Backing what works?

Social Impact Bonds and Evidence-Informed Policy and Practice

Summary

Social Impact Bonds (SIBs) offer an opportunity to explore the use of evidence to inform public policy and commissioning decisions in both discursive and practical terms in what are frequently highly politicised contexts. We identify three potential mechanisms by which SIBs may promote evidence use and explore these through empirical findings drawn from a three-year evaluation of SIBs applied to health and social care in the English NHS.

Keywords

Social Impact Bonds; Payment-for-Performance; Outcomes-based-contracting; Evidence-informed policy; Evaluation; NHS

Impact statement

It is important to cut through the ambiguities and paradoxes that permeate the literature devised by SIB practitioners and proponents. We highlight three mechanisms by which SIBs may encourage evidence-informed policy-making. Firstly, the ability of SIB financing to promote specific interventions for which a 'positive evidence base' already exists. Secondly, the opportunities that SIB financed programmes offer for the promotion of evidence use through improved local data collection practices. Thirdly, the opportunities that SIB financed interventions offer for formal evaluation. We empirically test these mechanisms and discuss the implications of our findings for policy-makers, public managers, and other interested parties.

Introduction

Globally, governments are exploring innovative ways of procuring public services to improve effectiveness. A high-profile example of this trend over the past decade is the development of Social Impact Bonds (SIBs). SIBs are pay-for-performance schemes in which private for-profit or social investors (who seek a blend of financial return and social good) provide some up-front finance towards the delivery of a public service and subsequently may receive an outcomes-based rate of return. A key attraction of the SIB model for governments is that they should only pay for 'what works' (Mulgan et al, 2011). A concern for 'what works' builds on advocacy of evidence-informed policy and practice (EIPP) directed at policy makers, practitioners and researchers over recent years (Davies et al, 2000).

In this paper, we explore the relationship between SIBs and EIPP. This is important because this relationship is somewhat ambiguous (Maier et al, 2017). Whilst some SIB proponents emphasise the promise that SIBs hold for furthering evidence-informed interventions or practices (Mulgan et al, 2011; Roman et al, 2014), other authors have highlighted potential epistemological (Warner, 2013), ethical (Roy et al, 2017; Berndt & Wirth, 2018), and practical (Edmiston & Nicholls, 2017) concerns around the relationship between SIBs and evidence. These conflicting viewpoints reflect the competing narratives discerned in the literature more broadly. Proponents emphasise the promise of SIBs as a 'win-win-win' policy tool (i.e. one that delivers better social outcomes for service users, cost-savings to government, and a return to investors). In contrast, critics caution about the potentially damaging implications of the SIB concept (Fraser et al, 2016).

The paper presents findings from a three-year evaluation of the first SIBs focused on health and social care in the English NHS (Authors, 2018). We identify three potential mechanisms by which SIBs may promote EIPP and explore these through our evaluation findings. In theoretical terms, we situate SIBs within wider debates linked to both the discursive and practical use of evidence in policymaking. We argue that SIBs are a particularly useful lens

for understanding evidence use in policy because evidence is strikingly central to the claims made by SIB proponents and their critics.

Evidence use & SIBs

Maier et al (2017) conducted a review of practitioner reports on SIBs and identified two key paradoxes found in these – one was 'cost-saving risk transfer to private investors' and the other was 'flexible but evidence-based services' (Maier et al, 2017). The first is paradoxical because SIBs have high transaction costs (beyond those of traditional commissioning) and SIB-financed initiatives that are rational choices for governments are unlikely to be attractive to investors (and vice versa) (Giacomantonio, 2017). 'Evidence-based flexibility' is also paradoxical as it suggests both fidelity to an evidence-based model whilst implying malleability which may run counter to model fidelity. Of the fifty-one practitioner reports reviewed by Maier et al (2017), thirty-four contained the paradox of 'evidence-based flexibility'. A strategy developed by practitioner report authors and identified by Maier et al to sidestep this apparent contradiction was to employ a very loose understanding of the terms 'evidence' and 'evidence-based'.

Beyond SIBs, the language of 'evidence' and 'evidence-based' change in policy making has a recognised discursive power aligned with positivistic, managerialist, and 'post-ideological' technocratic assumptions (Newman, 2001) since the mid-1990s. Use of an *evidential discourse* may highlight an intentionality on the part of SIB proponents that is worthy of deeper consideration. In policy terms, aligning SIBs with 'evidence' and 'evidence-based' interventions may be seen as an attempt to de-politicise and pre-empt some of the ideological and ethical criticism that has emerged about SIBs as part of a logic of 'financialisation' of human relations and social services (Warner 2012; Roy et al, 2017). Strategic attempts to intertwine SIBs within an 'evidence-based' discursive framing is a useful tactic to validate the policy and focus on the usefulness of the respective interventions whilst distancing these from the more controversial financial mechanisms inherent therein (Berndt & Wirth, 2018). In practical terms, the pervasive use of the

discourse of 'evidence' allied to SIBs may be seen as a strategy to down-play the risk of programme failure in the eyes of interested stakeholders, especially investors. This also permits SIBs to be framed as a technical tool to close the classic 'research-practice gap' through improved implementation of 'proven' interventions, thereby delivering better services and outcomes.

