



Ministry of  
**JUSTICE**

# **The feasibility of conducting an impact evaluation of the Dedicated Drug Court pilot**

Full report

**McSweeney, T., Meadows, P., Metcalf, H., Turnbull, P.  
and Stanley, C.**

**National Institute of Economic and Social Research  
and the Institute for Criminal Policy Research, Kings  
College, London**

Full report to Ministry of Justice Research Summary 2/10

# **The feasibility of conducting an impact evaluation of the Dedicated Drug Court pilot**

Full report

**McSweeney, T., Meadows, P., Metcalf, H., Turnbull, P. and Stanley, C.  
National Institute of Economic and Social Research and  
the Institute for Criminal Policy Research, Kings College, London**

This information is also available on the Ministry of Justice website:  
[www.justice.gov.uk/publications/research.htm](http://www.justice.gov.uk/publications/research.htm)

## **Disclaimer**

The views expressed are those of the authors and are not necessarily shared by the Ministry of Justice (nor do they represent Government policy).

© Crown Copyright 2010.

Extracts from this document may be reproduced for non-commercial purposes on condition that the source is acknowledged.

# Contents

Summary	i
Acknowledgements	vi
Glossary and abbreviations	vii
1 Introduction	1
1.1 Aims of the study	1
1.2 Dedicated drug courts	3
1.3 Evaluation approaches	4
1.4 Other design considerations	7
1.5 Method	8
1.6 Report structure	12
2 The effect of drugs courts	13
2.1 Evidence on the impact of drug courts	13
2.2 Impact of drug courts on recidivism	13
2.3 Impact of drug courts on substance misuse	14
2.4 Impact of drug courts on sentence completion and compliance	15
2.5 Implications for the expected impact of ddcs	15
3 Ddcs: policy and practice	17
3.1 Introduction	17
3.2 Exclusivity, continuity and dedicated sessions	17
3.3 Training	18
3.4 Inter-agency working and information sharing	18
4 Practice in jurisdictions without a ddc	19
4.1 Introduction	19
4.2 Exclusivity, continuity and dedicated sessions	19
4.3 Training	19
4.4 Inter-agency working and information sharing	19
4.5 Data collection and monitoring	20
5 Key issues: comparators	21
5.1 Introduction	21
5.2 The policy counterfactual	21
5.3 Experimental design (random assignment)	21
5.4 Between-area comparison	22
5.5 Within-ddc area comparison	22
5.6 Historical comparisons	22
6 Key issue: sample size	23
6.1 Introduction	23
6.2 Minimum detectable effect and sample size	23
7 Key issues: data	25
7.1 Introduction	25
7.2 Main sources of data	25

8	Potential impact evaluation designs	28
8.1	Potential designs	28
8.2	Rct design	28
8.3	Between-area comparison design	33
9	Conclusions	37
9.1	Introduction	37
9.2	Expected impact	37
9.3	Data availability	37
9.4	Ddc and courts in jurisdictions without a ddc	37
9.5	Detectable effect	38
9.6	Possible evaluation approaches	38
	References	40
Appendix 1	ddc policy and practice: details	46
Appendix 2	policy and practice in jurisdictions without a ddc	52
Appendix 3	costs	58
Appendix 4	propensity score matching	61

# Summary

## Aims

The aim of the study was to establish whether a robust impact evaluation of the Dedicated Drug Court (DDC) pilot were possible and, if so, the requirements, including costs, for such an evaluation. The study was to examine robust evaluation methods only, i.e. those which identify not only what has happened to offenders who have gone through DDCs, but provide a means of estimating what might have happened to them if they had been dealt with by other courts. Other evaluation methods exist (e.g. before and after designs) which could provide useful information for policy makers and practitioners. However, these approaches did not form part of the remit of the study.

DDCs have the same powers as other magistrates' courts, including sentencing, dealing with sentence breaches and reviewing Community Orders with drug rehabilitation requirements (DRRs). Although DRRs form a major part of DDCs' work, DDCs have a wider remit and the study was to explore the potential for a robust evaluation of the full role of the DDC.

## Dedicated Drug Courts

The DDC pilot aims to improve the processes and effectiveness of the magistrates' courts in dealing with drug-misusing offenders, aiming to reduce drug use and reoffending and improve sentence completion and compliance. It provides a new framework in magistrates' courts for dealing with drug-misusing offenders who commit low-level 'acquisitive' crime to fund their addiction. It does not introduce new treatment or sentencing options.

The key elements of the DDC pilot are to:

1. provide continuity of judiciary from sentence through to completion, and if necessary (and where possible), breach;
2. list cases for sentence, breach and DRR reviews together in dedicated court sessions;
3. provide additional training for the judiciary and other staff dealing with these cases to improve awareness and understanding of drug addiction issues; and
4. improve inter-agency working and information sharing processes.

Pilot DDCs agreed to adhere to this basic framework but the way in which they implemented the model and details of operation could vary.

## Evaluation approaches

To assess reliably the programme outcomes, one needs to know what happened to those exposed to the programme and what would have happened to them otherwise (the counterfactual). The difference between these is the effect of the programme. There were, potentially, four main ways of doing this:

1. randomly assigning relevant offenders<sup>1</sup> to the DDC and to other courts (Random Control Trial, RCT);<sup>2</sup>
2. comparing offenders dealt with by the DDC with similar offenders in jurisdictions without a DDC (between-area comparisons);
3. in the jurisdictions with a DDC, comparing offenders dealt with by the DDC with similar offenders in the same jurisdiction who were dealt with by other courts (within-area comparison);
4. in the jurisdictions with a DDC, comparing offenders dealt with by the DDC with similar offenders in the same jurisdiction prior to the introduction of the DDC (historical comparison).

It is important that the offenders subjected to the DDC and those who represent what would have happened to these offenders if they had not gone before the DDC are similar in terms of their characteristics such as age, gender, educational and employment background and offending history. Otherwise, differences in outcomes may be due to differences between the groups and not to the DDC. Random assignment, if done properly, should normally ensure that the two groups are similar. For the other approaches, statistical techniques can adjust for some dissimilarity between groups.

## Method

To identify possible designs, the researchers needed to assess:

- the extent to which DDC and non-DDC practice differed;
- the availability of relevant data; and

---

1 DDCs deal with defendants who intend to plead guilty, as well as convicted offenders (those sentenced by other courts and those having their Community Orders reviewed). As the former become convicted offenders as soon as their plea is accepted, 'offender' is used to refer to both offenders and to defendants, in relation to the DDC.

2 The subjects of the intervention being evaluated (in this case people who meet the criteria for referral to the DDC) are divided at random into two or more groups. Those assigned to the control group do not receive the new intervention (in the DDC evaluation, they would have their case heard in an ordinary magistrates' court). Those assigned to the experimental/treatment group receive the new intervention (in the DDC evaluation, they would have their case dealt with by a DDC). Provided the allocation of individuals to the two groups is entirely random, and the sample size is large enough, under the right conditions any differences in the average outcomes between the treatment and control groups can be attributed to the intervention.

- the potential impact of the DDCs on outcomes (to estimate the sample size required to detect effects).
- For an RCT, the researchers needed to consider the way offenders are allocated to the DDC in order to assess whether the processes involved would permit genuinely random allocation of cases between DDCs and other courts.

Information was gathered through interviews with the six pilot DDCs, with courts and probation in jurisdictions which did not have a DDC, collation of existing activity data on DDCs, a literature review of evidence on the effectiveness of drug courts and a review of existing health and criminal justice datasets.

### **Findings: the potential impact of the DDC pilot**

International drug courts' evaluations provided a poor indication of the likely impact of a DDC pilot. This was because the drug courts evaluated elsewhere differed substantially from the DDC model. In particular, many other drug courts involved not only a change of process but also recourse to enhanced drug treatment for offenders. What evidence there was related almost exclusively to recidivism. Taking into account differences in programmes, the potential impact on reoffending of a different **process** alone could be expected to be a reduction of around five percentage points.

### **Findings: DDC and non-DDC practice**

Practice in DDCs varied and none of the key characteristics of the DDC model were consistently defined across the six pilot sites or systematically measured and monitored. The target groups and the extent to which offenders were referred to the DDC for sentencing varied. However, all DDCs sought a degree of continuity of a magistrate and reviewed all DRRs (with the exception of those for some offenders who lived outside their jurisdiction and were sent to their home court for DRR reviews post-sentence). This meant that a within-area comparison design was not feasible.

The main distinguishing feature between DDCs and courts in jurisdictions without a DDC was continuity of magistracy, although this did not always extend to sentencing in the DDCs in all cases. The researchers could find no evidence of differences in reported levels of training and inter-agency working between DDC and non-DDC areas.

Historically, some of the DDCs had been operating as quasi-drug courts prior to the pilot and it was not possible to establish when the practice had changed. This meant that a historical comparison design was not considered feasible.



## Findings: data

There were no data constraints affecting potential designs. However, the data on continued illicit drug use of offenders were not comprehensive which would affect the ability to assess drug misuse outcomes.

The data required for the evaluation could be assembled through matching existing administrative datasets. Relevant datasets include:

- for characteristics, nature and extent of need and reoffending, the Home Office Police National Computer (HOPNC);
- for current and continued illicit drug use, police test on arrest data and the Drug Interventions Record (DIR), Treatment Outcomes Profile (TOP), the Offender Assessment System (OASys) and probation case management systems (e.g. Delius);
- for sentence completion, probation case management systems, Enforcement Tracker and the Interim Accredited Programmes Software (IAPS);
- for sentence compliance, probation management systems and Enforcement Tracker.

## Findings: allocation of offenders to the DDC

Offenders were referred to the DDC by a range of people: police, magistrates (on sentencing or referral for sentencing) and the court administration (through their listings). This would make it difficult to randomly allocate offenders to the DDC or another court and any system of random allocation would run a high risk of implementation failure.

## Implications for an evaluation design

Within-area comparison and historical comparison designs were not feasible.

A Random Control Trial is feasible in theory. However, in practice, it would run a high risk of failure for a number of reasons, including practical difficulties of randomisation and gaining the compliance of all in the criminal justice system (CJS). A difficulty is that DDC magistrates and District Judges (DJs) also sit in other courts. If their DDC experience affects their treatment of drug-misusing offenders in the other courts, the evaluation would underestimate the effectiveness of DDCs. An effective evaluation based on an RCT would require a sample size of at least 2,800, in order to be able to measure the likely impact on reoffending. Because of the limited number of people appearing before DDCs, a sample of this size would take some years to assemble. With follow-up, the study would take between eight and a half and eleven years. It would cost between £1,000,000 and £1,500,000.

Between-area comparison, comparing similar offenders in jurisdictions with DDCs with those in jurisdictions without DDCs, might be feasible. It could compare DDCs with courts nationally or courts in selected jurisdictions without a DDC. A minimum sample size of 1,600 in the DDCs and 1,600 in other jurisdictions would be needed in order to measure the likely impact. It would take between four and five years for DDCs to process 1,600 cases. The national comparison would cost £260,000 to £300,000. The selected between-area comparison would cost £280,000 to £325,000. The cost would be higher because a larger sample size would be needed to take account of differences in approach.

Both designs assume a minimum impact of five percentage points (e.g. the difference in rates of reconviction between DDC and non-DDC groups). This is based on the outcomes achieved by drug courts in other countries, while recognising that the DDCs in England and Wales do not differ from other courts in either their sentencing or their treatment options. If the effect is smaller, the sample sizes above would be too small to detect the effect. Given the lack of robust evidence on the likely effect and the small degree of difference between DDCs and non-DDCs, this is a high risk to the evaluation.

The national between-area comparison seems the most cost-effective and least risky approach. However, as with the random allocation design, the sample size is based on ensuring the ability to detect an estimated minimum five percentage point reduction in reoffending. If the effect is smaller than this it will not be detectable even with these relatively large sample sizes.

## Acknowledgements

We would like to thank the court and probation staff who gave so generously of their time to assist us in this study. We would also like to thank Tim Millar from the National Drug Evidence Centre at the University of Manchester for his advice on data linkage and DRR treatment modalities and Sara Skodbo from the Home Office for her advice on data linkage. Finally, we would like to thank Jocelyn Green, Jessica Haskins and Jenny Patterson at the Ministry of Justice for their management of, and assistance with, the study.

## Glossary and abbreviations

CJS	Criminal Justice System
Comparator group	The name of the group who do not receive the treatment in a quasi-experimental design
Control group	The name of the group who do not receive the treatment in a RCT (a specific type of comparator group created through random assignment to the treatment and comparison groups)
Counterfactual	What would have happened to those processed by the DDCs if they had been dealt with by the normal courts
DIR	Drug Intervention Record
DDC	Dedicated Drug Court
DRR	Drug rehabilitation requirements
DTTO	Drug treatment and testing order
Experimental designs	Experimental designs randomly assign offenders to the treatment group (the DDC) and to the control group (other courts) and compare the outcomes for the two groups
Experimental group	The group who receive the policy treatment (in this case the DDC) in RCTs
HOPNC	Home Office Police National Computer
Model evaluation	An evaluation designed to assess the effectiveness of a specific model compared with lack of that model, i.e. abstracting as far as possible from variations in practice in implementation of the model and from similarities to the model in non-model practice. In this case, how does a pure DDC compare with courts that exhibit as little DDC practice as possible
Pilot evaluation	An evaluation designed to assess the effectiveness of a pilot against practice by those not involved in the pilot, i.e. it is accepted that the pilot may not be implemented perfectly and that those not involved in the pilot may exhibit practices similar to the pilot
Quasi-experimental designs	Quasi-experimental designs compare the outcomes for the treatment group (the DDC) with those of the comparison group, where the comparison group has been constructed to be similar in to the treatment group, in terms of their characteristics. This may have been done through selection or statistical
Random Control Trial	See Experimental designs
RCT	Random Control Trial. See Experimental designs.
TOP	Treatment Outcomes Profile
Treatment group	The group who receive the policy treatment (in this case the DDC). Also known as the experimental group in RCTs

# 1 Introduction

## 1.1 Aims of the study

The aim of the study was to establish whether a robust impact evaluation of the Dedicated Drugs Court pilots were possible and, if so, the requirements, including costs, for such an evaluation. The impacts of interest were the DDC pilot objectives of:

- reducing reoffending by drug-misusing offenders; and
- reducing drug misuse.

Intermediate outcomes<sup>3</sup> were also to be considered, such as attendance at court and probation appointments and numbers of recorded breaches (i.e. sentence compliance and completion).

The study had the following objectives:

- to identify the most suitable methods for carrying out a robust impact analysis of the DDC model, including the conditions which must be satisfied for their effectiveness and their robustness; and
- to identify the resource implications of these impact analysis methods, in terms of staff and time required, their broad cost and the likely level of robustness of conclusions.

DDCs do not conduct trials. Otherwise, they have the same powers as other magistrates' courts, including sentencing, dealing with sentence breaches and reviewing Community Orders with drug rehabilitation requirements.<sup>4</sup> Although DRRs form a major part of DDCs work, DDCs have a wider remit and the study was to explore the potential for a robust evaluation of the full role of the DDC. Thus, the designs considered (and their samples) cover all offenders<sup>5</sup> appearing before the DDC, irrespective of whether they receive a DRR or not.

---

3 A wide range of other outcomes have been assessed in other evaluations (e.g. skills, social competence and psychological well-being) (Latimer *et al.*, 2006). These outcomes have normally related to the provision of the drug court under evaluation. Many would be difficult to assess, other than through surveys of offenders with consequent problems of response bias.

