modelling. Other frequently reported facilitators include: adequate in house capacity (n = 6); use of an independent evaluator (n = 6); and additional staff responsible for monitoring and data collection (n = 6).

***Box 1: Example***

A key finding from the NSW Social Benefit Bonds Trial, was that capacity building will be vital for future bonds. The bonds require agency engagement and staff with specific technical skills related to measurement (KPMG 2014).

*"I advise NGOs that they need to be prepared to make an investment equal to around 2 FTEs to get a SIB off the ground. They also need access to skills covering financial modelling and experimental design, plus detailed knowledge of the current government system and how NGO services operate on the ground." —Jocelyn Bell, The Benevolent Society, AU*

*Data and analytics*

Factors related to data and analytics were frequently reported as both barriers to and facilitators of rigorous evaluation. Many studies noted that data is often difficult to access or non-existent (n = 12). Data access issues are also closely linked to problems of data quality (n = 5), with various studies highlighting that data is often of a poor quality. In many instances, it is designed to support casework, rather than to measure outcomes. Other frequently

reported barriers included: data systems issues (n = 5); and lack of independent performance data (n = 4). The most frequently reported facilitator related to data and analytics, was 'access to required data', including sufficient data from government and non-government providers (n = 11). Access to data was perceived as being important for measuring outcomes *and* unintended consequences. In addition to data access, many studies highlighted the importance of: adequate systems for reporting, data collection and performance management (n = 7); and data sharing and transparency (n = 6).

*Contact and Collaboration*

The most frequently reported barrier in this category was 'differing stakeholder incentives, views and needs' (n = 8). Many studies emphasised that it can be challenging to balance the conflicting agendas of each party. For example, outcome metrics need to be: suitably robust for commissioners; achievable for service providers; and understandable for investors. Another frequently reported barrier was the 'challenge of reaching consensus between stakeholders'. Many studies (n = 6) noted that it a time consuming and analytically complex process to develop outcome measures that have the confidence of *all* parties involved. The most frequently reported facilitator was 'early and ongoing collaboration to refine and test outcome metrics' (n = 10). Studies highlighted the importance of ensuring operational stakeholders are involved in identifying outcome indicators and evaluation goals. Another facilitator was regular monitoring and performance management (n = 9). In the absence of regular performance management, studies suggested that impact evaluations may provide crucial information too late. Frequent monitoring was perceived as being important for building a continuous feedback loop, thus allowing stakeholders to track outcomes and identify unintended consequences. Many studies (n = 7) also emphasised the importance of clarity on shared objectives and roles, particularly with respect to measurement issues.

*Measurement Practices*

Barriers and facilitators related to 'measurement practices' were regularly cited. Many studies (n = 13) reported 'attribution challenges' as a significant barrier, due to issues such as: small sample sizes; time lag between intervention and outcomes; potential overlap of services; lack of a counterfactual; and/or unable to attribute changes directly to SIB funding (no comparison site using alternate funding model). Another frequently reported barrier was 'challenges in identifying accepted outcome metrics' (n = 13). Studies highlighted the

complexity associated with defining outcome metrics; in some policy areas (e.g reducing loneliness), it may be difficult to quantify outcomes. There may be considerable debate about whether to use a single outcome or a 'basket' of outcomes. Other frequently reported barriers included: trade-off between simplicity and robustness (n = 7); difficult to standardise evaluation approaches (n = 6); ethics or practicalities of randomised approaches (n = 5); broad spectrum for 'acceptable' methodological practices (n = 5); and concern about use/suitability of proxy measures (n = 5). The most frequently reported facilitator was 'use of comparable counterfactual/control group'. Many studies (n = 10) noted that control groups are important for ensuring rigour and controlling for external factors. Another facilitator was the use of 'easily quantifiable and measurable outcomes'; some practitioners prefer to use binary (easy to understand) outcome measures (n = 8). 8 studies highlighted 'adequate safeguards against perverse incentives' as being particularly important in light of evidence of perverse incentives in PbR models. Other frequently reported facilitators included: clear causal link between intervention and result (n = 7); streamlined measurement methodology (n = 6); and clear theory of change (n = 6).

***Box 2: Example***

Agreement of outcome metrics is crucial in SIB negotiations, but it is a complex process, particularly as it is challenging to place monetary figures on social outcomes. In the Peterborough SIB, stakeholders described this process as being analytically complex and time consuming. Noting that it is usually 'impossible' to create perfect measures, they eventually decided to use a proxy measure as it was deemed to be operationally feasible and easy to measure (Disley et al., 2011).

*"It was a balance between operational feasibility and measurement because ... we had originally looked at ... short-sentence prisoners being discharged to Cambridgeshire. The trouble was that population was too small to be statistically significant, and so it was a back and forth of changing the cohort terms and definition to be statistically significant, but then also assessing the operational feasibility of whether we could make that work, what the cost of delivering that would be". (Interviewee 5, Director, Social Finance)*

*3.5.2 Sensitivity analysis*

To test the influence of study quality on results, a sensitivity analysis was conducted; 'low quality' studies were excluded and results were compared. The top five cited barriers remained the same (Table 16).

***Table 16 - sensitivity check – top reported factors operating as barriers (n = number of studies cited in)***

| Top reported barriers (all studies) | Top reported barriers (high quality and moderate quality only) |
|---|---|
| Attribution challenges (e.g small sample size, overlapping service provision, no theory of change etc) (n = 13) | Attribution challenges (e.g small sample size, overlapping service provision, no theory of change etc) (n = 11) |
| Challenges in identifying accepted outcome metrics (n = 13) | Challenges in identifying accepted outcome metrics (n = 10) |
| Lack of access to required data (n = 12) | Lack of access to required data (n = 9) |
| Lack of financial resources for evaluation (n = 9) | Lack of financial resources for evaluation (n = 7) |
| Differing stakeholder incentives, views and needs (n = 8) | Differing stakeholder incentives, views and needs (n = 7) |

The top reported facilitators changed slightly (Table 17); 'regular monitoring and performance management' was ranked fourth instead of fifth and 'clear causal link between intervention and result' was cited as the equal fifth facilitator.

***Table 17 - sensitivity check - top reported factors operating as facilitators (n = number of studies cited in)***

| Top reported facilitators (all studies) | Top reported facilitators (high quality and moderate quality only) |
|---|---|
| Access to expertise in evaluation design and data management (n = 13) | Access to expertise in evaluation design and data management (n = 8) |
| Access to required data (n = 11) | Access to required data (n = 8) |
| Availability of comparable counterfactual/control group (n = 10) | Availability of comparable counterfactual/control group (n = 7) |
| Early and ongoing collaboration with stakeholders (n = 10) | Regular monitoring and performance management (n = 7) |
| Regular monitoring and performance management (n = 9) | Early and ongoing collaboration with stakeholders (n = 6); clear causal link between intervention and result (n = 6) |

Excluding the low quality studies had no effect on the descriptive themes. The order of the main categories remained the same for barriers and facilitators, suggesting they are robust in this analysis (Table 18 and Table 19).

***Table 18 - sensitivity check - top reported descriptive themes - barriers (n = number of studies cited in)***

| Top reported themes - barriers (all studies) | Top reported themes - barriers (high quality and moderate quality only) |
|---|---|
| Factors associated with measurement practices (n = 23) | Factors associated with measurement practices (n = 15) |
| Factors associated with data and analytics (n = 17) | Factors associated with data and analytics (n = 13) |
| Factors associated with resources and capacity (n = 16) | Factors associated with resources and capacity (n = 11) |

| | |
|---|---|
| Factors associated with contact and collaboration (n = 12) | Factors associated with contact and collaboration (n = 9) |

***Table 19 - sensitivity check - top reported descriptive themes - facilitators (n = number of studies cited in)***

| **Top reported themes - facilitators (all studies)** | **Top reported themes - facilitators (high quality and moderate quality only)** |
|---|---|
| Factors associated with measurement practices (n = 20) | Factors associated with measurement practices (n = 13) |
| Factors associated with resources and capacity (n = 20) | Factors associated with resources and capacity (n = 13) |
| Factors associated with data and analytics (n = 18) | Factors associated with data and analytics (n = 10) |
| Factors associated with contact and collaboration (n = 16) | Factors associated with contact and collaboration (n = 9) |

*3.5.3 Contribution of individual studies*

The reviewer also analysed how each study contributed to the construction of the final themes of factors operating as barriers and facilitators. The tables below suggest that some studies were over-represented or under-represented in the thematic framework. Regardless, the sensitivity analysis suggested no clear association between study quality and groups of themes; future studies could further explore associations between specific populations and themes.

***Table 20: Contribution of individual studies on the identified themes of barriers***

| **Study** | Factors related to measurement practices | Factors related to data and analytics | Factors related to resources and capacity | Factors related to contact and collaboration | Total (n = number of themes study contributed to) |
|---|---|---|---|---|---|
| Cooper et al (2013) | ✓ | ✓ | × | × | 2 |
| CRESSI (2016) | ✓ | × | × | × | 1 |
| DCLG (2014) | ✓ | ✓ | ✓ | ✓ | 4 |
| DCLG (2015) | ✓ | ✓ | × | × | 2 |
| Disley et al (2011) | ✓ | ✓ | ✓ | ✓ | 4 |
| Disley and Rubin (2014) | ✓ | ✓ | × | × | 2 |
| Edmiston and Nicholls (2017) | ✓ | × | ✓ | ✓ | 3 |
| Galley et al (2014) | ✓ | ✓ | × | ✓ | 3 |
| Gustafsson-Wright et al (2015) | ✓ | ✓ | ✓ | × | 3 |
| Gustafsson-Wright & Gardiner (2016) | ✓ | ✓ | ✓ | × | 3 |

| | | | | | |
|---|---|---|---|---|---|
| Hughes and Scherer (2014) | ✓ | ✓ | ✓ | ✓ | 4 |
| James (2016) | ✓ | × | × | × | 1 |
| KPMG (2014) | ✓ | ✓ | ✓ | ✓ | 4 |
| Murray et al (2012) | ✓ | ✓ | × | × | 2 |
| Reeder et al (2014) | ✓ | ✓ | ✓ | ✓ | 4 |
| Reeder et al (2015) | ✓ | ✓ | × | ✓ | 3 |
| Roberts and Cameron (2014) | ✓ | ✓ | ✓ | × | 3 |
| Ronicle et al (2014) | ✓ | ✓ | ✓ | × | 3 |
| Ronicle and Stanworth (2015) | ✓ | × | ✓ | ✓ | 3 |
| Rotheroe (2014) | ✓ | × | ✓ | × | 3 |
| Rudd et al (2013) | ✓ | × | × | × | 1 |
| Tan et al (2015) | ✓ | ✓ | ✓ | ✓ | 4 |
| Tomkinson (2012) | ✓ | × | ✓ | × | 2 |
| Vennema (2016) | × | × | ✓ | ✓ | 2 |
| Warner (2013) | × | ✓ | ✓ | ✓ | 3 |
| **Total**<br><br>***(n = number of studies that contributed to final theme)*** | 23 | 17 | 16 | 12 | |
| *This table reflects how each study contributed to the construction of the final themes of factors operating as facilitators. Note: each study can contribute with multiple factors to the same theme. Shading indicates studies that were rated to be of low quality.* | | | | | |

