Payment by results and social impact bonds in the criminal justice sector: New challenges for the concept of evidence-based policy?

Chris Fox and Kevin Albertson

Criminology and Criminal Justice 2011 11: 395 originally published online 8 August 2011
DOI: 10.1177/1748895811415580

The online version of this article can be found at:
http://crj.sagepub.com/content/11/5/395

Additional services and information for Criminology and Criminal Justice can be found at:

Email Alerts: http://crj.sagepub.com/cgi/alerts
Subscriptions: http://crj.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.com/journalsPermissions.nav
Citations: http://crj.sagepub.com/content/11/5/395.refs.html
What is This?
Payment by results and social impact bonds in the criminal justice sector: New challenges for the concept of evidence-based policy?

Chris Fox and Kevin Albertson
Manchester Metropolitan University, UK

Abstract
Payment by results allows the government to pay a provider of services on the basis of the outcomes their service achieves rather than the inputs or outputs the provider delivers. Social impact bonds (SIBs) are a form of payment by results which allow the financing of social outcomes via private investment. It is suggested that payment by results and SIBs will drive greater efficiency, innovation and impact in tackling social problems through focusing reward on outcomes and providing minimal prescription as to how these outcomes should be achieved. It is suggested that this may be achieved while also reducing risk for government. Here we set out the challenges likely to arise in developing payment by results models and SIBs in the criminal justice system of England and Wales. These include the uncertainty arising from defining outcomes, estimating the potential impact of interventions, measuring and attributing change, valuing benefits, demonstrating a fiscal return and getting interventions to scale. We conclude that, to a government trying to deliver ‘more for less’, payment by results may offer an attractive solution in some parts of the public sector. However, the case for this approach in the criminal justice sector, where the evidence base is contested and potential savings difficult to quantify and realize, is not yet proven.

Keywords
criminal justice, evidence-based policy, payment by results, social impact bond
Introduction

In 2010 the Conservative Party and the Liberal Democrats agreed to form the United Kingdom’s (UK) first formal coalition government for 55 years. The first bullet point under the ‘justice’ heading in the coalition agreement between reads: ‘We will introduce a “rehabilitation revolution” that will pay independent providers to reduce re-offending, paid for by the savings this new approach will generate within the criminal justice system’ (Her Majesty’s [HM] Government, 2010a: 23). Using the ‘payment by results’ mechanism in the criminal justice system was not the brainchild of the new coalition government. The work to implement a prisoner resettlement project at Her Majesty’s Prison (HMP) Peterborough, based on a social impact bond (SIB) – a form of payment by results – was started by the previous government. However, the idea of using payment by results as a mechanism to bring about change in the delivery of criminal justice services has been gathering pace ever since and the concepts of ‘payment by results’ (along with ‘social impact bonds’) feature regularly in politicians’ speeches (see, e.g., Clarke, 2010, and Blunt, 2011).

The idea of only paying for outcomes that are achieved is particularly attractive at the moment because of the state of public sector funding in the UK, which will be constrained for the next few years. Partly as a result of the global financial crisis, and partly because of the policies of the previous government, the UK has a budget deficit which is high by historical standards. In the financial year 2009–10 the UK government net borrowing was £159.8 billion, equivalent to 11.4% of gross domestic product (GDP). The total UK government debt at March 2010 was £1,000.4 billion, equivalent to 71.3% of GDP (Office for National Statistics, 2010). The total budget for the Ministry of Justice for 2009–10 was just over £10 billion (Ministry of Justice, 2009a). For 2010–11 it is £8.9 billion and by 2014–15 it will be £7.9 billion (Ministry of Justice, 2010a). The mantra from politicians is ‘more for less’. In this context the idea of commissioning for outcomes and specifically ‘payment by results’ is attractive.

Payment by results (PbR) raises interesting questions about the new government’s approach to criminal justice. What does the introduction of PbR tell us about the ‘culture of control’ in the UK (Garland, 2001) and does it represent a break with New Labour’s public sector modernization project in general (see, e.g., 6 and Peck, 2004) and specifically its strategy of ‘modernization through managerialization’ (McLaughlin et al., 2001) which some have argued characterized its approach to criminal justice reform? It is probably too soon to try to answer these questions. Rather, in this article we restrict ourselves to examining the practical challenges which might arise in the implementation of payment by results mechanisms in the criminal justice system of England and Wales. We place our examination in the context of the ongoing debate about the nature and limits of evidence-based criminal justice policy (see, e.g., Sherman, 2009; Hough, 2010) and the still-emerging area of criminal justice economics.

In the remainder of this section we provide a brief outline of what is meant by ‘payment by results’ and ‘social impact bonds’. We highlight previous and current examples of relevant initiatives in the UK public sector. In the following section we outline some of their main attractions to government. We move on to look at current and planned payment by results models in the English and Welsh criminal justice system.
before going on to discuss in detail what we see as the main challenges to making them work. In conclusion, we suggest the key challenges are, in essence, challenges to the implementation of evidence-based policy.