The discursive practices identified by Maier et al (2017) around the presentation of SIBs as 'flexible yet evidence-based' aligns with deeper definitional issues related to the SIB concept. Indeed, the idea of what a SIB is, or should be, is imbued with 'chameleonic' characteristics (Smith, 2013). SIBs demonstrate a high degree of 'strategic ambiguity' (Smith, 2013; Eisenberg, 1984). This is to say that SIBs are amenable to being framed in different ways for different audiences. More broadly, a feature of a SIB is that it may be framed as a 'social' innovation to those with a primarily social ethos. At the same time, a SIB may be framed as a 'financial' innovation for those who wish to emphasise the potential to deliver a financial return to investors who wish to engage in 'good works' (Authors, 2018). At different times, the 'evidence-based' potential of SIBs or their potential for 'flexibility' may be emphasised (Maier et al, 2017) by different actors for different purposes.

It is important to cut through some of the ambiguities and paradoxes characterise the relationship between evidence and SIBs in particular. This may be achieved by being clearer about *how* SIBs may promote or inhibit the use of evidence through empirical research of SIB projects. We highlight three mechanisms by which SIBs may be expected to demonstrate evidence-informed policy-making. Firstly, the ability of SIB financing to promote specific interventions for which a 'positive evidence base' already exists (Maier et al, 2017). Secondly, the opportunities that SIB financed programmes may offer for the promotion of evidence use through improved local data collection practices (Cox 2011; Liebman 2011; Stoesz, 2014). Thirdly, the opportunities that SIB financed interventions offer for formal evaluation (Fox and Morris, 2017).

Methods

This paper presents findings from a three-year evaluation of the SIB Trailblazers in Health and Social Care (Authors, 2018) funded by the Department of Health in England. Nine projects - collectively known as the SIB 'Trailblazers' - received seed funding in 2013 to explore whether to commission a service locally through a SIB, and, if so, how to set it up. The nine projects proposed SIB financed interventions targeted at a diverse set of population groups (in both geographical and size terms). Likewise, the strength of the evidence behind the respective interventions was heterogeneous – as described in Table 1.

INSERT TABLE 1 ABOUT HERE

The evaluation described and assessed the development of these projects over time with a view to considering whether, and if so, how, SIB financed services might deliver better outcomes than alternative financing mechanisms. We drew on comparative qualitative case study methods (Yin, 2013) to explore the decisions as to whether or not to commission services through a SIB financing mechanism across the nine SIB Trailblazers. Qualitative case studies are an appropriate method for exploring issues related to policy implementation (Pope and Mays, latest edition in press, 2018), exploring 'how' and 'why' questions about phenomena through detailed contextualised accounts of cases (Yin, 2013). We undertook qualitative analysis of documents (both local and national) and conducted interviews with relevant actors across the Trailblazers including interviews *before and after* the decisions were made not to initiate a SIB for those sites that eventually chose not to initiate SIB financed services. For those Trailblazers which did initiate SIB-financed programmes, we compared each of these qualitatively with sites elsewhere in the country that had the same or similar interventions (e.g. social prescribing, or specialist foster care services) serving similar populations provided by the same or similar organisations but not financed through a SIB mechanism. This comparison, though not perfect, sought to illuminate how the presence of SIB financing might have affected the management and delivery of services.

We conducted 177 interviews with 199 informants overall across all sites between June 2014 and May 2017 until 'data saturation' (Glaser, 1978). We purposively sampled informants to include commissioners (n=38 with 32 informants), providers (n=123 with 109 informants), intermediaries (n=23 with 13 informants), investors (n=9 with 10 informants) and others (n=5), e.g. central government, data analysts or consultants. Most interviews lasted an hour and were done face to face, though a number of interviews were also done over the telephone (n=27). Many interviews were conducted by two members of the research team together and a small number of interviews were conducted with more than one informant.

Interview transcripts were coded using specialist qualitative research software. Two members of the research team analysed data collaboratively to ensure inter-coder reliability and interrogated the data repeatedly in order to understand key issues in relation to the Trailblazers. These included the decision to initiate a SIB financed project or not; early implementation challenges where SIBs were commissioned; impacts of performance management and contract management decisions and service delivery upon different actors; and overall views about potential strengths and weaknesses of SIB financing mechanisms from staff as they developed and delivered SIB-financed projects. The interviews in the non-SIB comparison projects explored similar questions with the goal of attempting to tease out the main differences between delivering services with and without a SIB. The research generated a large volume of data. In this paper we draw on a subset of the data drawn from interviews across all sites (where SIBs were initiated, where they were not, and the non-SIB comparison sites) that focus specifically on the three aspects of evidence use introduced earlier in the paper.