4 Drug treatment and testing orders (DTTOs) were introduced by the Crime and Disorder Act 1998. Their main aim was to reduce 'drug-related' offending by using structured treatment to tackle substance misuse. Provisions of the Criminal Justice Act 2003 have since subsumed the English and Welsh DTTO under a new Community Order. These generic penalties now enable the courts to impose a community order with a drug rehabilitation requirement. Essentially, the DRR is equivalent to a DTTO, but with a greater degree of flexibility when it comes to supervision and management: attendance requirements now range from one contact to 15 hours of supervision each week depending on the needs, risks and seriousness of an offence. (The DTTO set a blanket 20-hour a week requirement during the early stages of the order).

5 DDCs deal with defendants who intend to plead guilty, as well as convicted offenders (those sentenced by other courts and those having their Community Orders reviewed). As the former become convicted offenders as soon as their plea is accepted, 'offender' is used to refer to both offenders and to defendants, in relation to the DDC.

Moreover, DRRs are available to all courts, not just DDCs. Thus, an assessment of the impact of DDCs has to take into account the need to measure outcomes over and above those achieved by DRRs. This presents a challenge since evaluations of drug courts in other countries have evaluated both the court process and the sentence as a single package. The availability of DRRs to other courts and to offenders who are not dealt with by DDCs means that this approach is not feasible in England and Wales.

### 1.1.1 'Robust' evaluations

In looking at 'what works' in the prevention of reoffending, the Home Office used two criteria to assess whether or not an evaluation is 'robust'. The first criterion was whether the sample size was large enough to measure the key outcomes the evaluation was addressing, and the second was whether it was based on a design which aimed to provide an estimate of what would have happened in the absence of the intervention (Harper and Chitty, 2005). The terms of reference for this study required the identification of 'robust' evaluation approaches which would enable the reliable identification of the costs and benefits of DDCs. In identifying 'robust' methods, these two criteria were used as the starting point for the study. Thus the key questions were: was it possible to measure the impact drug courts have on offenders who pass through them, and what would the outcomes for those offenders have been if they had been dealt with by the normal courts?

There is an ongoing debate among evaluators about appropriate methods to use in the evaluation of social interventions, particularly those involving a bundle of services, and where the process of delivery, including inter-agency relations, may have an impact on the outcomes achieved. While randomised controlled trials and other 'robust' methods are able to eliminate potential bias on the part of evaluators and practitioners, they only answer the question 'Does it work?' and fail to address the equally important questions "How does it work?" and "Under what circumstances does it work?" (Greenberg & Shroder, 2004; Shadish *et al.*, 1991; Pawson & Tilley, 1997; Meadows, 2007).

Other evaluation methods exist (particularly before and after designs), which have been successfully used in similar circumstances. For example, the National Treatment Outcomes Research Study was a cohort study without a comparison group which measured changes in both offending and drug use (Gossop *et al.*, 2006; 2005; 2003). However, studies of this type do not meet the Home Office definition of 'robustness' because they do not include the measurement of what would have happened in the absence of the intervention, and therefore fall outside the terms of reference of this study. In addition, qualitative studies can develop understanding of why something might be working, which is particularly valuable where an intervention includes a bundle of different practices (HM Treasury, 2007). Thus, even where a 'robust' evaluation approach is not feasible (perhaps because of inadequate numbers of participants), other approaches may be possible.

## 1.2 Dedicated Drug Courts<sup>6</sup>

The drug court model first emerged in the United States during the late 1980s in response to increasing numbers of drug-related court cases and spiralling prison populations (Bean, 2004). Drug courts were introduced in two Scottish sites (Glasgow and Fife) during 2001/02. Building on existing arrangements available through the drug treatment and testing order and DRRs, a pilot model for England was launched in 2005 in Leeds and London. A further four pilots (in Barnsley, Bristol, Cardiff and Salford) started operations in 2009.

The DDC pilots in England and Wales aim to reduce illicit drug use and reoffending amongst drug-misusing offenders who commit low-level crime to fund their addiction. The DDC model introduces a new framework in magistrates' courts for dealing with such offenders. Sessions are set aside in existing magistrates' courts for dedicated panels of magistrates or particular District Judges to sit for sentencing. A drug-misusing offender who is convicted of a low-level 'acquisitive' offence may be referred to the DDC for sentencing. The same District Judge or panel of magistrates who sentence the offender provide continuity in reviewing the offender's progress on Community Orders with a DRR so the sentencer and offender have continuity of contact and dialogue where progress is rewarded and non-compliance sanctioned. The rationale is that continuity of judiciary can help to improve offenders' motivation to stay in treatment and complete their sentence, leading to reduced drug misuse and related offending.

In addition, magistrates and court staff involved in the DDC are provided with specialist training focusing on drug addiction issues and how to work effectively with drug-misusing offenders. The model relies heavily on strong inter-agency working and effective information sharing between the judiciary, court staff, probation staff and treatment providers.

The key elements of DDCs are as follows.

**Exclusivity:** the DDC exclusively handles cases relating to drug-misusing offenders from (and including) sentence, through review to completion and, where possible, breach of their orders.

**Continuity:** the DDC will ensure sustained continuity of magistrates' bench or District Judge throughout the period an offender comes before the DDC.

**Training:** the judiciary and court staff receive additional training on working with drug-misusing offenders and the DDC model.

**Processes:** processes are designed to ensure all necessary information is before the court when required.

---

<sup>6</sup> Information provided by the MoJ.

**Partnership:** the DDCs are designed to ensure effective multidisciplinary working with other criminal justice system agencies and professionals, in particular between the court, the probation service and those providing drug treatment or other related services.

The model only involves changes to business processes and makes use of current services and treatment provision.<sup>7</sup> Pilot DDCs agreed to adhere to a basic framework although decisions on how best to implement it is a local decision, as is the level of judicial continuity to be achieved and other additional elements.

At each site a Local Steering Group meets regularly to agree the local model, review progress and discuss any issues arising in the DDC, helping to ensure the effective operation of the model and quality partnership working. In addition, a National Steering Group oversees developments in the project and resolves any issues escalated from a local level.

### 1.2.1 PSA targets

DDCs feed into a number of departmental Public Service Agreements (PSAs). The focus on tackling 'drug-related' crime by addressing the root causes of offending contributes to *PSA 23: Making communities safer*, while reducing the rate of drug-related offending contributes to *PSA 25: Reduce the harm caused by Alcohol and Drugs*. In addition, this project feeds into *PSA 24: Deliver a more effective, transparent and responsive CJS for victims and the public*, focusing on working effectively with the private and voluntary sector to increase efficiency and quality of service and to foster innovation. The DDC project now forms a part of the Government's drug strategy, *Drugs: Protecting families and communities Action Plan 2008 – 2011*. Developing and embedding problem-solving courts also forms an important commitment in the Business Plans for the Ministry of Justice (MoJ) and Her Majesty's Courts Service (HMCS).

## 1.3 Evaluation approaches

### 1.3.1 Overview of design approaches

To assess the **impact** of DDCs, i.e. whether DDCs make any difference to outcomes, one needs to know not just what happens to offenders who go through the DDCs (whether for sentencing or DRR review), but what would have happened if they had not. The impact is the difference between the two. Thus it is necessary to establish what would have been the outcome in the absence of DDCs, i.e. the counterfactual.

This can be done in a number of ways, the most accurate of which use experimental and quasi-experimental evaluation designs. Experimental designs<sup>8</sup> randomly assign offenders to the treatment group (the DDC) and to the control group (other courts) and compare the outcomes for the two groups. Quasi-experimental designs construct a comparison group in other ways (through selection or statistical methods) and compare outcomes between this group and the treatment group.

---

<sup>7</sup> On the assumption that probation national standards are being consistently applied.

<sup>8</sup> Also known as random assignment and random control trials.



The main possible types of design are:

1. experimental design: randomly assigning relevant offenders to the DDC and to other courts (Random Control Trial);
2. quasi-experimental design of three types:
  - a. 'between-area comparisons', comparing outcomes for offenders dealt with by the DDC with similar offenders in magistrates' courts without a DDC;
  - b. 'within-DDC area comparison', in the magistrates' courts with a DDC, comparing the outcomes for offenders dealt with by the DDC with similar offenders in the same magistrates' court who were dealt with by other courts outside the DDC; and
  - c. 'historical comparison' (or baseline comparisons), in the magistrates' courts with a DDC, comparing offenders dealt with by the DDC with similar offenders in the same magistrates' courts prior to the introduction of the DDC.

Both RCT and quasi-experimental methods could identify the outcomes discussed above. Each main method is described in more detail below.

### **1.3.2 Experimental design (random assignment) (RCT)**

Offenders eligible for DDCs are randomly assigned to the DDC (the treatment group) or another court (the control group). The difference between the average (mean) outcome for the treatment group and for the control group indicates the effect of the DDC.

RCTs are thought by many to be the gold standard of methods (HRDC, 1998). Others are more critical, arguing that quasi-experimental methods can be as effective, suffer fewer practical difficulties and offer more useful impact measures, both generally (Heckman and Hotz, 1989; Heckman and Smith, 1995; Cartwright, 2007) and in the context of criminal justice settings (Hollin, 2008; Hedderman, 2007; Hedderman and Hough, 2005). (RCT can only identify differences in mean outcomes between groups; quasi-experimental methods can examine other differences, such as median differences and the percentage benefiting.) Evaluation of DDCs by RCT may also raise ethical and legal issues (for example, it might provide grounds for judicial appeal).

### **1.3.3 Comparison between magistrates' courts with and without a DDC**

Outcomes (such as rates of reoffending) for those dealt with by the DDC are compared with outcomes for offenders in magistrates' courts without DDC arrangements. Differences are taken to indicate DDC effects.

The difficulty with this method is that differences (or lack of them) may be due to variations between courts or offenders unrelated to the DDCs. Therefore, a major challenge for this design would be to ensure that other differences were dealt with (either statistically or through selection matching of jurisdictions and offenders). Courts and offenders might then be matched on criteria known to affect outcomes.

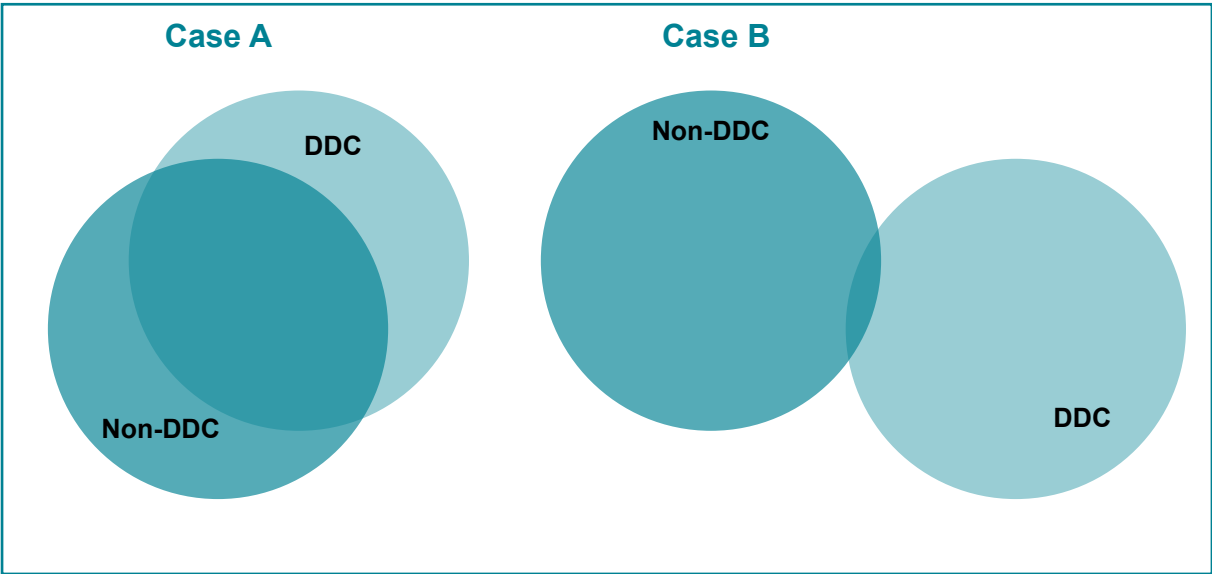
**1.3.4 Comparison within jurisdictions with a DDC**

An alternative to between-area comparison is to compare the outcomes of drug-misusing offenders who have committed a low-level acquisitive offence who went through the DDC with those in the same magistrates’ court who were dealt with by another court. This approach requires DDCs and other courts in the same magistrates’ courts to deal with similar offenders. In DDC areas, all eligible drug-misusing offenders should be referred to the DDC. However, if this model is not always followed, there may be at least some offenders dealt with by non-DDC courts who are similar to those dealt with by the DDC. These would provide a comparator group. Appropriate matching for key characteristics associated with reoffending would be required.

For this approach, an important issue is the degree to which DDC and non-DDC offenders overlap, as it is only those within the overlap who are similar and so can be compared without further adjustment. This is illustrated below.

In both Case A and Case B there is some overlap between the two groups (the shared area of the two intersecting circles). Case A has a substantial overlap and Case B a small overlap, but in neither case is the overlap complete, i.e. the two groups differ. However, without further adjustment, Case A offers the opportunity to evaluate the impact over a wider percentage of participants than does Case B.

**Figure 1.1 Characteristics of those who do and do not go through DDCs within a pilot area**



In principle, it is possible to adjust for this problem of selection bias via statistical methods using the approach developed by Heckman (see Heckman and Hotz 1989; Heckman and Smith, 1995; Heckman, Smith and Clements, 1997) to deal with heterogeneity. These approaches are designed to compensate for the imperfect overlap between the group which has received an intervention and the group which has not (i.e. the situation illustrated in Case A). In circumstances such as those in Case B, where there is little overlap between the two groups, there are insufficient data for these statistical techniques to be reliable.

Unless one can be certain that the pattern looks like Case A rather than Case B, a research design which is based only on comparing offenders who go through the DDC with those who do not will be incapable of generating any meaningful results, and would therefore not be worth undertaking. However, there might be ways of adjusting the design to provide a reliable estimate of impact. For example, it might be possible to boost the comparator sample within the DDC areas (i.e. those who appear to be within the target group but do not go through the DDC) with similar drug-misusing offenders in jurisdictions without a DDC. (Note this latter group provides a boost to allow statistical adjustment, it does not provide the sole comparator.)

### **1.3.5 Historical comparisons**

A third approach would be to compare outcomes within the DDC areas pre- and post-establishment of the DDC. The main potential problem with this lies in the ability to attribute any observed changes in outcomes to the existence of DDCs rather than to other changes in circumstances or practice that have taken place over the same time. It would also be necessary to take account of the fact that the introduction of the DDCs was staggered, and some had previously been operating as 'quasi-DDCs'. This means that the baselines for each court would have to differ to take account of these differences in timing. However, this might enable adjustments for other changes to be taken into account. Although not ideal, comparisons with baseline outcomes are generally feasible. Moreover, the validity of comparators might be able to be further strengthened through including before and after comparisons in other areas without DDCs.

## **1.4 Other design considerations**

Cross-cutting these technical approaches is the wider issue of what should be evaluated.

1. The DDC pilot, i.e. to assess the difference in outcomes between the pilot as a whole and the non-pilot situation as a whole, a 'pilot policy' evaluation. This would assess the impact of extending the pilot across other courts. The approach examines the impact of the pilot against the situation had the pilots not been established. It accepts that implementation in the DDCs may be variable and that DDC-type practices may exist in other courts. This approach would provide a realistic assessment of the changes which would occur if DDCs were rolled out over the country.

2. The DDC model, i.e. to assess differences in outcomes between courts which exemplify the DDC model and those which do not, a 'model evaluation'. This measures the difference between DDC and non-DDC models. It may exclude any DDCs which do not fully implement DDC practice, although it may choose to compare hybrids (e.g. non-DDCs which have some DDC practice). It better controls for the models being compared, but it would not indicate the impact of extending DDCs across the country.

