A structured review of PbR and SIB evaluation evidence: working paper

Prepared by Chris Fox and Chris O’Leary (Policy Evaluation and Research Unit, Manchester Metropolitan University) in Summer 2017


Methodology

Search strategy

Identifying PbR and SIB programmes

We constructed a list of Payment by Results contracts that have commenced over recent years in the UK. As a starting point we used a list compiled by the National Audit Office in Figure 4 of its 2015 report on Payment by Results. The NAO (2015) compiled its own list by liaising directly with a number of government departments known to operate Payment by Results schemes. Figure 1, below, is based on this list with some additional programmes from other central government departments included. It should be noted however, that, in some sectors, payment by results contracts have spread beyond those commissioned by national government and identifying these is beyond the scope of this book. For example, the UK government launched its Drug Alcohol Recovery Pilots in 2011 and in 2015 DrugScope (2015), in its annual review of the substance misuse sector noted that PbR contracts were also being developed outside of the formal pilots, most likely local contracts between substance misuse service providers and local government commissioners of services.

We exclude National Health Service Payment by Results programme, because these are a particular type of payment by results programme that is more akin to ‘payment by outputs’ and falls outside the definition of payment by results adopted for this review. We also
exclude programmes commissioned by the Department for International Development that have a PbR component but are delivered outside the UK on the basis that the context for these programmes is significantly different to those programmes delivered in the UK.

We identified every live SIB in the UK as at April 2017. This was based on a database maintained by Social Finance\(^1\). Figure 1, below, is based on this database.

*Identifying evaluations*

A thorough search was undertaken for any published evaluation associated with each PbR and SIB programme identified. This included searching websites associated with the programmes, their funders, investors and service providers. In addition a structured search of the two databases was undertaken: ASSIA and Web of Science. The search was conducted in June 2017 with no time or language filters using the following search terms:

“social impact bond*” OR “pay for success bond*” OR “pay for success contract*” OR “social outcome invest*” OR “social impact invest*” OR “social invest*” OR “payment by results” OR “pay for success” OR “outcome-based commissioning” OR “outcom-based payment” OR “outcome-based funding” OR “outcome commissioning” OR “commissioning for outcomes” AND evaluat*

*Screening*

Papers identified through these two search strategies were entered into Endnote and de-duplicated. 811 papers were sifted for relevance based on reading titles and abstracts. Inclusion criteria were as follows:

- Reports on empirical evaluations of PbR and SIB programmes undertaken in the UK

Exclusion criteria were as follows:

- Reports on empirical evaluations of PbR and SIB programmes undertaken outside the UK. This includes reports on empirical evaluations of PbR and SIB programmes commissioned by the UK Department for International Development but delivered

\(^1\) [http://www.socialfinance.org.uk/database/]
outside the UK

- Papers that did not report empirical evaluation data
- Papers that report empirical evaluations of the UK NHS Payment by Results programme
- Papers that reported ex ante evaluations of PbR or SIB programmes that had yet to be implemented.

A broad interpretation of ‘evaluation’ was taken so that a number of ‘reviews’ and pieces of research published in peer reviewed journals that nevertheless had evaluative elements were included.

58 papers were retained. The full text of all retained papers were retrieved and a second sift for relevance, using the same criteria set out above, but this time based on a reading of the whole paper was undertaken. At this point 46 papers were retained for full analysis. The 46 papers and the PbR or SIB programme they are associated with are set out in Figure 1.

Quality

No papers were excluded based on methodological rigour. However, the methodologies of the 46 papers were then assessed. Qualitative evaluations were assessed using the set of quality standards for qualitative evaluation was drawn up by the UK government’s Cabinet Office (Spencer et al. 2003). These was chosen because they are widely used and many of the evaluations under consideration were commissioned by the UK government, some with involvement from the Cabinet Office. The framework is based around four guiding principles – that research should be (2003:20):

- **Contributory** in advancing wider knowledge or understanding about policy, practice, theory or a particular substantive field.
- **Defensible** in design by providing a research strategy that can address the evaluative questions posed.
- **Rigourous** in conduct through the systematic and transparent collection, analysis and interpretation of qualitative data.
- **Credible** in claim through offering well-founded and plausible arguments about the significance of the evidence generated.
The high-level framework is set out in Figure ?? (see Spencer et al. 2003 for more detail on quality indicators associated with each appraisal question).

<table>
<thead>
<tr>
<th>Domain</th>
<th>Appraisal questions</th>
</tr>
</thead>
</table>
| Findings | How credible are the findings?  
| | How has knowledge/understanding been extended by the research?  
| | How well does the evaluation address its original aims and purpose?  
| | Scope for drawing wider inference – how well is this explained?  
| | How clear is the basis of evaluative appraisal? |
| Design | How defensible is the research design? |
| Sample | How well defended is the sample design/ target selection of cases/documents?  
| | Sample composition/case inclusion – how well is the eventual coverage described? |
| Data collection | How well was the data collection carried out? |
| Analysis | How well has the approach to, and formulation of, the analysis been conveyed?  
| | Contexts of data sources – how well are they retained and portrayed?  
| | How well has diversity of perspective and content been explored?  
| | How well has detail, depth and complexity (i.e. richness) of the data been conveyed? |
| Reporting | How clear are the links between data, interpretation and conclusions – i.e. how well can the route to any conclusions be seen?  
| | How clear and coherent is the reporting? |
| Reflexivity and neutrality | How clear are the assumptions/theoretical perspectives/values that have shaped the form and output of the evaluation? |
| Ethics | What evidence is there of attention to ethical issues? |
| Auditability | How adequately has the research process been documented? |
The framework is designed to aid the informed judgement of quality, but not to be prescriptive or to encourage the mechanistic following of rules. The questions are open-ended to reflect the fact that appraisals of quality must allow judgement, and that standards are inevitably shaped by the context and purpose of assessment (Spencer et al. 2003).

The design of impact evaluations was assessed using Sherman et al.’s (1998) Scale of Scientific Methods (the Maryland Scale). This is a fairly ‘broad brush’ scale which prioritises causal inference and internal validity in impact evaluation design. Where papers were not designed primarily as evaluations, but nevertheless had been included because they had some evaluative elements where the standards set out above were relevant, they were applied. Where they were less relevant, professional judgement was used to assess overall methodological rigour of the paper.

**Results**

Of the 46 papers reviewed 30 relate to PbR programmes, 15 to SIBs and one covers both a PbR programme and a SIB (Ministry of Justice 2014). A summary of the papers and the programmes they relate to is set out in Table below. Four papers were published in peer reviewed journals, the remainder are ‘grey literature’ reports, the majority of which are published by the UK government departments that commissioned them. UK government evaluation reports are themselves subject to a peer review process.

The majority of papers (37) are primarily implementation evaluations that use either exclusively qualitative methods (typically semi-structured interviews and reviews of project documentation) or a mix of qualitative and quantitative methods (typically a survey of service providers or service users combined with semi-structured interviews). Assessed against Spencer et al.s (2003) framework for ‘quality in qualitative evaluation’ the majority of evaluations are reasonably strong on overall evaluation design, analysis and reporting. Consistent areas of weakness were initial development of a theory of change or logic model, an approach adopted by a handful of evaluations and only fully elaborated in one (Department for Communities and Local Government 2014b); sampling, which typically
claimed to be purposive but rarely include a clear rationale for case selection; and research ethics, addressed by only a handful of evaluations.

Eight papers include a quantitative impact evaluation. Some are interim or supplementary reports and so in total they relate to five different programmes, one of which is a SIB. Of these, none involves a randomised controlled trial (RCT). Almost all involve a quasi-experimental design, typically an intervention group matched to a control group using Propensity Score Matching. Several studies report significant limitations to the analysis due to data quality. These evaluations typically rely on existing, administrative data sets and often report challenges in accessing data or the poor quality of the data sets they are able to assemble prior to analysis (see for instance Bewley et al. 2015 and Newton et al. 2014). Some weaker designs are also used. For example, Ministry of Justice (2015) reporting on the final reconviction results for the HMP Doncaster Payment by Results pilot uses a historical comparator group.

Synthesis

Some initial themes were taken from previous reviews including Tan et al. (2015), Fraser et al. (2016) and National Audit Office (2015). Others were created during the analytical process to reflect themes emerging from the data. The final list of themes was:

- Commissioning process
- Role of supply chains
- Risk transfer
- Service quality
- Outcomes and intermediate outcomes
- (Re)alignment of social and financial outcomes
- The role of intermediaries, social investors and partnerships
- Innovation and the role of evidence
- Scaling up existing approaches
- Finance including innovation in finance
- The role of incentives
- Unintended outcomes and perverse incentives
Markets, new models of delivery and new providers
Performance management
Targeting, segmentation and differentiation of client groups
Other

Our synthesis of findings from these evaluations is reported under the following headings:

- designing and commissioning outcome-based models;
- implementing outcome-based commissioning models;
- markets;
- performance management;
- innovation;
- the role of incentives; and
- overall outcomes.