Findings

(1) The strength of the evidence behind an intervention financed by a SIB mechanism

Three proposed Trailblazer interventions with the most established research evidence underpinning them were all initiated – these were the Manchester TFCO-A programme, the Newcastle Ways to Wellness programme and the London Rough Sleeping SIB. There are a number of trials and a recent systematic review exploring Social Prescribing (Bickerdike et al, 2017) and academic research into interventions that aim to improved targeted adolescent behaviour including TFCO-A (Evans et al, 2017). Key elements of the rough sleeping intervention have been evaluated through quasi-experimental approaches (Pleace and Bretherton, 2013) and the 'Housing First' principles it draws on has also been subjected to systematic review (Fitzpatrick-Lewis et al 2011).

It is notable that the results of the reviews of these respective three interventions are somewhat mixed – and this may be significant – particularly with respect to Social Prescribing. Clinical champions for Social Prescribing suggested that a key aim of the programme (alongside social improvement and cost savings) was to add to the evidence base behind Social Prescribing at scale:

> *'[T]his is actually a research [project]… you have to be able to prove it works. And you really do need a cohort of control patients to say… is it making that much of a difference?'*
>
> Clinical champion

As part of this project, the counterfactual quasi-experimental evaluation including a control group from another part of the city performed a dual role. For the clinical champions, it sought to generate data on the effectiveness of the Social Prescribing intervention itself. For the commissioners, it sought to demonstrate changes that could be causally attributed to the intervention itself thereby justifying performance based payments. There were both experimental and managerial reasons to maintain a level of rigidity to these metrics for both the clinical champions and the commissioners. However, this view was not shared by the investors, who, in the light of implementation problems (that put their investment at risk), sought to 'change and flex' aspects of the intervention and this created tensions amongst the different parties.

In the case of the Manchester TFCO-A project, all parties (commissioners, providers and investors) valued the evidence informed nature of the evaluation and its track record. The local government commissioners stated that without SIB financing and the transfer of implementation risk away from the Local Authority that it represented, their finance colleagues in the Local Authority would not have been prepared to pay for the service:

> *'You know, it's not something that we as a local authority would have invested in because it's so difficult and so complex and so challenging in terms of making it work. But… the risk [is] shared [through the] social impact bond. And also the, the basis originally of this TFCO… was that it was really [effective for] offenders.'*
>
> Commissioner

Ring-fencing the budget for staffing the dedicated TFCO-A social work team team required achieve model fidelity required was too high in the context of financial austerity in the UK since 2010 for the Local Authority to fund itself. Indeed, elsewhere in the UK over this period, many TFCO-A teams financed through conventional local government funding were under immense financial strain – impeding the delivery of TFCO-A and leading to the closure of many services including our non-SIB comparison site. As with the Social Prescribing example above, the SIB financing mechanism was central to the initiation of TFCO-A by mitigating the implementation risk of an intervention which has numerous clinical champions and an emerging (if contested) evidence base.

In addition, two Trailblazer services lacking strong supportive research evidence were also commissioned. In the case of Shared Lives, the Local Authorities in each site had run 'in-house' Shared Lives services or had worked in close collaboration with local Third Sector Organisations (e.g. voluntary, community, not-for profit) to deliver the service to relatively static numbers of service users for many years. Here, the SIB offered them a way to 'scale up' the service and potentially realise cost-savings locally. So, whilst there is little academic research evidence on the 'effectiveness' of Shared Lives services in academic terms, there is well developed local experience of the promise of the programme in terms of service user

satisfaction and the promise of overall reductions in costs. The Worcester Reconnections project was a targeted intervention to reduce loneliness, and thereby lead to improvements in health outcomes (for example, a reduction in depression, or delayed onset of dementia). It was expected to benefit service recipients as less socially isolated individuals are expected to remain more active (therefore reducing likelihood of non-communicable diseases linked to sedentary lifestyles such as diabetes, stroke, coronary heart disease, and increased frailty and disability). This service intervention was intended as a proof of concept project to generate evidence about a personalised approach to combating social isolation and the harmful health effects associated with it.

Saliently, when we turn to the four Trailblazers that were not initiated, there was little academic or other research evidence supporting any of these interventions. Indeed, this was noted as a factor that contributed to the decision not to commission the services in two of these cases specifically (Leeds and Cornwall). The research evidence behind the proposed interventions in Sandwell and East Lancashire was also weak. In summary, it is the case that the SIB financing mechanism enables both interventions with and without research evidence behind them to be initiated. For those interventions without an evidence base, local experience of the service providing confidence in the minds of commissioners (Shared Lives), or a commitment to experimentation (Worcester Reconnections) appear to be significant in garnering support.