## 1.5 Method

### 1.5.1 Requirements for constructing the counterfactual

Other than for RCT, constructing the counterfactual requires knowledge of potential comparators (i.e. of drug-misusing offenders who commit low-level acquisitive offences who are not dealt with by DDCs), their treatment and their characteristics.

For within-area comparisons, this entails identifying whether, within magistrates' courts with a DDC, all eligible drug-misusing offenders are sentenced and reviewed by the DDC and, if not, whether those dealt with by the DDC differ from those dealt with by other courts. For between-area comparisons, the extent to which practice differs between magistrates' courts with and without a DDC needs to be identified, as well as the characteristics of the offenders being sentenced and supervised. If comparator offenders or areas differ in ways that might affect their outcomes (e.g. age or employment status) from the offenders processed by the DDCs or the DDC areas it is possible to use matching techniques to adjust for these.<sup>9</sup> Their development requires knowledge of the factors which might affect outcomes (as well as data with which to conduct the matching).

The extent to which practice differs between the DDCs and the courts from which comparator offenders are drawn also needs to be known. This is because if they differ little, the impact will be small and so larger sample sizes are required to identify any effect. In addition, for between-area comparisons, it needs to be decided which courts to pick for comparison: those that are dissimilar from DDCs or those which represent standard practice (which may incorporate elements of DDCs). Therefore, the study examined in detail differences in practice in relation to the key DDC characteristics between DDCs and non-DDCs.

### 1.5.2 The researchers' approach

The approach was iterative: considering the small number of quasi-experimental or experimental methods potentially appropriate for evaluating the DDCs and gathering information to indicate which, if any, methods were feasible in practice.

---

<sup>9</sup> The matching can be part of the process of selecting comparators (e.g. an area could be selected on the basis that their labour market and demographic characteristics were similar to those in a particular DDC area. They can also be based on statistical adjustment to compensate for differences. Propensity score matching is one of the more commonly used techniques to do this. Where treatment and comparison groups differ in their demographic characteristics they can still be matched on the basis of a composite score representing the extent to which they meet the criteria for referral to the DDC. See Appendix 4 for more details.

Information was gathered through fieldwork with the six pilot DDCs and with court and probation staff in jurisdictions without a DDC, together with collation of data on DDC activity, a literature review of evidence on the effectiveness of drug courts and a review of existing administrative datasets (e.g. the Drug Interventions Record). Further details of fieldwork follow.

### 1.5.3 The fieldwork

In order to identify the scope for a DDC impact evaluation it was important to gain an insight into how the DDC model was being implemented and how practice differed from non-DDCs. If DDC and non-DDC practice differed little then the impact of the DDCs would be small and so evaluation sample sizes would need to be larger. Moreover, if practice varied between DDCs or between non-DDCs, this could affect the type of evaluation conducted. Therefore, the study examined the way in which the key characteristics of the DDC pilot were implemented in the DDCs and the extent to which these characteristics were implemented in non-DDC magistrates' courts.

#### The DDC fieldwork

In order to develop a more rounded understanding of the scope for measuring the impact of the DDC approach it was important to gain an insight into how the DDC model was being implemented, developed and delivered. As well as completing in-depth structured interviews with representatives from the original DDC pilots in Leeds and West London, the researchers also sought the views and experiences of those working in the extended DDC pilot courts of Barnsley, Bristol, Cardiff and Salford.

Not only did this provide them with an important opportunity to assess, first-hand, the degree of consistency between **stated** DDC policy (as described by process maps and service level agreements made available to the research team) and **actual** practice, it enabled the researchers to develop a greater appreciation of the range of contextual and confounding factors which interact to shape DDC processes and outcomes.

Each of the six DDC co-ordinators in these sites completed an in-depth interview as part of the impact feasibility study. Additionally, in the two longer-established areas, other stakeholders were interviewed about their views and experiences of the DDC. These included magistrates (2), a Deputy Justice Clerk (1), a Legal Team Manager (1), District Judges (2) and probation staff (3). This approach ensured that a range of different backgrounds and perspectives were included (though others, such as the police, drug treatment providers and commissioners, and offenders themselves, were not).<sup>10</sup> Inevitably, some interviewees may have promoted particular insights or expressed views that conflicted with or contradicted those of others. The researchers have aimed in these instances to report and represent the variety of views held.

---

<sup>10</sup> These perspectives may feature more prominently in the DDC process study that was subsequently commissioned by the MoJ.

Most of the interviews were undertaken during May 2009 and all were completed by July 2009. The interviews were recorded for the purposes of transcription. The qualitative data generated from the 15 in-depth interviews were then subject to thematic analysis. This involved assigning codes and labels to the text from the interview transcripts allowing the researchers to organise and collate data under key headings. These were structured according to the core themes defined by the interview schedules.

Developed in consultation with the MoJ and HMCS staff, the in-depth interview schedule focused largely on the core features of the DDC approach:

- ensuring exclusivity and continuity;
- the provision of training;
- inter-agency working and information sharing; and
- monitoring and evaluation.

Rather than addressing the relative benefits or shortcomings associated with these characteristics, the aim of the interviews was to identify those confounding issues which were likely to have an influence on efforts to accurately measure impact.

### **The fieldwork in magistrates' courts without a DDC**

In order to inform the researchers' thinking about the scope for identifying suitable comparison sites, they undertook a short semi-structured telephone survey with representatives from 15 court areas to gauge the extent to which their current practice and existing arrangements were consistent with the key features associated with the DDC approach. When purposively selecting sites they deliberately included some of those areas which already shared characteristics of the DDC approach (e.g. four of the fourteen courts with powers conferred by section 178 of the 2003 Criminal Justice Act to provide for the periodic review of community orders),<sup>11</sup> but excluded those courts which had historically considered themselves to be drug courts (e.g. Wakefield Magistrates' Court). This approach, it was hoped, would facilitate the identification of what might be considered jurisdictions without a DDC and jurisdictions with a quasi-DDC.

Developed in consultation with the MoJ and HMCS staff, the survey focused on processes for:

- identifying and sentencing drug-misusing offenders;
- undertaking DRR reviews and ensuring continuity;

---

<sup>11</sup> In addition to the North Liverpool Community Justice Centre, this also included Birmingham, Bradford, City of Salford, Enfield, Haringey, Kingston-upon-Hull, Leicester, Merthyr Tydfil, Nottingham, Plymouth, South Western, Stratford and Teesside Magistrates' Courts.

- listing DRR cases in dedicated court sessions for sentencing and/or reviews;
- the provision of training;
- inter-agency working and information sharing; and
- monitoring and evaluation.

Data were gathered via telephone interviews from 20 respondents between May and July 2009, representing court (11) and probation (9) perspectives.<sup>12</sup> In seeking the views of probation staff serving five of these courts the intention was to assess the degree of consistency between the accounts being offered by the two key stakeholder groups in the same locations. There were some important inconsistencies observed in the accounts given by court and probation interviewees from the same areas. The most notable discrepancies arose in relation to the extent to which the lower courts unintentionally provided continuity of judiciary and listed DRR cases together for breach and review hearings, the effectiveness of existing inter-agency working and information-sharing processes, and the monitoring of training activity.

Court respondents in the 11 areas included the Deputy Justices' Clerk (6), the Clerk to the Justices (1), the Legal Advisor (2), the Principal Legal Advisor (1) and the Legal Team Manager (1). Probation interviewees included an Area Manager (1), a Deputy Director of Operations (1), District Managers (2), senior probation officers (4) and a probation court officer (1).

The magistrates' courts surveyed as part of this work were:

Highbury	Hull
Leicester	Luton
Norwich	Nottingham
Sheffield	Southampton
Teesside	Vale of Glamorgan
Wolverhampton	

The probation teams questioned included:

City of Manchester	Highbury
Maidstone	Newcastle
Nottingham	Plymouth
Sheffield	Vale of Glamorgan
Wolverhampton	

<sup>12</sup> Despite the researchers' efforts it was not possible to secure interviews in two jurisdictions without a DDC with representatives from probation in the time available.

Responses from the survey were entered onto an MS Excel spreadsheet in preparation for analysis.

## **1.6 Report structure**

The next chapter presents previous evidence on the impact of drug courts. The following two chapters describe practice in the DDCs (Chapter 3) and in a sample of jurisdictions without a DDC (Chapter 4). Chapters 5 to 7 discuss whether the key requirements for a robust impact evaluation can be met: in Chapter 5 whether comparators might be identified, in Chapter 6 the sample size required to detect an effect and in Chapter 7 the data required. Chapter 8 argues that two approaches were not feasible and discusses details of the remaining two approaches. The final chapter reports the researchers' conclusions.



## 2 The effect of drug courts

### 2.1 Evidence on the impact of drug courts

The literature review focused on evaluations of interventions that were described as being drug courts. Most drug courts in other countries combine a specialised judicial process, exclusive sentencing powers and the exclusive or enhanced ability to access drug treatment services. Evidence on the potential impact of drug courts comes from several countries: the USA (Government Accountability Office (GAO), 2005), Australia (Weatherburn *et al.*, 2008; Freeman, 2003; Wundersitz, 2007), Canada (Fischer, 2003), Puerto Rico (Wenzel *et al.*, 2001), the Republic of Ireland (Farrell, 2002), England (Matrix Knowledge Group, 2008) and Scotland (McIvor *et al.*, 2003), together with meta-analyses by Latimer *et al.* (2006) and Wilson *et al.* (2006).

Much of the early research has been criticised on methodological grounds. The GAO (2005), in its systematic review, identified 117 evaluations which examined recidivism, substance use relapse or sentence completion, but considered only 27 as robust (based on adequate sample sizes and the use of a comparison group). These 27 studies included five RCTs.

Most of the positive outcomes that have been observed (reoffending and drug use) are confined to periods while offenders are retained on programmes (Belenko, 1998, 1999, 2001; Huddleston *et al.*, 2005; GAO, 2005). In general this is because evaluations have not examined outcomes over longer periods. Consequently, little can be said with any certainty about the effectiveness of drug courts over the longer term in tackling substance misuse and improving individual and social functioning (Anderson, 2001; Belenko, 2002).

### 2.2 Impact of drug courts on recidivism

Two systematic reviews (GAO, 2005; Wilson *et al.*, 2006) and a meta-analysis (Latimer *et al.*, 2006) found evidence of reduced recidivism. This ranged from eight to 21 percentage points after one year (GAO, 2005), 11 percentage points<sup>13</sup> (Latimer *et al.*, 2006) and 24%<sup>14</sup> (Wilson *et al.*, 2006). An RCT evaluation of the New South Wales drug court found a 17% reduction in the reoffending rate (a five percentage point reduction) (Weatherburn *et al.*, 2008).

Where longer-term research follow-up took place, the differences in reoffending rates were larger (i.e. the drug courts had a long-term effect on recidivism). This means that in order to identify the impact of the DDC programme it is important to evaluate the effects over several years and not just over the treatment period and shortly afterwards.

---

13 This rate includes programme dropouts and non-completers, who should be included to gain a true assessment of the programme. Some studies exclude this group. Excluding this group, the improvement in recidivism rises to 14 percentage points (Latimer *et al.*, 2006).

14 This study looked at the reduction as a proportion rather than the percentage point approach used in the other studies.

Evidence on differences in impact by treatment, offences and offender characteristics has been problematic (GAO, 2005). Nevertheless, Latimer *et al.* (2006) found that programmes lasting 12 to 18 months were more effective than shorter or longer programmes, hypothesising that shorter programmes had too little time to take effect and longer programmes might lead to 'treatment fatigue'.<sup>15</sup> If DDCs lead to magistrates and District Judges better judging the appropriate length of DRRs, then this suggests a way in which the impact of DDCs may be higher.

Differences in how drug courts have been implemented and delivered have been shown to impact upon their effectiveness (Bouffard and Smith, 2005; Goldkamp *et al.*, 2001). Outcomes also vary according to the offender groups targeted and their characteristics, for example, by prior convictions and type of offence (Weatherburn *et al.*, 2008). There are also variations in outcomes depending on the treatment approaches used. For example,<sup>\*</sup> in contrast to court-based treatment in the UK, Australia and other jurisdictions, the US drug court model targets low-level or first-time offenders. This is likely to result in outcomes which are better than could be expected with more persistent offenders or those who have committed more serious offences. The vast majority also operate abstinence-based treatment philosophies (Bean, 2004); historically, US drug courts have made little use of <sup>\*</sup> cognitive behavioural therapy approaches (Bouffard and Taxman, 2004) or methadone maintenance (Peyton and Gossweiler, 2001). US drug courts appear to have enjoyed a greater degree of success in engaging and retaining offenders. However, as with court-ordered treatment completion rates in Britain and elsewhere, there is considerable variation<sup>\*\*\*</sup> in US drug court graduation rates: from 27 to 66% (GAO, 2005: 62).

\*\*

Measured recidivism rates differed by evaluation method. Latimer *et al.* (2006) pointed out that the RCT studies showed a smaller impact on recidivism (and one found no statistically significant effect).<sup>16</sup> They considered this might be due to inadequate adjustment for selection effects in non-RCT evaluations. In other words, if those going into drug court programmes are better motivated and have better social support, they would be more likely to have better outcomes than other offenders even in the absence of the drug courts. This emphasises the importance of ensuring that any matching process is able to take into account factors such as family circumstances that influence motivation.

### 2.3 Impact of drug courts on substance misuse

The evidence of the impact of drug courts on substance misuse is mixed and varies with the method of measurement (drug testing or self-report) (GAO, 2005). In some cases, there was evidence of reduced drug use whilst on the programme. However, no studies found evidence of reduced drug use post-programme (i.e. there was no statistically significant difference in

15 At the same time there is evidence to suggest that many of the benefits from drug treatment, whether 'coerced' or 'voluntary', are accrued during the early stages of an intervention (McSweeney *et al.*, 2007).

16 Throughout this report a difference is treated as statistically significant if there is a 5% or less probability of a result having come about by chance. This is sometimes referred to as the 95% confidence interval or  $p < 0.05$ .

drug use between treatment and comparison groups in post-programme follow-up). In part this reflects relatively short follow-up periods. It is also likely that studies that included testing may have found similar rates of positive tests, but this might still reflect lower frequency or intensity of drug use given the persistence of some drugs.

## 2.4 Impact of drug courts on sentence completion and compliance

The researchers found no evidence of the effect of drug courts on sentence completion or compliance. However, studies of DTTOs provide some evidence on the scale of effect which the DDCs might expect to achieve.

During 2004/05 completion rates for DTTOs – which preceded DRR arrangements – ranged from 10% in North Yorkshire to 52% in Dyfed/Powys (National Probation Service, 2005: 17). The initial DDC process evaluation only reported a DRR completion rate for Leeds (26%) (Matrix Knowledge Group, 2008). This was considerably lower than the national DRR average (44%) during this period (National Probation Service, 2007: 11). During 2008/09 the DRR completion rate in the west London DDC boroughs of Hammersmith and Fulham (42%) and Kensington and Chelsea (42%) were below the London average (50%). This is important given that the appropriateness of the treatment being delivered as part of such orders had been linked to improved retention and completion rates (Turnbull *et al.*, 2000). Completion of these and other substance misuse programmes is associated with reduced rates of reconviction (Hough *et al.*, 2003; Hollis, 2007).

Evidence from the evaluation of the Community Justice Initiatives (which seek to ensure judicial oversight of offenders' progress post-sentence) in north Liverpool and Salford suggest caution should be applied to the use of breach as a measure of sentence compliance. Liverpool and Salford had significantly higher breach rates<sup>17</sup> than a comparison area (Manchester). However, there were no statistically significant differences in rates of reoffending within the first year of the initiatives. This has led researchers to conclude that offender managers in the Community Justice Courts may be applying guidance on breaches more rigorously than other areas (Jolliffe and Farrington, 2009). The same may occur in DDCs, confounding the use of breach as a success measure.