Designing and commissioning outcome-based models

SIBs

A consistent finding across a number of SIB evaluations is around the complexity of contractual relationships with a dense web of contractual obligations being created between the various parties. For example, Disley et al. (2015) describe six contractual relationships created as part of the Peterborough SIB between: (i) Social Impact Partnership (the limited partnership set up by Social Finance which was the contracting entity in the SIB) and the Ministry of Justice; (ii) Social Impact Partnership and the Big Lottery Fund; (iii) Social Impact Partnership and the commissioned service providers; (iv) Social Impact Partnership and the investors; (v) The Ministry of Justice and Peterborough Prison Management Limited (the consortium which holds the private finance initiative contract for HMP Peterborough); (vi) The Ministry of Justice and the Independent Assessor.

Developing the Peterborough SIB took approximately 18 months from the point at which Social Finance initially discussed the concept with civil servants, to its launch (Disley et al. 2011). The Ways to Wellness SIB took three years (ATQ Consultants and Ecorys 2015), the
Reconnections SIB took ten months (ATQ Consultants and Ecorys 2016) and Roberts and Cameron (2014: 6) report that: “Commissioning the SIB in Essex was a long and complex process”

The commissioning process for most SIBs involved detailed design and development work to agree on a policy problem, identify a potential solution, design an interventions outcomes and payment schedule and evaluate funding options (see for instance Tan et al. 2015 on the Health and Social Care SIBs). Sometimes this included formal feasibility studies and consultation exercises (for example, the London Homelessness SIB as described by Department for Communities and Local Government 2015). One of the challenges in development stage is in the sequencing of different stakeholders in relation to the decisions that must be made (for example, see ATQ Consultants and Ecorys 2016 on the Mental Health and Employment SIB).

Public procurement rules add to the complexity and uncertainty because organisations that participate in the feasibility phase usually still have to enter into competitive tenders to provide the service meaning that the eventual provider may not be the one involved in feasibility discussions (Tan et al. 2015). A competitive, open procurement process was used in the Reconnections SIB allowing the commissioner to stipulate the shape and nature of the programme and encouraging providers to submit a more competitive offer. However, the open procurement process only produced one tender ATQ Consultants and Ecorys (2016). Factors such as these explain why evaluations consistently found that SIBs take a long time to design and commission and that the cost of establishing them is high. Several evaluations comment on the high costs of commissioning a SIB (for example, Sin and Roberts 2016 in the Essex Multi-Systemic Therapy SIB and the Department for Communities and Local Government 2014 in the London Homelessness SIB, but few quantify these costs. ATQ Consultants and Ecorys’s (2016) review of the Mental Health and Employment SIB reports that the cost of designing the SIB was £150,000 although other costs borne by stakeholders were found not to have been above and beyond the normal cost of commissioning a new service. The Reconnections SIB cost nearly £200,000 to develop (ATQ Consultants and Ecorys’s 2016). Costs included £90,000 to carry out a cost-benefit analysis of the project and understand some of the metrics that may suit a possible SIB; £64,000 to pay for a part-time Project Manager and extra research to help inform the development of the application; and
£35,000 in-kind contribution from the local authority linked to staff time to further develop the SIB.

**PbR**

A number of evaluations highlight the complexity of PbR commissioning models compared to comparable commissioning exercises. For example, the evaluation of the Supporting People PbR pilot found that there was a need to provide clarity around targets, how performance would be measured and the consequent impacts on income, thereby often requiring more detailed guidance to be provided at tender stage (Department for Communities and Local Government 2014). that was overly complex with a large number of metrics included.

A number of evaluations of PbR programmes found that insufficient time was allowed for setting up and commissioning the programme (for example, Department for Education 2014 in the national commissioning elements for PbR in Childrens’ Centres and Wong et al. 2015 in the Local Justice Reinvestment Pilot). Several evaluations identified that building performance management systems to support PbR took additional time and resource compared to comparable commissioning models (eg Wong et al. 2013 in relation to Local Justice Reinvestment Pilots and Murray et al. 2012 in relation to Doncaster Prison PbR pilot). However, establishing PbR contracts did not always take additional time. The Department for Communities and Local Government (2014) found that, in the case of the Supporting People PbR pilot the process of contracting Payment by Results terms followed the same process as other non-PbR contracts with overall timescales consistent with non-PbR commissioning processes.

The high cost of procurement in outcomes-based programmes is a recurrent theme. In PbR evaluations both prime contractors and their sub-contractors routinely report to evaluators that upfront investment in the commissioning process was higher than they had expected (eg Newton et al. 2014 in the Youth Contract, Department for Communities and Local Government 2014 in the Supporting People PbR pilot, Lane et al. 2013 in the Work Programme). In the Childrens’ Centre PbR programme trial grant funding was viewed as essential to the development of PbR in most areas (Department for Education 2014). A
study by Egdell et al. (2016) of the experiences of Third Sector Organisations (TSO) in the Work Programme in Scotland found that TSOs made considerable investments in building relationships with potential primes during the procurement process including providing considerable information to potential primes, often in different formats. A two-stage procurement process was used, meaning TSOs had to build relationships with numerous potential primes across both stages. But this activity was highly speculative since it was not known whether the primes that the TSOs built relationships with would be successful in getting contracts.

Different commissioning approaches

Where flexibility in commissioning processes was possible this seemed to improve the process of developing outcome-based commissioning. For example, in relation to the Supporting People PbR Pilot in many pilot areas, commissioners actively involved providers, and in some cases clients, in the development of Payment by Results models. The evaluators found that in those areas where providers have been heavily involved in the process, they appear to be more comfortable with the terms adopted (Department for Communities and Local Government 2014). Disley et al. (2014), based on their evaluation of the Peterborough SIB suggest that SIBs designed in a ‘bottom-up’ way, like the Peterborough pilot, are more likely to attract social investment. In that case investors were motivated by the desire to fund interventions to address identified needs of service users. The Peterborough SIB was perceived by investors to respond to a locally defined problem and to have been developed in collaboration with local agencies taking local context into account. In the London Homelessness SIB evaluators describe a ‘Competitive Dialogue’ process used to commission the SIB that might illustrate one approach to ‘bottom-up’ design. The competitive dialogue was new for a service intervention of this type, but important as it enabled a dialogue with providers about the complex issues in the SIB design. The evaluators reported that the process produced refinements rather than substantial changes, but those involved agreed that it enabled both commissioners and providers to develop confidence: for commissioners, that high quality tenders would be submitted; and for providers, that they could develop their initial plans to meet the requirements of commissioners (Department for Communities and Local Government 2014).
Underpinning debates about quantifying and measuring social outcomes are a range of political debates. For example, in the evaluation of the Health and Social Care SIBs the evaluators reported that in some sites health commissioners expressed concerns about involving private investors because of the legacy of PFI and the common view that the public sector paid out excessively large returns to the private sector (Tan et al. 2015)

**Implementing outcome-based commissioning models**

A number of factors seem to consistently inform implementation of PbR and SIB programmes.

*Relations between prime providers and ‘supply chains’*

Different approaches to sharing risk with supply chains and delivery organisations are evident.

In some of the earlier and larger PbR programmes it was common for risk to be passed down the supply chain. In the Work Programme evaluators (Lane et al. 2013) found that almost all first tier subcontractors were paid on roughly the same outcomes basis as prime providers or a modified version of this model. The evaluators reported that primes typically deducted a 10 to 20 per cent ‘management fee’ from payments made to first tier subcontractors. They conclude that:

> “It seems that the risk in the outcome-based funding model is to a large extent being handed down the supply chain and shared by Tier One subcontractors” (Lane et al. 2013: 35).

A similar approach was identified by evaluators of the Youth Contract PbR programme (Newton et al. 2014).

Risk transfer to sub-contractors was not always a feature of PbR programmes. For example, in the Doncaster Prison PbR pilot the arrangement between the private sector prime contractor and the voluntary sector sub-contractor did not put the voluntary sector provider at risk. A proportion of the outcome-related payment would be paid to the voluntary sector provider if the outcome measure was achieved but there was no associated financial risk if
the outcome was not achieved. This risk was retained by the private sector prime contractor (Murray et al. 2012).