**What is payment by results?**

PbR allows the government to pay a provider of services on the basis of the outcomes their service achieves rather than the inputs or outputs the provider delivers. It is suggested, by focusing reward on outcomes and providing minimal prescription as to how these outcomes should be achieved, payment by results models will drive greater efficiency, innovation and impact in tackling social problems. This approach is being explored in different parts of the public sector, including the Department for Work and Pensions. PbR is seen as a key tool in reforming criminal justice services:

The principle of incentivising performance through payment by results, with success based on the absence of re-offending, should be introduced for prisons, the providers of community sentences and the providers of rehabilitation programmes – whether in the public, private or voluntary sector. With devolved responsibilities and new incentives, we can create a revolution in how offenders are managed, and drive down re-offending. (Conservative Party, 2009: 49)

The coalition government proposes that rehabilitation services will be paid for in part according to how successful they are at reducing re-offending; a basic tariff will be paid to cover their costs and an additional tariff paid if targets to reduce re-offending are met. In delivering this they quote the ‘Peterborough social impact bond’ as a promising model (Clarke, 2010).

**Social impact bonds**

Because payment arises after outcomes are known, which might involve substantial time delays, a key challenge to the delivery of PbR models of service delivery is the challenge of raising working capital (Mulgan et al., 2010). This might be solved by delivery organizations raising capital through normal markets. So, for example, the large private sector providers who are envisaged to deliver the Department for Work and Pensions’ Work Programme via contracts worth £10–15 million pounds in each sub-region of the UK and based on PbR (Department for Work and Pensions, 2010) will, presumably, raise working capital through the money markets at commercial rates. The social impact bond presents a new method of financing social outcomes via private investment. It is envisaged that the SIB will be used to raise capital for social projects in the way bonds are used for investment projects; a branch of national or local government will agree to pay for a measurable, social outcome and this prospective income is used to attract new funds to meet the up-front costs of the activity. The new funds could come from the public sector, the private sector or a social investor (Mulgan et al., 2010). A social investor might be an individual philanthropist or a charitable trust. Mulgan and colleagues (2010) suggest that a philanthropic social impact bond will see funds raised from philanthropic sources invested through a special purpose vehicle that would
sub-contract with non-governmental organizations (NGOs) to deliver services to achieve the outcomes. A public sector social impact bond might involve a local authority borrowing or using existing revenues to finance a specific initiative (Mulgan et al., 2010). Commercial social impact bonds see private finance replacing philanthropic or public finance, thus creating a new asset class that can be invested in by banks, pension funds and others (Mulgan et al., 2010). Mulgan and colleagues (2010) suggest that this model will be suitable where there are proven models of intervention and reliable delivery partners with track records. Notwithstanding, they acknowledge that this situation will take 5–10 years to develop.

Payment by results in the UK

Perhaps the two best known, recent examples of PbR models in the UK come from the health sector where a PbR mechanism has been in operation for several years and the radical proposals by the Department for Work and Pensions (DWP) to introduce PbR as part of its new Work Programme.

PbR was introduced into the National Health Service (NHS) in the 2003–4 financial year. Under the scheme, health service providers are paid for the number and type of patients treated, in accordance with national rules and a national tariff (Audit Commission, 2008). Arguably, this does not have the same focus on outcomes as some schemes proposed for the criminal justice system. The aim is to improve the fairness and transparency of hospital payments and to stimulate provider activity and efficiency (Audit Commission, 2008). Implementation was phased over four years. At the end of this period the Audit Commission concluded of payment by results (PbR) that:

PbR has undoubtedly improved the fairness and transparency of the payment system. It has also, perhaps, had a positive effect on activity and efficiency . . . However, other policies have also encouraged such trends, particularly the need to meet waiting time targets, and detailed analysis suggests that other factors have also brought about the changes. We consider that PbR has at most contributed to these positive trends rather than driven them. Meanwhile, the negative impact on quality which some feared would result from PbR has not been realised. (Audit Commission, 2008: 2)

The DWP is proposing a PbR mechanism as part of its Work Programme. The Work Programme is described as ‘the centrepiece of the Government’s plans to reform welfare-to-work provision in the UK’ (Department for Work and Pensions, 2010: 2). The Work Programme brings together various programmes designed to help people back into work into a single scheme. The Universal Credit brings together current working age benefits and tax credits with a single welfare payment. It is envisaged that the Work Programme will be delivered by large, private sector providers delivering 40 Work Programme contracts across 18 areas with at least two providers in each area (Department for Work and Pensions, 2010). Each individual contract will be worth £10–50 million per year. These contracts will be based on a PbR model with three main payments: an attachment fee paid when a benefit claimant starts on the programme, a job outcome fee paid when claimants enter work, and sustainment payments paid for
keeping claimants in work (Department for Work and Pensions, 2010). It is the DWP’s intention that ‘a significant amount of the total amount we’re willing to pay will be in sustainment payments’ (Department for Work and Pensions, 2010: 6).