## 2.5 Implications for the expected impact of DDCs

It is possible that the effects of drug courts are statistically significant, but small.<sup>18</sup> Therefore, in terms of practical effectiveness (and cost-effectiveness) they may not be of great importance. Irrespective of the robustness of the evaluation, the measured impact varies. This is unsurprising: drug courts are not homogeneous. They differ from non-drug courts in a number of important ways, including the type of drug treatment offered and in their eligibility

<sup>17</sup> Significantly different at the 5% level.

<sup>18</sup> In statistical terms if the standard errors are sufficiently small it would be possible to detect smaller differences in recidivism than the five percentage points that the designs considered here are based on. The issue is whether a reduction in recidivism of say two percentage points would prove to be cost-effective.

criteria (e.g. many exclude those convicted of crimes of violence; some require guilty pleas). For evaluations, these differences are crucial: impact evaluations measure difference in outcomes between pairs of treatments; the more the treatments differ, the greater the potential effect on outcomes.

This is a problem for identifying the likely impact of DDCs. Previous evaluations do not always describe the treatment for both the treatment and comparator/control groups. However, it is apparent that many drug courts in other countries combine changes in court processes and procedures with different drug treatment programmes. It seems likely that the accessibility and suitability of drug treatment programmes, rather than the court procedures and processes may account for some or all of the better outcomes from drug courts.<sup>19</sup>

The DDC pilots in England and Wales largely change court procedures and processes. Although these changes may have some effect on the type of drug treatment each individual receives, they do not provide additional or enhanced treatment options.<sup>20</sup> A major element of the English and Welsh approach, the DRR, is available to all courts in England and Wales. This means that any evaluation would need to measure the impact of DDCs over and above any impact of DRRs. This is because it is likely that at least a minority, and possibly a large majority of comparators would have received a DRR. One would therefore expect the impact of DDCs to be relatively small compared with some of the impacts found elsewhere.

---

19 The design of most drug court interventions is based on the premise that all three elements: process, sentence and treatment, all contribute towards the outcomes. It is possible that sentencing and treatment do not contribute at all, but this is unlikely.

20 Although a District Judge in one site did negotiate access to detoxification on behalf of DRR offenders.

## 3 DDCs: policy and practice

### 3.1 Introduction

Fieldwork was conducted to identify whether the ways in which the DDCs operated affected possible evaluation designs. This focused on the way they addressed the key DDC characteristics (Section 1.2). Similar fieldwork was conducted in magistrates' courts without a DDC. The aim was to identify whether DDC practice differed from courts without a DDC (which has implications for identifying comparator courts and the expected size of effect); whether practice differed between DDCs (which has implications for the choice between a model or a pilot evaluation and the expected size of effect); and the extent to which practice in DDCs had changed once they had acquired formal DDC status. This chapter presents key findings on DDC practice and the next on a sample of courts without a DDC. The rest of the report draws on these chapters in its consideration of feasible designs.

The DDC pilot sites were asked to use a basic framework for sentencing drug-misusing offenders, review of any DRRs and breach (where possible), but to determine their own local model in agreement with local partners and stakeholders.

The researchers found that, with the possible exception of continuity, none of the key characteristics of the DDC approach were consistently defined across the six DDC pilot sites or being systematically measured and monitored. For example, the target groups differed. One DDC included cannabis users and growers, whereas the others only included Class A drug users.

More information on the researchers' findings is given in Appendix 1.

### 3.2 Exclusivity, continuity and dedicated sessions

In relation to ensuring exclusivity and continuity of the judiciary the researchers found the following.

- All the pilot DDCs were offering a degree of exclusivity and continuity – in terms of reviewing DRR cases – long before they acquired their formal DDC status.
- Not all eligible drug-misusing defendants are exposed to the DDCs or sentenced by them.
- A number of offenders sentenced by the DDCs lived outside the DDC catchment area and had no ongoing contact with them post-sentence. This means that the court process for these offenders would in practice be no different than if they had been sentenced in an ordinary magistrate's court. They would therefore 'contaminate' the DDC population.
- There is not a single DDC model in terms of style of operation and approach to the review process. If either of these have an influence on outcomes it may be important for any evaluation to take these differences into account.

### **3.3 Training**

Regarding the provision of training for DDC staff and judiciary the following was found.

- The nature and extent of training varied between DDCs; training was tailored to address locally identified gaps in skills and knowledge.
- The extent to which the impact of DDC training and levels of satisfaction with it were monitored varied across the sites – from no formal monitoring to the regular use of evaluation sheets.
- There was an acknowledgement from some that training could be a more prominent and regular aspect of the DDC.

### **3.4 Inter-agency working and information sharing**

Respondents across the six DDCs were unanimous in believing that the DDC had delivered improvements to inter-agency working and information sharing. This did not mean that respondents believed inter-agency working could not be improved. Examples for improvement included: more involvement in arrangements for throughcare and aftercare following completion of a DRR; and closer ties with mental health practitioners and specialists working with powder cocaine users.

Finally, the target groups differed. One DDC included cannabis users and growers, whereas the others only included Class A drug users. This variation in the characteristics of those going through DDCs has implications for the selection of a comparator sample. If all comparators are Class A drug users this might result in misleading conclusions if some of the DDC offenders are cannabis users.

## 4 Practice in jurisdictions without a DDC

### 4.1 Introduction

Fieldwork with 20 respondents from 15 courts areas was conducted to identify the extent to which DDC and non-DDC practice differed, as this affects evaluation possibilities. This focused on the extent to which courts in jurisdictions without a DDC exhibited the key DDC characteristics and other initiatives which could confound a between-area evaluation and data collection.

More information on the researchers' findings is given in Appendix 2.

### 4.2 Exclusivity, continuity and dedicated sessions

The practice in the magistrates' courts without a DDC differed from the DDC model in that the former, reportedly, did not offer exclusivity, continuity or dedicated sessions.

In none of the 15 magistrates' courts without a DDC did the courts or probation report collectively listing, sentencing or handling cases involving drug-misusing offenders in dedicated sessions (excluding DRR reviews). There were a varied set of arrangements in place for monitoring and reviewing progress and compliance with DRRs imposed by the courts.

None of the court or probation-based respondents reported that there was an intention to provide continuity of **magistrate** from the point of DRR sentence through to completion, and if necessary (and where possible) breach hearings. However, in three jurisdictions courts or probation commented that a greater degree of continuity was offered by District Judges.

### 4.3 Training

It was unclear whether training differed from that in the DDCs. All 20 respondents across the 15 jurisdictions without a DDC reported that training had been provided for members of the Local Justice Area bench and other court staff to improve their awareness and understanding of drug-related offending or addiction issues. This training varied in its form, frequency and intensity.

### 4.4 Inter-agency working and information sharing

It was unclear whether inter-agency working and information sharing differed from that in the DDCs.

Most of the 20 respondents described the inter-agency working and information-sharing processes between the court, local probation and other agencies providing interventions to drug-misusing offenders as either very (9) or quite (10) effective. This may have been due to other recent initiatives or developments (in addition to DRR arrangements) aimed at drug-misusing offenders which had contributed towards improved inter-agency working and information-sharing processes (reported by 17 respondents). Examples included the drug interventions programme (DIP) and priority and other prolific offender schemes (PPOs).

## 4.5 Data collection and monitoring

Almost all (19) jurisdictions without a DDC reported that, on average, they routinely collated data on two-thirds (30) of the 44 variables recorded by the six DDCs as part of the central DDC pilot monitoring; see Table 4.1. This could be useful for sample identification in a between-area comparison design.

However, for identifying comparator courts in relation to their practices, there would seem to be little scope for accurately defining and measuring most of the key features associated with the DDC model (e.g. inter-agency working, information sharing and training) in jurisdictions without a DDC using existing systems and data sources.

**Table 4.1 Data routinely collected by the six pilot DDC**

Unique ID	First name	Surname
D.O.B.	Gender	Ethnicity
PNCID	Number of previous convictions	Number of previous offences
Educational qualifications	Court file reference	Section 178 review
Offence details	Date of offence	DRR hearing date
DRR hearing attendance	DRR hearing outcomes	DRR expiry date
Plea	Date of guilty plea/convicted	Sentence
Date of sentence	Community order details	Chairman/DJ
Magistrate 2	Magistrate 3	Legal advisor
Drug misuse details	Drug test date	Drug test result
Breach of order	Date of initial breach hearing	Date of other breach hearing
Breach outcome	Name that dealt with the breach	OASys score (Tier/Risk level)



## 5 Key issues: comparators

### 5.1 Introduction

Key to the evaluation is whether, in principle, a comparator may be identified. This chapter uses the evidence of the operation of the DDCs and other courts to discuss which, if any, of the four designs might yield a comparator.

### 5.2 The policy counterfactual

Key characteristics of the DDC model have been described above. Chapters 3 and 4 showed that, in practice:

- there were differences in implementation across DDC sites so that there is no 'single' DDC model; and
- there were similarities in practice in inter-agency working, information sharing and training between DDCs and courts in jurisdictions without a DDC.

The implications are that a 'model' evaluation would reveal little about the impact of extending the pilot across the country. It could identify the impact if the pilot policy were changed to ensure a more prescriptive and uniform implementation of the key elements of DDCs. A 'pilot policy' evaluation, by contrast, would identify the impact of both the pilot and the policy were it to be extended further.

### 5.3 Experimental design (random assignment)

The fieldwork findings are generally of little relevance in relation to the feasibility of identifying a control for an RCT: the experiment itself creates control and treatment groups. However, the researchers had been concerned about potential ethical or legal obstacles<sup>21</sup> to random assignment (that the treatment received by offenders on sentencing and subsequently should not be random). Not all cases involving drug-misusing defendants within a given DDC's jurisdiction were dealt with by the DDC. In some cases this was due to capacity constraints and cases were dealt with by the DDC on a 'first come, first served' basis in court listings. From an ethical or judicial point of view, this is no better than random assignment and so suggests that randomisation should be acceptable. However, all eligible DRRs, even if passed outside the DDC, were referred to the DDC for subsequent review. Thus, even if random assignment proved to be feasible, it would be difficult, if not impossible, to control the subsequent treatment of offenders in the control group who had received DRRs, which would contaminate the experiment. Therefore, the argument that current practice was no different ethically or judicially than randomisation does not hold.

---

<sup>21</sup> For example, one issue might be whether offenders who have been randomly assigned received a sentence which was apparently more severe than the one they would have received otherwise, and this might give rise to grounds to appeal.

## 5.4 Between-area comparison

There are similarities in practice between DDCs and courts in areas without a DDC and there is a lack of consistency across the DDCs. This makes a comparison between DDC areas and other areas more complicated. However, the existence of **some** differences between DDCs and courts in other areas provides scope for conducting the evaluation through between-area comparisons. This could be an evaluation either of the pilot policy as a whole or of the DDC models themselves.

Adjustments would need to be made for differences, other than the DDC, between jurisdictions which might affect outcomes. This includes, for example, differences in the operation and delivery of the DRR, drug treatment services, socio-economic factors, and access to supporting services such as housing and education.

## 5.5 Within-DDC area comparison

The advantage of within-DDC area comparisons is that offenders are presented with the same policy, social and service environment, i.e. there is no need to adjust for these.

Within-DDC area comparisons are only feasible if the DDC does not deal with all cases within its magistrates' courts and if the characteristics of those not dealt with by the DDC include the same type of offenders as are dealt with by the DDC.

In some magistrates' courts with a DDC, not all appropriate offenders were dealt with by the DDC. However, most relevant offenders were dealt with by the DDC and even where other courts sentenced DDC-eligible offenders their DRR reviews were dealt with by the DDC. This means that within DDC areas it is not possible to ensure that comparator offenders with DRRs are never dealt with by the DDC. Therefore, a within-DDC area comparison would be restricted to offenders not receiving a DRR, an unacceptable restriction for an evaluation of DDCs. This approach is not considered further.

## 5.6 Historical comparisons

Several DDCs had been operating as quasi-DDCs prior to the pilot and some had done so for many years. Therefore, historical comparisons would need to identify a baseline prior to quasi-DDC operation. However, no data were available to identify when this might have been and it did not appear feasible to identify this with much accuracy through discussion. (Moreover, the evaluation would need to adjust for other changes (e.g. in the criminal justice system, drug use and drug treatment) which had taken place between the baseline and DDC operation. Given the length of time, this adequacy with which this could be done is doubtful. For these reasons this approach is not considered further.

## 6 Key issue: sample size

### 6.1 Introduction

A key consideration of the design is the size of the sample required to ensure that, if the DDC has an impact, that impact is identified and detected with statistical significance. The smaller the effect of the DDC, the larger the sample size required to detect any effect. Therefore, the likely size of effect and the sample size needed to detect this effect must be estimated. To do this, the researchers drew on the evidence of the impact of other drug courts (reported in Chapter 2), together with actual rates of recidivism (outside DDCs). As the main evidence of the impact of drug courts is on recidivism, their estimates are for measuring the impact of recidivism.

One of the limitations on the sample size is the number of offenders who pass through DDCs each year. In a typical month, between 70 and 90 people pass through the six DDCs in England and Wales. In theory, therefore, an evaluation could draw on around 500 cases if six months' worth of inflow through the system was followed up. However, at least one DDC is dealing with offenders who fall outside the criteria used by the other DDCs (e.g. cannabis defendants). Because of the substantial difference in target groups and the otherwise general focus on Class A drug users, it may be that either this DDC or its offenders who were not Class A drug users should be excluded from the evaluation. Exclusion of this site would result in the loss of up to 400 offenders over a six-month period, and 800 over 12 months. This is the throughput assumed in the remainder of the report. (Including all offenders dealt with by this DDC would only alter the time for sample build-up and hence the total time for the study.)

### 6.2 Minimum detectable effect and sample size

One-year reconviction rates – the new standard measure adopted by the Ministry of Justice – for DTTOs in England and Wales fell from 79% in 2002 to 70% in 2005 (there were reductions in the frequency of offending during this period too) (Ministry of Justice 2008: 21). There are, however, no published data on regional differences in DTTO reconviction rates.

The 12-month reconviction rate for the DTTO cohort in 2005 is lower than the recent figures reported for both users accessing mainstream treatment as part of the National Treatment Outcome Research Study (74%) (Gossop *et al.*, 2006) and subsamples of drug-misusing offenders supervised by the prison and probation services (74-75%) (Howard, 2006; May, Sharma and Stewart, 2008). In fact, the rate at which drug-misusing offenders are reconvicted appears to have changed little since the early 1990s (May, 1999). The researchers are not aware of any published statistics on rates of reconviction for DRRs.

A reduction in reoffending of 11 percentage points, as suggested by the evidence of Latimer *et al.* (2006) would suggest a reduction in reoffending to around 62%. A reduction on the

scale found by Wilson *et al.* (2006) would suggest a much greater fall in reoffending rates to around 53%. Weatherburn *et al.* (2008) found that 63% of those who went through New South Wales drug courts reoffended, compared with 68% of the control group.

In general, larger sample sizes are more reliable than smaller ones, but there is clearly a trade-off to be made in terms of cost and practicality. As a general principle, larger sample sizes are needed to detect smaller effects, while smaller sample sizes are sufficient to detect larger effects. The trade-off is formalised by a statistical power calculation. The calculation uses standard formulae to estimate the sample size necessary to detect a particular effect, or conversely the size of effect that could be reliably detected with a particular sample size.