A number of PbR evaluations highlighted the extension of ‘marketisation’ into the management of contracts. For example, the Department for Work and Pensions (DWP) managed the provider ‘market’ so that prime contractors in its Work Programme could compete for market share based on their performance. This happened through a process of ‘market share shifting’, under which better-performing providers were rewarded by being allocated additional clients to work with, while poorer-performing providers (who nevertheless remained above a minimum quality threshold) received fewer clients (Department for Work and Pensions 2014). Prime contractors then applied the same approach to managing their supply chains (Department for Work and Pensions 2014).

In SIBs risk has generally not been transferred to the front-line delivery organisation. For example, in the Peterborough SIB the intermediary and organisations invested in the SIB felt that a strength of the model was that commissioned service providers from the voluntary and third sector did not bear outcome risk dependent on results. They were paid upfront or on a fee-for-service basis (Disley et al. 2015).

Role of social investors and intermediaries

Where social investors were involved evaluations generally reported that they were committed to securing good social returns for interventions as well as ensuring that they were financially viable (eg Griffiths et al. 2016 in the Department for Work and Pensions Innovation Fund, Department for Communities and Local Government 2014 in the London Homelessness SIB). Several evaluations noted that social investors often gave significant expertise and ‘hands-on’ support to the interventions they had invested in and that this went beyond financial support (eg Griffiths et al. 2016 in the Department for Work and Pensions Innovation Fund and Roberts, Tan et al. in the Health and Social Care SIBs and Cameron 2014 in the Essex Multi-Systemic Therapy SIB).

Several evaluations also noted that intermediaries were often key to the success of the SIB, whether the intermediary functions were undertaken by a separate body or managed in-house. Intermediaries were particularly important in helping to manage increased
workloads in areas such as finance and performance management (eg Disley et al. 2015 in the Peterborough Prison SIB and Griffiths et al. in the Department for Work and Pensions Innovation Fund). Roberts and Cameron (2014), based on their evaluation of the Essex Multi-Systemic Therpay SIB go as far as to suggest that the success or otherwise of the SIB approach may not be due not simply to the SIB model but to the behaviours and levels of engagement of different investors in specific SIBs.

Throughputs

A common finding in evaluations of both PbR programmes and SIBs is that the throughput of clients estimated during the commissioning process turned out to be lower in practice than commissioners had suggested it would be (eg Newton et al. 2014, evaluating the Youth Contract) or long-term downward trends are not fully understood until the commissioning process is underway (Roberts and Cameron 2014). Newton et al 2014 report that prime providers in the Youth Contract stated that the level of their own and subcontractors’ upfront investment was higher than expected. Combined with what they claimed to be an overestimate of the eligible population in the specification, the low cost base, slow flows of young people onto the Youth Contract and the back-loading of funding on sustained outcomes, this was said to have ultimately undermined their financial models.

Scale is therefore important. Wong et al. (2015) in their evaluation of the Justice Reinvestment Pilots contrast Greater Manchester where a decision was made to run the pilot across ten local authorities with pilots in London that were typically in in one London Borough. Greater Manchester benefited from economies of scale associated with its larger ‘footprint’ and higher throughput of service users.

‘Bid candy’

A concern within the Third Sector has been that voluntary sector organisations have sometimes been included in bids for PbR programmes to enhance the perceived quality of the bid, but have then not received any or as much business as was promised at the bid stage. Such organisations have been referred to as ‘bid candy’ [REF]. This issue was addressed explicitly during the evaluation of the Work Programme. The evaluators (Lane et
al. 2013) found that there was evidence of some providers dropping out of supply chains between bids and contracts starting, but they found little to support the idea that subcontractors had been named in proposals purely to help prime contractors secure the contract and were subsequently dropped from supply chains. Egdell et al. (2016) in their study of Third Sector Organisations in the Work Programme in Scotland found that many such organisations reported only a small number of client referrals, which created uncertainty for staff and income flows and Lane et al. (2013) report that in some cases, lower than expected referrals had led to staff being laid off or kept on zero hours contracts. Lane et al. (2013) found that sub-contractors not receiving referrals were often unhappy about this but tended to feel that this was due to other factors such as insufficient referrals volumes rather than lack of intent on the part of primes. However, later in the evaluation of the Work Programme evaluators found evidence that some prime contractors who used a ‘referral volume shift’ model to manage their subcontracts, whereby referrals were shifted from lower to higher performing generalist end-to-end providers in their supply chain, were also maintaining or expanding their own referral volume share even where some of their subcontractors were outperforming them (Department for Work and Pensions 2014).

Markets

In the UK, the introduction of PbR has often been explicitly linked to (re)shaping the market for services in a particular sector. A number of PbR evaluations document how the introduction of PbR has made a significant impact upon a market:

Purvis et al. (2013) in their evaluation of the Work Choice Specialist Disability Employment Programme identified significant changes to the structure of the specialist disability employment provider market. The introduction of the prime provider model meant that DWP managed fewer contracts with a smaller number of providers. A number of providers entered the market; most of these had previously delivered other DWP contracts though a small number were new to DWP contracting and were typically small disability-focused charities. Some of the new prime providers had previous experience as prime providers on DWP programmes but had limited experience of specialist disability provision. A large number of local authorities exited the market. At a local level, within the supply chains of prime providers a number of the providers told evaluators that they believed there had
been a loss of expertise, local knowledge and employer relationships with the move to the current market due to reduction in number of specialist providers and their reduced role.

By contrast, the evaluation of the Work Programme found that, while there had also been significant change within the market, the market was not concentrated when compared to other markets such as the UK supermarket sector or the UK domestic gas and electricity market. At prime provider level, the evaluators found that the top four prime providers delivered around 54 per cent of the Work Programme. They therefore described the market as an “unconcentrated, competitive oligopoly, which has remained fairly stable over time” (Department for Work and Pensions 2014: 30). When they took supply chains (circa 800 organisations) into account their analysis shows even less concentration, although market concentration was increasing over time (Department for Work and Pensions 2014). The evaluation of PbR in Children’s Centres also found no indications that PbR had had adverse effects on the types of providers willing to operate in the market (Department for Education 2014).

A few evaluations found mixed effects for PbR on the structure of markets. For example, Newton et al. (2014) in their evaluation of the Youth Contract report that some national stakeholders commented that the providers were not typical of those operating in education and would as a result, bring new blood into it. However, the evaluation also identified a concern that small organisations in the voluntary, community and social enterprise (VCSE) sector were squeezed out of the national model because they were unable to risk upfront investment since payments were weighted towards later outcomes. Suggestions were made that their exclusion had stifled innovation and the delivery of specialist support.

**Performance management**

Many of the evaluations reviewed identify a greater focus on performance management and the development of new or enhanced performance management systems. More evaluations tended to identify positive effects associated with the increased focus on performance management, in particular evaluations of SIBs.

*Advantages and disadvantages of a stronger focus on performance management*
SIB evaluations tended to report the most unequivocally positive effects associated with performance management. So, for example, Griffiths et al. (2016) in their evaluation of the Department for Work and Pensions Innovation Fund noted that the amount of data required and the intensity of performance monitoring needed to successfully manage a 100 per cent outcome funded project came as something of a ‘culture shock’ to many deliverers, but also documented a great deal of direct and ‘hands-on’ involvement in projects from social investors in the areas of performance management, client tracking, and outcomes-profiling systems. They concluded that the SIB funding model appeared to have created a high intensity of focus on performance across all projects. Similar findings are reported by Roberts and Cameron (2014) in relation to the evaluation of the Essex Multi-Systemic Therapy SIB, Tan et al. (2015) in relation to Health and Social Care SIBs and ATQ Consultants and Ecorys (2016) in their review of the Mental Health and Employment Partnership SIB.

Some PbR evaluations also identified predominantly positive effects arising from a greater emphasis on performance management. The Evaluation of the Troubled Families Programme found that, while the financial framework presented many challenges to local teams, including imposing a high degree of accountability, requiring a certain level of multi-agency co-operation and effective data sharing (Day et al. 2016), this approach was instrumental in raising the quality and capacity of local data management systems (White and Day 2016). Similar findings are reported in the evaluation of the Supporting People PbR Pilot (Department for Communities and Local Government 2014) the evaluation of PbR Childrens’ Centres (Department for Education 2014) and the evaluation of the Youth Justice Reinvestment Custody Pathfinder (Wong et al. 2015).