***Table 21: Contribution of individual studies on the identified themes of facilitators***

| **Study** | Factors related to measurement practices | Factors related to resources and capacity | Factors related to data and analytics | Factors related to contact and collaboration | Total (n = number of themes study contributed to) |
|---|---|---|---|---|---|
| Cooper et al (2013) | × | × | ✓ | ✓ | 2 |
| CRESSI (2016) | ✓ | × | × | ✓ | 2 |
| DCLG (2014) | ✓ | ✓ | ✓ | ✓ | 4 |
| DCLG (2015) | ✓ | ✓ | × | × | 2 |
| Disley et al (2011) | ✓ | ✓ | ✓ | × | 3 |
| Disley and Rubin (2014) | ✓ | ✓ | ✓ | ✓ | 4 |
| Edmiston and Nicholls (2017) | × | × | × | × | |
| Galley et al (2014) | × | ✓ | ✓ | ✓ | 3 |
| Gustafsson-Wright et al (2015) | ✓ | ✓ | ✓ | ✓ | 4 |
| Gustafsson-Wright & Gardiner (2016) | ✓ | ✓ | ✓ | ✓ | 4 |
| Hughes and Scherer (2014) | ✓ | ✓ | ✓ | ✓ | 4 |
| James (2016) | ✓ | × | × | × | 1 |
| KPMG (2014) | ✓ | ✓ | ✓ | ✓ | 4 |
| Murray et al (2012) | × | × | ✓ | ✓ | 2 |
| Reeder et al (2014) | ✓ | ✓ | × | × | 2 |
| Reeder et al (2015) | ✓ | ✓ | × | × | 2 |
| Roberts and Cameron (2014) | ✓ | ✓ | ✓ | ✓ | 4 |
| Ronicle et al (2014) | ✓ | ✓ | ✓ | × | 3 |

| | | | | | |
|---|---|---|---|---|---|
| Ronicle and Stanworth (2015) | ✓ | ✓ | ✓ | × | 3 |
| Rotheroe (2014) | ✓ | ✓ | ✓ | ✓ | 4 |
| Rudd et al (2013) | ✓ | ✓ | ✓ | ✓ | 4 |
| Tan et al (2015) | ✓ | ✓ | ✓ | ✓ | 4 |
| Tomkinson (2012) | ✓ | ✓ | ✓ | ✓ | 4 |
| Vennema (2016) | ✓ | ✓ | ✓ | ✓ | 4 |
| Warner (2013) | × | ✓ | × | × | 1 |
| **Total**<br><br>***(n = number of studies that contributed to final theme)*** | 20 | 20 | 18 | 16 | |
| *This table reflects how each study contributed to the construction of the final themes of factors operating as facilitators. Note: each study can contribute with multiple factors to the same theme. Shading indicates studies that were rated to be of low quality.* | | | | | |

### *3.5.4 Analytical themes*

The reviewer identified three overarching analytical themes that describe the implications of the above findings (each discussed below).

**1) SIB practitioners often have differing opinions as to what constitutes a 'rigorous' impact evaluation and the importance of rigorous evaluation**

The term 'rigorous impact evaluation' is contested in this field. Practitioners often have differing opinions regarding the degree of evaluation rigor needed within a SIB transaction. This was exemplified by many of the reported barriers and facilitators. Some argue that an indication of causality is *not* required in the context of SIBs, as outcome payments may be triggered through validation of service provider or government administrative data (e.g number of days spent in out-of-home-care). Others argue that quasi-experimental or experimental evaluation is worthwhile for providing evidence that the intervention caused the outcomes, thus contributing to the knowledge base of 'what works'. To move the discussion forward, it may be necessary to further develop the notion of "appropriateness" of evaluation rigour in this field. Rigour and quality will arguably always remain important; however, it is important to note that SIBs are typically evaluated in complex environments where simple causal relationships are difficult to establish. Thus, various types of evidence (not merely evidence of intervention effect) may be required. Multiple research methods are likely to be important (e.g experiments, observations, interviews etc) and each should be underpinned by the relevant standards of quality and validity. Therefore, rather than following a single hierarchy or approach, an 'appropriate' level of evaluation rigour would take into account the research question and practical

constraints, with the quality assessment being derived from the relevant research tradition/methodology.

**2) Traditional frameworks of causality may be over-restrictive for practitioners and inappropriate for this emerging field**

While RCTs provide the best causal evidence for intervention effectiveness, they are not always feasible or ethical, particularly in SIB transactions that have a complex mix of stakeholders and resource constraints. Many of the frequently cited barriers and facilitators related to challenges with establishing causality and creating an ideal counterfactual scenario. Some academics have critiqued the standard of evaluation in this field as a result, suggesting that it is not sufficiently rigorous. A counter argument is that the standard of evidence is not necessarily poor; it simply does not conform to the traditional 'causal inference' movement that has become dominant in the 21st century. This movement largely promotes RCTs for assessing causality, at the expense of other (potentially useful) methods (Vandenbroucke et al., 2016). Consequently, evidence may be prematurely rejected due to perceived lack of rigour. The 'counterfactual framework' is merely one way of assessing causality. There are many other methods which may be more inclusive; for example, using triangulation, which involves 'strengthening causal inferences by integrating results from several different approaches where each approach has different key sources of potential bias' (Lawlor et al., 2016). Rather than viewing non-statistical evidence as inferior, researchers should aim to use a wide variety of non-statistical evidence of mechanisms (e.g case studies, expert testimony, interviews) in conjunction with statistical evidence of correlation (Clarke et al., 2013). The barriers and facilitators identified in this paper suggest that it is important to use the totality of (complementary) evidence available in each situation; this allows researchers to search for converging sources of evidence, rather than relying on a single (sometimes unattainable) source of evidence which "will fail significantly often" (Clarke et al., 2012: 747). In light of the regularly cited practical constraints and complexity involved with establishing attribution in SIBs, it may be important to reconsider and loosen the traditional (narrow) notions about causality.

**3) The complexity of the systems in which SIB-funded interventions are implemented effects efforts to evaluate impact**

Many practitioners highlighted barriers and facilitators related to the complex systems in which interventions are implemented. SIBs are implemented in complex adaptive systems, where interventions are likely to have "diverse, far reaching and non-linear effects" (Shiell et al., 2008). Attention has been drawn to such issues in recent public health debates and literature relating to complex systems models of evidence (Rutter et al., 2017). Within the field of public health, the evidence underpinning responses to many contemporary public health challenges has been generated using tools and methods that were developed to answer questions about the effectiveness of clinical interventions (Shiell et al., 2008). Therefore, the tools and methods are based on linear models of cause and effect; evaluation of outcomes often involves measuring changes at an individual level and summing those to capture the 'social' impact. The properties of a complex system cannot simply be predicted from the elements within it, as they are "more than just the sum of its parts" (Rutter et al., 2017: 1). Consequently, a complex systems model of public health: 1) conceptualises health inequities as outcomes of 'a multitude of interdependent elements within a connected whole'; 2) recognises that these elements effect each other in subtle, often non-linear ways; and 3) uses a broad spectrum of methods to evaluate interventions in recognition of this complexity (Rutter et al., 2017: 1). Within the context of SIBs, it is also challenging to use a simplistic model of cause and effect, as the 'intervention' is not necessarily a single event resulting in a clearly discernible outcome. SIBs involve a complex mix of stakeholders, commissioning processes, social interactions, conditions and environments. Thus, it is important to consider the ways in which the processes and outcomes drive change, at all points within the system. Further exploring how to effectively evaluate systems level change would be worthwhile in this emerging field. It would be valuable as many existing evaluations have focused purely on the outcomes of individuals who have received a SIB-funded intervention, thus overlooking the complexity and systems level interactions.

# CHAPTER 4 – DISCUSSION

This systematic review and thematic synthesis was informed by an Expert Advisory Panel of 10 experienced practitioners. This discussion section will highlight some of the perspectives of panel members. It will also outline study strengths and limitations, and implications for practice and research.

## 4.1 Perspectives of panel members

In completing this thesis, it was beneficial to receive continuous feedback from panel members, who each contributed perspectives drawing on their area of expertise. Firstly, it is worth noting that many practitioners felt that academic writing on this subject is overly critical and not always informed by empirical findings. Some academics have suggested that evaluation practices are limited and the success of SIBs is largely determined by their high profile and innovativeness (e.g Roy et al., 2017). However, many panel members felt that there are some examples of quality evaluations in this field and evaluations are often more detailed than those conducted for comparable non-SIB funded interventions (many of which are not evaluated at all). This perspective was not unanimous, as one panel member noted that many SIB evaluations have ignored the possibility that outcomes may have improved without a SIB; outcomes are often measured without an appropriate counterfactual. This panel member suggested that it may not be possible to properly demonstrate the social value that has been delivered, without a Cochrane styled investigation.

With regards to the first analytical theme, differing perspectives emerged regarding the necessity of conducting 'rigorous' evaluation. For example, one panel member emphasised that 'sophisticated' evaluation is often of limited interest to the payee (i.e the commissioner) for financial reasons. Apart from any funding obtained from the Life Chances Fund or Commissioning Better Outcomes Fund, the commissioner provides the only source of 'income' (outcome payments) in a transaction. Thus, any costs to the SIB (including evaluation costs) ultimately increase the price paid per outcome. Another panel member had similar perspectives, noting that evaluation costs need to be practical and proportionate to the scale of the SIB contract; in some instances, (e.g if an intervention has already been tested with the target group), the additional cost of using a robust method may not be worthwhile. In relation to the second analytical theme, various panel members agreed that it would be worthwhile to challenge the traditional frameworks of assessing causality, as evidence in this

field is sometimes dismissed due to factors such as small sample sizes and lack of a counterfactual. With regards to the third analytical theme, panel members broadly agreed on the importance of measuring change, at all points within the system. Two panel members strongly emphasised the importance of assessing the impact of the SIB-funding, rather than merely assessing intervention effects. As a SIB adds costs to the intervention (both transactional costs and in terms of time and capacity), they argued that it is important for the commissioner to understand whether the SIB funding has delivered added value (compared to a non-SIB funding option). Another panel member emphasised that systems level evaluations are required due to the complexity of the systems in which SIBs are implemented. Overall, the feedback provided by panel members was integral as it assisted the reviewer to better understand the differing perspectives between different groups of stakeholders.

### 4.2 Reflexivity

This review was primarily conducted by a single researcher (with guidance from an Expert Advisory Panel) who had existing professional experience working on SIBs in Australia. The researcher needed to consider their own prejudices and biases about factors such as what 'rigorous evaluation is' and the importance of 'rigorous' evaluation. The reviewer did not convey particular opinions to panel members and documented all decisions when coding themes, to ensure that prejudices or biases did not influence results. While the reviewer was initially more familiar with the grey literature relating to Australian SIBs, they expanded the search beyond these sources by searching reference lists, hand searching journals, and contacting experts. Thus, the reviewer is confident that the majority of the relevant (English) documents were retrieved, covering various jurisdictions that have launched SIBs. The quality appraisal of studies was also important for mitigating potential biases.

### 4.3 Strengths and weaknesses of this study

This systematic review is exhaustive and it was conducted following PRISMA guidelines. The reviewer followed a protocol which was distributed to panel members at the outset. Panel members provided input in relation to the research question and methods; included and excluded studies; and the author's interpretation of findings. They also answered questions and provided clarifications as required. This process ensured that various viewpoints were taken into account and it minimised potential biases. To overcome

limitations with ‘vote counting’ (discussed below), the author conducted a thematic synthesis and organised the factors into descriptive and analytical themes; and conducted extensive quality appraisal to explore the trustworthiness of the findings.

Regardless, there are also a number of limitations. As with most reviews of this type, the study incorporated a form of ‘vote counting’, as the factors reported as barriers and facilitators were initially aggregated, to identify the top cited barriers and facilitators. While this can be a useful initial step, there are several limitations associated with ‘vote counting’. For example, it does not take into account factors such as study size, study quality or the weighting of each factor. Another limitation is that the search was limited to English-language documents (for practical reasons); however, there are a number of potentially relevant non-English publications (as SIBs have now been launched in 20 countries). For pragmatic reasons, it was not possible for studies to be double-screened and double-coded. There were inherent limitations in the included studies, which may limit the generalisability of the findings. For example, different types of stakeholders often have different perspectives regarding the barriers to and facilitators of rigorous evaluation. However, studies did not always clearly specify how many practitioners of each type were included in the sample; or clearly attribute perspectives to particular practitioner groups. The included studies were not specifically focused on identifying barriers to or facilitators of rigorous evaluation (although they did discuss either barriers or facilitators). The reviewer therefore had to use considerable judgement and analysis, to create codes and themes. It is possible that the themes may have been different if double-coded. Further, the inclusion criteria focused on barriers and/or facilitators as perceived by practitioners. However, it may be possible to identify factors affecting evaluation practices, without relying on the perceptions of research participants (e.g ethnographic studies). Thus, some relevant evidence may have been overlooked.