(2) The opportunities that SIB financing can offer for evidence generation through a specific intervention locally

In the five sites with interventions financed by SIBs, informants emphasised that local administrative and descriptive data were routinely drawn upon, analysed, and used to guide local decision-making. Indeed, this enhanced use of data was cited by informants as a central advantage of SIB-financed work compared with their prior experience. This is a consistent finding across UK SIB research (DWP, 2014; Disley et al, 2015). Furthermore,

interviews with the non-SIB financed comparator sites revealed that staff at these sites drew less upon administrative and descriptive data than the SIB-financed sites:

> *'[T]here's a lot of data processing that needs doing there and we haven't got the capacity to do it, and probably not the knowledge to do that.'*
>
> Provider (non-SIB financed comparison site)

These findings align with key arguments of SIB proponents who posit that the SIB mechanism encourages reflective practice and improved capacity for active oversight of programmes through enhanced data collection techniques, management systems and sophisticated governance arrangements - often including the input of investors (Mulgan et al, 2011; HM Government, 2011; 2013).

Nevertheless, the picture with respect to this issue is more nuanced than espoused by SIB proponents. A key issue identified in the SIB-financed Trailblazer sites that was not found in the conventionally financed comparator sites related to how locally produced administrative or management data was used to inform decisions about payments amongst the respective parties (as opposed to local learning and reflective practice). This inflated the importance of these data and led to conflict between different parties in some of the Trailblazers (Authors, 2018). There is an assumption amongst SIB proponents that the goals of all parties can be aligned and that more active and extensive use of data will be intrinsically beneficial for all (Mulgan et al, 2011; HM Government, 2011). However, much of the increased local data collected in these sites did not relate directly to service user *outcomes* – but rather service provider *processes* – with important implications for performance management.

We found a range of managerial approaches to missed targets in the Trailblazers. In a number of cases, these had serious financial implications for service provider organisations should they fail to achieve targets detailed in the contracts between the service providers and investors (or intermediary organisations):

> *'We had a target of getting I think it was seven [new clients] I think by the end of March. So it was all gung-ho to try and get that through. We missed it by one I think.*

Now because of that I think we lost £600,000 worth of investment. So obviously that has had a big impact on everybody really… That was going to come from the investors. But they wouldn't give us it.'

Provider

Evidently, the goals of investors may not necessarily be aligned with those of service providers, service users, and commissioners. Withholding finance from service providers in the light of data that showed missed process targets protected the investors from further potential losses but did not necessarily further the development of the intervention locally, leading instead to inter-organisational turbulence amongst the different partner organisations and significant financial strain for the provider. It should be noted that in another Trailblazer, missed process targets triggered a change management process that enabled the reorganisation of service delivery in ways that were welcomed by most sub-contractors. It also led to a contract renegotiation between the commissioners and providers and revised targets that reflected what was possible for the remainder of the intervention rather than penalties for sustained underperformance.

This link between increased process measurement and organisational performance through the use of administrative data pressurised some staff working on the SIB-financed interventions. In interviews, some informants recounted that the financial goals of the SIB as linked to local data collection conflicted with their professional goals and responsibilities to service users (and intervention fidelity where applicable):

'I would have liked to see [a service user] sit on the programme for maybe a few months more. But from a financial perspective and from the investors' perspective we had to [terminate the process]. In many ways, that was okay, but having not had that SIB there, that side of things there, I would have been advocating or pushing further for a few months on the programme. So that's probably a really key example of where the clash is.'

Provider

For some informants working on SIB financed programmes, increased data collection attributed to the SIB mechanism was interpreted as a disciplinary device to focus service provider staff on the achieving outcomes-related rewards for the provider organisation as opposed to a collectively devised method to refine service delivery through innovative approaches that reflected on long term intervention outcomes. We found some examples of 'gaming' in the Trailblazers – as have other SIB research teams (Edmiston and Nicholls, 2017; DWP, 2014). However we also found contrasting data from provider staff committed to avoid gaming practice:

> *'[B]ecause we are a […] charity, we've been able to just ignore the potential issues with payment by results, which are that you cherry-pick and you don't work with the most in need. We have, anyway, just because we see that as our role. Reputationally, it'd be rubbish for us to just say, well, we're going to work with these easy people, and morally – why would you work for an organisation like this if you're going to do that?'*
>
> Provider

In some Trailblazer sites, we found that the data collected by service providers and local commissioners became highly politicised with debates amongst different stakeholders about the appropriate methodologies to analyse and interpret the counterfactual evidence. Once more, this was linked to the increased financial stakes related to these data as they were used as evidence to trigger payments. Whilst there is the possibility for adverse behaviours with other forms of financing, it is notable that there were no such issues in the non-SIB financed comparator sites.