**Potential advantages of payment by results**

Various potential advantages are envisaged for a PbR model:

**Greater efficiency**

It is suggested, by focusing reward on outcomes and providing minimal prescription as to how these outcomes should be achieved, that PbR models (including the SIB) will drive greater efficiency in tackling social problems. A system that focused on the delivery of outcomes might look very different to the current system with, perhaps, a wider range of service providers, more variation in models of service delivery and a much-reduced role for ‘national standards’.¹ One mechanism by which greater efficiency will be achieved is through the correction of misaligned incentives. For instance, an organization responsible for delivering a service has not always shared in the benefits of that activity (Mulgan et al., 2010). For example, a local authority which provides services to divert young people away from crime does not necessarily share in the savings realized by the criminal justice sector arising from reduced prison numbers (Mulgan et al., 2010).

A second mechanism by which greater efficiency will be achieved is by allocating resources to where they will achieve the most impact. For example, it is suggested that SIBs are particularly appropriate as a mechanism to focus attention on preventative measures (Mulgan et al., 2010). The recent review of Early Intervention by Allen (2011) provides numerous examples of interventions which fall into this category. Many of these have a long-term objective of reducing offending.

**More innovation**

The focus on outcomes, which is a part of the PbR initiative, and the associated reduced focus on commissioners ‘micro-managing’ the processes put in place to achieve those outcomes seems to be a rejection of the prevailing ‘New Public Management’ approach which has dominated the public sector² and, increasingly, the voluntary sector services over recent years. Proponents of PbR argue that ‘freeing up’ providers to deliver services in different ways will encourage greater innovation.

**Transfer of risk**

PbR transfers risk away from the branch of government commissioning the service and towards the service provider. Given the need to reduce public sector spending (see earlier), the transference of risk is an attractive proposition for government.
Encouraging new market entrants

Implicit in the concept of more innovation and the transfer of risk is the potential of PbR models to promote more market testing of public services and encourage new market entrants, particularly from the private and voluntary sectors. Thus the Ministry of Justice Green Paper suggests that:

The payment by results approach will encourage innovation and bring out the diverse skills from all sectors. We must ensure our commissioning model harnesses the creativity and expertise that independent providers can bring. This includes the small and specialist voluntary providers and social enterprises. (Ministry of Justice, 2010b: 41)

Payment by results in the English and Welsh criminal justice system

Several initiatives which are either direct attempts to implement PbR models or which have established the policy context within which PbR will be implemented have already been undertaken in the criminal justice system. For example, in its green paper the Ministry of Justice states that:

Significant amounts of public money have been spent on rehabilitating criminals without properly holding services to account for the results they achieve. We will move to a new approach where providers are increasingly paid by their results at reducing re-offending. (2010b: 10)

Proposed pilots of various PbR models are then described. As another example, the Youth Justice Investment Pathfinder Initiative will see a portion of the central youth custody budget invested in ‘pathfinders’ where consortia of local authorities or a single local authority use the funding to commission and deliver their own responses to reduce levels of youth custody and youth re-offending in their area. Pathfinders will share the financial risks if the custody rate increases and keep the funding if custody numbers are kept low.

A social impact bond initiative has recently begun at HMP Peterborough. This is the first pilot of models of this kind in the English and Welsh criminal justice system. The Ministry of Justice has signed a contract with Social Finance (2010) to attempt to reduce the re-offending of a cohort of 3,000 adult males who are discharged from HMP Peterborough having served sentences of less than 12 months in custody.

Disley and colleagues (2011) report that investors have put £5 million in social impact bonds to fund the rehabilitation work and that they could earn a return of up to £8 million from the government and the Big Lottery Fund if re-offending among three cohorts, each of a 1,000 offenders, falls by 10% or if the rate of re-offending for all 3,000 offenders falls by at least 7.5%. If a reduction in re-offending beyond 7.5% is delivered then investors receive an increasing return capped at 13% over an eight year period (Social Finance, 2011: 3). The British Broadcasting Corporation (BBC, 2010) suggests that a comparable rate of return on investment in a conventional bond-market is 7.5% per year. The project at Peterborough has only just started, but the first output from the...
independent evaluators describes the planning and early implementation of the project. There is evidence that the Peterborough SIB has introduced some new funding into the delivery of criminal justice services, thus transferring risk away from government to non-governmental investors. The commissioning process was complex with six different contracts identified. The Ministry of Justice has contracted with Social Finance who, in turn have contracts with investors and service providers. The National Offender Management Service’s contract with HMP Peterborough (a private prison) has also been varied (Disley et al., 2011). The result of the commissioning process is a new scenario in which government has no direct control over the choice of service providers – this is the choice of the intermediary (Disley et al., 2011). The outcome measurement framework was designed to reduce the risk of ‘cherry-picking’ whereby the service provider concentrates on clients more amenable to change. Thus outcomes are measured for all offenders discharged from the prison, rather than just those who engage with SIB-funded services and frequency of re-conviction is measured rather than a binary measure of whether offenders were re-convicted or not (Disley et al., 2011: iv).

Some estimates of the return to society from such programmes is of the order of £100 for every £1 invested (BBC, 2010; see also Strickland, 2010: 3). It is, therefore, clear that the government will realize substantial savings if the scheme pays off. If, however, offending does not fall, investors stand to lose all their money, while the government has little downside risk.