The advantages of enhanced performance management associated with PbR were less clear-cut in some PbR evaluations. For example, in the evaluation of the Work Programme there were many reports of both primes and their sub-contractors improving their performance management systems over the lifetime of the contracts. However, the competing perspectives of contract and performance managers in the Department for Work and Pensions illustrated what the evaluators suggest appeared to be an underlying tension between offering flexibility to providers to deliver services as they see best and driving up performance by prescribing delivery methods that have been found to be effective. Thus, some DWP contract managers identified benefits from the rigorous contract management
that accompanied PbR, including that it provided DWP with more ‘power’ over prime providers and made them easier to manage, that it focused providers, and that it was more cost effective. Others however also questioned whether the new regime was too intensive and/or prescriptive (Department for Work and Pensions 2014).

In a number of PbR evaluations the effects of enhanced performance management were mostly negative. For example, in the evaluation of the Drugs and Alcohol Recovery PbR Pilot Programme evaluators found PbR had encouraged services and practitioners to place a greater emphasis on monitoring and reviewing the progress of those in treatment, but this had the potential to alter and distort aspects of practice and risked undermining the responsiveness of services. Challenges included: a lack of engagement and dialogue with providers prior to design and implementation; greater than anticipated data requirements and related costs for modelling and evidencing outcomes; local design of payment models and indicators; and concerns about the accuracy of estimates being produced for the complexity of treatment caseloads and the impact of this on the potential funding available to providers via PbR (Mason et al. 2015).

Newton et al. (2014) in their evaluation of the Youth Contract noted that some national stakeholders stressed that a downside to a ‘black box’ approach was that it could be difficult to access detailed information about delivery including the methods that were being tested by providers, emerging effective practice, or issues which would indicate any need for changes or flexibilities to the design of the programme. Within supply chains some primes reported that their more inexperienced sub-contractors struggled to use software systems supplied by primes and so under-reported outcomes. Pearce et al (2015) in their evaluation of the Doncaster Payment by Results pilot found that there was little confidence in the data management system among staff. This resulted in duplicating information across databases and obtaining limited management information.

**Complexity and cost**

A consistent finding across many evaluations is that enhanced performance management was more complex and hence more costly than in comparable contracts.
Evaluations noted that measuring social impact is still in its infancy. Disley et al. (2011) found that the ability of investors and markets to deal in social outcomes risk was currently underdeveloped: few agreed that metrics exist for quantifying social outcomes, whether positive or negative; conventional finance markets do not price social value creation; and consequently, there is a lack of comparable performance information to support the creation of a new or modified social investment marketplace. Similar points were made by ATQ Consultants and Ecorys (2015) in relation to the Ways to Wellness SIB. Tan et al. (2015) in relation to Health and Social Care SIBs describe the use of cost-benefit analysis to develop outcome metrics and the development of bespoke information management systems that, in some cases, require new or specialised staff. They note that performance measurement, outcome payment thresholds and values, are rarely transferable between SIB-funded initiatives. Wong et al. (2015) describe how the large number of metrics in the Local Justice Reinvestment Pilots were perceived as challenging because they made it harder for sites to work out what interventions to implement. The complexity of metrics was also one factor that they attribute to some pilot sites undertaking little or no performance management activity. Conversely, other evaluations of PbR in the criminal justice sector noted some challenges in adopting an overly simplistic, single, binary outcome measures in a PbR programme (for example, Foster et al. 2013 in the Employment and Reoffending Pilot and Pearce et al. 2015 in the Doncaster Payment by Results pilot).

Gosling (2016: 527) takes a more critical view of the challenges inherent in defining and measuring social outcomes and argues, in relation to the use of PbR in a Therapeutic Community that: “PbR creates a clear dichotomy between the achievement of a successful outcome and demonstration of a recovery journey.” It may be that progress has been made in this area over the last 6 or 7 years and in a review of one of the most recent SIBs with aims to address loneliness ATQ Consultants and Ecorys (2016) note that the SIB quantified in a rigorous way the costs and benefits of loneliness and put forward a stronger ‘case for investment’ for this issue.

In the Work Programme a number of primes reported an increasing administrative burden around performance management, in particular around providing evidence for sustainment payments (Lane et al. 2013). The Department for Communities and Local Government’s (2014) evaluation of the Supporting People Payment by Results pilots reports that, although
commissioners and providers were unable to quantify the cost of establishing and running Payment by Results services, all agreed that the costs were higher than for non-Payment by Results contracts due to the more intense monitoring requirements under PbR contracts. The evaluation of the London Homelessness SIB reported that the Greater London Authority had created a dedicated post for monitoring the two SIB contracts. This is in contrast to other contracts for homelessness services, all of which are covered by a single post.

Innovation

Two broad areas of innovation are common themes within evaluations of outcome-based commissioning: innovation in service design and delivery and innovation in financing: innovation in service design and delivery and financial innovation.

Innovation in service design and delivery

Only two evaluations made strong claims for extensive innovation within the programme they evaluated. Day et al. (2016) in their evaluation of the Troubled Families programme are clear that the Programme encouraged innovation, and the desire to trial new ways of working. They report local authorities adopting more evidence-based approaches, experimenting with early interventions and new modes of delivery. However, there is little evidence offered to demonstrate that innovations observed were the result specifically of the PbR element of the programme. ATQ Consultants and Ecorys (2015) in their review of the Ways to Wellness Social Impact Bond suggest that the intervention is truly innovative and that stakeholders were consistent in emphasising its innovative nature.

Primarily in relation to SIBS a common theme was that while programmes tended to implement well-established, evidence-based interventions, there were sometimes elements of innovation around the mode of delivery. The evaluation of the London Homelessness SIB notes that the service delivered is not in itself particularly innovative although the SIB framework allows for some innovation in mode of delivery, particularly around flexibility to deliver more individualised solutions (Department for Communities and Local Government 2015). Disley et al. (2015) note that while the service developed at Peterborough Prison was not particularly innovative, its extension to short-term prisoners who would not previously
have received an intervention was, to some extent, an innovation. Similarly the Essex Multi-
Systemic Therapy SIB extended an established model to a group not previously prioritised
for this intervention. Several SIB evaluations describe models that involve ‘scaling-up’
established service models (eg Tan et al. 2015 in relation to the Health and Social Care SIBs,
Day, et al. 2016 in relation to the Troubled Families programme and ATQ Consultants and
Ecorys 2016 in relation to the Mental Health and Employment SIB).

There were also a few isolated examples of innovation arising from some evaluations. One
prime provider in the Work Programme conducting a personal budgets pilot, another
piloting new approaches to in-work support, and another that had commissioned specialists
to develop tools for use by their whole supply chain. The evaluators report that the most
common examples of innovation in the Work Programme had occurred in ICT. For example,
two prime providers mentioned using Cloud technology, one had rolled out a new online
client portal and a further one was developing e-learning (Department for Work and
Pensions 2014). In the Doncaster Prison PbR Pilot processes presented as an innovation
included transferring staff from the private sector partner to the Third Sector partner as well
and elements of personalisation in case management and client engagement. The
commissioning of new services sometimes involves staff from an old service being
transferred from their old organisation to the winning contractor using the Transfer of
Undertakings (Protection of Employment) Regulations 2006\(^2\), commonly known as TUPE.

While such transfers might be regarded as innovative by some they are not particular to PbR
programmes and in this case the evaluators noted that although the TUPE transfer was
regarded as innovative by senior stakeholders, delivery staff stated that they found the
transition challenging and would have liked more communication, support and training.

Overall, evidence from the evaluations suggests that innovation in the design and delivery
of services was limited. For example:

- The evaluation of the Supporting People PbR Pilot found that: "True examples of
innovation in service delivery have not materialised as a direct consequence of Payment
by Results" (Department for Communities and Local Government 2014: 28).

\(^2\) As amended by the "Collective Redundancies and Transfer of Undertakings (Protection of
Employment) (Amendment) Regulations 2014
• The evaluation of Peterborough SIB noted some potentially innovative elements of service delivery but concluded that they were not necessarily a result of SIB funding and had long precedents in England and Wales (Disley et al. 2015).

• The evaluation of the Essex Multi-Systemic Therapy SIB notes that the intervention is well-established and not therefore innovative (Roberts and Cameron 2014).

• The review of the Mental Health and Employment SIB concludes that the service is not particularly innovative (ATQ Consultants and Ecorys 2016).