Finally, we found significant issues in relation to data access. In one instance, it was impossible for all parties to audit and validate the ways in which data were collected and used due to NHS data governance issues. This led to suspicion amongst the different parties and breakdowns of trust. In another instance, a proposed outcome payment was linked to reduced unplanned hospital admissions for a defined cohort of service users. Again, due to NHS data sharing rules, it proved impossible to access such data – in the absence of which, the commissioner paid the investor without any evidence that unplanned hospital

admissions had been achieved (this was a secondary outcome that comprised 5% of the total contract value). In SIB financed interventions (in the NHS at least), there may sometimes be overly ambitious assumptions around what is achievable in terms of increased data collection and local evidence generation due to access issues. Furthermore, whilst it is possible that local data are used as a collaborative learning device, they may also be mobilised as disciplinary device. Additionally, the Trailblazers highlight that different parties may contest these data and the financial implications of this evidence may render it problematic.

(3) The opportunities SIBs offer for formal programme evaluation.

There is a growing set of empirical studies commissioned by the UK government that evaluate SIB programmes (e.g. Disley et al, 2015; DCLG, 2015). However, it is still the case that there is very little rigorous counterfactual comparison of SIBs versus alternative methods of finance to deliver the same service to the same type of users, and thus a lack of evidence of costs and benefits compared with the alternative approach to procurement (Fraser et al, 2016). Almost all of these UK evaluations have reported qualitative findings and lack data about quantitative outcomes and costs. We distinguish between overarching evaluations that seek to generate comparative data across a number of SIBs such as the Trailblazer evaluation (Authors, 2018), and focused local impact and process evaluations of individual SIBs – such as the Peterborough SIB evaluation work (Disley et al, 2015; Joliffe and Hedderman, 2014; Anders and Dorsett, 2017).

Our focus in this section is on local evaluations within the respective Trailblazer sites. It is unclear sometimes how evaluations will be paid for – this can inhibit the development of local evaluations:

> *'All of the, the commissioners' money that we've got is going into reward payments to make that as big a pot as possible to get the most outcomes. So no money was kept aside for management for evaluation.'*

Commissioner

Of the five Trailblazer projects that were commissioned, Newcastle Ways to Wellness and the London Rough Sleeping SIB commissioned local impact evaluations (assessing programme effectiveness against a counterfactual). Questions of cost-effectiveness are linked to these impact evaluations. As with the Peterborough evaluation (Disley et al, 2015), there have been some contested issues related to data collection and interpretation in Newcastle. In some Trailblazers we found an implicit assumption that any and all outcomes were attributable to the SIB intervention regardless of the fact that no local attempt has been made to prove this through any impact evaluation (though Worcester Reconnections has commissioned a local evaluation). This runs counter to the early discourse of SIB proponents who pointed to the rigour of the evaluation of the Peterborough project (Mulgan et al, 2011, HM Government, 2011). The original SIB model that the UK government and others promoted included an independent evaluation as routine to ensure that the public purse would only pay for outcomes attributable to the interventions financed by the SIB mechanism (Authors, 2018). Empirical experience in the UK would suggest that this is the exception rather than the rule and that attribution is *assumed* as opposed to independently *proven*. Additionally, cost-effectiveness data are lacking from all UK SIBs.

Discussion

We found further evidence of the 'strategic ambiguity' (Eisenberg, 1984) within the SIB concept – it can be applied to the development of both interventions with and without a 'positive evidence base' (Maier et al, 2017). Evidence from the Trailblazers suggests the three proposed interventions with some research evidence support were initiated, and most of those without research evidence were not initiated. Empirically, the Trailblazers demonstrate that SIBs can indeed promote evidence-informed programme implementation (i.e. programmes with a 'positive evidence base'). Our data suggest that the SIB financing element may bring 'added value' for an intervention like Social Prescribing as it is seen as an effective way to increase the existing evidence base. In the case of TFCO-A, informants felt it

was a good way to transfer risk and set-up costs from commissioners in a context of austerity. This study also highlights that SIBs can lead to the initiation of programmes for which research evidence does not (yet at least) exist – for the sake of evidence building and experimentation.