### 4.4 Strengths and weaknesses of this study in relation to other studies

Existing systematic reviews related to SIBs have not focused on the barriers to or facilitators of rigorous evaluation. Fraser et al (2016), aimed to bring together academic and grey literature on SIBs to explore themes related to public versus private values; outcomes contracting; and calculating and allocating risk. Tan et al (2015), aimed to examine what the empirical literature suggests in terms of what SIBs are, how the function and what the early results demonstrate. In comparison to recent systematic reviews, this paper offers a more

focused analysis of the factors affecting evaluation practices. In this context, it could be argued that simply identifying the main barriers and facilitators offers an original and useful contribution.

### 4.5 Implications for practice

The findings suggest that there are various ways to enhance evaluation practices and overcome barriers in this field. The top cited facilitator was 'access to expertise in evaluation design and data management' (n = 12); a number of studies also mentioned the importance of 'capacity building and skills development' (n = 7). Surprisingly, financial support was not regularly cited as a facilitator, apart from when studies mentioned the possibility of building alliances with bodies or organisations who can fund evaluation (n = 4). This suggests that it would not be sufficient to merely increase funding for evaluation. Instead, the analysis highlights the importance of improving support for undertaking rigorous evaluation, particularly in terms of access to experts and capacity building in areas such as evaluation design and data management. SIBs rely on the availability of high quality data from government and service providers; thus, steps should be taken to respond to challenges in accessing timely and accurate data. Many studies mentioned the challenges associated with identifying accepted outcome metrics (n = 13), suggesting that there may be a need for more commonly accepted outcome indicators, along with evaluation guidelines that could be used by practitioners. While it is difficult to standardise many aspects of evaluation due to the diversity and complexity of SIB transactions, it may be possible to develop further tools or guidelines to assist practitioners.

Another frequently identified barrier (n = 6) in rigorously evaluating SIBs was the perceived difficulty in standardising evaluation processes or practices, or comparing the results of SIB-funded interventions, as they are often implemented in different contexts, for different populations. This can result in a perception that rigorous evaluation is not possible in the complex landscape of SIBs. However, Hawe et al (2009), suggest a new approach, in which an intervention is conceptualised dynamically as an event (or series of events) within a system; the process and sequence of change can remain the same in all sites, but the form (i.e specific program components) may differ. When evaluating future SIBs, practitioners could adopt this reconceptualised notion of standardisation; theory-based evaluations could be used

to test causal mechanisms and processes within an individual SIB contract and to evaluate and compare interventions across contexts.

### 4.6 Implications for future research

The findings of this review and thematic synthesis could be used by research centres such as the Government Outcomes Lab (Go Lab) to better understand how the academic community can contribute to meeting the challenges associated with conducting more robust impact evaluations. Importantly, there are several ways in which future research could build upon the findings of this review. While 'practitioners' were defined broadly, future research could expand the inclusion criteria to include studies that identify barriers or facilitators without relying on the perceptions of practitioners. This may include studies that produce observational data about evaluation practices and cultures. Rather than explicitly focusing on barriers and facilitators (which could be regarded as restrictive), future studies could look more broadly on the experiences of practitioners when undertaking impact evaluations. While this study focused on impact evaluations at the level of an individual SIB transaction (i.e whether the SIB-funded intervention produced intended effects), future studies should also explore the barriers to and facilitators of evaluating the impact of SIB funding (i.e the impact of the SIB funding mechanism compared to non-SIB funding mechanisms). As a number of the included studies had serious methodological flaws (e.g small sample size, poor reporting), this study suggests that a qualitative investigation could be particularly worthwhile. This could expand the availability of empirical evidence focused on understanding evaluation practices. Practitioner groups that are currently under represented in the literature (e.g evaluators and academics) should be included in the investigation, along with other stakeholder groups. A qualitative investigation could focus on understanding how and why evaluation practices differ between different jurisdictions, policy areas and stakeholder groups.

# CHAPTER 5 – CONCLUSION

This stakeholder-informed systematic review and thematic synthesis contributes towards better understanding the barriers to and facilitators of undertaking or being involved in rigorous impact evaluations of SIBs. The review focused on individual scheme level evaluation (i.e efficacy of individual programmes that have been funded using a SIB structure). Importantly, it is the first systematic review specifically focused on identifying factors that either hinder or support rigorous evaluation; and it incorporated (often overlooked) perspectives of practitioners. To minimise biases and enhance research relevance, this review was informed by an Expert Advisory Panel. This review found that practitioners face considerable challenges in terms of: adequately assessing attribution; identifying accepted outcome metrics; accessing required data; funding complex evaluations; and balancing differing stakeholder's incentives, views and needs. Conversely, the review found that there are a number of facilitators which can enable rigorous evaluation, including: access to appropriate expertise; access to required data; availability of comparable counterfactual group; early and ongoing stakeholder consultation to refine and test outcome metrics; and regular monitoring and performance management. Existing empirical studies have not systematically categorised the most commonly experienced barriers and facilitators. Thus, identifying the barriers and facilitators (using a thematic synthesis) is an integral and original result, given the lack of academic research in this area and the diversity of the available grey literature. Importantly, the findings of this thesis will be presented at an upcoming conference ("Progressing the debate on Social Impact Bonds") hosted by the Newcastle University Business School; London School of Hygiene and Tropical Medicine (Policy Innovation Research Unit); Blavatnik School of Government (Government Outcomes Lab) and RAND Europe. Ultimately, the findings may be used to inform future efforts to incentivise more rigorous impact evaluations and to remove associated barriers.

**Sources of support**

No internal or external funding was obtained for this report. It was completed in partial fulfilment of the MSc in Evidence-Based Social Intervention and Policy Evaluation.

## REFERENCES

### Included studies


Cooper, C., Graham, C., O'Dwyer, B. (2013). Social impact bonds: can private finance rescue public programs? Accounting, Organizations and Society Conference on "Performing Business and Social Innovation through Accounting Inscriptions", Galway, Ireland. Available at: https://pure.strath.ac.uk/portal/en/publications/social-impact-bonds(a376dd30-45f8-4b6b-a440-713a63130edc)/export.html.

CRESSI. (2016). CRESSI Policy Roundtable Summary: Social Innovation and Social Impact Bonds: The Role of Private Capital in the Delivery of Welfare Services, CRESSI Deliverable 6.3. Available at: https://www.sbs.ox.ac.uk/sites/default/files/research-projects/CRESSI/docs/cressi-policy-roundtable-summary-SIBs.pdf.

Department for Communities and Local Government (DCLG). (2014). Qualitative Evaluation of the London Homelessness Social Impact Bond: First Interim Report, London. Available at: https://www.gov.uk/government/publications/qualitative-evaluation-of-the-london-homelessness-social-impact-bond-first-interim-report.

Department for Communities and Local Government (DCLG) (2015). Qualitative Evaluation of the London Homelessness Social Impact Bond: Second Interim Report, London. Available at: https://www.gov.uk/government/publications/qualitative-evaluation-of-the-london-homelessness-social-impact-bond-second-interim-report.

Disley et al (2011). Lessons learned from the planning and early implementation of the Social Impact Bond at HMP Peterborough, Ministry of Justice. Available at: https://www.gov.uk/government/publications/social-impact-bond-at-hmp-peterborough.

Disley and Rubin (2014). Phase 2 report from the payment by results Social Impact Bond pilot at HMP Peterborough, Ministry of Justice Analytical Series. Available at: https://www.gov.uk/government/publications/phase-2-report-from-the-payment-by-results-social-impact-bond-pilot-at-hmp-peterborough.

Edmiston, D., & Nicholls, A. (2017). Social Impact Bonds: The Role of Private Capital in Outcome-Based Commissioning. *Journal of Social Policy*, 1-20. doi:10.1017/S0047279417000125

Galley, A., McIsaac, E., Ymeren, J. (2014). From Impact to Investment: The NFP Experience with Social Impact Bonds, Mowat Research. Available at: https://mowatcentre.ca/from-investment-to-impact/.

Gustafsson-Wright et al (2015). The Potential and Limitations of Impact Bonds: Lessons from the first five years of experience worldwide, Brookings: Global Development and Economy. Available at: https://www.brookings.edu/wp-content/uploads/2015/07/impact-bondsweb.pdf.

Gustafsson-Wright and Gardiner (2016). Using Impact Bonds to Achieve Early Childhood Development Outcomes in Low- and Middle-Income Countries, Brookings: Global Development and Economy. Available at: https://www.brookings.edu/research/using-impact-

bonds-to-achieve-early-childhood-development-outcomes-in-low-and-middle-income-countries/.

Hughes, J., Scherer, J (2014). Foundations for Social Impact Bonds: How and Why Philanthropy Is Catalyzing the Development of a New Market, Social Finance. Available at: https://www.missioninvestors.org/tools/foundations-for-social-impact-bonds-how-and-why-philanthropy-is-catalyzing-the-development-of.

James, D. (2016). Understanding the Capacity and need to take on Investment within the Social Sector: Case Studies, NCVO. Available at: https://www.ncvo.org.uk/images/documents/policy_and_research/funding/investment-within-the-social-sector-2016-case-studies.pdf.

KPMG (2014). Evaluation of the Joint Development Phase of the NSW Social Benefit Bonds Trial, Government Advisory Services. Available at: http://www.osii.nsw.gov.au/assets/office-of-social-impact-investment/files/Evaluation-of-the-Joint-Development-Phase.pdf.

Murray, D., Jones, J., Pearce, S., Hichens, E (2012). Findings and lessons learned from the early implementation of the HMP Doncaster payment by results pilot. Ministry of Justice Research Series 18/12. Available at: https://www.gov.uk/government/publications/findings-and-lessons-learned-from-the-early-implementation-of-the-hmp-doncaster-payment-by-results-pilot.

Reeder, N., Rocyn Jones, G., Loder, J., Colantonio, A. (2014). Measuring Impact: Preliminary insights from interviews with impact investors, LSE Cities, London School of Economics and Political Science. Available at: https://files.lsecities.net/files/2014/05/Measuring-Impact-Report-V3.pdf.

Reeder, N., Colantonio, A., Loder, J., Rocyn Jones, G. (2015). Measuring impact in impact investing: an analysis of the predominant strength that is also its greatest weakness, *Journal of Sustainable Finance & Investment*, 5:3, 136-154. Available at: DOI: 10.1080/20430795.2015.1063977.

Roberts, L & Cameron, G. (2014). Evaluation of the Essex Multi Systemic Therapy Social Impact Bond. Interim Evaluation Report, OPM, UK. Available at: https://www.opm.co.uk/wp-content/uploads/2015/10/Interim-report-Essex-MST-SIB-Evaluation.pdf.

Ronicle, J., Stanworth, N., Hickman, E., Fox, T. (2014). Social Impact Bonds: The State of Play (Full Report), Commissioning Better Outcomes Evaluation. Available at: https://www.biglotteryfund.org.uk/research/social-investment/publications.

Ronicle, J & Stanworth, N. (2015). Ways to Wellness Social Impact Bond: The UK's First Health SIB: Deep Dive Report, Commissioning Better Outcomes Evaluation. Available at: https://www.biglotteryfund.org.uk/research/social-investment/publications.

Rotheroe (2014). Lessons and opportunities: perspectives from providers of social impact bonds, New Philanthropy Capital (NPC). Available at: http://www.thinknpc.org/publications/lessons-and-opportunities/.

Rudd, T., Nicoletti, E., Misner, K., Bonsu, J. (2013). Financing Promising Evidence-Based Programs: Early Lessons from the New York City Social Impact Bond, MDRC. Available at: http://files.eric.ed.gov/fulltext/ED545464.pdf.

Tan, S., Fraser, A., Giacomantonio, C., Kruithof, K., Sim, M., Lagarde, M., Disley, E., Rubin, J., Mays, N. (2015). An evaluation of Social Impact Bonds in health and social care: interim report, Policy Innovation Research Unit (PIRU) Publication 2015 – 12, Department of Health Services Research and Policy London School of Hygiene and Tropical Medicine, and RAND Europe. Available at: http://researchonline.lshtm.ac.uk/2391521/.