**Challenges**

Analysis of the underlying concepts behind PbR together with early experiences from approaches being piloted in the UK suggests a number of challenges which must be faced if payments by results mechanisms are to work effectively, particularly in a criminal justice context. In particular we consider:

1) Scale of change: The first challenge is to do with the scale of change that is possible, and the level which is required for statistical significance. Evidence (discussed later) suggests even successful criminal justice interventions bring about only relatively small changes in outcomes. This creates problems for investors looking for a decent return – and for commissioners and evaluators who want to be sure any change which is observed is the result of the intervention, rather than simple random variation.

2) Evidence of impact: As Mulgan et al. (2010) note, a key challenge for investors considering SIBs is the relative weakness of the evidence base: it is not always clear *a priori* which interventions are most likely to deliver the desired results. This leads to difficulties for investors making rational investment decisions. Later we further break down this second group of challenges into the challenge of defining outcomes, estimating impact and measuring impact.

3) Economic and fiscal returns: The third group of challenges arises from the process of estimating the economic value of outcomes, the more problematic estimation of which public sector organizations will accrue the benefits from an intervention and, more problematically still, deciding on whether there is a fiscal benefit to those organizations.
Scale of change

What level of impact might programmes in the criminal justice system be expected to achieve? As we have seen, at least a 7.5% reduction in re-offending is required in the Peterborough SIB before investors will receive any payout at all from the government (Social Finance, 2011: 3). In order to achieve a rate of return comparable with a regular bond market, a reduction in re-offending of at least 10% will be required. However, as Social Finance (2011: 1) notes, ‘The SIB is not a traditional “bond” because all of the investors’ capital is at risk’. On the other hand the return investors may expect is ‘capped at a maximum of 13% per year’ (Social Finance, 2011: 3) The downside risk is, therefore, much greater than the upside risk.

Standard economic theory suggests that the SIB investors will wish to be compensated for their taking on a risk which is arguably higher than that available in the regular bond market. That is, they are likely to want a reduction of re-offending in excess of 10% to break even (in opportunity costs terms) and be compensated for the (arguably) high risk of the SIB. The Council on Social Action’s estimate of the impact of the St Giles’ Trust intervention suggests that they should be optimistic:

The St. Giles worker patiently develops relationships with his clients whilst they are still in prison. When they leave prison, he works intensively with them, helping them to settle back into society, sever their ties with previous bad associations, find accommodation and get into education, training or employment. The usual re-offending rate for this group is 70–75%. The re-offending rate for the St. Giles Trust group is 10%. (Council on Social Action, 2008 [quoted in Strickland, 2010: 2])

However, no-one working in the sector would expect to see 60 percentage point reductions in re-offending as a matter of course, particularly across fairly large groups of offenders. Mulgan et al. (2010) argue prudent impacts would be in the range of 10–20%. This is based on Lipsey and Cullen (2007) which summarizes a number of meta-analyses of rehabilitation. But Lipsey and Cullen (2007) might not agree with Mulgan et al.’s interpretation. Commenting on Table 2, Lipsey and Cullen suggest that:

The global question of whether rehabilitation treatment works is thus answered affirmatively by the favourable mean effects on recidivism found by every meta-analyst who has conducted a systematic synthesis of a broad sample of the available experimental and quasi-experimental research. (Lipsey and Cullen, 2007: 306).

But, having summarized further meta-analyses of specific rehabilitative interventions they finally conclude that:

The research on rehabilitation treatment reviewed here provides an encouraging indication of the relatively large effects that might be attainable in actual practice, but cannot be interpreted as evidence that current practice has such effects or, indeed, that it has any positive effects at all. (Lipsey and Cullen, 2007: 315, emphasis added)

Lipsey and Cullen’s (2007) concern stems from the wide range of treatment effects represented by average effects expressed in each meta-analysis. Further, evidence from
criminal justice evaluations indicates even well-executed projects and programmes often bring about only relatively small changes, a point made by Berman and Fox (2010) in their thought-provoking examination of success and failure in criminal justice policy reform. They argue there is a strong case made for the limited effect of even the best thought out and most carefully implemented policies and programmes. A quote from their interview with Carol Weiss sums this up nicely:

‘Realistic expectations are important’, says researcher Carol Weiss. ‘With criminal justice programmes it’s hard, slow work. It’s a little odd that people expect so much from them. When you run an advertising campaign for Toyota, changing sales by a percentage point or a two is considered a huge success. The same is true in running a big election campaign. Why is that different in criminal justice?’ (Berman and Fox, 2010: 118–19)

When commissioning impact feasibility studies in recent years, the Ministry of Justice research team seem to have used 5% reductions in re-offending as a default figure for power calculations unless there is clear evidence to the contrary. One particular issue, well known to evaluators, is the drop-off in impact which often occurs between a pilot project and widespread roll-out. This might in part be due to the difficulty of maintaining ‘construct validity’ (Farrington, 2003). For this reason, in their economic modelling of the effect of implementing criminal justice interventions, Aos et al. (2006) discount the findings from their meta-analysis of effect studies by 25% to allow for ‘pilot effect’.

Modest outcomes, or variable outcomes, will provide little security for service providers taking on the risk of delivery under a PbR mechanism or investors in a SIB. The margin between a profitable outcome and an unprofitable outcome will be narrow.