Importantly, there are epistemological questions around how we judge what a 'positive evidence base' is, and wider debates about what counts as 'good' evidence for policy and practice. At different times, policy makers, practitioners and service users may need to draw on different forms of knowledge and ways of knowing, depending upon the questions they seek to answer (Nutley et al, 2013). An interest in a 'positive evidence base' may orient SIB proponents towards academic research and interventions that have been developed and evaluated through established research mechanisms such as Randomised Control Trials and systematic reviews. There are advantages in policymakers carefully considering the evidence underpinning different interventions. Where their primary question is 'what works' (i.e. the question is one of relative effectiveness), there are well established 'hierarchies of evidence' based on study design (Nutley et al, 2013). Such approaches categorise evidence strength and quality based on criteria that privilege quantitative study design and value internal validity. More problematically however, hierarchies based on study design exclude important forms of evidence and underrate the value of good observational studies – moreover they fail to develop programme theory – i.e. how and why interventions may work, and a disregard for the importance of local context (Nutley et al, 2013). The prioritisation of quantitative evidence over qualitative evidence in SIB financed interventions, whilst understandable given the need to measure relative effectiveness in order to pay the investors, has been criticised on epistemological and ethical grounds. Such approaches limit the potential for programme learning, may stifle innovation, may increase pressure on provider staff and create incentives for 'creaming'. (Warner, 2013; Roy et al, 2017).

Academic research is just one type of evidence that has an impact on policy and practice (Weiss, 1979, Oliver et al, 2014). An important form of evidence is locally produced

administrative or descriptive data. A claim made by SIB proponents is that the SIB financing mechanism and the increased rigour it brings may deliver enhanced data monitoring techniques and skills to third sector providers – historically these providers were seen as lacking in this regard (Callanan and Law 2012). The importance of extensive, ongoing performance monitoring and concurrent independent evaluation by external actors is emphasised by SIB proponents as a way of ensuring that outcome payments are earned in a valid and attributable way (Cox 2011; Burand 2012). SIBs potentially offer the opportunity to draw on more (and better quality) administrative, descriptive and management data. Paying for outcomes might be expected to encourage increased and improved local data collection (Cox 2011; Liebman 2011). This in turn might increase transparency of practice for commissioners and third sector providers and increase accountability of programmes overall (Stoesz 2014).

Implementing evidence-informed interventions is highly complex, relies on the development of valued relationships over time, and assumes shared conceptions as to what evidence is (Oliver et al, 2014). Empirically, whilst the Trailblazers promoted increased data collection compared to non-SIB comparison sites, this sometimes led to increased financial pressure on provider organisations and increased managerial pressure on provider staff with potentially detrimental implications. We identified a danger that because performance data become so closely related to payment, they may become a focus for disputes between different parties and are thus counter-productive in that they remove organisational focus from service user outcomes and may introduce perverse incentives and damage inter-organisational relationships (Warner, 2013; Roy et al, 2017). Our findings support empirical data from other studies into SIBs in the UK which highlight that (as with other forms of Payment by Results) SIB financed programmes can lead to 'gaming' (DWP, 2014; Edmiston & Nicholls, 2017) which may weaken the validity of locally produced administrative data in SIB programmes. These insights align with both 'Campbell's Law' i.e. 'the more any quantitative social indicator is used for decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.' (Campbell, 1969) and 'Goodhart's Law' that suggests 'any measure

used for control is unreliable' (Goodhart, 1984). It is important to note that we did also find evidence of provider resistance to such pressure.

SIB proponents highlight the potential SIBs hold for wider learning about what works in terms of preventative policy-making and evaluation (Mulgan, 2011; Liebman, 2011; HM Government, 2011; 2013). Because SIB-financed interventions promise to pay a return to investors, it is important that public stakeholders are assured that any outcomes associated with SIB interventions are attributable to the interventions themselves. The opportunity for in depth evaluation makes the SIB concept of particular interest to the academic community and evaluation specialists (Fox and Morris, 2017). There are opportunities to combine (quantitative) impact evaluations with (qualitative) process evaluations and cost-effectiveness studies in SIBs thereby furthering knowledge of 'what works, why, when and for whom' – delivering research which transcends traditional 'hierarchies of evidence' of effectiveness and closer to a comprehensive matrix approach to evaluation (Nutley et al, 2013) that encourages collection of qualitative as well as quantitative data. Worryingly, only two of the Trailblazers included impact evaluations. In the remaining three cases, payments were linked to performance targets assessed over time as opposed to counterfactual impact evaluation. This finding aligns with broader trends identified in UK SIBs and payment by results programmes (Fox and Morris, 2017). The lack of impact and cost-effectiveness evaluation is problematic as it runs counter to aspects of the original SIB concept – that it delivers savings to government, and that government only pays for 'what works'.

Finally, we reflect on the implications of this study for practitioners. It is important that practitioners (whether commissioners or providers) are clear and explicit about their aims and expectations if considering SIB financed interventions. Our data highlight that the focus on evidence, data and evaluation in SIBs is more than just rhetorical and maybe of benefit in the design and delivery of services targeted at vulnerable populations. However, practitioners may wish to prepare for the tensions that linking evidence and data tightly to investment and outcome (re)-payments may bring.

