

## **Payment by Results (PbR) Drug & Alcohol Recovery Pilot Programme: a note of advice to the Department of Health (DH) on the proposed evaluation**

### **1. Background**

The DH, along with a number of other government departments and the National Treatment Agency for Substance Misuse, are developing a number of pilots to test new approaches to commissioning and delivery of drugs recovery systems that reward achievement of outcomes, or payment by results (PbR). The aim is to incentivise and reward providers that support individuals to recover from their drug dependence. Eight local areas have been chosen to pilot the drugs recovery PbR model, beginning in April 2012 and running for two years. The DH and its expert partners are working closely with the eight pilot areas to develop the details of the PbR model (eg, the performance outcomes) during a “co-design” period in the run-up to the April 2012 start date. DH is commissioning an independent evaluation of the pilot areas during the summer 2011.

PIRU has been asked by the DH Policy Research Programme to provide advice on the design of the pilots (eg, on the outcomes used for performance payments) and on their evaluation. PIRU has set up a team of three individuals to provide this advice: Bob Erens (LSHTM), Martin Roland (RAND Europe), and Martin Knapp (LSE). While PIRU have been able to read the relevant documentation on the PbR Drugs Recovery pilots, attend two co-design meetings in late May and July, and monitor the web forum set up by DH to communicate with the pilot sites, PIRU has had no input into the performance outcome measures for a number of reasons, eg: they are largely being driven by Ministers and from DH centrally; there are many interested parties across different government departments involved in setting the measures; and the detailed work on the outcome measures was well-advanced by the time PIRU became involved.

Our comments are based on our experience from working on other evaluations involving PbR/P4P, as well as a knowledge of the general PbR/P4P literature (and three reviews have been attached to this note: Christianson, Sutherland, Leatherman 2009; Conrad and Perry 2009; Van Herck et al 2010).

### **2. PbR outcomes measures**

#### *PbR for Drugs Recovery outcome measures*

The broad aim of this pilot programme is to explore how providers can be incentivised to deliver on recovery outcomes. This is probably the first pilot in the world where funding for drug treatment and recovery will depend solely on the outcomes achieved in specified domains. Payments will depend on providers achieving outcomes in four domains that have been identified by the government:

- 1 free from drugs of dependence
- 2 employment
- 3 offending
- 4 health & wellbeing.

The specific outcomes within each of these domains, which have been agreed by Ministers and the co-design group, are as follows:

**1 Free from drugs of dependence**

**Initial outcome:** Abstinent from all presenting substances (opiates, crack, cocaine, alcohol, cannabis, amphetamines) recorded on last two TOP reviews, and still in treatment.

**Final Outcome:** Planned discharge from structured treatment and no re-presentation to either treatment or CJ systems in following 12 months

**2 Employment**

**Initial outcome:** Planned discharge from structured treatment – ongoing substance misuse no longer a barrier to employment.

**Final outcome:** No firm measure proposed yet.

**3 Offending**

**Initial outcome:** Reduction in proven offending for either a group (measured either by number of proven offences or number of offenders), or for each individual in first 6 months in structured treatment.

**Final outcome:** Reduction in proven offending for either a group (measured either by number of proven offences or number of offenders) in the 12 months from beginning structured treatment, or for each individual over a period of 12 months (initially from the beginning of structured treatment, but with opportunity to ‘restart the clock’).

(\* Pilots are able to choose whether to use a group or individual measure. The strong steer use a group measure, as this minimises the risk of paying for outcomes that may have occurred anyway. Areas with a small number of drug using offenders will need to carefully balance the difficulties of using a group measure against the risks associated with using an individual one).

**4 Health & wellbeing**

**Injecting Initial outcome:** For those injecting at start of treatment, a recording of 0 days injecting at last 2 TOP reviews in last 12 months

**NFA/Housing Problem Initial outcome:** Of those NFA or with a housing problem at start of treatment, recording of **NO** housing problem at last 2 TOP reviews

**Hep B Initial Outcome:** Those who, having been assessed as requiring it, complete a course of Hep B vaccinations in last 12 months.