**Defining outcomes**

Key to PbR models is a clear definition of the outcome which is being purchased. In the English and Welsh criminal justice sector this is not straightforward. Consider, for example, the seemingly straightforward result of reducing re-offending: re-offending will generally be measured via the proxy of re-convictions (i.e. ‘proven re-offending’). Thus it would appear to matter little to the PbR model if offenders cease to offend, or merely become less easy to catch. Further, even this proxy can be defined differently. As noted in the recent Ministry of Justice (2010c) consultation on offender statistics, at present there are six different measures of re-offending. If outcomes are difficult to define and measure, the result is likely to be that PbR models in the criminal justice system will tend to concentrate on the narrow set of outcomes which is easiest to define and measure.

There is a further layer of complexity to be considered when defining outcomes. Take, for example, drug treatment. As Nutt (2010) notes, the recently-published drug strategy (HM Government, 2010b) concentrates on treatment outcomes such as abstinence and, implicitly, social and economic outcomes such as employment. Nutt (2010) contrasts two competing understandings of addiction: (1) a medical model which sees addiction as a chronic relapsing disease; and (2) the other – seemingly favoured by the government in the drug strategy – sees addiction as a lifestyle choice. In essence, this is a debate between competing mid-level theories of addiction. The implication for PbR models may be that,
where interventions are not adequately theorized, intervention specifications will be poorly thought-out, inappropriate outcome measures will be chosen and evaluation will be more difficult and more costly. This raises the overhead costs for service providers and/or investors.

**Prospective estimates of impact**

To implement a PbR model it is necessary to have a robust estimate of the size of impact of an intervention or programme before implementation commences. Organizations contracting to deliver a service on the basis of PbR or investors contemplating a SIB will need such an estimate to ensure they fully understand the level of risk implied by the transaction. Early findings from the Peterborough SIB suggest that: ‘The development of a methodologically-robust outcome measure, which had the confidence of all stakeholders, was a time-consuming and analytically-complex process’ (Disley et al., 2011: iii). There are various ways in which a reliable estimate might be derived. One option would be to apply the PbR model only to an intervention which an organization has previously delivered and evaluated; this would stifle innovation. Also, where an evaluation does suggest that an intervention is effective, replication is notoriously difficult (see, e.g., Pawson and Tilley, 1994; Berman and Fox, 2010).

Very often, therefore organizations delivering, investing in or commissioning PbR models will need to draw on the wider evaluation literature. In its guidance on Social Return on Investment models, the Cabinet Office (2009) suggests drawing on a broad range of previous research and experience; this might include information from membership organizations, government departments, market research firms, consulting companies, and partner organizations as well as published research. Given the need for a high degree of confidence that observed effects can be attributed to the intervention in question – and the need to be able to estimate with a known level of confidence the likely effect of an intervention – the preferred strategy will generally be to make use of an existing systematic review or commission a new one. Systematic reviews favour evaluation designs with high levels of internal validity: social experiments and quasi-experimental designs and, where possible, they will include a meta-analysis of study results.

An immediate issue will be that most of the studies with high internal validity – those likely to be included in systematic reviews of criminal justice interventions – will come from the United States (US). This will raise issues about transferability to the UK context. Sherman (2009) advocates the UK government investing in more multi-site randomized trials and the production of more systematic reviews. If his advice were taken up, a rapid increase in the number of UK experiments in the criminal justice system might address this issue in due course, but in, the current economic climate seems unlikely.

In any event, the value of systematic reviews is contested. Focusing on the use of meta-analysis Pawson (2002a) puts forward a number of objections. Specifically, his arguments concentrate on the melding of programme mechanisms (the danger of grouping together interventions which are dissimilar), oversimplifying programme outcomes (the outputs of a meta-analysis are expected outcomes of averages of means) and the
concealment of programme contexts (the danger of not taking sufficient account of subjects and situations). Hough (2010), focusing particularly on the evidence base for interventions which reduce re-offending, is concerned that ‘traditional’ systematic reviews fail to recognize that ‘work with offenders is a highly reflexive process in the sense that the meanings attributed to the process by those involved in it will affect the outcomes’ (Hough, 2010: 14). Systematic reviews that look for “clinching evidence” based on randomized control trials (RCTs) and similar evaluative methods can result in an accumulation of generalizable hard knowledge about what works in reducing re-offending’ (Hough, 2010: 14) but fail to recognize that the effectiveness of interventions will be highly context-specific. Thus: ‘what is missing from systematic reviews in this field – as currently practised – is the development of “middle level theories” to explain why some things work and others don’t’ (Hough, 2010: 14). Partly in response to concerns such as these, a number of methods have been developed for reviewing and synthesizing qualitative studies (see, e.g., Pawson, 2002b; Barnett-Page and Thomas, 2009). Notwithstanding, to date there is no widely-agreed strategy for bringing together the output from these qualitative syntheses with those from traditional, quantitative reviews and meta-analyses.

Before drawing out the implications of this debate for PbR it is worth considering how the impact of a PbR intervention which has in fact been implemented might be measured. Such measurement will, of course, add to the body of literature which will inform future SIB investors.