Tomkinson, E. (2012). Perspectives from the Social Finance Forum 2012: An Australian Snapshot - Social Impact Bonds, the Centre for Social Impact (CSI). Available at: http://www.csi.edu.au/research/project/social-impact-bonds-australian-snapshot/.

Vennema, B. (2016). Evaluating Social Impact Bonds: Developing a generic evaluation tool for Social Impact Bonds. Available at: uuid:7c0d195f-3d7c-41b9-a199-0a24f5d5a9ed.

Warner, M. (2013). Private finance for public goods: social impact bonds, Journal of Economic Policy Reform, 16:4, 303-319. Available at: http://dx.doi.org/10.1080/17487870.2013.835727.

**Excluded studies**

Addis, R., McLeod, J., Raine, A. (2013). IMPACT–Australia Investment for social and economic benefit, Australian Government, Department of Education, Employment and Workplace Relations. Available at: https://docs.employment.gov.au/system/files/doc/other/impact-australia_nov_2013_2.pdf

Addis, R., Bowden, A., Simpson, D. (2014). 'Delivering on Impact: The Australian Advisory Board Breakthrough Strategy to Catalyse Impact Investment, IMPACT Investing Australia. Available at: https://impactinvestingaustralia.com/wp-content/uploads/0109Delivering_on_impact.pdf

Aylott, M., & Shelupanov, A. (2011). Social Impact Bonds in criminal justice: from interesting idea to business as usual, *Prison Service Journal*, 195, May 2011, pp.3-8, Her Majesty's Prison Service of England and Wales. Available at: http://www.scie-socialcareonline.org.uk/social-impact-bonds-in-criminal-justice-from-interesting-idea-to-business-as-usual/r/a1CG0000000GTgEMAW

Hideaki, B., Tsukamoto, I., Yoshikoa, T. (2016). A Comparative Study on Impact Measurement of Social Impact Bonds for Social Service Innovation, Paper submitted to the 12th International conference of the International Society for Third Sector Research, Ersta Sköndal University College, Stockholm, Sweden, June 28 - July 1, 2016. Available at: http://baba-hi72.up.seesaa.net/image/ISTR_baba_etal_20160630.pdf.

Baliga, S. (2013). Shaping the success of social impact bonds in the united states: Lessons learned from the privatization of U.S. prisons, *Duke Law Journal*, [Vol. 63:437]. Available at: http://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=3403&context=dlj.

Belinsky, M., Eddy, M., Lohmann, J., Georgea, M. (2014). The application of social impact bonds to universal health-care initiatives in South-East Asia. *WHO South East Asia J Public Health*. 2014 Jul-Dec;3(3):219-225. DOI: 10.4103/2224-3151.206744.

Bengo, M., & Calderini, M. (2016). New development: Are social impact bonds (SIBs) viable in Italy? A new roadmap. Public Money & Management, 06 June 2016, Vol.36(4), p.303-306. http://dx.doi.org/10.1080/09540962.2016.1162999.

Bertha Centre (2014). Final Policy Paper: Exploration of Social Impact Bonds for SME Development. Available at: http://www.socialfinance.org.uk/wp-content/uploads/2014/05/Bertha-Policy-Paper.pdf.

Breuls, L. (2014). The Quest for Welfare State Sustainability: Managing Youth Unemployment by Social Impact Bonds? Utrecht University, Faculty of Social and Behavioural Sciences Theses. Available at: https://dspace.library.uu.nl/handle/1874/296991.

Brown, A., Norman, W. (2011). Lighting the touchpaper. Growing the Market for Social Investment in England, Boston Consulting Group and the Young Foundation, London, 2011. Available at: https://www.bcg.com/documents/file92199.pdf.

Callanan, L., Law, J., Medonca, L. (2012). From Potential to Action: Bringing Social Impact Bonds to the US. McKinsey & Company. Available at: http://www.mckinsey.com/industries/social-sector/our-insights/from-potential-to-action-bringing-social-impact-bonds-to-the-us.

Castellas, E., Findlay, S., Addis, R. (2016). Benchmarking Impact - Australian Impact Investment Activity and Performance Report 2016. IMPACT Investing Australia. Available at: https://impactinvestingaustralia.com/wp-content/uploads/Benchmarking-Impact.pdf.

Cooper, C., Graham, C., Himick, D. (2016). Social Impact Bonds: The securitization of the homeless. *Accounting, Organizations and Society*, Volume 55, November 2016, Pages 63-82. https://doi.org/10.1016/j.aos.2016.10.003.

Costa, K. (2014). Investing for Success Policy Questions Raised by Social Investors, Center for American Progress. Available at: https://www.americanprogress.org/issues/general/reports/2014/03/03/85099/investing-for-success/.

Costa, K., & Tomasko, L. (2014). Networking for Success: Building Networks Is Essential to Investment in Social Impact Bonds, Center for American Progress. Available at: https://www.americanprogress.org/issues/general/reports/2014/03/03/85106/networking-for-success/ .

Crowley, M (2014). The Role of Social Impact Bonds in Pediatric Health Care. *Pediatrics*, August 2014, Volume 134 / Issue 2. Available at: http://pediatrics.aappublications.org/content/134/2/e331.

Dagher Jr, P. (2013). Social Impact Bonds and the Private Benefit Doctrine: Will Participation Jeopardize a Nonprofit's Tax-Exempt Status? *Fordham Law Review*. Volume 81: 3479. Available at: http://ir.lawnet.fordham.edu/flr/vol81/iss6/10.

Dandala, G. (2015). Exploration of Social Impact Bonds to fund social development in South Africa. Gordon Institute of Business, University of Pretoria. Available at: http://repository.up.ac.za/handle/2263/52374.

Deloitte (2014). Social Impact Bonds in Canada: Investor insights. Available at: https://www2.deloitte.com/ca/en/pages/financial-services/articles/social-impact-bonds-investor-insights.html.

Dembek, K., Madhavan, D., Michaux, F., Potter, B. (2016). Impact Investing Australia: 2016 Investor Report, Impact Investing Australia Ltd. Available at: https://impactinvestingaustralia.com/wp-content/uploads/Impact-Investing-Australia-2016-Investor-Report.pdf.

Department for Work and Pensions. (2016). Qualitative evaluation of the DWP Innovation Fund: final report. Research Report No 922. Available at: https://www.gov.uk/government/publications/qualitative-evaluation-of-the-dwp-innovation-fund-final-report.

Disley, E., Giacomantoni, C., Kruithof , K., Sim, M. (2015). The payment by results Social Impact Bond pilot at HMP Peterborough: final process evaluation report, Ministry of Justice. Available at: https://www.gov.uk/government/publications/social-impact-bond-pilot-at-hmp-peterborough-final-report.

Dowling, E. (2016). In the wake of austerity: social impact bonds and the financialisation of the welfare state in Britain, *New Political Economy*, Volume 22, pp 294-310. Available at: http://dx.doi.org/10.1080/13563467.2017.1232709.

Drew, R., & Clist, P. (2015). Evaluating Development Impact Bonds: A Study for DFID. DFID, London, UK (2015) 62 pp.

Dugger, R., & Litan, R. (2012). Early Childhood "Pay-For-Success" Social Impact Finance: A PKSE Bond Example to Increase School Readiness and Reduce Special Education Costs, A Report of the Kauffman Foundation - ReadyNation Working Group on Early Childhood Finance Innovation. Available at: http://humcap.uchicago.edu/RePEc/hka/wpaper/Dugger_Litan_2012_early-childhood-pay.pdfc.

Ebrahim, A., & Rangan, V. (2014). What Impact? A Framework for Measuring the Scale & Scope of Social Performance. California Management Review 56, no. 3 (Spring 2014): 118–141. Available at: http://www.hbs.edu/faculty/Pages/item.aspx?num=47515.

Ecorys UK., & ATQ Consultants. (2016). Reconnections Social Impact Bond: reducing loneliness in Worcestershire, Case Study Report. Available at: https://www.bigsocietycapital.com/latest/type/research/reconnections-social-impact-bond-reducing-loneliness-worcestershire.

Edmiston, D. (2016). CRESSI Practitioner Seminar Summary: Social Impact Bonds: Opportunities and Challenges for Social Innovation, CRESSI Deliverable 6.4. Available at: https://www.sbs.ox.ac.uk/sites/default/files/research-projects/CRESSI/docs/cressi-practitioner-seminar-summary-sibs-d6.4.pdf.

Evenett, R., & Richter, K. (2011). Making Good in Social Impact Investment: Opportunities in an Emerging Asset Class, The Social Investment Business, TheCityUK. Available at: http://www.engagedx.com/downloads/Making_Good_In_Social_Impact_Investment.pdf.

EY. (2016). Social impact investing research: Department of Social Services for the Prime Minister's Community Business Partnership - Final Report. Australia. Available at: http://www.communitybusinesspartnership.gov.au/wp-content/uploads/2016/11/social_impact_investing_research_report.pdf.

Fels, G. (2015). Development and Evaluation of Risk Mitigation Techniques for Social Impact Bonds. Universidade Católica Portuguesa. Available at: http://repositorio.ucp.pt/bitstream/10400.14/18382/1/Master%20Thesis_Gerwin%20Fels_Final.pdf.

Fischer, R., & Richter, F. (2017). SROI in the pay for success context: Are they at odds? *Evaluation and Program Planning*, Volume 64, pp 105-109. https://doi.org/10.1016/j.evalprogplan.2016.11.012.

Fitzgerald, J. (2013). Social impact bonds and their application to preventive health. Aust Health Rev. 2013 May;37(2):199-204. DOI: 10.1071/AH12238.

Fox, C., & Albertson, K. (2011). Payment by results and social impact bonds in the criminal justice sector: New challenges for the concept of evidence-based policy? *Criminology & Criminal Justice*. Vol 11, Issue 5, pp 395 – 413. https://doi.org/10.1177/1748895811415580

Fox, C and Albertson, K (2012). Is payment by results the most efficient way to address the challenges faced by the criminal justice sector? *Probation Journal* Vol.59(4), pp.355-373.

Fraser, A., Tan, S., Lagarde, M., and Mays, N. (2016) Narratives of Promise, Narratives of Caution: A Review of the Literature on Social Impact Bonds. *Social Policy & Administration*, DOI: 10.1111/spol.12260.

Freireich, J., Fulton, K. (2009). Investing for Social & Environmental Impact: A design for catalysing an emerging industry, Monitor Institute. Available at: http://monitorinstitute.com/downloads/what-we-think/impact-investing/Impact_Investing.pdf.

Giacomantonio, C. (2017). Grant-Maximizing but not Money-Making: A Simple Decision-Tree Analysis for Social Impact Bonds, *Journal of Social Entrepreneurship*, Volume 8, 2017 - Issue 1, pp 47-66.

Glaenzel, G., & Scheuerle, T. (2016). Social Impact Investing in Germany: Current Impediments from Investors' and Social Entrepreneurs' Perspectives. *Voluntas: International Journal of Voluntary & Nonprofit Organizations*, Volume 27, Issue 4, pp 1638–1668. DOI: 10.1007/s11266-015-9621-z.

Gelsthorpe, L., & Hedderman, C. (2012). Providing for women offenders the risks of adopting a payment by results approach. *Probation Journal*. 59(4) 374–390. DOI: 10.1177/0264550512458477.

Godeke, S., & Resner, L. (2012). Building a Healthy & Sustainable Social Impact Bond Market: The Investor Landscape, Godeke Consulting. Available at: https://www.missioninvestors.org/tools/building-a-healthy-sustainable-social-impact-bond-market-the-investor-landscape.

Government Outcomes Lab (GO Lab) (2017). Children's Services Workshop: A Summary Report. Available at: http://golab.bsg.ox.ac.uk/sites/golab.bsg.ox.ac.uk/files/2017-02/GoLab-ChildrensServicesPostWorkshopReport_0.pdf.