**HWB Initial Outcome:** Client achieves a norm of health and social functioning in the last 2 TOP reviews.

*PIRU comments on outcome measures*

A number of characteristics of the payment mechanism are likely to influence both the effectiveness of the scheme and the potential for gaming. Relevant to the PbR Drugs Recovery outcome measures are the size and nature of the payments in relation to overall provider income. For the PbR Drugs Recovery pilots, the government’s goal is for 100% of provider income to be based on meeting the outcomes set out above. This is in stark contrast to most PbR schemes, where generally only around 5% to 10% of overall income is regarded as suitable for PbR payments for providers. Increasing the percentage of income related to performance may increase the effectiveness of the incentive but will also increase the incentive for gaming. Also, large incentives tend to crowd out intrinsic motivation and encourage providers to alter their behaviour to meet the outcome measures required for payment, even though this can result in perverse consequences. (An example of this is

how the incentive introduced for GPs in 2004 to offer appointments in 48 hours resulted in patients being unable to book further ahead, even though the target was met as patients could book appointments within 48 hours.) With income dependent upon 100% PbR, and the incentive for gaming (eg, cream skimming) this will encourage among providers, it will be very important: a) for pilot sites to carefully balance the payments across the different domains of the outcome measures and across initial versus final outcomes; b) ensure that tariffs are set independently and in such a way that cream skimming will be discouraged; c) achievement of outcomes can be clearly defined, and are capable of accurate, and unambiguous, measurement.

Our clear advice from past research is that tying such a high proportion of income to results is not necessary to improve performance and highly likely to produce unintended negative consequences. Tying 100% of income to performance is most likely to have perverse consequences when the outcome is not fully under the control of the provider, and this is precisely the case for drug usage – e.g. see comments below to do with Offending outcome.

There is also likely to be a tension between the cash-flow needs of the providers and the measures of the scheme's success. The key outcome measures are those identified as "final", since they indicate the lasting recovery of the individuals; however, data to measure some of these final outcomes (eg, reduction in offending 12 months from start of treatment) may not be available for one year or more until after treatment is completed, introducing a very long potential delay into the final payment to the provider. Several pilot sites have expressed concern about the potential length of time that providers may have to wait until they are paid their tariff, as many providers, especially the smaller ones, are not able to wait very long before payment without encountering significant cash-flow problems. This could lead to pilot sites loading payments on the interim outcomes, which in itself would be an adverse outcome since these outcomes are not as meaningful as the final ones in terms of measuring the overall success of the scheme.

In terms of the individual outcome measures:

**Freedom from drug dependence:** This will be monitored using TOP reviews, which is essentially self-report, and on clients not re-presenting for treatment within 12 months of completion of their treatment. To the extent that this assessment relies on self-report, there could be strong incentives for providers to encourage clients to downplay their reports of drug use, and potentially to inform clients who re-present for treatment within the 12 month period that they cannot be seen until after the 12 month period has passed (eg, because they are "fully booked"). If the providers are responsible for the TOP reviews, they may have discretion to decide when to carry out a review, choosing times when they expect to get a favourable outcome; and if they don't get the expected outcome, they may simply not declare that a review has been done. The most reliable way to measure this outcome would be verifiable testing, eg, based on samples.

**Employment:** We understand that this outcome domain has been dropped.

**Offending:** The pilot sites are being given a strong steer towards using group, rather than individual, measure of crime. Whichever measure is used, it is important that wider trends in crime in the areas in which the pilots are located are monitored (eg, there could be changes in levels of reported crime due to changes in policing in the area, or in reporting crimes, or even changes in levels of employment which could alter opportunities to find work). Such trends could have a much bigger influence on crime rates, or on an individual's likelihood of committing a crime, than the PbR Drugs Recovery pilot. This affects not only whether or not the performance outcome is met, but also evaluating the impact of PbR overall.

Health & wellbeing: Since many of these outcomes are based on TOP reviews, ie, self-reports, they depend on how accurate the self-reports are and whether they are open to manipulation.

Other issues to be aware of include:

Cream-skimming, or only taking on clients who are likely to achieve the outcome measures. The pilots are handling this in two ways: firstly, adjusting the tariffs so that more difficult cases attract higher payments than less complex clients; secondly, having an independent Local Area Single Assessment and Referral System (LASARS) assess and refer clients and set their tariffs. What is unclear is what will happen if certain providers refuse to take more difficult clients, or say they are too busy to take on a particularly difficult client. Longstanding drug users become well known within local drug services, and it may be relatively easy for providers to identify clients who they regard as “too difficult” for this system. Providers can use a number of ways to avoid the most difficult cases, eg, being generally unhelpful or unsympathetic to the client so that they default from attendance, thus avoiding committing resources to a client whom they judge likely to fail.

Fraud: Tying 100% of income to results increases the possibility of fraud, and robust measures should be put in place to detect fraud. However, the DH needs to be aware that some fraud will be very difficult to detect. For example, if abstinence is based on self-report and a client reports taking drugs ‘only occasionally’, it will be hard to know if the staff member records the response as ‘drug free’ in the case notes.

LASARS: The PbR system places considerable onus on having an independent LASARS which can carry out accurate assessment of client needs, set the tariff according to needs, and refer the client to an appropriate provider. Very much depends on how a LASARS is constituted, what skills it requires, and who it includes, because providers could suffer (or benefit) if a LASARS continually makes incorrect assessments. If a LASARS in an area perceives the need to keep providers afloat, might they direct “easy” clients to providers who may be struggling to make their targets, and “difficult” clients to providers who may be outperforming others because the consequences of them failing a difficult client will have fewer consequences.