**Measuring impact**

Given the need to make precise estimates of the scale of impact achieved – and to be confident that this impact can be attributed to the intervention – evaluation designs with high levels of internal validity (experiments and quasi-experiments) will be preferred to those with lower levels of internal validity. So, for example, the Ministry of Justice and the Big Lottery Fund will only pay investors in the Peterborough social impact bond so long as there is a measured reduction in re-conviction ‘relative to the experience with a control group’ (Social Finance, 2011: 2). This raises some practical evaluation challenges. Small outcomes will require large cohorts to ensure statistically-significant results from any subsequent outcome evaluation which is undertaken (Mulgan et al., 2010). Also, many of the interventions of interest in reducing criminal justice expenditure, those which might be subject to PbR, involve some level of multi-agency working and are envisaged to have multiple outcomes. Both elaborating a coherent theory of change (Weiss, 1995) and ensuring construct validity (Farrington, 2003) during an intervention will be challenging. If we adopt this approach to the evaluation of PbR models, the result, as Mulgan et al. (2010) suggest, will be the widespread use of evaluations which make use of comparator groups, and this will raise the overhead costs of PbR substantially. That is, PbR and SIB investors will be required to raise funds to cover the intervention, and a further tranche of funds to cover a costly evaluation.

More fundamentally, the likely preference of PbR models for estimates of impact derived from (quasi-) experiments raises difficult questions about the limitations of evaluation and of evidence-based policy. For example, scientific realists have been critical of
RCTs arguing that ‘[t]he central problem [with RCTs] lies in the deficient and defective conception of the programme which is built into the methodology’ (Pawson and Tilley, 1994: 297). Scientific realists therefore argue greater attention must be paid to ‘context-mechanism-outcome’ configurations (Pawson and Tilley, 1997).

Taking a narrow view of this critique one could, at a minimum, concede impact evaluations based on experimental and quasi-experimental designs should be supplemented by evaluation methods which study implementation for specific groups in specific contexts. This will be particularly important if PbR interventions are to be delivered to large cohorts across broad geographic areas. As Farrington (2003: 64), a supporter of the experimental model, observes: ‘few evaluation researchers would disagree with Pawson and Tilley’s argument that contexts and mechanisms (or, more generally, moderators and mediators) should be investigated’. Such an evaluation strategy will further raise the overhead costs for PbR models; perhaps further than Mulgan et al. (2010) anticipate.

Embracing the scientific realist critique fully suggests further serious problems for the validation and onward development of PbR models. The implication of the scientific realist position is that, because people in the social world are active, recreate their social worlds and act reflexively (Pawson and Tilley, 1997), scientific realist evaluations cannot provide a definitive, generalizable statement about whether an intervention works. Rather, knowledge of what works will be contingent and ‘the grand evaluation payoff is thus nothing other than improved theory, which can then be subjected to further testing and refinement, through implementation in the next programme. And so the cycle continues’ (Pawson and Tilley, 1998: 89–90). If this view of evaluation is accepted, the implication for PbR is not only that extensive evaluation both of outcomes and implementation in different contexts will be required, but also that, every time an intervention is implemented, there will be an ongoing need for such evaluation.11 Overhead costs are therefore likely to remain.

Clearly PbR provides an interesting test of the limits of evidence-based policy in the criminal justice sector. Some of the strongest challenges to the implementation of PbR models stem from limitations within the existing evidence base and highly credible arguments about the limits of experimentally-driven crime policies. These are not debates that potential investors and commissioners can afford to ignore. They suggest that reliable estimates of impact may be hard to reach and that reliance upon them as providers and investors seek to roll out successful interventions in different areas or with different client groups might be problematic.

**Valuing outcomes**

Putting a value on the outcomes resulting from a PbR model will be important if investors and commissioners are to judge whether the rate of return is worth the investment. Estimating the costs of crime will be important for any intervention where crime is an outcome. Given the challenges involved in estimating the costs of crime, existing estimates published by the Home Office in 2005 will generally be used (Home Office, 2005). However, these estimates are increasingly problematic because they are 2003–4 values. Looking at the total costs set out in the 2005 publication, half of the costs identified are the physical and emotional impact on victims (Home Office, 2005: chart 4.2). It may be
the case that costs associated with the physical and emotional impact of crime on individual victims will not change substantially over time, and a simple process of compounding will be sufficient to calculate their net present value (NPV). However, this is an unsupported assumption.

The other half of the costs are a mix of public and private sector costs; for example, the costs to the criminal justice system and the costs to employers of lost output. As these figures relate to a period increasingly remote in time from the present, it is likely that the underlying costs used in these estimates will have become out-of-date. A simple process of compounding will not account for this. As an example, as real GDP changes, the costs of lost output will change.