Goodall, E. (Bridges Ventures). (2014). Choosing Social Impact Bonds: A Practitioner's Guide. Available at: http://www.longfinance.net/images/reports/pdf/bvbaml_sib_2014.pdf.

Gosling, HJ (2016) 'All this is about is money and making sure that heads are on beds. ' Perceptions of Payment by Results in a Therapeutic Community. *Probation Journal*: the journal of community and criminal justice, 63 (2). pp. 144-152. ISSN 1741-3079. Available at: http://researchonline.ljmu.ac.uk/3124/.

Grannemann, T. W., & Brown, R. S. (2017). Adapting Evaluations of Alternative Payment Models to a Changing Environment. *Health Serv Res*. doi:10.1111/1475-6773.12689.

Gregory, D., Hill, K., Joy, I., Keen, S. (2012). Investment Readiness in the UK, Big Lottery Fund. Available at: https://www.biglotteryfund.org.uk/research/social-investment/publications.

Hatry, H., Eldridge, M., Kreeger, A., Jordan. R. (2017). Making Sound Cost Decisions in Pay for Success Projects. Urban Institute. Available at: http://www.urban.org/research/publication/making-sound-cost-decisions-pay-success-projects.

Hazenberg, R., Seddon, F., Denny, S. (2015). Intermediary Perceptions of Investment Readiness in the UK Social Investment Market. VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations, June 2015, Volume 26, Issue 3, pp 846–871. Available at: https://link.springer.com/article/10.1007/s11266-014-9473-y.

Hill, K. (2011). Investor Perspectives on Social Enterprise Financing, Report prepared for the City of London Corporation, ClearlySo, UK. Available at: https://www.cityoflondon.gov.uk/business/economic-research-and-information/research-publications/Documents/research-2011/Investor%20Perspectives%20on%20Social%20Enterprise%20Financing.pdf.

Howard, E. (2014). Challenges and opportunities in social finance in the UK, CICERO Group. Available at: http://nsig.org.uk/index.php/2014-03-20-11-52-06/publications/send/4-policy-and-strategy-documents/15-challenges-and-opportunities-in-social-finance-in-the-uk.

Humphries, K. (2014). Not Your Older Brother's Bonds: The Use and Regulation of Social-Impact Bonds in the United States, 76 *Law and Contemporary Problems* 433-452 (2014) Available at: http://scholarship.law.duke.edu/lcp/vol76/iss3/15

Jackson, E., & and Harji, K. (2012a). Unlocking Capital, Activating a Movement: Final Report of the Strategic Assessment of The Rockefeller Foundation's Impact Investing

Initiative. Available at: https://assets.rockefellerfoundation.org/app/uploads/20150904105231/Impact-Investing-Evaluation-Report-20121.pdf.

Jackson, E., & and Harji, K. (2012b). Accelerating Impact Achievements, Challenges and What's Next in Building the Impact Investing Industry. Available at: https://assets.rockefellerfoundation.org/app/uploads/20120707215852/Accelerating-Impact-Full-Summary.pdf.

Jackson (2013). Evaluating social impact bonds: questions, challenges, innovations, and possibilities in measuring outcomes in impact investing, *Community Development*, 44:5, 608-616, <http://dx.doi.org/10.1080/15575330.2013.854258>

Jackson, E., & and Harji, K. (2014). Assessing Impact Investing: Five Doorways for Evaluators, the Rockefeller Foundation Evaluation Office. Available at: https://assets.rockefellerfoundation.org/app/uploads/20140701194516/Impact-Investing-Thematic-Briefs-FINAL.pdf.

King, J. (2017). Using Economic Methods Evaluatively. *American Journal of Evaluation* 2017, Vol. 38(1) 101-113. DOI: 10.1177/1098214016641211.

Koekoek, K & Zwietering, H. (2015). Social Impact Bonds – October 2015 Opportunities and challenges in the Netherlands, ABN AMRO. Available at: https://www.abnamro.com/en/images/Documents/040_Sustainable_banking/ABN_AMRO_Rapport_Social_Impact_Bonds.pdf.

Lantz, P., Rosenbaum, S., Ku, L., Iovan, S. (2016). Pay for Success and Population Health: Early Results from Eleven Projects Reveal Challenges and Promise. *Health Affairs*. November 2016 vol. 35 no. 11 2053-2061.

Marty, F. (2016). Social Impact Bonds: PPPs reloaded ? - [Les obligations à impact social: une nouvelle génération de PPP pour les politiques sociales?] Discussion Paper Series, EPPP DP No. 2016-3. Available at: http://chaire-eppp.org/wp-content/uploads/2016/11/Marty_2016b.pdf

Mason, P., Lloyd, R., Andrews, M., Henry, N. (2014). Learning from the Development and Commissioning of the London Homelessness SIB: A summary report from the qualitative evaluation, ICF Consulting Services. Available at: https://data.gov.uk/sib_knowledge_box/sites/default/files/140818%20ICF%20Learning%20from%20the%20%20London%20Homelessness%20SIB%20Summary%20Report%20FINAL%20July%2014.pdf.

Massey, M., Walsh, K., Milner, J., Derrick-Mills, T. (2016). Pay for Success: Early Childhood Education – Project and Performance Management, Urban Institute. Available at: http://www.urban.org/sites/default/files/publication/86616/payforsuccesstoolkit_project_and_performance_management_0.pdf.

Maynard, A., Street, A. and Hunter, R. (2011). Using 'payment by results' to fund the treatment of dependent drug users—proceed with care! Addiction, 106: 1725–1729. doi:10.1111/j.1360-0443.2011.03503.x.

McHugh, N., Sinclair, S., Roy, M., Huckfield, L. & Donaldson, C. 2013. 'Social impact bonds: a wolf in sheep's clothing?'. *Journal of Poverty and Social Justice*, 21, 247-257.

Moreau, S., Hornsby, A. (2015). Oranges & Lemons: The State of Play of Impact Measurement among UK Social Investment Finance Intermediaries, A report by Investing for Good. Available at: https://www.investingforgood.co.uk/news/orangesandlemonsreport.

Mosenson, S (2013). New models of private sector financing of not-for-profit corporations, Northeastern University. Available at: http://hdl.handle.net/2047/d20004861.

Nicholls, A., and Daggers, J. (2016a). Big Society Capital: The World's First Social Investment Wholesale Bank – Case Study, Saïd Business School Case, University of Oxford. Available at: https://www.sbs.ox.ac.uk/sites/default/files/Skoll_Centre/Images/Research/BSC-Case-Study.pdf.

Nicholls, A., and Daggers, J. (2016b). The Landscape of Social Impact Investment Research: Trends and Opportunities. Saïd Business School Case, University of Oxford. Available at: https://www.sbs.ox.ac.uk/sites/default/files/research-projects/CRESSI/docs/the-landscape-of-social-impact-investment-research.pdf.

O'donohoe, N., Leijonhufvud, C., Saltuk, Y., Bugg-levine, A., Brandenburg, M. (2010). Impact Investments: An Emerging Asset Class. Global Impact Investing Network (GIIN). Available at: https://thegiin.org/assets/documents/Impact%20Investments%20an%20Emerging%20Asset%20Class2.pdf.

Ogundeji, Y. (2015) Pay-for-performance for Health Service Providers: Effectiveness, Design, Context, and Implementation. PhD thesis, University of York. Available at: http://etheses.whiterose.ac.uk/9036/.

Olsen, S., & Galimidi, B. (2008). Catalog of Approaches to Impact Measurement: Assessing Social Impact in Private Ventures, Social Venture Technology Group. Available at: http://www.midot.org.il/Sites/midot/content/Flash/CATALOG%20OF%20APPROACHES%20TO%20IMPACT%20MEASUREMENT.pdf.

Ormiston, J., Charlton, K., Donald, S., & Seymour, R. (2015). Overcoming the Challenges of Impact Investing: Insights from Leading Investors. *Journal of Social Entrepreneurship* Vol. 6, Iss. 3,2015. http://dx.doi.org/10.1080/19420676.2015.1049285.

Parthenon Group., Bridges Ventures., & GIIN Network. (2010). Investing for Impact: Case Studies across Asset Classes. Available at: https://www.missioninvestors.org/system/files/tools/investing-for-impact-case-studies-across-asset-classes-the-parthenon-group-bridges-ventures-and-the-global-impact-investing-network-giin.pdf.

Pathik P., Pratik, D (2014). Social return on investment: three technical challenges, *Social Enterprise Journal*, Vol. 10 Issue: 2, pp.91-104, https://doi.org/10.1108/SEJ-06-2012-0019.

Pauly, M., Swanson, A. (2013). Social Impact Bonds in Nonprofit Health Care: New Product or New Package? National Bureau of Economic Research (NBER), Working Paper No. 18991. Available at: http://www.nber.org/papers/w18991.

Paydar, N., Moody, M.., & Littlepage, L. (2015). Measuring Social Return on Investment. *Nonprofit Management and Leadership*, 26: 19–37. doi:10.1002/nml.21145.

Reeder, N., Mulgan, G., Aylott, M., Bo'sher, L. (2010). Social Impact Investment: the challenge and opportunity of Social Impact Bonds, The Young Foundation. https://youngfoundation.org/wp-content/uploads/2012/10/Social-Impact-Investment-The-opportunity-and-challenge-of-Social-Impact-Bonds-March-2011.pdf.

Rees, J. (2014). Public sector commissioning and the third sector: Old wine in new bottles? *Public Policy and Administration* 2014, Vol. 29(1) 45–63. DOI: 10.1177/0952076713510345.

Ronicle, J., Fox, T., Stanworth, N. (2016). Commissioning Better Outcomes Fund Evaluation: Update Report, Ecorys Consulting, UK. Available at: https://www.biglotteryfund.org.uk/research/social-investment/publications.

Saltman, K. (2017). The Promise and Realities of Pay for Success/ Social Impact Bonds, *Education Policy Analysis Archives*, Volume 25, Number 59. DOI: http://dx.doi.org/10.14507/epaa.25.2640.

Schinckus, C. (2017). Financial innovation as a potential force for a positive social change: The challenging future of social impact bonds. *Research in International Business and Finance* 39 (2017) 727–736. https://doi.org/10.1016/j.ribaf.2015.11.004.

Sheehy, S., Taylor, A., Clark, A., Banders, A. (2015). Social impact bonds and homelessness: Questions from the field [online]. *Parity*, *Vol*. 28, No. 9, Nov 2015: 39-40. Available at: http://search.informit.com.au/documentSummary;dn=605035340140743;res=IELHSS. ISSN: 1032-6170.

Sheil, F & Breidenbach-Roe, R. (2014). Payment by Results and the voluntary sector 'Future of Public Services' Services, NCVO. Available at: https://www.ncvo.org.uk/images/documents/about_us/media-centre/payment-by-results-and-the-voluntary-sector-april-2014.pdf.

Smeets, D (2017). Collaborative Learning Processes in Social Impact Bonds: A Case Study from the Netherlands. *Journal of Social Entrepreneurship*, Vol. 8, Iss.1, 2017. http://dx.doi.org/10.1080/19420676.2017.1299034.

Social Impact Investment Taskforce (2014). Measuring Impact: Subject paper of the Impact Measurement Working Group. Available at: http://www.socialimpactinvestment.org/reports/Measuring%20Impact%20WG%20paper%20FINAL.pdf.

Social Finance (2014). The Essex Social Impact Bond: A Year in Review. Available at: http://www.socialfinance.org.uk/wp-content/uploads/2014/11/Essex_A_year_in_review.pdf.

Stump, E., & Johnson, F. (2016). An Examination of Using Social Impact Bonds to Fund Education in Maine, Maine Education Policy Research Institute, University of Southern Maine. Available at: https://usm.maine.edu/sites/default/files/cepare/Examination_of_Using_Social_Impact_Bonds_to_Fund_Education_in_Maine.pdf.

Trupin, E., Weiss, N., Kerns, S. (2014). Social Impact Bonds: *Behavioral Health Opportunities*. JAMA Pediatr. 2014;168(11):985-986. doi:10.1001/jamapediatrics.2014.1157.