Also, it seems that, in at least two pilot sites (Wakefield and Wigan) for reasons of costs, the LASARS will not be independent of the providers, and this could open the system up to potential abuse, as providers sitting on LASARS will be able to set tariffs, allocate clients to particular services, decide favourable outcomes, etc.

### 3. Evaluating PbR Drugs Recovery pilots

DH is currently tendering for an independent, robust evaluation of the PbR Drugs Recovery pilot programme. The evaluation was expected to commence in October 2011 and be conducted over 36 months. However, with the delay in the start of the pilots until April 2012, there is less pressure to have the evaluation team in place. We would suggest the evaluation team is in place by the end of 2011, and certainly before the start of the pilots. The evaluation will attempt to measure the impact of a PbR approach to drugs recovery on provider behaviour, provider/market stimulation, client experience and recovery outcomes.

The overall objectives of the evaluation as set out in the ITT are to:

- Assess the effectiveness of the PbR Drugs Recovery programme against key outcome measures in the broad domains of treatment, employment, offending, and wider health and well-being. This should include the process and impacts (including unintended impacts) of the LASARS.
- Undertake an economic evaluation of the individual pilots and of the PbR Drugs Recovery programme as a whole.
- Identify the key lessons for ensuring the quality, effectiveness and efficiency of the PbR Drugs Recovery models in the future.

The evaluation team will undertake a programme-wide process evaluation and a robust impact evaluation across a number of sites, in order to draw reliable conclusions that can be transferred to practice in other sites. This should include an economic evaluation of the pilot schemes, as well as a descriptive analysis of the different pilot schemes.

The *process evaluation* is expected to identify the key lessons for ensuring the effectiveness of PbR schemes in future. This part of the evaluation will examine: different PbR models and provider partnerships, including the underlying “logic” model; services provided; assessment and referral processes; take-up of services; clients’ and stakeholders’ views of the services; cross-sectoral partnership between commissioners, providers and care professionals; what types of eligible people use the services and why types do not, with reference to cream-skimming, etc; and how much it costs to set up and run a PbR drugs recovery scheme, including the cost of LASARS.

The *impact evaluation* is to assess the effectiveness of the pilots, compared with standard treatment services, in terms of: recovery from drug or alcohol dependence; re-integration into their communities (eg, committing fewer crimes, finding employment); and their health and well-being. What are the (positive and negative) unintended consequences of the pilots? Randomisation within a pilot site is not going to be possible, so non-experimental approaches are required to assess impact. Not all 8 pilot sites need be included in the impact evaluation, but different types of pilots should be included.

The *economic evaluation* will assess the extent to which the pilots represent good value for money in terms of impact realised. It should also assess the impact of a PbR approach on stimulation of the drugs recovery provider market, and on the economic behaviour of the providers. It should be carried out for all pilot sites. The evaluation must examine how much the scheme costs to set up and run, and what other costs there may be such as healthcare, criminal justice, and economic productivity costs.

#### *PIRU comments on the evaluation*

While not a comprehensive list, we highlight below some issues that pertain to the overall evaluation of the PbR Drug Recovery pilots.

In order for a robust assessment of outcomes, it will be essential for the evaluators to be able to access data from comparator sites. If this is not possible, then the ability of the evaluation to report robustly on the outcome of the project will be severely limited.

This is not a randomised experiment, either overall or within pilot sites. This will make it difficult to compare one pilot with another, or pilot areas with control areas. Moreover, while the outcomes are centrally defined, because of the localism agenda, different pilot sites may operationalise the outcomes in different ways, tariffs could differ for each outcome, and pilot areas may add more outcomes or other features to their scheme. To the extent these actions take place, it will be difficult to make comparisons between pilots and to disentangle what is affecting the outcomes (if there is any effect).

Another potential confounder for looking at the effect of PbR is that, during the course of setting up the new pilot schemes, the pilot areas may be using new providers and/or new treatments (indeed PbR is meant to encourage this), so it may be difficult to determine the extent to which improvements in outcomes are due to PbR or to the new providers/treatments used. This also raises the question of whether these new providers/treatments could have been accessed before the introduction of PbR.

Given the difficulties just described, the evaluation team will need to work with each pilot site to develop a logic model for their scheme, in order to assist with identifying which factors in the pilot area may be contributing to any observed improvements. (The evaluation team, and DH, should be aware, however, that the pilot sites will have a strong incentive to report positive progress to the evaluators.)