A further limitation of existing Home Office cost estimates is that they don’t take account of the impact of crime on ‘quality of life’. We can identify two aspects of quality of life. First, there is the impact of crime on potential victims. This will manifest itself as fear of crime (Dolan and Peasgood, 2007). Estimates for fear of crime have been calculated (Dolan and Peasgood, 2007) but are not included in the Home Office costs. Second, there is the broader impact on quality of life in a community. The impact of crime may be greater than the sum of fear of crime experienced by individuals. These costs will not be captured in an approach to costing crime that focuses on the costs of criminal justice agencies such as the police, courts and prisons and the physical and emotional costs of individual victims (Cohen, 2007). One approach to generating such estimates is the further development and wider use of willingness-to-pay surveys – also referred to as a contingent valuation survey (Cohen, 2007). While this is a method used extensively in fields such as environmental policy (Cohen, 2007) there have only been a handful of contingent valuation crime studies to date, and only one in the UK. An extensive programme of research will be required to address this gap (Cohen, 2007). It should also be noted that willingness to pay studies have their own limitations. The amount a respondent might pay to avoid certain death might conceivably be more than 1,000 times greater than the amount the same agent might pay to reduce their likelihood of dying by one millimort.12

For new cost of crime estimates to be calculated each time a PbR model is estimated is unrealistic. Not only would this add substantial overhead costs to each PbR contract, but inevitable inconsistencies between the methods used for different estimates would reduce the ability of commissioners to take a strategic view of the desirability of different PbR options. The Home Office decision not to regularly provide updates of the estimated cost of crime13 is becoming an increasingly important obstacle to economic evaluation in the criminal justice system.

Identifying and allocating fiscal benefits to public agencies

If PbR models are to become widespread they will rely on the organization commissioning the service being able to realize a monetary saving from which to pay for the results commissioned. Arguably the Peterborough social bond initiative sidesteps the issue of identifying fiscal benefits in two ways. First, the bond is underwritten by the Ministry of Justice, not local commissioners. The bond appears to be based only on broad economic returns to the Ministry of Justice which would result from a reduction in re-offending,
rather than specific fiscal savings local commissioners, such as the Regional Offender Manager, would have required. The evaluators of the Peterborough SIB conclude that ‘The Peterborough SIB is too small to deliver substantial ‘cashable’ savings (monetized benefits)’ (Disley et al., 2011: iv). Their suggestion is that to find out whether cashable savings can be identified the SIB model would need to be implemented on a larger scale (Disley et al., 2011: iv).

Second, even the basic premise that less re-offending results is a large, direct saving for the Ministry of Justice (Mulgan et al., 2010) is itself debatable. Prison numbers are not a direct product of re-offending, they are the product of criminal justice policy decisions made by governments and the sentencing decisions made by individual sentencers. For example, the Ministry of Justice (2009b) have demonstrated the recent rise in prison numbers is due to longer, harsher sentences and an increase in breaches of community sentences. There is no guarantee that a reduction in re-offending will reduce demand for prison places.

It should further be borne in mind, not all reductions in re-offending (or, more accurately, re-convictions) bring realizable savings. Where the costs of existing services are largely fixed, the identification of fiscal savings might be difficult. For example, looking at the Home Office (2005) costs of crime, 20% are criminal justice costs and a further 6% are National Health Service costs. Fiscal savings will largely have to come from this 26% of the social and economic costs of crime. However, not all such costs are marginal – that is, they will not decline proportionally as re-conviction rates fall. The intervention impact which might allow a whole wing of a prison to close might well be beyond the scope of a single PbR model. A broader PbR approach that works across a whole area – along the lines envisaged in the Ministry of Justice pilots in Manchester and London14 (Ministry of Justice, 2010b) might address this. The new Communities Fund might provide a mechanism that would allow a number of public providers to come together and jointly commission such a model. However, such a model will be complex and further exacerbate the challenges of measuring impact described above.

The evaluators of the Peterborough SIB found that development of a payment model required ‘considerable analytical resources and relied upon the availability of Ministry of Justice data about the cost of re-conviction events’ (Disley et al., 2011: iv). By way of comparison, consider the DWP’s ‘Work Programme’. While not without its own complexities, relevant welfare claimants can be numbered in the millions and the tens and hundreds of thousands in each region or sub-region. The key outcome – of getting a
welfare claimant into work and off welfare is – relatively easy to conceptualize and measure and generates a quantifiable fiscal benefit for government (a saving in welfare payments). Although some of the government’s provision for welfare claimants is fixed and would require large numbers of claimants to move into work before a marginal cost could be realized (e.g. the infrastructure to administer benefits or job centres providing support to the unemployed) the welfare payments themselves are a marginal cost and a saving can be made even if the employment status of only one individual changes (see Mulgan et al., 2010, for further discussion of this contrast).

The need to generate fiscal savings is a major challenge for the viability of PbR models. Scale will be one important factor. Models which work with large cohorts are more likely to generate substantial savings, a proportion of which might be monetizable. But, as discussed earlier, delivering substantial changes in entrenched criminal behaviour will be particularly challenging when attempted with large cohorts and accurate measurement of such interventions will be challenging. Evidence available (Home Office, 2005) suggests that savings arising from a reduction in re-offending largely accrue to society at large. Yet only those savings accruing to government agencies are (potentially) realizable. Thus PbR providers will be paid on the basis of a lower-bound estimate of the benefit of their work.