A statistical power calculation reveals that in a random assignment evaluation, a sample size of 433 in each group would be sufficient to detect differences on the scale suggested by Latimer *et al.* (2006) (i.e. the study towards the lower end, but not the smallest in the reported reviews) with a 95% confidence interval and 80% statistical power.<sup>22</sup> Thus, if these are the scale of effects likely to be found in DDCs in England and Wales the number of offenders processed each year by the DDCs would be sufficient to detect them.

A comparison group (either within-DDC areas or between areas) would be expected to result in larger measurement errors than a random assignment would. This would mean a larger sample size would be necessary. Because the scale of the errors is unknown it is not possible to perform an exact statistical power calculation. However, it is likely that sample sizes of 600 in each group would be sufficient to ensure that differences were measured with accuracy.

However, as discussed above, a further challenge is that it is possible that some of the effects attributable to drug courts in international evaluations are actually due to the fact that drug courts offer a route into drug treatment (or more effective treatment) rather than to the court process itself. This might suggest that one can reasonably expect differences in reconviction rates between DDCs and non-DDCs to be fairly modest. Assuming a five percentage point difference in reconviction rates between DDC (65%) and non-DDC (70%) groups (i.e. a similar pattern to that found in New South Wales), a combined sample of over 2,800 (with 80% power) would be required to detect such a difference (at 95% confidence) – 1,416 in both experimental and control groups. Given current DDC throughputs it would take four years to assemble such a sample.

A quasi-experimental design would require a larger sample, both because of increased measurement errors and also because statistical adjustment is likely to be necessary to compensate for any differences in characteristics between the treatment group and the

---

22 See Footnote 16 for an explanation of the 95% confidence interval. Statistical power is the probability that a sample of a given size will be able to detect an effect of the expected size. It is generally regarded as good practice to use 80% (or 0.8) as the minimum statistical power when designing research studies.

comparison group. For these reasons, it may be necessary to have up to 1,600 in each group. Given current DDC throughputs it would take two years to assemble such a sample. However, offenders already dealt with by the DDC could form part of the sample and so this two-year period is already underway.

## 7 Key issues: data

### 7.1 Introduction

Data on the following are required or desirable:

- outcomes (recidivism,<sup>23</sup> drug-misuse, sentence completion and sentence compliance);
- participant characteristics affecting recidivism, drug-misuse and sentence completion (for matching in quasi-experimental design);
- local characteristics which may affect outcomes (for between-area comparison design);
- court practice (for model evaluation design – advisable for pilot evaluation design and RCT).

For an RCT, only outcome data are required. However, it is advisable to check that the characteristics of those in the treatment and control groups are similar and so data on offender characteristics are required.

In quasi-experimental evaluations, the treatment and comparator group are matched to ensure they are similar in terms of factors which affect outcomes (e.g. recidivism, drug-misuse and sentence completion). Extensive work has been conducted by the Home Office and now the Ministry of Justice to identify predictors of reoffending (see, for example, Cunliffe and Shepherd, 2007).

For between-area comparisons, adjustments would need to be made for differences, other than the DDC, between areas which might affect outcomes (Section 5.4), e.g. operation and delivery of the DRR, drug treatment services, socio-economic factors and access to supporting services.

### 7.2 Main sources of data

#### 7.2.1 Administrative data

Data sources for administrative data are given in Table 7.2Table 7.2.

For outcomes, data are comprehensive for recidivism. Sentence completion data cover completion of Community Orders but not failure to pay fines. Sentence compliance

<sup>23</sup> Recidivism might be measured as reconviction within a set period, period to reconviction and number of convictions within a set period. It might also measure the severity of offending.

**Table 7.2 Administrative data sources**

Data source	Individual identifier	Relevant group covered	Relevant data	Use		
				outcomes	treatment/comparator matching	comparator identification
HOPNC	PNC numbers	convicted offenders since 1995	recorded convictions, reprimands, cautions and warnings; demographics	,	,	,
Drug Intervention Record, DIR	Initials, date of birth, gender	all testing positive for heroin or cocaine on arrest or charge	demographics, current substance use	,	,	,
Police test on arrest data	Initials, date of birth, gender	DIP intensive areas only: all testing positive for heroin or cocaine on arrest or charge	demographics, current substance use	,	,	,
OASys	PNC numbers	Many offenders assessed by the correctional services	current and previous drug-misuse	,	,	,
Treatment Outcomes Profile	Initials, date of birth, gender	some people receiving NHS treatment for drugs misuse in England	drug use over each of the preceding four weeks	,	,	
probation case management system e.g. Delius	PNC numbers	All dealt with by probation	Community Order completion; attendance at court and probation appointments; recorded breaches	,	,	
Interim Accredited Programmes Software (IAPS)	Initials, date of birth, gender		Community Order completion	,	,	
Enforcement Tracker	Initials, date of birth, gender		Community Order completion; attendance at court and probation appointments; recorded breaches	,		

a Plus initials, date of birth, gender.

b For all offenders, an estimate of the likelihood of reoffending is made using the Offender Group Reconviction Scale (OGRS). This score would be useful for matching (though it fails to take into account dynamic risk factors such as drug-misuse).

data cover compliance with Community Orders (e.g. attendance at court and probation appointments and numbers of recorded breaches). Data on drug misuse outcomes are problematic. Subsequent drug-misuse is not fully covered by any dataset and would require searching and matching probation case management systems, the Drug Intervention Record and Treatment Outcomes Profile to capture drug outcomes. However, outcomes would only be identifiable at the time when an individual is undergoing treatment or is tested on arrest.

For matching treatment and comparison samples (so that offenders in both groups are similar), data are good for recidivism (Home Office Police National Computer) . Matching in respect of the likelihood of continued substance-misuse would be less rigorous, due to data limitations. However, data on the nature and extent of drug- misuse could be derived from DIR forms with reasonable accuracy and consistency.<sup>24</sup> For between-area comparisons, adjustments for special programmes and some socio-economic factors are relatively simple using existing data sources (e.g. probation datasets such as OASys and IAPS), although the number of adjustments may be problematic leading to loss of statistical power. Differences in the nature and quality of service provision, including DRR drug treatment would be extremely difficult to factor in, although important data could be gleaned from existing sources such as the National Drug Treatment Monitoring System (NDTMS) and/or the DIR.

Data sources would need to be matched and not all have a unique individual identifier (viz the PNC number). However, matching should be feasible using attributer codes (initials, date of birth and gender) (Skodbo *et al.*, 2007; Millar *et al.*, 2008). Other variables, such as date of drug test, could also be used to reduce duplicate matches against the PNC. Matching is required to provide outcome measures for sentence completion (probation case management systems and IAPS), sentence compliance (probation case management systems and Enforcement Tracker) and for drug-misuse (probation case management systems, test on arrest data, TOP and DIR). For quasi-experimental designs, it would be necessary to produce a dataset containing both sample matching data and outcomes.

Not all data are available in an analytically friendly form and would require sorting and structuring.

### 7.2.2 Other data

For a model evaluation, comparison areas would be selected for their difference from the DDC in their court practice (e.g. lack of consistency in magistracy hearing drugs cases). This would require qualitative research to identify appropriate comparators.

---

24 Note the drug information is good for matching and poor for measuring outcomes. This is because the data for matching are[?] drug use when appearing before the DDC (when DIR data are available), but data for outcomes relate to a later period, when the offender may no longer be receiving treatment. It is only those with longer treatment or who subsequently receive treatment who will be captured on these data sets for outcomes. [should this be p.31? – if so change Contents page]

The study considered the use of survey data to complement the administrative outcome data. However, it is particularly problematic to collect reliable information on illegal drug use (or any form of unlawful behaviour) through surveys. Those who volunteer to take part in such surveys tend to be those who are less involved in problem behaviour, so that such surveys give a more optimistic picture of outcomes than is really the case. Given the more reliable coverage of the administrative data, the survey approach was rejected.

## 8 Potential impact evaluation designs

### 8.1 Potential designs

The evidence above suggests that two designs are unfeasible.

- Historic comparisons: the DDCs had previously operated as quasi-DDCs and it would be impossible to identify the comparison period (i.e. when they had been operating as a standard court).
- Within-DDC area comparisons: all those receiving DRRs were reviewed by the DDC and so no offenders receiving DRRs could have been included in the evaluation.

The following designs cannot be entirely ruled out, but carry substantial risks.

- An experimental design, if it is considered legally and ethically acceptable.
- A between-areas comparison design.

This chapter examines these last two designs in more detail.

### 8.2 RCT design

#### 8.2.1 Introduction

The basic RCT design is simple: in the jurisdictions with a DDC, those allocated to the DDC by the courts are randomly assigned to the treatment group or to the control.<sup>25</sup> The impact of the DDC is the difference in average (mean) outcome between the two groups. RCTs have the advantage of being analytically simple and the results are widely accepted. They provide a clear answer to the question: 'Does the treatment make a difference to the outcome?'

The challenge is implementation. Two issues in particular have represented challenges to random assignment experiments: ensuring that subjects are **randomly** assigned and that the treatment and control groups do receive different treatment. Below, implementation for a DDC evaluation is discussed.

---

<sup>25</sup> Assessment covering selection criteria would: identify those eligible under imposed selection criteria and randomly allocate these to the treatment and control groups; separately, courts would judge which offenders should go to the DDCs; those in the treatment group judged by the courts eligible for the DDC would go to the DDC, all others would not.



### 8.2.2 The process of randomisation

The practicalities of randomisation need to be considered. This is informed by the court process, which is described first.

The point at which randomisation might take place depends on the processes for allocation to courts, as it is this allocation (to a DDC or other court) which needs to be randomised. Offenders are entering DDCs via several routes:

- those admitting guilt will be bailed, in the police station, to appear at court and may be referred to a DDC for their hearing or not;
- those not bailed to the DDC may then be remanded by the non-DDC court to the DDC for sentencing or, if given a DRR, for DRR reviews;
- those not admitting guilt will be tried outside the DDC and, at the hearing, either sentenced or remanded to the DDC for sentencing.

The fieldwork did not investigate who allocated individuals to DDCs or other courts. However, it is apparent that a range of people are involved in setting a date for hearings (or the allocation to the DDC): the court staff and magistrates (at hearings) and, possibly, the police and the Crown Prosecution Service (CPS). Given that DDCs do not sit every day (and so offenders cannot be randomly assigned once their hearing date has been set), random assignment to the DDC must take place prior to the individual being told the date of their hearing. The system would also need to allow for those randomly allocated to the DDC to continue to be allocated to the DDC for further relevant offences.

It would be extremely difficult to establish a system of random assignment which could be operated effectively across these different environments. The decision to assign to the DDC has first to be made (by the magistrate, the court etc.). Randomisation then takes place and the offender is either allocated to the DDC or not. It may also be seen as unacceptable interference with the judicial process for a randomisation check to be made at this point (e.g. when magistrates have retired and decided guilt and decided to remand for sentencing, they may be unhappy to have a check made to randomise, prior to announcing the verdict).

In these circumstances it would be essential that randomisation is done on demand by independent, experienced individuals. However, there would be difficulties of enforcement as the randomiser would not be able to alter or get any misallocations altered.

### 8.2.3 Compliance and commitment

There is likely to be resistance to an RCT. Farrington (2003), who is broadly supportive of RCTs in criminal justice settings, documents the controversy among researchers and practitioners following a successful RCT undertaken in Leeds by Berg and colleagues involving the sentencing of juvenile truants. However, the experiment involved a single bench of magistrates

so there was no randomisation process prior to the sentencing hearing (Farrington, 2003). As in that case, the RCT may be seen as denying useful treatment to offenders, adding to the workload and interfering with the role of magistrates. Moreover, compliance cannot be enforced. This poses major risks to the RCT. It would, therefore, be essential to gain commitment to the RCT from those who can affect the randomisation process.

Information and education on the RCT and how to operate it successfully would be essential for all those involved in the court process, sentencing and DRRs.

#### **8.2.4 Contamination**

If an intervention is successful in improving outcomes this may not be detected in an evaluation because of contamination between the treatment and control groups. Contamination means that some members of the control group receive some of the intervention treatment and therefore gain some of the improvements in outcome that the treatment provides. This will mean that the measured difference between the treatment and control groups will underestimate the effectiveness of the intervention. This may mean that an evaluation concludes that an intervention is not effective when in fact it is.

Contamination is likely to occur:

1. through interaction between DDC and non-DDC magistrates;
2. through improved inter-agency working around the DDC leading to improvements to other courts;
3. through DDC magistrates sitting in other courts and hearing control group cases;
4. if the randomisation process results in control group cases being listed together (this would happen if DDC-eligible cases were assigned to the day(s) when the DDC sits to facilitate change between control and treatment group because of misallocation or to allow randomisation to take place after the hearing date has been set).

The first three would be extremely difficult to prevent or minimise. The fourth is a factor to be avoided by the random allocation protocol and should be checked through monitoring of the courts which control group members attend.

#### **8.2.5 Monitoring**

It is important in RCTs to monitor implementation, to ensure effective randomisation and to

prevent contamination. A system would be required for reporting to the research team that the court was attended by all in the treatment and the control groups and for monitoring the data and reporting systems. Monitoring data would include offender characteristics, illicit drug use and current and previous offences, to examine whether there appeared to be any bias. This would need to be on-going in order to investigate and address any apparent misallocation.

### **8.2.6 Sample**

Based on the assumption of a minimum five percentage point reduction in reoffending, a sample size of at least 2,800 would be required (across the treatment and control groups). This may still be optimistic and so an evaluation with these sample sizes still presents a high risk of not detecting any effect.

The researchers estimate that 800 useable cases flow through the DDCs annually. Therefore, it would take more than four years to assemble an adequate sample.

### **8.2.7 Data matching**

The DDC courts would provide details for all those appearing before the DDC and information on drug misuse and offence. The DDC sample would be matched: to the HOPNC, using name, gender, date of birth and the PNC number, to provide data on recidivism; to the DIR and TOP (using attributor codes comprising initials, date of birth, gender and date of drug test) to provide outcome data on continued drug misuse; to probation case management systems and IAPS to provide sentence completion measures; and to probation case management systems and Enforcement Tracker to provide measures of compliance. Matching on these outcomes<sup>26</sup> would be conducted on at least two occasions to identify short and longer term impacts.

### **8.2.8 Outcome measures**

The outcomes have been described above. An RCT measures the impact on average (mean) outcomes only and, in this respect, is more limited than a between-area comparison design.

### **8.2.9 Timescale and costs**

An RCT would take a minimum of eight to nine years to identify the impact of the DDC on short-period outcomes over six months. It would take six months to two years longer to identify the impact over one or two years, Table 8.1. Further details of costs are given in Appendix 3.

The main reason for the length is the fieldwork period. This is due to the time required to build a large enough sample to detect an effect, given the throughput of the DDCs.

Fieldwork costs include provision for a person to be available at all times to randomly assign study participants. Further details of costs are given in Appendix 6.

---


<sup>26</sup> The Home Office is currently developing a data warehouse comprising some of the key datasets described here in order to facilitate linkage between health and criminal justice data with a view to monitoring outcomes for drug misusing offenders.