Several evaluations explored whether the ‘black box’ approach to commissioning led to innovation. Generally, they concluded that it did not. Thompson et al. (2011) in their evaluation of the Work Choice Specialist Disability Employment Programme note that commissioners intended that the ‘black box’ model of commissioning would drive innovation, but that providers felt that prescriptions about the amount and intensity of client contact constrained innovation. Purvis et al. (2013) in a later report from the same evaluation surveyed the 56 organisations delivering Work Choice about impact the commissioning model had on aspects of their activity and found they generally reported negative responses on all aspects including innovation. The Work Programme evaluation found that:

"The black box model allows for a provider to innovate within the limits of the contract in order to test out new ways of working, however, there appeared to be limited evidence of innovation in service design and delivery at the prime provider or subcontractor level.” (Department for Work and Pensions 2014: 121)

Factors potentially limiting innovation were not fully explored in the evaluation, but the evaluators suggest they may be linked to perceived limitations in contract flexibility or financial elements of the contract. In an early evaluation report it was noted that commissioners and providers tended to have a different understanding of the ‘back box’ principle with commissioners only applying it during the commissioning phase while many providers believed they had continuing freedom to deliver in flexible ways once the programme went live (Lane et al. 2013).

Many national stakeholders interviewed as part of the Youth Contract highlighted the potential for ‘black box’ commissioning to encourage innovation (Newton et al. 2014). However, the evaluation also identified a concern that small organisations in the voluntary,
community and social enterprise (VCSE) sector were squeezed out of the national model and that this had stifled innovation and the delivery of specialist support. Also, a survey of local authorities suggested that novelty and innovation in the targeting of young people was limited.

More typically, improvements in service design were attributed to continuous improvement stimulated by an increased focus on performance management. For example, Wong et al.’s (2015) evaluation of the Youth Justice Reinvestment Custody Pathfinder gives little sense that innovation was a significant factor in achievement of outcomes. Rather, improvements are explained in terms of marginal gains drive by a focus on key performance metrics. Similar findings emerge from Griffiths et al.’s (2016) evaluation of the Department for Work and Pensions Innovation Fund and Wong et al.’s (2013) evaluation of the Local Justice Reinvestment Pilot.

Financial innovation

Several evaluations, mostly of SIBs, report that one of the impacts of introducing a PbR framework was in financial innovation. For example:

- In relation to Health and Social Care SIBs Tan et al. (2015) found some service providers were highly motivated to take part in a SIB because SIBs were seen as offering financial freedom from shorter grant-giving cycles and process-measure-driven contracts. Specifically mentioned were SIB projects that focused on preventive interventions that were seen as attractive because they would allow time for interventions to become established and for outcome metrics to determine whether the intervention was working.

- ATQ Consultants and Ecorys’s (2016) review of the Mental Health and Employment Partnership SIB allowed public sector organisations to carry on commissioning Voluntary, Community and Social Enterprise (VCSE) providers of services where there would otherwise have been cuts in budgets and to the scale of delivery. Service providers benefited from the certainty of 3 year contracts.

- Providers reported new investment associated with participation in Work Choice PbR contracts and typically focussed on the infrastructure to allow providers to operate the programme effectively including ICT, premises and new staff/staff training (Thompson
• Disley et al. (2015) note that the service created by the Peterborough SIB was perceived to be more flexible and agile than other interventions, allowing for new providers to be commissioned and new ways of working to be implemented throughout the pilot. Similar points are made by Griffiths et al. (2016) in relation to the SIBs funded by the Department for Work and Pensions Innovation Fund.

• Disley et al. (2014) in their evaluation of the Drugs and Alcohol Recovery PbR pilot programme note that when asked to identify the most effective aspects of the PbR funding models, interviewees tended to emphasise how these had helped to ensure a greater focus by commissioners and providers on the nature and complexity of the needs of local treatment populations and had sharpened the focus on particular issues, such as throughcare and aftercare. A similar finding comes from Newton et al.’s (2014) evaluation of the Youth Contract where most national stakeholders thought that PbR was crucial to achieving sustained outcomes and would increase providers’ focus on young people’s needs and White and Day in relation to the use of PbR in the Troubled Families programme.

The role of incentives

In general, evaluations suggest that providers tend to be risk averse (eg Disley et al. 2014 in the Drugs and Alcohol Recovery PbR Pilot) and that taking on increased risk has led to providers looking for bigger financial incentives, although Griffiths et al. (2016) in their evaluation of the Department for Work and Pension’s Innovation Fund in which a number of SIBs were funded, found that that paying a high rate for an outcome will not incentivise intended behaviour if investors perceive the risk of pursuing that outcome to be too great. In this scenario providers will tend not to enter the market. In the Work Programme evaluation (Lane et al. 2013) the main reasons given by potential primes that were successful in getting onto the bidding framework, but did not then bid for contracts, were primarily financial concerns around the viability and the level of risk associated with the Work Programme.

Evaluators of the London Homelessness SIB identified that as a result of having to compensate for the transfer of risk:
“There is concern amongst providers and some investors that SIB investment models result in ‘leakage’ of funding away from provision through the payment of return on investment.” (Department for Communities and Local Government 2014: 95)

Rates of return

The majority of investors in SIBs were charitable organisations. Several evaluations note that one of reasons investors gave for investing in the SIB was their desire to invest money more ethically and in ways that were more aligned with their missions (eg Disley et al. 2011 in their evaluation of the Peterborough Prison SIB). Tan et al. (2015) in their evaluation of the Health and Social Care SIBs found that some informants felt it was important to establish rates of return that were attractive enough to interest private or social investors. Nevertheless, they also must be sufficiently modest to avoid antagonising public commissioners and local politicians who may harbour suspicions of the motivations of those potential investors.

In the Work Choice PbR programme Thompson et al. (2011) reported that providers felt that incentives were set at an appropriate level for the client group, but most did not expect to make a significant profit from the programme. Third sector providers reported being motivated more by the opportunity to fulfil their organisational aims than by profit.

The effect of incentives on organisational behaviour

A number of evaluations consider the effect of incentives on organisational performance. Some SIB evaluations found that the financial model had driven behaviour, for example, Thomas and Griffiths (2014) and Griffiths et al. (2016) in relation to the SIBs funded under the Department for Work and Pensions Innovation Fund where outcome-based commissioning was widely seen as having incentivised the achievement of outcomes with projects being highly dynamic and a 'continuous improvement' approach being incentivised. Two PbR evaluations also reported positive effects linked to incentive structures. The Youth Justice Reinvestment Pathfinder evaluators (Wong et al. 2015) reported that the use of a commissioning model with upfront funding and a ‘claw back’ mechanism in the event of
under-performance seemed to provide an effective way of incentivising sites to achieve their targets. The New Homes Bonus where a PbR scheme that sought to incentive local councils to increase the number of homes available was found to provide a clear financial incentive (Department for Communities and Local Government 2014).

However, more evaluations, and most PbR evaluations, that considered the issue found that incentives had little or no effect:

- The Department of Work and Pensions (2014) evaluation of the Work Programme reports a provider survey in which providers report that the PbR commissioning model had small effects or no effect on their providers were asked about effect of commissioning model on their financial turnover, efficiency and ability to deliver services.
- Thompson et al.’s (2015) evaluation of the Work Choice funding model that pays providers a 70% service fee, with a further 15% being paid when a client progresses into supported employment, and a final 15% when a client progresses into unsupported employment found that providers were generally comfortable with the principle and level of outcome-based funding, feeling that this had helped to generate a focus on progressing clients through the programme towards unsupported employment. However, providers reported that although incentives were set at an appropriate level for the client group, most did not expect to make a significant profit from the programme.
- In the Supporting People PbR pilot, both commissioners and providers felt that Payment by Results contract terms had potential to offer value for money. However, in terms of service delivery, most providers said they had not taken any particular steps to achieve value for money (Department of Communities and Local Government 2014).
- The evaluation of PbR in Children’s Centres found that impacts of the national PbR programme on local thinking and decision-making were mostly driven by the trial per se and by the national performance measures rather than the financial aspect of the scheme. There was broad local approval of much of the detail on the design of the national payment mechanism. However, effectiveness was limited by the low level of reward payment amounts; a lack of national ring-fencing of rewards and core budgets; and a lack of financial resources to invest in children’s centres due to the economic climate. (Department for Education 2014).
• The evaluation of the Local Justice Reinvestment Pilot notes that stakeholders reported that there were insufficient incentives (the lack of upfront funding and lack of a penalty if the target was not met) for the sites to substantially change practice and invest in new initiatives (Wong et al. 2015).

• In the Troubled Families programme the evaluators reported that the financial framework and targets were contentious in many local areas and were thought to have made it more difficult for local teams to work with the families that they were most concerned about (White and Day 2016). They also incentivised a rapid throughput of families from an early stage in the programme even though some areas were still at an early stage in testing their model, leading to negative impacts on quality (Day et al. 2016).