Discussion: Payment by results and the challenge for evidence-based policy

Payment by results models have a place in the provision of ‘public’ services. The transference of risk is attractive to government. It is not, of course, so attractive for the private sector SIB holder. Indeed, it might be supposed that private investors are less able to absorb risk than national governments. Moreover, if the SIB in Peterborough is an indicative example, the risk structure may be highly skewed: returns to investors are capped by government, but there is a likelihood of the loss of all the invested capital if the intervention does not produce the specified outcome. Economic theory suggests that the sources of private capital will require compensation – through a higher expected rate of return – for undertaking risky projects, yet the rate of return on the Peterborough SIB is capped at relatively modest levels. Thus the estimation of effect sizes is of crucial importance.

In some parts of the public sector where the potential client group is large, measurement of outcomes is straightforward and the fiscal return on achievement of outcomes is easy to establish, the likely return on investment might be attractive to private investors. In the case of interventions in the criminal justice system, outcomes are relatively difficult to measure and value. Fiscal benefits are shared across a number of local providers (police, courts, probation, prison and the NHS to name but a few) and existing evidence on what works is still relatively patchy, particularly if relying on UK evaluation data. There is a vibrant debate within the evaluation field about the validity of the experimental and quasi-experimental evaluation designs, evidence from which payment by results models favour. The result of these uncertainties is that overhead costs are likely to be relatively high and risks for providers, investors and commissioners relatively uncertain.
Given this, it is likely that early entrants to this new investment market are likely to be social investors for whom a specified return on investment may be less critical. However, even where preliminary forecasts of intervention effects are reliable, and benefits are realized and monetized, these benefits are likely to be well below those which accrue to society. Social entrepreneurs are likely to see their interventions undervalued.

It would appear that every link in the chain, from the establishing of an intervention’s likely effects through to a fair and uncontentious payment, is subject to much debate and, we would suggest, challenge. The weakness of the chain arises not from its weakest link but from the sum total of the weakness of every link. Simple economics would suggest such uncertainty and the associated risk to investors will lead to a sub-optimal provision of interventions in the criminal justice sector. The development of payment by results models in the criminal justice system is at an early stage. Different versions of the basic concept are possible, but all will need to address the issues raised in this paper if sub-optimal provision is to be avoided.

Notes

1. PbR envisages flexibility for the providers of services over how services are delivered. This would represent a complete change of approach to that followed in recent years. For instance, over recent years the work of probation services has been tightly prescribed by national standards and a national framework for training Probation Officers and Probation Support Officers. It is worth noting that the most recent national standards, published by the coalition government (Ministry of Justice, 2011) represent a substantial relaxing of central government direction and might be consistent with preparations for greater market testing of probation services.


3. Our reading of Disley et al. (2011) is that none are from the private sector.

4. Interesting questions are raised about the sources of charitable funding with some investors using their grant-giving mechanisms, but others investing their endowment capital on the basis that the SIB is a financial instrument that offers a return on investment as well as meeting an organization’s charitable mission (Disley et al., 2011).

5. Mulgan et al. (2010) set out a table in Annex One that is based on Table 2 in Lipsey and Cullen (2007). However, Mulgan et al.’s (2010) table has the addition of a line that appears to be a meta-analysis of the meta-analyses summarized by Lipsey and Cullen, but we are unclear how this additional meta-analysis was calculated.

6. Construct validity refers to the adequacy of the operational definition and measurement of the theoretical constructs that underlie the intervention and the outcome (Farrington, 2003).

7. An issue we return to below.

8. Alternatively a rapid evidence assessment might be more appropriate, given likely time and resource constraints.

9. Disley et al. (2011) describe how, for the Peterborough SIB, the size of impact required to trigger payment was driven, in part, by working out what size of impact could be identified in a cohort such that the finding would be statistically significant.

10. Disley et al. (2011: 8) also point out that outcome measurements relying on comparison with a control group cannot be rolled out nationally because, for example: ‘If every short-sentenced prisoner in England and Wales could be part of a SIB, there would be no control group with which to compare.’

11. Investors will not only need to invest heavily in evaluation, they will also have a strong interest in seeing positive findings. Will this raise some of the challenges seen in other fields such
as medical research? A systematic review by Lexchin et al. (2003) found 30 separate studies examining the impact of funding on the outcomes of trials. Overall, Lexchin et al. found that studies funded by a pharmaceutical company were four times more likely to give results that were favourable to the company than independent studies (cited in Goldacre, 2009: 211).

12. That is, the probability of a one in a thousand of dying from a particular activity.

13. In the 2005 publication the Home Office stated that the estimates presented in that publication were part of ‘an ongoing programme of research, which will generate regular and periodic updates of the estimates in future’ (Home Office, 2005: 17)

14. These payment by results models concentrate on outputs (for example, reducing the number reaching court, reducing the number on community sentences and reducing the number receiving custody) rather than outcomes (for example, reducing the rate of re-offending).

15. There may also be longer term outcomes that flow from this such as increasing tax revenues or better health.

References


**Biographies**

**Chris Fox** is Professor of Evaluation and Director of the Policy Evaluation and Research Unit at Manchester Metropolitan University, UK. He has undertaken many evaluations in the criminal justice system.

**Kevin Albertson** is Acting Head of the Department of Economics and Deputy Head of the Policy Evaluation and Research Unit at Manchester Metropolitan University. They are joint authors of *Crime and Economics: An Introduction* published by Routledge.