Vo et al (2016). Understanding evaluation practice within the context of social investment, *SAGE Journals*, Evaluation, Vol. 22(4) 470–488. Available at: http://evi.sagepub.com/content/22/4/470.abstract.

Warner. (2015). Profiting from Public Value? The Case of Social Impact Bonds. pp 143-160 in Creating Public Value in Practice, ed. By Bryson, J., Crosby, B., & Bloomberg, L, New York: CRC Press, Taylor and Francis, 2015.

White, R. (2014). Application of Social Impact Bonds in Built Infrastructure Sustainability Projects. Doctoral dissertation, Texas A & M University. Available at: http://hdl.handle.net/1969.1/152577.

World Economic Forum (2013). From Ideas to Practice, Pilots to Strategy: Practical Solutions and Actionable Insights on How to Do Impact Investing, A report by the World Economic Forum Investors Industries. Available at: http://www3.weforum.org/docs/WEF_II_SolutionsInsights_ImpactInvesting_Report_2013.pdf.

**Additional references**

Adam. (2006). Evaluating Social Fund Impact: A Toolkit for Task Teams and Social Fund Managers, Social Protection Discussion Paper No 0611, World Bank.

Adams, J., Hillier-Brown, F., Moore, H., Lake, A., Araujo-Soares, V., White, M., Summerbell, C. (2016). Searching and synthesising 'grey literature' and 'grey information' in public health: critical reflections on three case studies, *Systematic Reviews* (2016) 5:164. doi: 10.1186/s13643-016-0337-y.

Bastian, H. (1994). The Power of Sharing Knowledge: Consumer Participation in the Cochrane Collaboration. Oxford: UK Cochrane Centre.

Boyle, D. (2011). The pitfalls and perils of payment by results. *Local Economy*, 26, 627-634.

Bridges Ventures. (2016), Better Outcomes, Better Value: The Evolution of Social Impact Bonds in the UK, London: Bridges Ventures LLP.

Brouwers, M., Kho, ME., Browman, G, Burgers, JS., Cluzeau, F., Feder, G., Fervers, B., Graham, ID., Grimshaw, J., Hanna, S., Littlejohns, P., Makarski, J., Zitzelsberger, L. (2010). AGREE II: Advancing guideline development, reporting and evaluation in healthcare. *Can Med Assoc J*. 2010. Available at: http://www. agreetrust.org/

Cabinet Office. (2017), The Social Impact Bond Knowledge Box, London; Cabinet Office.

Chalmers, I. (2005). If evidence-informed policy works in practice, does it matter if it doesn't work in theory? Evidence & Policy: *A Journal of Research*, Debate and Practice, 1(2), 227-242.

Charity Commission. (2012). *Charities and Public Service Delivery – An Introduction and Overview*. London: Charity Commission.

Clarke, B., Gillies, D., Illari, P., Russo, F., Williamson, J. (2013). The evidence that evidence-based medicine omits, *Preventive Medicine*, 57 (6) (2013), pp. 745-747. Available at: http://www.sciencedirect.com/science/article/pii/S0091743512005452.

Clarke, B., D. Gillies, P. Illari, F. Russo, and J. Williamson. (2014). Mechanisms and the evidence hierarchy. *Topoi* 33(2): 339–360. Available at: https://link.springer.com/article/10.1007/s11245-013-9220-9.

Conservative Party. (2010). Big Society not Big Government, London: Conservative Party.

Cooke, A, Smith, D. & Booth, A. (2012). Beyond PICO: The SPIDER tool for qualitative evidence synthesis. *Qual Health Res* 2012(10):1435-43.

Critical Appraisal Skills Programme. (2014). CASP Checklists. Oxford. Retrieved from: http://media.wix.com/ugd/dded87_29c5b002d99342f788c6ac670e49f274.pdf.

Davison, R. (2013) *Does Social Finance Understand Social Need?* Available at: http://www.cancook.co.uk/wp-content/uploads/2013/01/Does-Social-Enterprise-Understand-Social-Need.pdf.

Dickersin, K., Scherer, R., Lefebvre, C. (1994). Systematic Reviews: Identifying relevant studies for systematic reviews *BMJ* 1994; 309:1286. https://doi.org/10.1136/bmj.309.6964.1286.

Doak, CM., Visscher, TL., Renders, CM., Seidell JC. (2006). The prevention of overweight and obesity in children and adolescents: a review of interventions and programmes. *Obesity Reviews*, 7(1): 111–136.

Fox, C. and Albertson, K. (2011). 'Payment by results and social impact bonds in the criminal justice sector: New challenges for the concept of evidence-based policy?', *Criminology and Criminal Justice*, 11, 5, 395–413.

Fraser, A., Tan, S., Lagarde, M., and Mays, N. (2016). Narratives of Promise, Narratives of Caution: A Review of the Literature on Social Impact Bonds. *Social Policy & Administration*. doi: 10.1111/spol.12260.

Gough, D., Oliver, S., & Thomas, J. (Eds.). (2012). An introduction to systematic reviews. London: Sage Publications.

Guyatt, G., Oxman, AD., Akl, EA., Kunz, R., Vist, G., Brozek, J., Norris, S., Falck-Ytter, Y., Glasziou, P., deBeer, H., Jaeschke, R., Rind, D., Meerpohl, J., Dahm, P., Schünemanna. H. (2011). GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. Journal of Clinical Epidemiology. 2011;64: 383–394. https://doi.org/10.1016/j.jclinepi.2010.04.026.

Harlock, J. (2013). Impact measurement practice in the UK third sector: a review of emerging evidence, Third Sector Research Centre, Working Paper 106.

Hawe, P., Shiell, A., Riley, T. (2004). Complex interventions: how far 'out of control' should a randomised controlled trial be? *BMJ* 328:1561–63 68.

Hawe, P., Shiell, A., Riley, T. (2009). Theorising interventions as events in systems. Am. *J. Community Psychol*. 43:267–76.

Hawe, P. (2015). Lessons from Complex Interventions to Improve Health. *Annual Review of Public Health*, 36, 307-323.

Higgins, J. P., & Green, S. (Eds.). (2011). Cochrane handbook for systematic reviews of interventions (Vol. 5.1.0). Chichester: Wiley-Blackwell.

Hopewell, S., Clarke, M., Lefebvre, C., Scherer, R. (2007). Handsearching versus electronic searching to identify reports of randomized trials, Cochrane Database of Systematic Reviews, 2007: 2, John Wiley & Sons, Ltd. 10.1002/14651858.MR000001.pub2.

Hudson, M., Phillips, J., Ray, K., Vegeris, S., Davidson, R. (2010). The influence of outcome-based contracting on provider-led pathways to work, Research Report No. 638, London: Department for Work and Pensions. Available at: http://www.psi.org.uk/site/publication_detail/480/.

Joy, M., J, Shields. ((2013). Social Impact Bonds: The Next Phase of Third Sector Marketization? *Canadian Journal of Non-profit and Social Economy Research*, Vol. 4, No 2, pp. 39–55. Available at: http://anserj.ca/index.php/cjnser/article/view/148.

Lagarde, M., Wright, Michael., Nossiter, Julie., Mays, N. (2013). *Challenges of payment-for-performance in health care and other public services – design, implementation and evaluation.* Technical Report. Policy Innovation Research Unit. Available at: http://researchonline.lshtm.ac.uk/2478790/.

Lawlor, DA., Tilling, K., Davey Smith, G. (2016). Triangulation in aetiological epidemiology. *International Journal of Epidemiology*, Volume 45, Issue 6, 1 December 2016, Pages 1866–1886. DOI: https://doi.org/10.1093/ije/dyw314.

Lewin, S., Glenton, C., Munthe-Kaas, H., Carlsen, B., Colvin, CJ., Gülmezoglu, M., Noyes, J., Booth, A., Garside, R., Rashidian, A. (2015). Using qualitative evidence in decision making for health and social interventions: an approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual). *PLOS Med*. 2015;12: e1001895. pmid:26506244.

Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on psychological science*, 2(1), 53-70.

Littell, J., J. Corcoran, and V. Pillai. 2008. Systematic reviews and meta-analysis. New York: Oxford University Press.

Mahood, Q., Van Eerd, D., Irvin, E. (2014). Searching for grey literature for systematic reviews: challenges and benefits. *Research Synthesis Methods*. 2014;5 (3):221–34.

McCord, J. (2003). Cures that harm: Unanticipated outcomes of crime prevention programs. *The Annals of the American Academy of Political and Social Science*, 587(1), 16-30.

McHugh, N., Sinclair, S., Roy, M., Huckfield, L. and Donaldson, C. (2013), 'Social impact bonds: a wolf in sheep's clothing?', *Journal of Poverty and Social Justice*, 21, 3, 247–257.

Mhaskar, R., Emmanuel, P., Mishra, S., Patel, S., Naik, E., & Kumar, A. (2009). Critical appraisal skills are essential to informed decision-making. *Indian Journal of Sexually Transmitted Diseases*, 30(2), 112–119.

Moher D., Liberati A., Tetzlaff J., Altman DG & The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* 6(7): e1000097. doi:10.1371/journal.pmed1000097

Moher, D., Schulz, K. F., Simera, I., & Altman, D. G. (2010). Guidance for developers of health research reporting guidelines. *PLoS Med*, 7(2), e1000217.

National Audit Office (NAO) (2015), 'Outcome-based payment schemes: government's use of payment by results', London: National Audit Office
Available at: https://www.nao.org.uk/wp-content/uploads/2015/06/Outcome-based-payment-schemes-governments-use-of-payment-by-results.pdf.

Nicholls, A. (2009). 'We Do Good Things, Don't We?': 'Blended Value Accounting' in Social Entrepreneurship', *Accounting, Organisations and Society* 34 (6–7): pp. 755–769

Nicholls, A. and Tomkinson, E. (2015), 'The Peterborough Pilot Social Impact Bond', in A. Nicholls, R. Paton and J. Emerson (eds.), Social Finance, Oxford: Oxford University Press.

Oliver, K. et al. (2014). 'A systematic review of barriers to and facilitators of the use of evidence by policymakers', *BMC Health Services Research*, 14 (2).

Pauly, M., Swanson, A. (2013). Social Impact Bonds in Nonprofit Health Care: New Product or New Package? National Bureau of Economic Research (NBER) Working Paper Series, Cambridge. Available at: http://www.nber.org/papers/w18991.pdf.

Petrosino, A., Turpin-Petrosino, C. & Buehler, J., (2003). Scared Straight and Other Juvenile Awareness Programs for Preventing Juvenile Delinquency: A Systematic Review of the Randomized Experimental Evidence. *The Annals of the American Academy of Political and Social Science*, 589 (1), pp 41-62. doi https://doi.org/10.1177/0002716203254693

Roy, M. (2017). Social Impact Bonds – Evidence-based policy or Ideology? In: Handbook of Social Policy Evaluation, Publisher: Edward Elgar Publishing, Editors: Bent Greve.
Rutter, H., Savona, N., Glonti, K., Bibby, J., Cummins, S., Finegood, T., Greaves, F., Harper, L., Hawe, P., Moore, L., Petticrew, M., Rehfuess, E., Shiell, A., Thomas, J., White, M. (2017). The need for a complex systems model of evidence for public health. *The Lancet*. https://doi.org/10.1016/S0140-6736(17)31267-9.

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC Medicine*, 8(1), 1.

Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). Foundations of program evaluation: Theories of practice. Sage.

Shadish, William R. Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Wadsworth Cengage learning.

Shiell, A., Hawe, P., & Gold, L. (2008). Complex interventions or complex systems? Implications for health economic evaluation. *BMJ : British Medical Journal*, 336(7656), 1281–1283. http://doi.org/10.1136/bmj.39569.510521.AD.

Social Finance (2017). Impact Bond Global Database
< http://www.socialfinance.org.uk/database/>

Summerbell CD, Waters E, Edmunds LD, Kelly S, Brown T, Campbell KJ (2005) Interventions for preventing obesity in children. *Cochrane Database of Systematic Reviews*. Issue 3. Art. No.: CD001871.