• The evaluators for the Employment and Reoffending Pilot reported that the additional attachment fee for pilot participants did not appear to have provided sufficient incentive for providers to engage with prisoners pre-release, however low referral volumes may also have affected their ability to achieve economies of scale (Foster et al. 2013).

A small number of PbR evaluations reported adverse effects resulting from incentives structures. Mason et al. (2015) found that an increased likelihood of service users declining to continue with drug and alcohol treatment in pilot areas was a probable unintended consequence of the PbR scheme (see below for further discussion) and Pearce et al. (2015) that a binary re-offending outcome measure in the Doncaster Prison PbR Pilot did not capture the frequency or severity of reoffending. The evaluators concluded that this resulted in community-based services being withdrawn from individuals who had reoffended within a cohort year. White and Day (2016) found that in the Troubled Families Programme the financial incentives model created some perverse incentives, where local teams sought to claim outcomes for families at the margins of the programme to ensure that targets were met. By contrast SIB evaluations reported that perverse incentives had not been identified.

_The effects of incentives on individual behaviours_
A few evaluations considered the possible effect of outcomes-based payment on the behaviour of individuals within provider organisations and most found little effect. The evaluators of the PbR in Children’s Centres concluded that:

"There is some doubt about whether the managers and staff of children’s centres are motivated by financial rewards and would respond to the financial incentives inherent in PbR. Motivation is seen to be driven primarily by a desire to make a difference for children and families”. (Department for Education 2014: 13)

The evaluators of the Work Programme identified three types of performance pay and incentives within provider organisations: individual performance pay and bonuses linked directly to achievement of targets; group bonuses; and promotion. They found that a close link between pay and participant outcomes was rare and there was no clear evidence that the support provided to participants was conditioned by payments systems in any but a handful of the 56 providers interviewed in this study. They concluded that the main way in which there appeared to be scope for a performance-pay link to affect participant support decisions was through the promotion process. In a very small number of cases promotion was linked to the achievement of targets. In these, it seemed that the focus on specific actions (intermediate targets) and outcomes was likely to affect provision (Newton et al. 2012).

In the evaluation of Day One Mandation of Prison Leavers to the Work Programme the evaluators (George et al. 2014) found some evidence of incentives within a PbR programme influencing individual behaviours in some organisations. Overall, the evaluators conclude that differences in support for claimants may have been driven by tailoring, rather than by outcome-based funding. However, in some prime providers and their subcontractors, there was evidence of outcome based funding having an impact. Thus, the evaluators identify evidence of an awareness of the higher payments, including amongst some advisers, and report that this affected the level of support and development of specialist support for this group. In some cases, advisers’ performance measures reflected the higher payments. However, set against this evidence are findings that, for some providers, the higher cost of gaining outcomes for PG9 claimants was judged to outweigh the higher sustainment payments. This meant that some ran their services at a loss, in some cases specialist support was not developed and some providers had actually reduced specialist support.
One exception is the evaluation of the New Homes Bonus which sought to incentivise local authorities to increase the number of homes available. The evaluators (Department for Communities and Local Government 2014) report that almost half of all planning officers agreed the Bonus is a “powerful incentive”, with around 40% also agreeing that the Bonus had resulted in officers and their elected members being more supportive of new homes.

**Segmentation and gaming**

Avoiding perverse incentives is a challenge of all performance management systems because, by necessity they encourage service providers to concentrate on achieving those outputs and/or outcomes that are included in the reward system at the expense of other outcomes even where these are also important (Hoverstadt 2011). A persistent critique of PbR schemes in the early years of the UK Coalition Government was that they increase the risk of ‘gaming’ and in particular ‘creaming’ or ‘cherry-picking’ whereby providers might select ‘easy’ cases to work with and ‘parking’ where they ignore ‘hard’ cases; thereby maximising their returns without accepting fully the transfer of risk or undertaking the intervention envisaged by the public sector commissioner (Battye and Sunderland 2011).

It is clear from the evaluations reviewed that constructing incentive structures is difficult, an issue highlighted in the evaluation of the Department for Work and Pensions European Social Fund Support for Families with Multiple Problems:

"The initial payment model adopted had significant negative effects for both providers and those accessing support. This reinforces the vital nature of a well designed, well structured and achievable approach to setting triggers for outcome payments. While it is reasonable for payment by results (PbR) approaches to be challenging and stretching for providers, it is equally important that they should not compromise the quality of support or the degree to which it can be effectively delivered." (Department for Work and Pensions 2016: 126)

Some programmes used differential pricing structures to incentivise providers to work with client groups that had more complex needs or that were harder to engage. For example, the Work Programme used a ‘differential pricing’ structure under which providers were paid at
different rates for outcomes achieved by different target. However, the evaluation suggests that differential pricing had little impact in driving provider behaviour in how they segmented customers and prioritised support. Providers reported that a key reason that differential payment groups did not drive their customer segmentation and delivery, was the heterogeneity in levels of support need within and between payment groups. In practice they also felt it was not possible to manage delivery and performance against the complexity of the payment model (Department for Work and Pensions 2014). Egdell et al. (2016) in a study of Third Sector experience of the Work programme in Scotland note that some providers were concerned that gaming might be taking place, but provide limited evidence to support this.

In the Drug and Alcohol Recovery PbR programme an independent system was established to assess and refer clients in order to prevent ‘gaming’. However, the evaluators noted that the need to attend an assessment prior to being allocated to a treatment provider represented an additional step in the users treatment journey, which might reduce the likelihood of users entering treatment. The evaluators did not find evidence of increased gaming in sites that opted not to have an independent assessment and referral service, however, in some sites where there was such a service they also noted some positive impacts, such as shorter waiting times for treatment, improved data collection, enhanced integration of drug and alcohol treatment services and the assessment and referral service fulfilling a user advocacy role (Disley et al. 2014).

A number of other evaluations of PbR and SIB programmes explicitly looked for evidence of gaming, but none found evidence that it was taking place (for example, Roberts and Cameron 2014 in relation to the Essex Multi-Systemic Therapy SIB, Department for Communities and Local Government 2014 in relation to the Supporting People PbR, Murray et al. 2012 in relation to the Doncaster Prison PbR Pilot, Disley et al. 2014 in relation to the Peterborough Prison SIB).

**Overall outcomes**

There is only limited evidence on whether outcome-based commissioning has an effect on the quality of services delivered or the outcomes services achieve.
Impact of outcome-based commissioning on service quality

A recurring theme of PbR evaluation reports is that outcomes-based commissioning does have an impact on organisational behaviours. Generally, the effect has been to increase risk-averse behaviours, reduce specialist service delivery and potentially drive down quality.

- In the Work Programme evaluation (Lane et al. 2013) found that some providers who were successful in securing Work Programme contracts felt that PbR had the potential to enable them to deliver good quality services to participants but that the interaction between the payment model, the economic climate (increased total referrals) and the differing referral profile (fewer ‘high value’ referrals) had forced them to change the services they delivered – potentially to a reduced quality. Generalist end-to-end providers dominated Work Programme delivery, and the majority of these delivered all support in-house with low levels of onward referral to specialist support, particularly where this involved formal contracting arrangements (Department of Work and Pensions 2014).

- Similarly, the evaluation of the Work Choice Specialist Disability Employment Programme (Purvis et al. 2013) found that the use of specialist non end-to-end providers within Work Choice delivery was limited and this appeared, to some extent, to be linked to a perceived reduction in contract income. There was some evidence that limited use of these specialist providers has had a negative impact on service delivery and the range of participants that the programme can adequately support. The evaluation noted examples where outgoing provider organisations had made staff redundant or faced closure, and also gives an example of the loss of staff with specialist expertise. Overall this indicates a potentially negative impact on specialist service delivery and implications for service diversity and future supply chain competition. Purvis et al. (2013) found that prime contractors had a strong focus on outcome target monitoring and management and suggest this was possibly at the expense of a focus on quality. Providers and non-providers were concerned about the quality of support for participants that had specific needs.

- In the evaluation of the Youth Contract, Newton et al. (2014) noted that the commissioning process was said by prime providers to have emphasised cost over quality and that primes claimed that this had implications for the nature of support that could be configured.
In the evaluation of the Troubled Families Programme the evaluators (White and Day 2016) reported that several of the case study local authorities had taken the decision to cascade the PbR formula to commissioned service providers by the local authority, so that they were paid on the basis of demonstrating outcomes for families against the financial framework. This put some pressure on smaller organisations – principally VCS providers – to maintain a throughput of families and to close cases relatively quickly, as they were more reliant on being able to draw down the funding allocated to them through the programme and in a less favourable position to cross-subsidise with other funding sources. This fast pace of implementation also meant that some areas were at an early stage with testing their delivery model when the first cohort of families was supported.