**Table 8.1 RCT timescale and costs**

	1	2	3	4	5	6	7	8	9	10	11	
	Elapsed time											
	Cost											
Design, including identification of point of randomisation and system for randomisation	£60,000-90,000	1 year										
Set up, including set up of randomisation process and discussions/education/training with parties to gain commitment to randomisation	£170,000-250,000											
Fieldwork (including on-going monitoring and adjustments of randomisation)	£500,000-780,000	4-5 years										
Period over which outcomes are measured for six- month outcomes plus period for any new offences to appear on the HOPNC <sup>a</sup>		2 years from last case										
Data matching providing outcomes measured at six months from first DDC contact <sup>a,b</sup>	£50,000-75,000	6 months										
Analysis (six-month outcomes) and reporting	£80,000-130,000	6 months										
Period over which outcomes are measured for one or two-year outcomes plus period for any new offences to appear on the HOPNC <sup>a</sup>		2-3 years from last case										
Data matching providing outcomes at one or two years from first DDC contact <sup>a,b</sup>	£30,000-45,000	0 for outcomes after 1 year, 6 months for outcomes after 2 years										
Analysis (one or two-year outcomes) and reporting	£90,000-140,000	6 months										
<b>Total</b>	<b>£1-£1.5 million</b>	<b>8 to 11 years</b>										

a Assumes a maximum 18-month delay for offences to appear on the HOPNC.

b Note this means that some offenders may not have completed their sentence.

 minimum time required

 maximum time required

### 8.2.10 Risks

An RCT seems highly risky.

- The sample size is based on estimates of the minimal detectable effect, but because of lack of evidence on comparable DDCs, these estimates may prove optimistic and so there is a relatively high risk that no effect will be detected even if one exists. This risk could be reduced by increasing the sample size and, hence, the fieldwork period.
- The randomisation requires the compliance of a large number of people across disparate roles. Their support cannot be guaranteed. In particular, magistrates may not be willing to send eligible offenders to non-DDCs and errors may be made in allocation. The number and range make compliance particularly risky.

### 8.2.11 Cost

The researchers estimate that an RCT would cost between £1,000,000 and £1,500,000. This includes an estimate for statistical analysis where there appear to be systematic differences in characteristics between the treatment group and the control group (see Appendix 3 for further details).

## 8.3 Between-area comparison design

### 8.3.1 Design

The between-area comparison design would compare the outcomes for the treatment group (offenders dealt with by the DDCs) with a comparison group of similar offenders in other jurisdictions.

The between-area comparison can be a model or a pilot evaluation. For both, the comparison courts should be drawn from jurisdictions without a DDC, but for the model evaluation, as far as possible, the comparison courts should exhibit little of the DDCs characteristics but should be similar otherwise (both in terms of CJS practice and the locality). In both cases, the number of comparison courts (or jurisdictions) need not equal the number of DDCs (i.e. there does not need to be one-to-one matching of courts or jurisdictions). Indeed, the best approach for the pilot evaluation would be to draw the offender sample nationally, as this allows full comparison with the range of courts without a DDC.

In the jurisdictions without a DDC a sample of offenders would be drawn who would have been eligible for a DDC if there were one (in terms of being a drug-misuser and having committed a low-level acquisitive offence). This sample would be matched to the DDC sample on factors which affect outcomes (e.g. previous convictions, gender, age, identified Class A drug-misuse need) to create a similar comparison group. Matching would be done statistically (e.g. using propensity score matching.)<sup>27</sup>

---

<sup>27</sup> See Appendix 4 for an explanation of propensity score matching.

A danger is that, if all major influences affecting outcomes are not adjusted for, the outcome differences between DDCs and courts in jurisdictions without a DDC will be falsely attributed to the DDC and not to these other differences. This problem can be minimised by close matching of jurisdictions with and without a DDC or by including a large number of jurisdictions without a DDC with varying characteristics. Consequently, the researchers describe two alternative designs: one, the model evaluation, with a small number of comparison courts, the other a pilot evaluation, using courts in all jurisdictions without a DDC for its comparator.

### **8.3.2 Data requirements**

A dataset would be built up as follows.

For the treatment group, the DDC courts would provide the PNC number for all those appearing before the DDC and information on drug-misuse and offence. The DDC sample would be matched to the HOPNC, using name, date of birth, gender and the PNC number, to provide standardised data on characteristics (for matching) and recidivism. It would be matched to the DIR and TOP (using attributor codes and date of drug test, initials, date of birth and gender) to provide outcome data on continued drug misuse; to probation case management systems and IAPS to provide sentence completion measures; and to probation case management systems and Enforcement Tracker to provide measures of compliance.

For the selected comparator jurisdictions, comparator offenders would be identified from HOPNC (eligible offence) and OASys/DIR (Class A drug-misuser). For other data, the sample would be matched as for the DDC sample.

### **8.3.3 Sample size**

Based on the assumption of a minimum five percentage point reduction in reoffending a minimum sample size of 1,600 in the DDCs and 1,600 in comparison in jurisdictions without a DDC would be required. A five percentage reduction in reoffending may be optimistic and so an evaluation with these sample sizes still presents a high risk of not detecting any effect.

### **8.3.4 Outcome measures**

The outcomes have been described above. Outcomes could be examined for the mean and median impact and for the percentage of offenders with improved outcomes.

### **8.3.5 Analysis**

The comparison group would be created through statistical matching and the difference in outcomes compared between the treatment and comparison group. A risk for matched comparison is that comparators are not found for some of the treatment group. This is unlikely to occur in this case as the personal characteristics of offenders who meet the DDC criteria are likely to have a very large degree of overlap across different areas. However, there is a risk that

**Table 8.2 Between-area comparisons: timescale and costs**

	Cost, £		Elapsed time	1	2	3	4	5	6
	Model evaluation	National pilot evaluation							
Period for 1,600 cases to appear before the DDC			2 years <sup>d</sup>						
Design	43,000-51,000	23,000-27,000	0 years <sup>c</sup>						
Measurement period for 6-month outcomes plus period for any new offences to appear on the HOPNC <sup>a</sup>			2 years from last case						
Collation and matching of data for the treatment and comparator samples, providing data on matching variables and outcomes at six months from first DDC contact <sup>b</sup>	74,000-87,000	74,000-87,000	9 months						
Analysis (six-month outcomes) and reporting	70,000-82,000	70,000-82,000	8 months						
Period over which outcomes are measured for one or two-year outcomes plus period for any new offences to appear on the HOPNC <sup>a</sup>			2-3 years from last case						
Data matching providing outcomes at one or two years from first DDC contact <sup>b</sup>	44,000-51,000	44,000-51,000	0 – 6 months						
Analysis (one or two year outcomes) and reporting	49,000-58,000	49,000-58,000	6 months						
<b>Total excluding period for 1600 cases to appear before the DDC<sup>a</sup></b>	<b>£280,000-£325,000</b>	<b>£260,000-£300,000</b>	<b>4 years - 5 years</b>						

a Assumes maximum 18-month delay for offences to appear on the HOPNC.

b Note this means that some offenders may not have completed their sentence.

c Design done during period for DDC to process 1,600 cases and outcome data to become available.

d The sample can include cases which occur prior to the evaluation start-up. Therefore the time required for the sample to build up depends on when the evaluation starts. If it started after the DDC had dealt with 1,600 cases, there would be no period awaiting sample build.

matches will be poor as these are based on factors which predict outcomes for drug-misusing offenders and these are difficult to predict. Poor matching would reduce the identified impact.

Matching, or other statistical adjustments, will be made for differences in characteristics between jurisdictions.

### **8.3.6 Timescale**

Once the DDC had dealt with 1,600 cases, a between-area comparison evaluation would take approximately three and a half years to identify the impact of the DDC on short-term outcomes over six months. It would take four to five years to identify the impact over one or two years, respectively. Timescale and costs are shown in Table 8.2 Further details of costs are given in Appendix 3.

### **8.3.7 Risks**

The between-area comparison faces some, potentially large, risks.

- The sample size is based on estimates of the minimal detectable effect. However, because of lack of evidence on comparable DDCs, these estimates may prove optimistic. Therefore, there is a relatively high risk that no effect will be detected even if one exists. This risk could be reduced by increasing the sample size and, hence, the fieldwork period.
- The quality of the estimates depends on the closeness of the matching, both of jurisdictions (i.e. to allow for differences other than the presence of DDCs which affects outcomes) and for offender characteristics affecting outcomes. If these can only be poorly predicted the estimates will be less robust.

### **8.3.8 Costs**

The researchers estimate that a between-area comparison pilot evaluation would cost between £260,000 and £300,000. A model evaluation would cost between £280,000 and £325,000 (see Appendix 3 for more details).



## 9 Conclusions

### 9.1 Introduction

The study has explored the factors which affect the feasibility of a robust evaluation of the impact of the DDCs on recidivism, substance-misuse, sentence completion and sentence compliance. Key requirements were to establish whether any impact would be identifiable (with an acceptable level of statistical confidence) and whether a control or comparator group which was similar to the treatment group (in terms of factors likely to affect outcomes) could be created (i.e. which did not suffer from selection bias).

This entailed the following.

- Investigating the extent to which DDCs conformed to one drug court model and the extent to which non-DDCs differed: the greater the similarity between DDCs and non-DDCs, the smaller the impact would be.
- A review of the international evidence of the impact of drug courts.
- Investigating the availability of administrative data for constructing similar treatment and comparator groups.
- Investigating the availability of administrative data for measuring outcomes.

### 9.2 Expected impact

The impact of drug courts on reoffending identified in many other studies varied. However, many of these studies were of drug courts which embodied not only court procedural changes (as does the DDC pilot) but changes in drug treatment provision for offenders (which the DDC does not). Much of the impact is almost certainly due to the drugs treatment and related support, rather than the court procedural changes. Therefore the effect of DDCs is almost certainly small.

### 9.3 Data availability

Data are available for evaluating the impact of the DDCs on recidivism (through the HOPNC), sentence completion and sentence compliance (via probation case management systems, Enforcement Tracker and IAPS). However, data on drug misuse are more problematic, as the datasets which would show continued drug misuse (e.g. TOPS and DIR) do not cover all subjects in the study.

### 9.4 DDC and courts in jurisdictions without a DDC

DDC practice varied: in the ways in which the four main elements of the DDC model (continuity of judiciary, list DRRs in dedicated sessions, additional training and improved inter-agency working) were implemented; in the offenders dealt with; and in the stages at which offenders were referred (for sentencing or DRR review). This would further diminish the impact of the DDCs and the ability to detect any effect.

Moreover, practice in courts in jurisdictions without a DDC varied and some appeared to differ little from DDCs; the only characteristic which appeared to differentiate DDC and courts in jurisdictions without a DDC was the intention to provide continuity of magistrate from the point of DRR sentence through to completion, and if necessary (and where possible) breach hearings. Thus, the difference in performance of the pilot DDCs compared with a similar (or representative) sample of courts in jurisdictions without a DDC would be particularly small. The difference in performance between the DDCs which best exemplify the DDC model and courts in jurisdictions without a DDC which least exemplified the DDC model would be greater, but still small. (The latter would also have a smaller sample size, reducing the likelihood of detecting effects.)

## **9.5 Detectable effect**

In Section 6.2, the researchers' estimates based on international research evidence of the detectable effect suggested that a reduction of five percentage points in reconviction rates between DDC and non-DDC groups (roughly half those found by Latimer et al., 2006 in their review) would be detectable with a sample of 1400 in both the treatment and comparison groups. The researchers believe that it is necessary to reduce the potential expected effect below the level found by Latimer et al. (2006) because the evaluations reviewed in that study involved a package of drug court procedures with treatment programmes. In addition, the varied application of the DDC model across the pilot sites and the significant overlap in both processes and sentences with courts in jurisdictions without a DDC means that only small effects would be expected to be found. It is possible that even a five percentage point difference in recidivism would be optimistic in these circumstances.

Furthermore, this estimate is of the detectable effect of similar samples (i.e. without any selection bias); a larger sample would be required for a quasi-experimental evaluation. The sample size required increases further for between-area comparisons, where differences between jurisdictions need to be taken into account. The sample would need to be collated over a long period. The researchers would expect that both treatment and comparison group samples would need to be around 1,600 or more.

For these reasons, the researchers think that the risk of an impact evaluation using any form of comparison being unable to detect any effect on either offending or drug-misuse is high. The minimum sample sizes they have estimated may still be over-optimistic, as they have had to make assumptions about the proportion of effect attributable to process rather than treatment.

## **9.6 Possible evaluation approaches**

The RCT and between-area comparison designs were examined in more detail.

The designs could estimate the effect on reoffending, drug misuse, sentence completion and sentence compliance. For the RCT, this would be limited to the difference in average (mean) outcomes for the treatment and control group. The between-area comparison outcomes

could also be measured in other ways, providing greater indication of the effectiveness of the DDCs. However, both designs have a high risk of failing to detect any effect (even if it were there) because the sample sizes may be too small to detect the size of impact. Increases in sample size are the only way to address this.

The RCT presents a high risk of failure, due to breakdown in the randomisation, as it requires compliance and assistance from a large number of practitioners, including those who value their independence. The between-area comparison runs the risk of falsely attributing outcome differences to the DDC when other factors are involved.

The timescales for the evaluations are long, between four and five years for the between-area comparison (once the DDC has dealt with 1,600 cases) and eight and a half to 11 years for the RCT.

An RCT would cost between £1,000,000 and £1,500,000. A between-area comparison pilot evaluation would cost between £260,000 and £300,000. A between-area comparison model evaluation would cost between £280,000 and £325,000.

The between-area comparison seems the most cost-effective and least risky approach. This would enable the evaluation to rely on national administrative datasets and probation case management datasets (i.e. court records would not need to be examined). Misattribution of effects to the DDC (rather than to other factors which differ between jurisdictions) would be minimised by the use of national comparisons and, if it were deemed necessary, the evaluation could be extended (as long as the pilots continue to operate) to allow an increase in sample size.

## References

- Anderson, J.F. (2001) 'What to do about "much ado" about drug courts?', *International Journal of Drug Policy*, 12 (5–6): pp 469–475.
- Bean, P. (2004) *Drugs and Crime (2nd edn)*. Cullompton: Willan.
- Belenko, S. (1998) 'Research on Drug Courts: A Critical Review.' *National Drug Court Institute Review*. Volume 1: pp 1–42.
- Belenko, S. (1999) 'Research on Drug Courts: A Critical Review: 1999 Update.' *National Drug Court Institute Review*. Volume 2: pp 1–58.
- Belenko, S. (2001) *Research on Drug Courts: A Critical Review: 2001 Update*. New York: National Center on Addiction and Substance Abuse at Columbia University.
- Belenko, S. (2002) 'The challenges of conducting research in drug treatment court settings', *Substance Use & Misuse*, 37 (12–13): pp 1,635–1,664.
- Bouffard, J.A. and Smith, S. (2005) 'Programmatic, counselor, and client-level comparison of rural versus urban drug court treatment', *Substance Use & Misuse*, 40 (3): pp 321–342.
- Bouffard, J. and Taxman, F. (2004) 'Looking inside the "lack box" of drug court treatment services using direct observations', *Journal of Drug Issues*, 34 (1): pp 195–218.
- Bryson, A., Dorsett, R. and Purdon, S. (2002) 'The Use of Propensity Score Matching in the Evaluation of Active Labour Market Policies.' *Department for Work and Pensions Working Paper no 4*. <http://research.dwp.gov.uk/asd/asd5/WP4.pdf>
- Cartwright, N. (2007) *Are RCTs the Gold Standard?* AHRC Contingency and Dissent in Science Project. London: London School of Economics and Political Science.
- Cunliffe, J. and Shepherd, A. (2007) 'Re-offending of adults: results from the 2004 cohort', *Home Office Statistical Bulletin*, 06/07, March.
- Dehijia, R. and Wahba, S. (1999) 'Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs'. *Journal of American Statistical Association*, 94(448): pp 1,053-1,062
- Dorsett, R., Hudson, M. and McKinnon, K. (2007) 'Progress2work and progress2work-LinkUP: an exploratory study to assess evaluation possibilities' *DWP Research Report No 464*. London: Department for Work and Pensions.

Farrell, M. (2002) *Final Evaluation of the Pilot Drug Court*. Dublin: Farrell Grant Sparks Consulting.