Taylor-Gooby, P. and Stoker, G. (2011) 'The Coalition Programme: A New Vision for Britain or Politics as Usual?' Political Quarterly, 82 (1). pp. 4-15.

Thomas, J., Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*. 2008; 8:45. https://doi.org/10.1186/1471-2288-8-45

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, 19(6), 349–357. http://doi.org/10.1093/intqhc/mzm042.

Tyndall, J. AACODS Checklist. Flinders University, 2010. Available from http://dspace.flinders.edu.au/dspace/

Vandenbroucke, J., Broadbent, A., Pearce, N. (2016). Causality and causal inference in epidemiology: the need for a pluralistic approach, *International Journal of Epidemiology*, Volume 45, Issue 6, 1 December 2016, Pages 1776–1786, https://doi.org/10.1093/ije/dyv341.

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. *Preventive Medicine*, 45(4), 247–251. http://doi.org/10.1016/j.ypmed.2007.08.012.

Weale, A., Clark, S. And Martin, N. (2010) [UCL] *Charities and Public Service Delivery.* Available at: http://www.ucl.ac.uk/spp/research/social-contract-deliberative-democracy-and-public-policy/Charities_and_Public_Service_Delivery_-_Charity_Commission_Seminar.pdf

Weisburd, D., Lum, C. M., & Petrosino, A. (2001). Does research design affect study outcomes in criminal justice? *The Annals of the American Academy of Political and Social Science*, 578(1), 50- 70.

Welsh, B. C., Peel, M. E., Farrington, D. P., Elffers, H., & Braga, A. A. (2011). Research design influence on study outcomes in crime and justice: A partial replication with public area surveillance. *Journal of Experimental Criminology*, 7(2), 183-198. DOI: 10.1007/s11292-010-9117-1.

## Appendix 1: Expert Advisory Panel – Participant Information Sheet

### 1. Study title

"Current barriers to and facilitators of rigorous evaluation in Social Impact Bonds (SIBs) – a systematic review"

### 2. Background and aims of the study

This study aims to answer the following question:

*"What are practitioners' perceptions about the main barriers to and facilitators of rigorous evaluation in the delivery of social impact bonds (SIBs)?"*

### 3. Why have I been invited to take part?

Given the absence of publicly available data regarding the effectiveness of SIBs, the review aims to identify the main barriers to and facilitators of rigorous evaluation, in order to understand how evaluation practices might be enhanced. An Expert Advisory Panel will be established to inform the systematic review, thus increasing its rigour and relevance to policy makers.

### 4. Do I have to take part?

No. You can ask questions about the study before deciding whether or not to participate in the advisory panel. If you do agree to participate, you may withdraw yourself from the study at any time, without giving a reason and without penalty, by advising the researchers of this decision.

### 5. What will happen in the study?

If you agree to take part in the study, you will be asked to provide informal advice in three stages, as outlined below. The project will conclude on 15 August 2017.

***Stage one:*** provide input on the research question and proposed methods, either in person, by phone or email

***Stage two:*** provide thoughts and opinions on studies retrieved, either in person, by phone or email

***Stage three:*** provide input on interpretation of findings, either in person, by phone or email

You will primarily be consulted individually, via email and phone. If required, occasional group meetings may also be organised (with Skype option available for members who cannot join in person). While your advice will inform important decisions in the review, the ultimate responsibility will remain with the research team. This will ensure that the required time commitment for advisory panel members is kept to a minimum.

**6. Are there any potential risks in taking part?**

There is no potential risk in taking part in this study.

**7. Are there any benefits in taking part?**

By taking part in the study, you will be able to use your expertise to inform the various steps of the systematic review, thereby increasing its rigour and relevance to both policy makers and practitioners.

**8. Expenses and payments**

There will be no payment for taking part in this study.

**9. Will the research be published?**

The University of Oxford is committed to the dissemination of its research for the benefit of society and the economy and, in support of this commitment, has established an online archive of research materials. This archive includes digital copies of student theses successfully submitted as part of a University of Oxford postgraduate degree programme. Holding the archive online gives easy access for researchers to the full text of freely available theses, thereby increasing the likely impact and use of that research.

If you agree to participate in this study, the research will be written up as a thesis. On successful submission of the thesis, it will be deposited both in print and online in the University archives, to facilitate its use in future research. The thesis will be published open access. An aim is also to publish the study in an academic journal.

**10. Who has reviewed this study?**

This study has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee Ref: SPI_C1A_16_12.1

**11. Who do I contact if I have a concern about the study or I wish to complain?**

If you have a concern about any aspect of this study, please speak to the researcher [Tel: XXX] or their supervisor [XXX@spi.ox.ac.uk], who will do their best to answer your query. The researcher should acknowledge your concern within 10 working days and give you an indication of how they intend to deal with it.

If you remain unhappy or wish to make a formal complaint, please contact the relevant chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter in a reasonably expeditious manner:

**Chair, Social Sciences & Humanities Inter-Divisional Research Ethics Committee**; Email: ethics@socsci.ox.ac.uk; Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD

## Appendix 2: Experts contacted and sample email

Various experts (n = 92) were contacted to identify additional (published or unpublished) articles of relevance. The table below includes the details of the experts who had responded to the request at the time of submission. In total, the experts suggested 86 articles and many commented on the state of evaluation practices in this field.

| # | Name | Organisation |
|---|---|---|
| **1** | Alberto Motta | UNSW |
| **2** | Anne Vo | University of Southern California |
| **3** | Aunnie Patton Power | University of Capetown |
| **4** | Brian van Es | Rotterdam University |
| **5** | Cam Donaldson | Glasgow Caledonian University |
| **6** | Catherine Clark | CASE |
| **7** | Cristopher Fox | Manchester Metropolitan |
| **8** | Darrick Jolliffe | University of Greenwich |
| **9** | David Rabinowitz | Deloitte |
| **10** | David Stoesz | University of Illinois-Springfield |
| **11** | Dominique Smeets | Utrecht University |
| **12** | Edward Jackson | Carleton University |
| **13** | Eleanor Carter | University of Oxford |
| **14** | Elizabeth Lower-Basch | The Center for Law and Social Policy |
| **15** | Emma Tomkinson | Community Insight Australia |
| **16** | Francisca Richter | Case Western Reserve University |
| **17** | Frédéric Marty | Université Côte d'Azur |
| **18** | Jade Wong | University of Oxford |
| **19** | Jarrod Ormiston | University of Sydney |
| **20** | Jasper Kim | Ewha Woman's University |
| **21** | Jennifer Rubin | King's College London |
| **22** | John Loxley | University of Manitoba |
| **24** | Joshua Sharfstein | Boston University |
| **25** | Julian King | University of Melbourne |
| **26** | Kevin W Humphries | Kirkland & Ellis |
| **27** | Leslie Huckfield | Glasgow Caledonian University |
| **28** | Mario Calderini | Politecnico di Milano |
| **29** | Mhairi Aylott | Ministry of Justice UK |
| **30** | Michael Roy | White & Case |
| **31** | Neil McHugh | Glasgow Caledonian University |
| **32** | Nick Henry | Coventry University |
| **33** | Peter Dagher | White & Case |
| **34** | Samantha Iovan | Michigan University |
| **35** | Stephanie Berzin | Boston College |
| **36** | Stephen Sinclair | Glasgow Caledonian University |

| | | |
|---|---|---|
| **37** | Suzanne Kerns | University of Denver |
| **38** | Tessa Hebb | Carleton University |
| **39** | Thomas Scheuerle | Heidelberg University |

***Sample email***

Dear _____,

I am a postgraduate student, undertaking the MSc. in Evidence-Based Social Intervention and Policy Evaluation at the University of Oxford. For my thesis, I am conducting a systematic review of the barriers of and facilitators to rigorous evaluation in Social Impact Bonds (SIBs).

I am collecting studies, which focus on the design and/or implementation of SIBs and discuss practitioner attitudes towards barriers or facilitators to evaluation. As you have been involved in social impact investment research, I am emailing to see if you would be willing to send me any relevant studies or reports (whether or not they are published) that you are aware of.

Thank you for your assistance.

Kind regards,

[student name]

## Appendix 3: Data Extraction Form

The reviewer adapted an existing data extraction tool for qualitative research (NICE – data extraction form for qualitative studies). The tool was modified slightly to ensure that sufficient data was extracted on the barriers to and facilitators of rigorous evaluation.

| |
|---|
| **Bibliographic details**<br>• Study ID (created by review author).<br>• Review author ID (created by review author).<br>• Citation and contact details<br>• Publication type |
| **Eligibility (if ineligible, note the primary reason for exclusion)**<br>Is the study either primary research (any study design) or a systematic review? **YES / NO**<br>Is the study about the design or implementation of at least one SIB **YES / NO**<br>Does the study consider barriers to and/or facilitators of rigorous evaluation? **YES/NO**<br>Does the study include a sample of practitioners with relevant experience? **YES / NO**<br>Does the study discuss practitioner attitudes regarding barriers or facilitators to rigorous evaluation? **YES / NO**<br>Is the study an English document? **YES / NO**<br>Is the study retrievable? **YES / NO** |
| **Study aims**<br>What was the aim of the study?<br>Is the study linked with any other studies? |
| **Key findings**<br>What are the key study findings? |
| **Evaluative Summary** |

| |
|---|
| *Draw together brief comments on the study and its strengths and weaknesses. Is further work required? What are its implications for policy, practice and theory?* |
| **Ethical standards**<br><br>Was ethical approval obtained?<br><br>Was informed consent obtained?<br><br>Does the study address ethical issues adequately?<br><br>Has confidentiality been maintained? |
| **Context**<br><br>Are the aims and purpose of the study clearly stated?<br><br>What specified SIB/s are discussed? |
| **Setting**<br><br>What is the setting for the study?<br><br>What is the rationale and appropriateness for this choice?<br><br>Is there sufficient detail about the setting?<br><br>Over what time period did the data collection take place? |
| **Sample**<br><br>Who was included in the study?<br><br>Who was excluded from the study?<br><br>How was the sample selected?<br><br>Were there any factors that influenced how the sample was selected?<br><br>What is the size of the sample and groups comprising the study?<br><br>Is the sample appropriate for the aims of the study, the depth of data it enables to be collected and its breadth? |
| **Data collection**<br><br>What data collection methods were used? Was the data collection adequately described and rigorously conducted? |

| |
|---|
| What is the role of the researcher within the setting? Are there any potential conflicts of interest?<br><br>Is the process of fieldwork adequately described? (if relevant?)<br><br>How are the data analysed? How adequate is the description of the data analysis? Is adequate evidence provided to support the analysis?<br><br>Are the researchers' own position, assumptions and possible biases outlined?<br><br>Are the findings substantiated by the data and has consideration been given to any limitations of the methods or data that may have affected results? |
| **Findings: Barriers and Facilitators**<br><br>What barriers to rigorous evaluation were identified in the study?<br><br>What facilitators of rigorous evaluation were identified in the study?<br><br>Was the prevalence of the identified barriers and facilitators among the surveyed or interviewed stakeholders reported? |
| **Policy and practice**<br><br>What is the country of the study? What is the policy area of focus? To what extent are findings generalisable?<br><br>What are the implications for policy?<br><br>What are the implications for practice? |
| **Miscellaneous**<br><br>Funding source?<br><br>Miscellaneous comments from study authors?<br><br>Miscellaneous comments by review author?<br><br>References to any other relevant studies? |
| **Decision**<br><br>Should this study be included in the final review? **YES NO UNCLEAR** |