Mason et al. (2015) found in the evaluation of the Drug and Alcohol Recovery PbR pilot that to trigger payment for a successful completion, a service user must have completed a planned treatment and be considered as free of dependency from their presenting substance in addition to crack and opiates. The evaluators noted that the introduction of the PbR scheme and the structure of the payments may have led providers to be more risk-averse in discharging service users from treatment and thus recording them as completed successfully, compared with previous financing arrangements. An independent piece of research in one Therapeutic Community undertaken by Gosling (2016) illustrated this tension. Gosling reports that staff believed that the initiative added a financial dimension to their workload, increased the amount of bureaucratic processes which surrounded key working residents and felt that the values and principles of PbR would dilute the therapeutic integrity of the programme because results payment were attached to the completion of each programme stage which reduced the extent to which individual treatment plans could be personalised and varied as client needs changed over time.

However, some programmes, most often smaller ones reported more positive impact on quality. Pearce et al. (2015) found that that in the Doncaster Prison PbR the introduction of the PbR contract resulted in the delivery of additional services to offenders in custody and the community to support their rehabilitation. Thompson et al. (2011) found in their evaluation of Work Choice that prime providers had invested in delivery infrastructure such
as IT systems, premises and staff recruitment and development. Both prime and sub-contractors commonly reported a strong ethos of partnership and cooperation within the programme. Primes had expanded their supply chains during the programme to bring in specialist sub-contractors, indicating a broadening of the services available to clients to reflect more closely individual needs. The Department of Work and Pensions (2016) evaluation of the European Social Fund Support for Families with Multiple Problems that that the delivery of effective support to those engaged was evidence across much of the provision reviewed. There were also some evaluations that suggested PbR had had little impact on delivery. For example, the Department of Communities and Local Government (2014) evaluation of the Supporting People PbR pilot found that the extent to which providers embraced PbR was variable with some adopting a ‘business as usual’ approach to delivery and only adapting their monitoring processes to meet contract reporting requirements.

Assessment of the impact of outcomes-based commissioning on quality of service delivery emerging from evaluations of SIBS is generally more positive than for PbR. Disely et al. (2015) found that in the Peterborough Prison SIB the introduction of the PbR contract resulted in the delivery of additional services to offenders in custody and the community to support their rehabilitation. The evaluation of the Innovation Fund, which funded a number of SIBs targeting disadvantaged young people found that the funding model was seen as having been a significant factor in driving-up performance and developing expertise. There was a widespread belief that projects had achieved better results than they would have done had the pilot been commissioned using more traditional methods and many schools that had bought into new services were reportedly seeking to maintain the new provision beyond the contract period (Griffiths et al 2016).

Impact of outcome-based commissioning on outcomes

Within the evaluations reviewed there is relatively little evidence on the effect of outcome-based commissioning on outcomes. This is consistent with previous commentary on PbR and SIB. Thus, the National Audit Office, in its review of current PbR programmes concluded that “Most operational PbR schemes have yet to finish so there is not yet enough evidence to evaluate the effectiveness of either individual schemes or the PbR mechanism itself”
In some cases evaluations of PbR have been cancelled or modified before being completed or evaluation plans became impractical because rapid public sector reform ‘contaminated’ potential control areas or policies were rolled out nationally making rigorous, comparative evaluation impossible (Webster 2016).

Regarding the social impact of programmes commissioned through a SIB, O’Flynn and Barnett highlight what they see as a paradox within impact investing: that the sector is concerned with “the prioritisation of ‘social impact’ without prioritising ‘impact evidence’” (O’Flynn and Barnett 2017: 3). There are various reasons for this lack of priority for impact evidence. O’Flynn and Barnett list cost considerations by investors (see also Tan et al. 2015); the administrative burden placed on the investee; that impact is implicitly assumed and so doesn’t need to be measured; and, that social outcomes might occur many years after the investment is made. To this we would add the complexity of designing evaluations that can reliably attribute social outcomes to programmes and debates about methodology within the evaluation sector (the so-called ‘paradigm wars’) that can be off-putting to commissioners. Thus, Social Finance (2016: 2) characterize the evaluation debate in the social investment sector as:

“increasingly polarized among those that maintain that only randomised control trials (RCTs) will do, and those that advocate less intensive approaches in order to accelerate the market.”

In our review eight impact evaluation reports were identified. They relate to only five separate programmes and not all are methodologically robust, therefore findings should be treated with caution. Three programmes were found to make a positive impact on outcomes, one had no clear impact and one had a negative impact.

Two of the evaluation reports reviewed (Newton et al. 2014, Nafilyan and Speckesser 2014) set out results of the rigorous, counter-factual impact evaluation of the Youth Contract PbR programme. This found that the Youth Contract increased substantially re-engagement in learning and also included a cost-benefit analysis showing that the intervention generated substantial net benefits by improving educational attainment. Two reports are from the impact evaluation of the first cohort for the Peterborough Prison SIB. The evaluation
employed a rigorous ‘matched pairs’ design and showed a significant improvement in the outcome, albeit not sufficient to trigger a payment to investors (Ministry of Justice 2015, Jolliffe and Hedderman 2014).

The impact evaluation of the first cohort for Doncaster Prison PbR pilot, employing a relatively weak evaluation design showed a successful outcome with the outcome target being achieved (Ministry of Justice 2014).

The evaluation of the Troubled Families Programme found that it did not have any significant or systematic impact across a range of outcomes that covered the key objectives of the programme. Some small, positive impacts were identified on some ‘intermediate outcomes’. The evaluation was subject to some significant limitations, in particular relating to data availability and quality (Bewley et al. 2016).

Mason et al. (2015: 1120) evaluating the impact of paying treatment providers for outcomes as part of the Drugs and Alcohol Recovery PbR pilot found that it had a negative impact on outcomes:

“In the first year of the pilot ‘Payment by Results for Drugs Recovery’ scheme in England, linking payments to outcomes reduced the probability of completing drug misuse treatment and increased the proportion service users declining to continue with treatment.”

During the review we did not identify any published impact evaluations associated with the UK SIBs that have concluded and paid out to investors. Subsequently, while completing the book the final impact of the Peterborough Prison SIB was published (almost two years late) and on the basis of this investors were paid.

A number of implementation evaluations that we reviewed, while not including a formal impact evaluation nevertheless included some evidence, usually qualitative, related to outcomes. Generally, this was positive (Wong et al. 2015, Department for Communities and Local Government 2014b, Department for Communities and Local Government 2015, Thompson et al. 2011, Department for Work and Pensions 2016, Wong et al. 2015). For example, Griffiths et al. (2016) in their qualitative evaluation of the Department for Work
and Pensions Innovation Fund under which a number of SIBs were funded document across a range of stakeholders from all ten pilots:

"a strongly expressed conviction that tangible, positive social impact was being achieved with young people and could be evidenced. Stakeholders pointed to a plethora of data and anecdotal evidence which they believed demonstrated positive impact on participating young people. The evidence offered ranged from ‘turn-around’ success stories of individual young people, to institutional-level changes in schools, to assessments of benefit across comparable cohorts of non-participants".

(Griffiths et al. 2016: 14)
Appendix One: list of SIB and PbR programmes included in the review