Farrington, D.P. (2003) 'British Randomized Experiments on Crime and Justice' *The ANNALS of the American Academy of Political and Social Science* 589:pp 150-167.

Fischer, B. (2003) 'Doing good with a vengeance: A critical assessment of the practices, effects and implications of drug treatment courts in North America', *Criminal Justice*, 3 (3): pp 227–248.

Freeman, K. (2003) 'Health and well-being outcomes for drug-dependent offenders on the NSW Drug Court programme', *Drug & Alcohol Review*, 22 (4): pp 409–416.

French, P. (2005) 'Why the UK needs drug courts', *Drugs and Alcohol Today*, 5 (3): pp 34–38.

Goldkamp, J.S., White, M.D. and Robinson, J.B. (2001) 'Do drug courts work? Getting inside the drug court black box', *Journal of Drug Issues*, 31 (1): pp 27–72.

Gossop, M., Trakada, K., Stewart, D. and Witton, J. (2006) 'Levels of Conviction Following Drug Treatment: Linking Data from the National Treatment Outcome Research Study and the Offenders Index.' *Home Office Research Findings* 275. London: Home Office.

Gossop, M., Trakada, K., Stewart, D, and Witton, J.(2005)'Reductions in criminal convictions after addiction treatment: 5-year follow-up.' *Drug Alcohol Depend.* 79(3):pp 295-302.

Gossop, M., Marsden, J., Stewart, D, and Kidd, T. (2003) 'The National Treatment Outcome Research Study (NTORS): 4-5 year follow-up results.' *Addiction*.98(3):pp 291-303.

Government Accountability Office (2005) 'Adult Drug Courts: Evidence Indicates Recidivism Reductions and Mixed Results for Other Outcomes.' *Report to Congressional Committees*. Washington, DC: US Government Accountability Office.

Greenberg, D. & Shroder, M. (2004) *The digest of social experiments*. (third edition). Washington DC: The Urban Institute Press.

Harper, G. and Chitty C. (eds.) (2005). 'The impact of corrections on re-offending: a review of 'what works.' *Home Office Research Study* 291. London: Home Office.

Heckman, J. and Hotz, V.J. (1989) 'Choosing among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs.' *Journal of the American Statistical Association* 84 (408)

Heckman, J. and Smith, J.A. (1995) 'Assessing the Case for Social Experiments', *Journal of Economic Perspectives* 9(2)

Heckman, J., Smith, J. and Clements, N. (1997) 'Making The Most out of Social Programme Evaluations and Social Experiments: Accounting for heterogeneity in programme impacts', *Review of Economic Studies* 64(4)

Hedderman, C. and Hough, M. (2005) 'Diversion from prosecution at court and effective sentencing', in Perry, A.E., McDougall, C. and Farrington, D.P.(eds.), *Reducing Crime: The Effectiveness of Criminal Justice Interventions*, 53–72. Chichester: Wiley.

Hedderman, C. (2007) 'Past, present and future sentences: What do we know about their effectiveness?', in Gelsthorpe, L. and Morgan, R. (eds.), *Handbook of Probation*, pp 459–484. Cullompton: Willan.

HM Treasury Government Social Research Unit (2007) *The Magenta Book: Guidance notes for Policy Evaluation and Analysis*. London: HM Treasury.

Hollin, C.R. (2008) 'Evaluating offending behaviour programmes: Does only randomization glister?', *Criminology and Criminal Justice*, 8 (1): pp 89–106.

Hollis, V. (2007) *Reconviction Analysis of Programme Data using Interim Accredited Programmes Software (IAPS)*. London: RDS/NOMS.

Holloway, K., Bennett, T. and Farrington, D. (2005) 'The Effectiveness of Criminal Justice and Treatment Programmes in Reducing Drug-Related Crime: A Systematic Review.' *Home Office Online Report 26/05*, London: Home Office.

Hough, M., Clancy, A., McSweeney, T. and Turnbull, P.J. (2003) 'The Impact of Drug Treatment and Testing Orders on Offending: Two Year Reconviction Results.' *Home Office Research Findings 184*. London: Home Office.

Howard, P. (2006) 'The Offender Assessment System: an evaluation of the second pilot.' *Home Office Research Findings 278*. London: Home Office.

Huddleston, C.W., Freeman-Wilson, K., Marlowe, D.B. and Roussell, A. (2005) *Painting the Current Picture: A National Report Card on Drug Courts and Other Problem Solving Court Programs in the United States. Volume 1, Number 2*. Virginia: National Drug Court Institute.

Joliffe, D. and Farrington, D.P. (2009) 'Initial evaluation of reconviction rates in Community Justice Initiatives. *Research Summary 09/09*.' London: Ministry of Justice.

Latimer, J., Morton-Bourgon, K. and Chrétien, J-A (2006) *A Meta-Analytic Examination of Drug Treatment Courts: Do They Reduce Recidivism?* Research and Statistics Division, Department of Justice, Canada. pp 06-7.

Lind, B., Weatherburn, D., Chen, S., Shanahan, M., Lancsar, E., Haas, M., & De Abreu Lourenco, R. (2002). *New South Wales drug court evaluation: Cost effectiveness*. Sydney: NSW Bureau of Crime Statistics and Research.

Marlowe, D.B., Festinger, D.S., Dugosh, K.L. and Lee, P.A. (2005) 'Are judicial status hearings a "key component" of drug court? Six and twelve month outcomes', *Drug & Alcohol Dependence*, 79 (2): pp 145–155.

Matrix Knowledge Group (2008) 'Dedicated Drug Court Pilots: A Process Report. *Ministry of Justice Research Series 7/08*. London: Ministry of Justice.

May, C. (1999) 'The role of social factors in predicting reconviction for offenders on community penalties.' *Home Office Research Findings.97*. London: Home Office.

May, C., Sharma, N. and Stewart, D. (2008) 'Factors linked to reoffending: a one-year follow-up of prisoners who took part in the Resettlement Surveys 2001, 2003 and 2004.' *Research Summary 5*. London: Ministry of Justice.

McIvor, G., Eley, S., Malloch, M. and Yates, R. (2003) 'Establishing Drug Courts in Scotland: Early Experiences of the Pilot Drug Courts in Glasgow and Fife.' *Research Findings 71*. Edinburgh: Scottish Executive.

McSweeney, T., Stevens, A., Hunt, N. and Turnbull, P.J. (2007) 'Twisting arms or a helping hand?: Assessing the impact of 'coerced' and comparable "voluntary" drug treatment options.' *British Journal of Criminology*, 47 (3): pp 470 – 490.

Meadows, P. (2007). 'The methodologies for the evaluation of complex interventions: an ongoing debate.' in Belsky, J., Barnes, J. and Melhuish, E. *The National Evaluation of Sure Start: Does area-based early intervention work?.* Bristol: Policy Press.

Millar, T., Jones, A., Donmall, M. and Roxburgh, M. (2008) *Changes in offending following prescribing treatment for drug misuse*. London: National Treatment Agency.

Mohr, L.B. (1995) *Impact analysis for program evaluation* Second edition, Thousand Oaks CA: Sage.

National Association of Drug Court Professionals (NADCP) (1997) *Defining Drug Courts: The Key Components. Office of Justice Programs*. Washington, DC: US Department of Justice.

National Probation Service (2005) *Performance Report 16 and Weighted Scorecard 2004/05*. London: Home Office.

National Probation Service (2007) *Annual Report 2006–2007*. London: National Offender Management Service.

Orr, L. (1999) *Social experiments: evaluating public programs with experimental methods*, Thousand Oaks CA: Sage.

Pawson, R. & Tilley, N. (1997). *Realistic Evaluation*. London: Sage.

Peyton, E.A. and Gossweiler, R. (2001) *Treatment Services in Adult Drug Courts: Report on the 1999 National Drug Court Treatment Survey, Executive Summary*. Washington, DC: US Dept. of Justice and US Dept. of Health and Human Services.

Rodriguez, N. and Webb V. (2004) 'Multiple measures of juvenile drug court effectiveness: Results of a quasi-experimental design', *Crime & Delinquency*, 50 (2): pp 292–314.

Rosenbaum, P. and Rubin, D (1983) The central role of the propensity score in observational studies for causal effects *Biometrika* 70:pp 41-50.

Rubin, D.B. and Thomas, N. (2002). Characterizing the effect of matching using linear propensity score methods with normal distributions. *Biometrika*, 79(4):pp 797-809.

Shadish, W.R., Cook, T.D. & Levinton, D.C. (1991). *Foundations of Program Evaluation*. Newbury Park CA: Sage.

Skodbo, S., Brown, G., Deacon, S., Cooper, A., Hall, A., Millar, T., Smith, J. and Whitham, K. (2007) 'The Drug Interventions Programme (DIP): Addressing Drug Use and Offending Through "Tough Choices"'. *Home Office Research Report 2*. London: Home Office.

Turnbull, P.J., McSweeney, T., Hough, M., Webster, R. and Edmunds, M. (2000) 'Drug Treatment and Testing Orders: Final Evaluation Report.' *Home Office Research Study 212*. London: Home Office.



Uchtenhagen, A., Stevens, A., Berto, D., Frick, U., Hunt, N., Kersch, V., McSweeney, T. *et al.*, (2008) 'Evaluation of therapeutic alternatives to imprisonment for drug dependent offenders: Experience from a comparative European multi-country study', *Heroin Addiction and Related Clinical Problems*, 10 (2): pp 5-10.

Weatherburn, D., Jones, C., Snowball, L. and Hua, J. (2008) 'The NSW Drug Court: A re-evaluation of its effectiveness' *Crime and Justice Bulletin*, NSW Bureau of Crime Statistics and Research., Number 121, September.

Wenzel, S.L., Longshore, D., Turner, S. and Ridgely, M.S. (2001) 'Drug courts – A bridge between criminal justice and health services', *Journal of Criminal Justice*, 29 (3): pp 241–253.

Wilson, D.B., Mitchell, O. and Mackenzie, D.L. (2006) 'A systematic review of drug court effects on recidivism' *Journal of Experimental Criminology* vol.2 pp 459-487.

Wundersitz, J. (2007) *Criminal justice responses to drug and drug-related offending: Are they working?* Canberra: Australian Institute of Criminology.

## Appendix 1: DDC policy and practice: details

### Introduction

This appendix provides more detailed information of the researchers' findings in the DDCs on which they based the summaries in the main body of the report. It focuses on the four key elements of DDCs:

1. continuity of judiciary from sentence through to completion, and if necessary (and where possible), breach;
2. list DRRs together in dedicated court sessions;
3. provide additional training for the judiciary and other staff dealing with these cases to improve awareness and understanding of drug addiction issues; and
4. improve inter-agency working and information-sharing processes.

The fieldwork examined each of these aspects in depth in the DDCs and in a sample of courts without a DDC in order to identify a number of issues relevant to evaluation design. These included: being able to identify whether DDC practice differed from courts without a DDC (which has implications for identifying comparator courts and the expected size of effect); whether practice differed between DDCs (which has implications for the choice between a model or a pilot evaluation and the expected size of effect); and the extent to which practice in DDCs had changed on becoming a DDC. This Appendix provides information on DDC practice and the next Appendix on practice in a sample of courts without a DDC. The report itself pulls together differences and similarities between the two.

### Exclusivity, continuity and dedicated sessions

The DDC approach aims to ensure exclusivity by dealing with drug-misusing defendants from the point of sentence, during review and (where necessary and possible) breach processes, and through to completion of a court order. Ensuring sustained continuity of both magistrates' bench and District Judge (DJ) throughout these stages is a defining feature of the approach. Table A1 describes how each of the six pilot sites sought to deliver these elements of the DDC model.

**Table A1 Arrangements for ensuring exclusivity and continuity across the six DDC sites**

DDC pilot site	Key arrangements for ensuring exclusivity and continuity
Barnsley	Defendant is bailed to the DDC on a Wednesday morning following arrest for a 'trigger' <sup>a</sup> offence and positive test for opiates and cocaine (Class A drugs), or following arrest for a Misuse of Drugs Act offence (e.g. possession of cannabis). The DDC panel is made up of nine magistrates who sit on the bench in groups of three, and a DJ. All have undertaken to sit in the DDC on a four weekly cycle to ensure continuity. A Listing Officer contacts at least one of the original sentencing bench to return to deal with any subsequent breaches outside this cycle.
Bristol	All adjournments for DRR assessments are referred to the DDC which sits every Tuesday morning. Sentencing and breaches take place in the morning while review hearings take place in the afternoon. The DDC panel consists of 24 magistrates who are split into four groups of six. Three of them will return on a four-weekly cycle, thus ensuring that at each subsequent review at least two sentencing magistrates will be present. In the event of a breach; probation staff obtain a hearing date from court listing in the next available DDC list.
Cardiff	There are ten slots available every Wednesday morning with the DDC: four are assigned for sentencing and five for breach hearings. Reviews take place in the afternoon. Sentencing slots are allocated to first-time defendants testing positive for Class A drugs following arrest for drug possession, shoplifting or taking without owner's consent offences (these are considered the most prolific offences by local police). DRR assessments are also adjourned to the DDC sentencing court. Once the ten DDC slots have been exceeded, cases are then listed in another court. Any DRRs imposed outside the DDC are then referred back for subsequent review. The DDC panel comprises 18 magistrates organised into groups of three operating on a four-week rota. There is also a dedicated 'floating' magistrate assigned to each bench to cover sickness and/or annual leave. If necessary, breaches will be heard by any of the DDC panel in order to ensure these are dealt with in a manner consistent with probation National Standards. <sup>b</sup>
Leeds	The DDC sits on a Monday and Friday. The morning sessions are devoted to dealing with DRR breaches and reviews. In the afternoon the DDC sentences those defendants adjourned to it for a DRR assessment. The DDC has a panel of 32 magistrates organised into groups of four. The bench members will sit every four weeks, with three of them sitting at any one time. This is achieved by rotating the fourth member. All DRR breaches are heard before the reviewing bench wherever possible.
Salford	All restriction on bail <sup>c</sup> and DRR assessment cases are adjourned to the DDC which sits on a Monday. Sentencing occurs in the morning while reviews are heard in the afternoon. Low intensity DRRs are not listed to appear before the DDC. Both DRRs and other drug-related programme requirements (using s.178 powers) imposed outside the DDC are adjourned to appear before the next review court. The panel consists of 20 magistrates plus one district judge. The former are divided into five teams of four magistrates and sit once every six weeks, adjourning cases for a further six weeks for another review, if required. Three of the benches will sit with a 'floating' magistrate to cover holidays and other commitments. Breaches are listed at the earliest opportunity before at least one of the sentencing panel, but this could be in any court.

DDC pilot site	Key arrangements for ensuring exclusivity and continuity
West London	The DDC sits every Tuesday. Sentencing (of 20 listed cases) takes place in the morning for first listed defendants arrested for a target trigger offence and testing positive for Class A drugs; those adjourned for a DRR assessment; and DRR breaches and further offences. DRR reviews are held in the afternoon. The DDC comprises twelve members: three DJs sitting alone and a panel of nine magistrates operating as three separate benches. The intention is to ensure that at least two magistrates return for each subsequent DRR review, which takes place at six-weekly intervals. In order to ensure compliance with probation National Standards breaches are heard by whichever DDC bench is sitting at an appropriate time.