The evaluation will need to examine how providers are coping with the new PbR system, especially in terms of cash flow issues (eg, some clients are in treatment for long periods delaying even interim payments, while final payments will not be made for 1 year or more after leaving treatment), and what changes they would like to see to improve their delivery of services.

How does PbR handle the complexities of clients coming into treatment, perhaps leaving for a short while and then returning?

There will be difficulties finding control areas to make comparisons with the pilot areas, so how will these be selected? The control areas should be similar in terms of drug users' characteristics as well as in other ways, such as rates of crime and unemployment. Also, what happens if a control area decides to introduce PbR while the evaluation is still under way? Will there be sufficient data available for the control areas?

There are complexities comparing case mix in pilot sites before/after PbR and with control sites. Even defining the baseline is difficult, as some clients in the pilot areas will have started their treatment before PbR was introduced.

Perhaps PbR works well for some clients, but less well for others (eg, those with more complex needs), and the cost implications of PbR may vary between clients for reasons that are either positive (eg, responsiveness) or negative (eg, gaming). Also, do some providers cope better with PbR than others?

The current outcomes all focus on the individual in treatment, but it will also be useful to examine broader outcomes, such as the effect on the client's family and community (eg, in terms of reduction in crime). And what are clients' views of the new system (are they even aware of the change)?

Similarly, the dependency-free outcome is based on presenting substances, but should the final outcome be more ambitious, eg free of all drug use?

The final outcomes are longer-term, and require following up clients for 1 year after treatment ends. What are the difficulties in doing this (eg, if clients use different names or move to another area).

The economic evaluation should have a number of elements. It will need to examine the cost consequences of PbR: how much the scheme costs to set up and run, and what other budgetary implications there may be, such as for health care, criminal justice, and social care systems. Overall, have costs gone up (eg, because outcomes are too easy to achieve, or PbR was too successful and there are many more successful outcomes), or does PbR help to contain or even reduce costs?

The evaluation should assess the links between costs and both individual and wider outcomes. In that way it should examine the extent to which PbR offers a cost-effective way to deliver recovery outcomes for individuals, and should do so from the perspective not only of providers, but from public sector and wider societal perspectives (and so including any employment/productivity impacts).

The economic evaluation should also assess the wider impact of a PbR approach, for example on stimulation of the drugs recovery provider market, and on the economic behaviour of the providers. Does PbR attract new providers or does it leave the sector with fewer providers (eg, because the scheme is too complex, outcomes are too difficult to achieve, or because of cash-flow problems)?

It should also look at links across agencies and systems (addiction services, other health care, criminal justice, benefits and so on).

Some potential unintended consequences the evaluation team may wish to consider: Among people who are eligible for treatment, are particular types not taking up the offer of treatment under the PbR scheme? What are the numbers, and what are the differences with the previous scheme? The DH should note that it will be particularly difficult for the evaluators to detect some types of perverse behaviour, as it will be in the sites' interests to conceal them.

How have waiting times for treatment been affected? And also time spent in treatment, eg, has this decreased in order to save costs? Has the type or content of treatment changed? What are the budgetary consequences? Has the number of individuals treated over a given period increased or decreased? Is undue pressure being put on clients in order to achieve outcomes?

How well are LASARS performing in terms of assessing client needs and setting the tariff, and in particular, what are the consequences (if any) of an incorrect tariff? And how frequently is the tariff set 'incorrectly'? If a client proves to be more expensive than the tariff they are being given, does the provider ask for more funding from LASARS, or do they give up on the client?

Clients are not meant to be told what their tariff is, but what are the consequences if clients manage to find out their own, or others', tariffs?

How might providers 'game' the system? The evaluation team may wish to discuss this with providers outside the pilot areas to gain insight into how the system could be manipulated.

What type of auditing is required to detect potential fraud?

Longer-term, does PbR lead to some initial improvements, but then “stagnate” because incentives are no longer changing? This may be outside the scope of the initial evaluation, but perhaps the team can discuss with providers about how the incentive system needs to adapt to ensure there are continual improvements in the system.

Finally, the evaluation team will likely have to balance positive and negative outcomes in order to judge whether PbR has been a “success” and is worth continuing with: eg, if there are no new providers after PbR is introduced and waiting times for treatment have increased, but outcomes show some improvement, is that a success?

In selecting an evaluation team, the DH should look for awareness of these problems among applicants and expect them to provide credible plans for how they will deal with them.

10th August 2011

Bob Erens, London School of Hygiene & Tropical Medicine

Martin Roland, University of Cambridge

Martin Knapp, London School of Economics & Political Science

On behalf of the Policy Research Unit in Policy Innovation Research (PIRU)

Disclaimer: This is an independent report commissioned and funded by the Department of Health. The views expressed are not necessarily those of the Department.