<table>
<thead>
<tr>
<th>Name of programme</th>
<th>PbR / SIB</th>
<th>Type of evaluation</th>
<th>Evaluation outputs included in this review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supporting People</td>
<td>PbR</td>
<td>The DCLG commissioned an impact and implementation evaluation. The impact evaluation was not possible because of lack of a robust comparator group.</td>
<td>DCLG (2014a)</td>
</tr>
<tr>
<td>Sure Start Children's Centres</td>
<td>PbR</td>
<td>DfE commissioned an implementation evaluation</td>
<td>DfE (2014)</td>
</tr>
<tr>
<td>Work Programme</td>
<td>PbR</td>
<td>The DWP commissioned an implementation evaluation of the main programme and the Day One Mandation of Prison Leavers programme.</td>
<td>Newton et al. (2012), Lane et al. (2013), Foster et al. (2014), DWP (2014), George et al. (2014)</td>
</tr>
<tr>
<td>Two independent, qualitative studies of aspects of Work Programme implementation with some evaluative elements were also identified.</td>
<td>Egdell et al. (2016), Ingold and Stuart (2014)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>European Social Fund Support for Families with Multiple Problems</td>
<td>PbR</td>
<td>The DWP commissioned an implementation evaluation</td>
<td>DWP (2016)</td>
</tr>
<tr>
<td>Drug and Alcohol Recovery Pilots</td>
<td>PbR</td>
<td>The DoH commissioned an implementation and impact evaluation. An interim evaluation report and an impact study have been published.</td>
<td>Disley et al (2014a), Mason et al. (2015)</td>
</tr>
<tr>
<td>An independent study of the effect of PbR in a therapeutic community that had some elements of an implementation evaluation has also been identified.</td>
<td>Gosling, H. (2016)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Homes Bonus</td>
<td>PbR</td>
<td>The DCLG commissioned an implementation and impact evaluation.</td>
<td>DCLG (2014b)</td>
</tr>
<tr>
<td>Program</td>
<td>Type</td>
<td>Description</td>
<td>Evaluation Sources</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Justice Reinvestment Pilots</td>
<td>PbR</td>
<td>The MoJ commissioned an implementation evaluation</td>
<td>Wong et al. (2013a), Wong et al. (2015a)</td>
</tr>
<tr>
<td>Youth Justice Reinvestment</td>
<td>PbR</td>
<td>Custody Pathfinder</td>
<td>Wong et al. (2013b), Wong et al. (2015b)</td>
</tr>
<tr>
<td>Employment and Reoffending Pilot</td>
<td>PbR</td>
<td>The MoJ commissioned an implementation evaluation.</td>
<td>Foster et al. (2013)</td>
</tr>
<tr>
<td>Community Work Placements</td>
<td>PbR</td>
<td>The programme was scrapped after one year. We are not aware that an evaluation was commissioned.</td>
<td></td>
</tr>
<tr>
<td>Transforming Rehabilitation</td>
<td>PbR</td>
<td>No evaluation commissioned.</td>
<td></td>
</tr>
<tr>
<td>New Horizons (part of the DWP</td>
<td>SIB</td>
<td>The DWP commissioned an implementation evaluation</td>
<td>Thomas and Griffiths (2014), Griffiths et al (2016)</td>
</tr>
<tr>
<td>and Pensions Innovation Fund)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Advance Programme (part of</td>
<td>SIB</td>
<td>The DWP commissioned an implementation evaluation</td>
<td></td>
</tr>
<tr>
<td>the DWP Innovation Fund)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Think Forward (part of the</td>
<td>SIB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DWP Innovation Fund)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

³ After the review was completed the MoJ published a final evaluation of the Peterborough SIB: Anders and Dorsett (2017). This was outside the cut-off date for this review.
<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nottingham Futures (part of the DWP Innovation Fund)</td>
<td>SIB</td>
<td></td>
</tr>
<tr>
<td>Links for Life (part of the DWP Innovation Fund)</td>
<td>SIB</td>
<td></td>
</tr>
<tr>
<td>Living Balance (part of the DWP Innovation Fund)</td>
<td>SIB</td>
<td></td>
</tr>
<tr>
<td>Teens and Toddlers (part of the DWP Innovation Fund)</td>
<td>SIB</td>
<td></td>
</tr>
<tr>
<td>3SC Capitalise (part of the DWP Innovation Fund)</td>
<td>SIB</td>
<td></td>
</tr>
<tr>
<td>Energise (part of the DWP Innovation Fund)</td>
<td>SIB</td>
<td></td>
</tr>
<tr>
<td>Prevista (part of the DWP Innovation Fund)</td>
<td>SIB</td>
<td></td>
</tr>
<tr>
<td>Essex Children in Care</td>
<td>SIB, Essex County Council commissioned an implementation and impact evaluation. Subsequently the impact evaluation has been discontinued.</td>
<td>Roberts and Cameron (2014), Sin and Roberts (2016), Sin (2016)</td>
</tr>
<tr>
<td>London Rough Sleepers (Thames Reach), London Rough Sleepers (St. Mungos)</td>
<td>SIB, The DCLG commissioned an implementation evaluation.</td>
<td>DCLG (2014c), DCLG (2015)</td>
</tr>
<tr>
<td>The SIB is also included in an implementation and outcome evaluation of SIBs in health and social care commissioned by the DoH. An interim implementation evaluation has been published.</td>
<td></td>
<td>Tan et al. (2015)</td>
</tr>
<tr>
<td>It's All About Me (IAAM) Adoption Bond</td>
<td>SIB, No evaluation of implementation or impact. The Cabinet Office states: “This cohort of children is very unlikely to have found a home in the absence of this intervention given the rates of adoption and their characteristics. Therefore we assume that none of the cohort would have been placed without IAAM, and deadweight is therefore nil.”</td>
<td></td>
</tr>
</tbody>
</table>

---

4 [https://data.gov.uk/sib_knowledge_box/node/183](https://data.gov.uk/sib_knowledge_box/node/183)
<table>
<thead>
<tr>
<th>Location</th>
<th>Activity</th>
<th>Description</th>
</tr>
</thead>
</table>
| Manchester                          | looked after      | No impact evaluation commissioned. The Cabinet Office states that: “Through following the young peoples’ placements over time it was established that children who enter residential placements in early adolescence have a very high likelihood of spending the remainder of their time in high cost placements. The SIB therefore assumes no deadweight.”

  | The SIB is also included in an implementation and outcome evaluation of SIBs in health and social care commissioned by the DoH. An interim implementation evaluation has been published. |
| Birmingham                          | looked after      | We are not aware of an evaluation.                                                                                                                                                                    |
| Birmingham Homelessness              | (part of the DCLG and Cabinet Office Fair Chance Fund) |                                                                                                                                                                                                       |
| West Yorkshire Homelessness          | (part of the DCLG and Cabinet Office Fair Chance Fund) |                                                                                                                                                                                                       |
| Gloucestershire Homelessness         | (part of the DCLG and Cabinet Office Fair Chance Fund) | An evaluation has been commissioned but no reports have yet been published.                                                                                                                              |
| Greenwich Homelessness               | (part of the DCLG and Cabinet Office Fair Chance Fund) |                                                                                                                                                                                                       |
| Leicestershire Homelessness          | (part of the DCLG and Cabinet Office Fair Chance Fund) |                                                                                                                                                                                                       |
| Liverpool Homelessness               | (part of the DCLG and Cabinet Office Fair Chance Fund) |                                                                                                                                                                                                       |

---

5 https://data.gov.uk/sib_knowledge_box/manchester-city-council-children-care-sib
<table>
<thead>
<tr>
<th>Programme</th>
<th>Evaluation</th>
<th>Notes</th>
<th>Authors/Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newcastle Homelessness (part of the DCLG and Cabinet Office Fair Chance Fund)</td>
<td>SIB</td>
<td>The Cabinet Office commissioned an implementation and impact evaluation. No reports have yet been published.</td>
<td></td>
</tr>
<tr>
<td>Ways to Wellness</td>
<td>SIB</td>
<td>Being evaluated as part of the Commissioning Better Outcomes Evaluation funded by the Big Lottery Fund. A review report with some evaluative elements has been published.</td>
<td>ATQ Consultants and Ecorys (2015)</td>
</tr>
<tr>
<td>Career Connect (part of the Youth Engagement Fund)</td>
<td>SIB</td>
<td>The Cabinet Office commissioned an implementation and impact evaluation. No reports have yet been published.</td>
<td></td>
</tr>
<tr>
<td>Prevista (part of the Youth Engagement Fund)</td>
<td>SIB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Futureshapers (part of the Youth Engagement Fund)</td>
<td>SIB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teens and Toddlers (part of the Youth Engagement Fund)</td>
<td>SIB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worcester Reconnections</td>
<td>SIB</td>
<td>Being evaluated as part of the Commissioning Better Outcomes Evaluation funded by the Big Lottery Fund. A review report with some evaluative elements has been published.</td>
<td>ATQ Consultants and Ecorys (2016a)</td>
</tr>
<tr>
<td>Mental Health and Employment Partnership Social Impact Bond</td>
<td>SIB</td>
<td>Being evaluated as part of the Commissioning Better Outcomes Evaluation funded by the Big Lottery Fund. A review report with some evaluative elements has been published.</td>
<td>ATQ Consultants and Ecorys (2016b)</td>
</tr>
<tr>
<td>Lambeth</td>
<td>SIB</td>
<td>We are not aware of an evaluation.</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: List of the 46 empirical evaluations of UK PbR and SIB programmes included in the review
Bibliography


MoJ (2014) *Peterborough Social Impact Bond HMP Doncaster: Payment by Results pilots – Final re-conviction results for cohorts 1*, London: MoJ


Wong et al. (2013b) Youth Justice Reinvestment Custody Pathfinder: Findings and delivery lessons from the first year of implementation