Conclusion

The development of SIBs over recent years offers a unique opportunity to explore the explicit use of evidence in both discursive and practical terms to inform public policy and commissioning decisions in what are frequently highly politicised contexts. The relationship between SIBs and evidence goes beyond questions of the strength and transferability of the academic research that supports specific interventions and includes issues of the conversion of local performance management data into financial rewards, and of the power disparities between different local actors. Finally, the opportunities that SIBs may offer for in-depth independent rigorous evaluation, questions of cost effectiveness and overall understanding of complex interventions appear to be underutilised – it is important that policy-makers and other interested parties reflect on these as further SIB-financed programmes are developed.

Table 1: Description of SIB Trailblazer interventions

Project	Aim of project	Type of services	Outcome metrics	Strength of evidence behind the intervention	Local counter-factual data collected	Local qualitative evaluation	Status as of Aug 2018
Manchester Foster Care	Multidimensional Treatment Foster Care Oregon for Adolescents™ programme (TFCO-A™) providing behavioural interventions for 95 children aged 11 to 14 years	Specialised foster care services providing behavioural interventions for foster children in family-based settings	Number of children moved from residential care to foster placements. 'Bonus' outcome metrics: improved school attendance, better behaviour, and wider wellbeing	Medium evidence base	No	No	Active since 2014
London Thames Reach Homelessness	Personalised service pathway for a cohort of 415 entrenched rough sleepers	Navigators monitor cohort closely. Personalised approach tailored to individuals (e.g. assist to find housing)	Reduction in rough sleeping, move to stable accommodation, sustained reconnection, reduced A&E admissions, progress to employment, education, or volunteering	Medium evidence base	Yes	Yes	Completed (2012-2015)
Newcastle Ways to Wellness	Better self-management of long-term conditions through social prescribing for 14248 people with Long Term Conditions (LTC) living in West Newcastle	Social prescribing (through Link workers) (i.e. non-medical interventions in the local community to foster sustained healthy behaviours)	Achieved improvement of the outcomes on the Wellbeing Star and savings for secondary care acute usage	Medium evidence base	Yes	Yes	Active since 2015
Shared Lives	Provide an alternative to care homes for approximately 150 people in need of intensive support in two sites over 3 years	An alternative to home care and care homes for people in need of support, with support instead provided through living with a host family.	Number of new Shared Lives care placements established	Weak evidence base	No	No	Active since 2015
Worcester Reconnections	Reduce social isolation among older people through one-to-one tailored support to engage with local community. 3000 people identified as lonely aged 50+ years; reduced to 1800 people after contract renegotiation in -Spring 2016	Personalised service packages to engage individuals in local community activities (e.g. befriending services, gardening club)	Reduction in self-reported loneliness (using R-UCLA 12 scale)	Weak evidence base	No	Yes	Active since 2015

Cornwall	Improve wellbeing and health outcomes for older people.	Early interventions for a cohort of 1000 frail older people at risk of emergency admission.	Reduced A&E admissions, improved well-being (Edinburgh and Warwick mental well-being scale)	Weak evidence base	N/A	N/A	Not commissioned
East Lancashire	Improve wellbeing and health outcomes for identified cohort	Provision of patient-specific tailored health and social care interventions to reduce isolation, unemployment, and poor quality of life	Not confirmed. Outcomes would have likely included reducing isolation and returning to work or education	Weak evidence base	N/A	N/A	Not commissioned
Leeds	Improved specialist neuro-rehabilitation nursing services	Setting up a 75-bed nursing facility and creating a community of care delivering nursing care to a mix of high-needs people.	Not specified in detail. “Complex metrics” used, many outcomes, including money saved for the government by the interventions	Weak evidence base	N/A	N/A	Not commissioned
Sandwell and Birmingham	Improved End of Life Care services	Integrated community end-of-life care services	Increase in proportion of patients dying in their usual place of residence; decrease in unplanned emergency admission rate in final month of life	Weak evidence base	N/A	N/A	Not commissioned

References

Authors, 2018

Anders & Dorsett (2017) *HMP Peterborough Social Impact Bond - cohort 2 and final*

cohort impact evaluation. Department of Justice, London

Berndt, C., & Wirth, M. (2018). Market, metrics, morals: The Social Impact Bond as an emerging social policy instrument. Geoforum, 90, 27-35.

Bickerdike, L., Booth, A., Wilson, P. M., Farley, K., & Wright, K. (2017). Social prescribing: less rhetoric and more reality. A systematic review of the evidence. BMJ open, 7(4), e013384.

Burand, D. (2012), Globalizing Social Finance: How Social Impact Bonds and Social Impact Performance Guarantees Can Scale Development. *NYUJL & Bus.*, *9*, 447-502.