## Appendix 4: Excluded studies and reason for exclusion

| Study | Exclusion Reason |
|---|---|
| Addis et al (2013) | Does not investigate barriers or facilitators |
| Addis et al (2014) | Does not investigate barriers or facilitators |
| Aylott and Shelupanov (2011) | Not retrievable |
| Baba et al (2016) | Does not investigate barriers or facilitators |
| Baliga (2013) | Wrong study design |
| Belinsky et al (2014) | Wrong study design |
| Bengo and Calderini (2016) | Wrong study design |
| Bertha Centre (2014) | Does not include perceptions of practitioners |
| Breuls (2014) | Does not investigate barriers or facilitators |
| Brown (2011) | Not about SIBs |
| Callanan et al (2012) | Does not investigate barriers or facilitators |
| Castellas et al (2016) | Not about SIBs |
| Cooper et al (2016) | Does not investigate barriers or facilitators |
| Costa (2014a) | Does not investigate barriers or facilitators |
| Costa (2014b) | Wrong study design |
| Crowley 2014 | Does not investigate barriers or facilitators |
| Dagher Jr (2013) | Wrong study design |
| Dandala (2015) | Does not include perceptions of practitioners |
| Deloitte (2014) | Does not include relevant sample of practitioners |
| Dembek et al (2016) | Not about SIBs |
| Department for Work and Pensions (2016) | Does not investigate barriers or facilitators |
| Disley et al (2015) | Does not investigate barriers or facilitators |
| Dowling (2017) | Wrong study design |
| Drew and Clist (2015) | Does not include relevant sample of practitioners |
| Dugger et al (2012) | Wrong study design |
| Ebrahim and Rangan (2014) | Wrong study design |
| Ecorys (2016) | Does not investigate barriers or facilitators |
| Edmiston (2016) | Does not include relevant sample of practitioners |
| Evenett and Richter (2011) | Does not investigate barriers or facilitators |
| EY (2016) | Does not include relevant sample of practitioners |
| Fels (2015) | Does not investigate barriers or facilitators |
| Fischer and Richter (2017) | Wrong study design |
| Fitzgerald (2013) | Not retrievable |
| Fox and Albertson (2011) | Wrong study design |
| Fox and Albertson (2012) | Wrong study design |
| Fraser et al (2016) | Does not investigate barriers or facilitators |
| Freireich and Fulton (2009) | Not about SIBs |
| Giacomantonio (2017) | Wrong study design |
| Glänzel and Scheuerle (2015) | Not about SIBs |
| Gelsthorpe and Hedderman (2012) | Wrong study design |

| | |
|---|---|
| Godeke & Resner (2012) | Does not include perceptions of practitioners |
| GO Lab (2017) | Does not include perceptions of practitioners |
| Goodhall (Bridge Ventures) (2014) | Does not investigate barriers or facilitators |
| Gosling 2016 | Not about SIBs |
| Grannemann et al (2017) | Wrong study design |
| Gregory et al (2012) | Not about SIBs |
| Hatry et al (2017) | Does not investigate barriers or facilitators |
| Hawken et al (2013) | Does not investigate barriers or facilitators |
| Hazenberg et al (2015) | Does not investigate barriers or facilitators |
| Hill (2011) | Not about SIBs |
| Howard (2014) | Not about SIBs |
| Humphries (2014) | Wrong study design |
| Jackson and Harji (2012a) | Not about SIBs |
| Jackson and Harji (2012b) | Not about SIBs |
| Jackson (2013) | Does not include perceptions of practitioners |
| Jackson and Harji (2014) | Does not investigate barriers or facilitators |
| Kim (2009) | Not retrievable |
| King (2017) | Wrong study design |
| Koekoek (2015) | Does not include relevant sample of practitioners |
| Lantz (2016) | Does not include relevant sample of practitioners |
| Marty (2016) | Wrong study design |
| Mason et al (2014) | Does not investigate barriers or facilitators |
| Massey et al (2016) | Does not investigate barriers or facilitators |
| Maynard et al (2011) | Wrong study design |
| McHugh et al (2013) | Wrong study design |
| Moreau and Hornsby (2015) | Not about SIBs |
| Mosenson (2013) | Does not include relevant sample of practitioners |
| Nicholls and Daggers (2016) | Does not investigate barriers or facilitators |
| Nicholls and Daggers (2016) | Does not investigate barriers or facilitators |
| O'donohoe et al (2010) | Does not investigate barriers or facilitators |
| Ogundeji (2015) | Not about SIBs |
| Olsen and Galimidi (2008) | Not about SIBs |
| Ormiston et al (2015) | Does not investigate barriers or facilitators |
| Parthenon Group, Bridges Ventures, GIIN Network (2010) | Not about SIBs |
| Pathak et al (2014) | Not retrievable |
| Pauly and Swanson (2013) | Wrong study design |
| Paydar et al (2015) | Not about SIBs |
| Rabinowitz (2015) | Not retrievable |
| Reeder et al (2012) | Wrong study design |
| Rees (2014) | Wrong study design |
| Ronicle et al (2016) | Wrong study design |
| Saltman (2017) | Wrong study design |
| Schinckus (2017) | Wrong study design |
| Sheehy et al (2015) | Does not include relevant sample of practitioners |
| Sheil and Breidenbach-Roe (2014) | Does not investigate barriers or facilitators |
| Smeets and Jozef (2017) | Does not include perceptions of practitioners |
| Social Impact Investment Taskforce (2014) | Does not include relevant sample of practitioners |

| | |
|---|---|
| Social Finance (2014) | Does not investigate barriers or facilitators |
| Stump (2016) | Does not investigate barriers or facilitators |
| Trupin et al (2014) | Does not investigate barriers or facilitators |
| Vo et al (2016) | Not about SIBs |
| Warner (2015) | Does not investigate barriers or facilitators |
| White (2014) | Wrong study design |
| World Economic Forum (2013) | Does not investigate barriers or facilitators |

## Appendix 5: Summary of CERQual Assessment Ratings (per finding)

| *Colour code used for ratings* | | | |
|---|---|---|---|
| **Substantial concerns** | **Moderate concerns** | **Minor concerns** | **No concerns** |

This table shows the ratings that were assigned to each finding in four areas: 1) adequacy of data; 2) relevance; 3) coherence; and 4) methodological limitations. Using each of these ratings, the reviewer determined an 'overall confidence in review finding' score. Due to word count, it was not possible to re-specify all themes; an identifying number was presented per theme (as assigned in Table 13).

| Finding | Adequacy of data | Relevance | Coherence | Methodological limitations | Overall CERQual Assessment of Confidence |
|---|---|---|---|---|---|
| **RESOURCES AND CAPACITY** | | | | | |
| ***BARRIERS*** | | | | | |
| **1** | | | | | High confidence |
| **2** | | | | | High confidence |
| **3** | | | | | Moderate confidence |
| **4** | | | | | High confidence |
| **5** | | | | | Moderate confidence |
| **6** | | | | | High confidence |
| **7** | | | | | Low confidence |
| ***FACILIITATORS*** | | | | | |
| **8** | | | | | High confidence |
| **9** | | | | | Low confidence |
| **10** | | | | | Moderate confidence |
| **11** | | | | | Moderate confidence |
| **12** | | | | | High confidence |
| **13** | | | | | Moderate confidence |
| **14** | | | | | Low confidence |
| **15** | | | | | Low confidence |

| | | | | | |
|---|---|---|---|---|---|
| **16** | | | | | Low confidence |
| **17** | | | | | Moderate confidence |
| **DATA AND ANALYTICS** | | | | | |
| ***BARRIERS*** | | | | | |
| **18** | | | | | High confidence |
| **19** | | | | | High confidence |
| **20** | | | | | High confidence |
| **21** | | | | | Low confidence |
| **22** | | | | | Moderate confidence |
| **23** | | | | | Moderate confidence |
| **24** | | | | | Moderate confidence |
| **25** | | | | | Moderate confidence |
| **26** | | | | | Very low confidence |
| **27** | | | | | Moderate confidence |
| ***FACILITATORS*** | | | | | |
| **28** | | | | | High confidence |
| **29** | | | | | Low confidence |
| **30** | | | | | Low confidence |
| **31** | | | | | Low confidence |
| **32** | | | | | Low confidence |
| **33** | | | | | Very low confidence |
| **34** | | | | | Low confidence |
| **35** | | | | | Low confidence |
| **CONTACT AND COLLABORATION** | | | | | |
| ***BARRIERS*** | | | | | |
| **36** | | | | | High confidence |
| **37** | | | | | Low confidence |
| **38** | | | | | Low confidence |
| **39** | | | | | Very low confidence |
| **40** | | | | | Very low confidence |
| **41** | | | | | Very low confidence |

| | | | | | |
|---|---|---|---|---|---|
| **42** | | | | | Very low confidence |
| ***FACILITATORS*** | | | | | |
| **43** | | | | | Low confidence |
| **44** | | | | | High confidence |
| **45** | | | | | Moderate confidence |
| **46** | | | | | Very low confidence |
| **47** | | | | | Low confidence |
| **48** | | | | | Very low confidence |
| **49** | | | | | Moderate confidence |
| **MEASUREMENT PRACTICES** | | | | | |
| ***BARRIERS*** | | | | | |
| **50** | | | | | High confidence |
| **51** | | | | | High confidence |
| **52** | | | | | Moderate confidence |
| **53** | | | | | High confidence |
| **54** | | | | | High confidence |
| **55** | | | | | Low confidence |
| **56** | | | | | Moderate confidence |
| **57** | | | | | Low confidence |
| **58** | | | | | Low confidence |
| **59** | | | | | Low confidence |
| **60** | | | | | Low confidence |
| **61** | | | | | Moderate confidence |
| **62** | | | | | Low confidence |
| ***FACILITATORS*** | | | | | |
| **63** | | | | | Low confidence |
| **64** | | | | | High confidence |
| **65** | | | | | Low confidence |
| **66** | | | | | Moderate confidence |
| **67** | | | | | Very low confidence |
| **68** | | | | | Very low confidence |

| | | | | | |
|---|---|---|---|---|---|
| **69** | | | | | High confidence |
| **70** | | | | | Very low confidence |
| **71** | | | | | Very low confidence |
| **72** | | | | | Moderate confidence |
| **73** | | | | | Very low confidence |
| **74** | | | | | Moderate confidence |
| **75** | | | | | Very low confidence |

## Appendix 6: PRISMA Checklist

| Section/topic | # | Checklist item | Reported on page # |
|---|---|---|---|
| **TITLE** | | | |
| Title | 1 | Identify the report as a systematic review, meta-analysis, or both. | 1 |
| **ABSTRACT** | | | |
| Structured summary | 2 | Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number. | 4 |
| **INTRODUCTION** | | | |
| Rationale | 3 | Describe the rationale for the review in the context of what is already known. | 15 |
| Objectives | 4 | Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS). | |
| **METHODS** | | | |
| Protocol and registration | 5 | Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number. | 16 |
| Eligibility criteria | 6 | Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale. | 20 |
| Information sources | 7 | Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched. | 22-24 |
| Search | 8 | Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated. | 22-24 |
| Study selection | 9 | State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis). | 28 |
| Data collection process | 10 | Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators. | 28 |
| Data items | 11 | List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made. | 28 |
| Risk of bias in individual studies | 12 | Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis. | 28 |
| Summary measures | 13 | State the principal summary measures (e.g., risk ratio, difference in means). | 29 |

| | | | |
|---|---|---|---|
| Synthesis of results | 14 | Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., $I^2$) for each meta-analysis. | 29 |
| Risk of bias across studies | 15 | Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies). | 30 |
| Additional analyses | 16 | Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified. | 30 |
| **RESULTS** | | | |
| Study selection | 17 | Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram. | 31 |
| Study characteristics | 18 | For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations. | N/A |
| Risk of bias within studies | 19 | Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12). | N/A |
| Results of individual studies | 20 | For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot. | N/A |
| Synthesis of results | 21 | Present results of each meta-analysis done, including confidence intervals and measures of consistency. | N/A |
| Risk of bias across studies | 22 | Present results of any assessment of risk of bias across studies (see Item 15). | N/A |
| Additional analysis | 23 | Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]). | N/A |
| **DISCUSSION** | | | |
| Summary of evidence | 24 | Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers). | 57-61 |
| Limitations | 25 | Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias). | 58 |
| Conclusions | 26 | Provide a general interpretation of the results in the context of other evidence, and implications for future research. | 62 |
| **FUNDING** | | | |
| Funding | 27 | Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review. | 62 |