Callanan, L., & Law, J. (2012), Will social impact bonds work in the United States? McKinsey on Society. Washington DC.

Campbell DT. (1969) Assessing the impact of planned social change. *Evaluation and Program Planning*;**2** (1): 67–90.

Cox, B. R. (2011), Financing Homelessness Prevention Programs with Social Impact Bonds. *Rev. Banking & Fin. L.*, 31, 959-985.

Davies, H., Nutley, S. M., & Smith, P. C. (2000). What works? Evidence-based policy and practice in public services, Bristol. Policy Press

DCLG (2015), Qualitative Evaluation of the London Homelessness Social Impact Bond: Second Interim Report. London.

Disley, E., Giacomantonio, C., Kruithof, K., & Sim, M. (2015). The payment by results Social Impact Bond pilot at HMP Peterborough: final process evaluation report. Ministry of Justice (UK)

DWP (2014), Innovation Fund Pilots Qualitative Evaluation: Early Implementation Findings. London.

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

Eisenberg, E. M. (1984). Ambiguity as strategy in organizational communication. Communication monographs, 51(3), 227-242.

Evans, R., Brown, R., Rees, G., & Smith, P. (2017). Systematic review of educational interventions for looked-after children and young people: Recommendations for intervention development and evaluation. British Educational Research Journal, 43(1), 68-94.

Fitzpatrick-Lewis, D., Ganann, R., Krishnaratne, S., Ciliska, D., Kouyoumdjian, F., & Hwang, S. W. (2011). Effectiveness of interventions to improve the health and housing status of homeless people: a rapid systematic review. *BMC Public Health*, *11*(1), 638.

Fox, C. & Morris, S. (2017) Evaluating outcome-based payment programmes: what can we learn from recent methodological debates in the evaluation field? Presentation at the Second International Conference on SIBs. London. September 12, 2017

Fraser, A., Tan, S., Lagarde, M., & Mays, N. (2016). Narratives of promise, narratives of caution: A review of the literature on Social Impact Bonds. *Social policy & administration.*

Giacomantonio, C. (2017). Grant-Maximizing but not Money-Making: A Simple Decision-Tree Analysis for Social Impact Bonds. *Journal of Social Entrepreneurship*, 1-20.

Goodhart, C. E. (1984) "Problems of monetary management: the UK experience". In: Goodhart C. E. *Monetary theory and practice*. London,

Glaser, B. (1978) *Theoretical sensitivity* The Sociology Press, San Francisco

HM Government (2011), Growing the Social Investment Market: A vision and a strategy,

London

HM Government (2013), Open Public Services, London

Jolliffe, D. and Hedderman, C. (2014), Peterborough Social impact bond: final report on cohort 1 analysis, London: Ministry of Justice.

Liebman, J. B. (2011), Social Impact Bonds: A promising new financing model to accelerate social innovation and improve government performance. Center for American Progress, Washington D.C.

Mulgan, G., Reeder, N., Aylott, M., & Bo'sher, L. (2011), Social impact investment: the challenge and opportunity of social impact bonds. Young Foundation. London.

Maier, F., Barbetta, G. P., & Godina, F. (2016). Paradoxes of Social Impact Bonds. Social Policy & Administration.

Newman, J. (2001) Modernising governance, new Labour, policy and society Sage, London

Nutley, S., Powell, A., & Davies, H. (2013). What counts as good evidence? Provocation paper for the alliance for useful evidence. London: Alliance for Useful Evidence.

Oliver, K., Lorenc, T., & Innvær, S. (2014). New directions in evidence-based policy research: a critical analysis of the literature. Health Research Policy and Systems, 12(1), 34.

Pleace, N., & Bretherton, J. (2013). The case for Housing First in the European Union: A critical evaluation of concerns about effectiveness. *European Journal of Homelessness*, *7*(2).

Pope, C. & Mays, N. (forthcoming). Qualitative Research in Health Care, Fourth Edition, Sage

Roman, J. K., Walsh, K. A., Bieler, S., & Taxy, S. (2014). Five Steps to Pay for Success.

Roy, M. J., McHugh, N., & Sinclair, S. (2017). 14. Social impact bonds–evidence-based policy or ideology?. Handbook of Social Policy Evaluation, 263.

Smith, K. (2013). Beyond evidence-based policy in public health: The interplay of ideas. Springer.

Stoesz, D. (2014), Evidence-based policy reorganizing social services through accountable care organizations and social impact bonds, Research on Social Work Practice, 24, 2: 181–5.

Warner, M. E. (2013). Private finance for public goods: social impact bonds. Journal of Economic Policy Reform, 16(4), 303-319.

Weiss, C. H. (1979). The many meanings of research utilization. Public administration review, 39(5), 426-431.

Yin, R. K. (2013). Case study research: Design and methods. Sage publications.