- a Trigger offences encompass a range of offence provisions covered by the Theft Act (1968), Misuse of Drugs Act (1971), Fraud Act (2006), Criminal Attempts Act (1981) and Vagrancy Act (1824).
- b These stipulate that breaches should be dealt with within 25 working days of the date of a second unacceptable absence.
- c The Criminal Justice Act 2003 amended the Bail Act 1976 to provide for a restriction on bail (RoB) for adults who have tested positive for heroin or cocaine. At an initial bail hearing a defendant can be asked to undergo an assessment of their drug problem (a relevant assessment) and will have to agree to participate in any follow-up recommended by the assessor. If the defendant agrees, they will, in most cases, be released on conditional bail. However, if they refuse, the normal presumption for bail is reversed and the court will not grant bail unless satisfied that there is no significant risk of a further offence being committed whilst on bail.

An important observation with regards to the feasibility of undertaking historical comparisons is that all the pilot areas appear to have been offering a degree of exclusivity and continuity – at least in terms of sentencing and reviewing DRR cases – long before they acquired their formal DDC status:

*'I think it certainly became, it evolved into the dedicated drug court when we had the sort of formal opening, but it certainly existed prior to that...I don't think anything has changed dramatically other than the fact that it's now set up in a more formalised way'*

*'I think we've had a drug court since...2000, I think it was.'*

*'Q: [The court] has been up and running as a formal DDC since [date]...But it's had, if you like, drug court qualities prior to that?'*

*A: Yes, since 2002'.*

The accounts from respondents across all six pilot DDC areas indicated that, to a large extent, continuing to develop and refine the exclusivity and continuity elements under the DDC model had been a relatively smooth and successful process. This was attributable to both previous arrangements and the considerable commitment and enthusiasm shown by all the partners involved.

Not all drug using offenders committing low-level acquisitive crimes were seen by the drugs courts:

*'If somebody was going to get a fine on a Thursday, we wouldn't adjourn their case until Monday for them to get a fine, we would refer them voluntarily to the drug services... [if] they deserve more than six months imprisonment, we would send them directly to the Crown Court, we wouldn't send them to the drugs court for that decision to in some way be reviewed, we would send them to the Crown Court.'*

However, the extent to which this happened may vary across DDCs. Limited capacity within some of the DDCs meant that at least some drug-misusing defendants would not be referred to it for sentencing (there were 44 DRRs imposed outside of the DDCs in four of the pilot areas between February and March 2009, while data on DRR commencements in London during 2006/07 indicate that only one-third (60) of all DRRs imposed across the four west London boroughs in which the court operated during this period (197) were sentenced by the DDC), while there continue to be a number of offenders from outside the DDC catchment areas being sentenced by the DDCs:

*'You get a reasonable number of people from outside the area coming in to commit the offences, even if they test positive then they're not going to be there for the DDC to review post-sentence.'*

Conversely, in at least one of the pilot sites a DJ exercised his/her prerogative to retain the review function for all DRR offenders he/she sentences from outside the court's catchment area.

There was also acknowledgment of the considerable variation in approaches to the review process and how this could impact on outcomes:

*'Obviously part of the review process is to praise progress, but we haven't moved beyond that I confess. We don't have anything particularly innovative in terms of developing motivation beyond what the review process is designed to deliver.'*

## **Training**

The provision of regular training involving all local stakeholders features prominently in each of the six pilot DDC areas' service level agreements. The scope and content of this training was often discussed and agreed through the local DDC steering groups. This helped ensure that training was considered to be both responsive and specifically tailored to address locally identified gaps in skills and knowledge. Furthermore, training of this sort has also been an important vehicle for increasing awareness and understanding between the key partners about their respective roles and responsibilities:

*'Well what the panel are going to do is they're going to get together, I think, in practice once every four months because of the rotas that they're on, to discuss progress, but there will also be a training element to each one of those meetings, whether it be speakers from providers...In the forthcoming one we've established this voluntary referral group for cocaine and that would be a great, you know, let's find out more about that. They also do visits and things of that nature so that, you know, many people like to have a nose around the service providers and get a feel for what's going on, those sorts of things.'*

For the purpose of comparing how the provision of training differed from jurisdictions without a DDC, the extent to which the impact of DDC training and levels of satisfaction with it were monitored varied across the sites – from no formal monitoring to the regular use of evaluation sheets:

*'Well we did handouts and evaluation sheets... we had the police doing something on drugs and showing them what the drugs look like and what the offenders do, and we had the [providers] talking about how they deal with them and what they do in the treatment programmes, as well as the court going through what happens in court, so we did have a lot of positive feedback in relation to that..'*

*'I confess that a 30 minute session at a panel meeting, or a visit, would not be subject to evaluation other than the informal going to the elected representatives of the panel and saying, was that worthwhile, you know, or did it not work, what shall we do differently, but there's no formal monitoring and evaluation.'*

And while the provision of training is a key feature of the DDC approach, there was an acknowledgement from some that it could be a more prominent and regular one:

*'So it is something perhaps that in the past [this DDC]...haven't given enough attention to, and that the panel members have been on there for so long they know what they're doing. Therefore I haven't really done a lot of up-to-date training with them, only when new members come on, but I think perhaps it is something we should be doing in [this area] annually.'*

### **Inter-agency working and information sharing**

Effective inter-agency working and information sharing between the courts, probation, drug treatment providers and other support services are key to the success of the DDC approach. In all the sites there appeared to be a consensus that this had been achieved to a large degree through a combination of historical links between the agencies involved and the development of new structures implemented specifically to support and reinforce the delivery of the DDC model (e.g. local steering groups):

*'Well you see the thing is in [this area]...we have fantastic inter-agency working, fantastic. We have superb relationships with the agencies, including CPS, the police and probation...We've always talked together. If we'd wanted something doing...we have learnt that the best way is to get people on board and work as a team, so we've got that historically here. And these people were so enthused; they could see that the court was trying to help these people get off drugs.'*

Closer links between health and criminal justice agencies within the six sites had also been greatly facilitated by joint training events. These served to increase knowledge and awareness, not just of substance-misuse issues, but how the different partners involved in the endeavour operate, and the specific challenges they face:

*'We had a presentation about the changing nature of drug abuse in our city, and until then nobody had sent me anything, or any member of the judiciary had not been sent anything around that, and you know, we suddenly find out that nobody under 30 has tested positive for heroin in this city in 18 months, those sorts of things that perhaps are blindingly obvious to policy people, and people working in the services, were completely unknown to us. So now we are having more constructive discussions... about cocaine and its link with criminality...six months ago I was completely unaware of this, and now we're having intelligent discussions with professionals in the field in the city about how we can support other people's work better.'*

However, practice was not uniform across the DDCs. While at least one of the pilot sites benefits from the specialist input from an addictions psychiatrist, others noted that this was an area that needed to be tackled far more consistently within the DDC model through more effective inter-agency working with mental health professionals:

*'I just think perhaps mental health should be something that the DDC starts looking at.'*

## Appendix 2: Policy and practice in jurisdictions without a DDC

### Introduction

This Appendix provides more detailed information of researchers' findings in the jurisdictions without a DDC on which they based the summaries in the main body of the report. It focuses on the four key elements of DDCs:

1. continuity of judiciary from sentence through to completion, and if necessary (and where possible), breach;
2. list together in dedicated court sessions for DRR reviews and cases likely to be sentenced to a DRR;
3. provide additional training for the judiciary and other staff dealing with these cases to improve awareness and understanding of drug addiction issues; and
4. improve inter-agency working and information sharing processes.

### Exclusivity, continuity and dedicated sessions

No respondents in the 15 jurisdictions without a DDC reported collectively listing, sentencing or handling cases involving drug-misusing offenders in dedicated sessions (excluding DRR reviews).<sup>28</sup>

There were a varied set of arrangements for monitoring and reviewing progress and compliance with DRRs imposed by the court, as illustrated in Table A2 and described by the following quote:

*'DRRs [breaches] are all listed in court with Community Order breach cases once a week. Any written reports are presented at monthly reviews. There's no real input from DRR worker or drug agencies and sporadic attendance of probation. There's no real continuity – only occasional when a District Judge reserves reviews to himself but amongst DJs this is not common practice as some are of the view that we're all equally capable of managing issues on review. But many DRRs are six months without reviews – as recommended by probation.'*

---

<sup>28</sup> Though one court noted that they had done so in the past in response to a request by the local police following an enforcement operation.



**Table A2 Key aspects of existing arrangements for monitoring and reviewing DRRs, as described by respondents (N=20)**

	Dedicated review sessions	Sits weekly	Sits fortnightly	Sits monthly	Dedicated panel	Magistrates involved	District Judge(s) involved	Breaches heard in same review sessions
City of Manchester (probation)	✓	✓			✓	✓ (n=20)		
Highbury(court & probation )	✓			✓		✓	✓	✓
Hull (court)	✓	✓					✓	✓
Leicester (court)	✓	✓			✓	✓ (n=18)		✓
Luton (court)		✓				✓	✓	✓
Maidstone (probation)	✓	✓				✓		
Newcastle (probation)								
Norwich (court)	✓			✓	✓	✓ (n=8)		✓
Nottingham (court & probation)	✓	✓			✓	✓ (n=15)	✓	
Plymouth (probation)	✓			✓	✓	✓ (n=4)	✓	
Sheffield (court & probation)	✓	✓				✓	✓	✓
Southampton (court)	✓		✓				✓	✓
Teesside (court)	✓	✓				✓		
Vale of Glamorgan (court & probation)	✓			✓	✓	✓		
Wolverhampton (court & probation)	✓					✓	✓	✓

Four-fifths of the areas described dedicated review sessions for DRRs (13). Most benches reviewed these cases on a rotating weekly basis (8), but with fewer doing so using a dedicated panel (6). DRR reviews were conducted by lay magistrates (6), District Judges (2) or a combination of the two (6). In half the areas, DRR breaches were also heard during these review sessions (8).

However, none of the court or probation-based respondents reported that there was an intention to provide continuity of magistrate from the point of sentence through to completion of any DRRs, and if necessary (and where possible) breach hearings. Respondents from three areas commented that some degree of continuity was offered by DJs. In one of these areas the DJ was supported by four magistrates, but the former was 'responsible for 90% of DRR reviews'. The researchers also asked the 20 court and probation staff in these 15 areas to what extent they thought their lower courts unintentionally provided continuity of judiciary during the course of a DRR. (A court may inadvertently ensure a degree of continuity if, for instance, a DJ expressed a preference to sit in a certain court, at a particular time on a given day.) There were mixed views between the court and probation respondents in this regard: while all responses from the courts indicated that this could happen 'to some extent' (9), probation interviewees felt this kind of unintentional continuity was 'not at all' likely to occur (7).<sup>29</sup>

Only one respondent reported that arrangements for monitoring and reviewing progress and compliance (including with DRRs) differed in other courts/areas they covered.<sup>30</sup> In this additional area it was possible to have dedicated DRR reviews because of the smaller number of DJs and magistrates involved and the smaller DRR caseload sizes.

### **Listing DRR review and breach cases in dedicated court sessions**

Eight of the 11 court respondents described an intention within their courts to list DRR cases together in dedicated court sessions. While none listed DRR cases for sentencing purposes, they would do so for regular reviews and to a lesser extent breach hearings (4). The accounts from two of the remaining three court respondents, however, were inconsistent with their probation colleagues who stated that DRR cases were listed together. Three of the remaining four 'non-matched' probation areas also reported an intention to list DRR cases together for the purposes of review. Those areas not listing DRR cases in this way again tended to identify logistical and capacity issues as the main barriers preventing them from doing so.

Eleven of the 15 areas stated that there were some magistrates who were more likely to sit on DRR cases than others. This appeared to happen more by accident than design:

---

<sup>29</sup> Another stated 'don't know'.

<sup>30</sup> 18 of the 20 respondents stated 'not applicable'.

*'Because some magistrates sit on the same day of the week every week - that may coincide with the same day as the review and breach court. It really depends on availability of magistrates and arrangements of the rota.'*

## Training

All 20 respondents across the 15 jurisdictions without a DDC reported that training had been provided for members of the Local Justice Area bench and other court staff to improve their awareness and understanding of drug-related offending or addiction issues.

This training varied in its form, frequency and intensity, from lunchtime slots and half-day events to structured briefing sessions, written information and newsletters. One respondent reported that there was an intention to deliver substance-misuse training up to four times a year. Others commented that training of this sort had happened but was too infrequent.

Probation played a key role in facilitating and delivering this training, though the drug interventions programme, the police, local drug treatment providers and drug action teams (DATs) were also active participants in a number of areas:

*'When selected to sit on the magistrate panel, Justices receive half-day training from the probation substance-misuse team. All new panel members visit probation...to see how they work. Plus we have four meetings every year and invite all magistrates to a one hour presentation on drugs. It's delivered by a variety of agencies such as DIP, the DAT, prisons...[it covers the] effects of Class A drugs and how these agencies deal with the issue'.*

*'DRR teams and police provide fairly regular magistrates court training looking at local stats, trends, levels of DRRs in the area. There's a very good rapport with the judiciary and good feedback on successes etc. Probation funds the training. They will come in during morning and fit it around other court work.'*

Nine of the 11 court respondents reported that the quality, impact or satisfaction with this training was monitored in some way (though in only two of the five areas which involved perspectives from both court and probation staff was there agreement on this). All four non-matched probation areas also sought the views of staff on the utility of this training. This monitoring of training typically took the form of evaluation and feedback forms completed by participants at the end of sessions and/or verbal feedback. On occasion the results of this feedback were discussed at various committee/meeting structures in order to inform future training activity.

## Inter-agency working and information sharing

Most of the 20 respondents described the inter-agency working and information-sharing processes between the court, local probation and other agencies providing interventions to drug-misusing offenders as either very (9) or quite (10) effective. While the proportion of court and probation interviewees describing these arrangements as very effective was similar, in

three of the five matched areas (i.e. where the researchers interviewed both a probation and court representative) there was disagreement on the extent to which this happens.

The most commonly cited examples of effective practice across the 15 areas included:

- the quality and consistency of reports from probation;
- co-location of partner agencies at court;
- regular meeting structures involving all stakeholders;
- the commitment of all stakeholders; and
- lines of communication and information exchange.

Suggestions for further improving provision tended to focus on the need for:

- quicker DRR assessments;
- greater consistency regarding the timing and quality of DRR reviews;
- the provision of information from treatment providers;
- greater governance and accountability regarding DRRs (see quote below); and
- closer links and more of an emphasis on ancillary support (i.e. housing/benefits/employment).

Most respondents (17) commented that besides DRR arrangements, there had been other recent initiatives or developments aimed at drug-misusing offenders which had contributed towards improved inter-agency working and information-sharing processes between the court, probation and other agencies in their area. Respondents referred to a range of local initiatives including DIP (10), PPO schemes (7), integrated offender management approaches (3), mental health liaison (1), a bail support scheme (1) and the input from a local criminal justice group (1).

## **Data: monitoring and evaluation**

Eight interviewees (seven of them court respondents from seven different areas) stated that their court routinely collated figures on court throughputs and activity.<sup>31</sup> These data appear to have been collected on different groups across the court process, including all defendants (4), just those convicted (2), or on a 'weighted' caseload (1). This kind of monitoring and performance management data were being collected for different local (management groups, criminal justice board), regional (government office) and national (HMCS) purposes.

---

<sup>31</sup> The remaining respondents said such arrangements did not exist (no = 3) or they were not aware (don't know = 9).

The researchers asked respondents whether different aspects of provision within the court were being monitored in any way. They were keen to assess the extent to which the courts in jurisdictions without a DDC were collecting data on 44 key variables now being routinely gathered by the six dedicated DDC sites. The responses<sup>32</sup> indicate that the average (mean) number of variables for which data were being routinely collected by these jurisdictions without a DDC was 30. There were no notable differences in the average number of variables court (30) and probation (31) staff were collecting and recording.

However, in responding to the researchers' semi-structured questionnaire, few (3) indicated that reviewing progress and compliance through continuity of contact with a sentence was routinely being monitored. Seven respondents from five areas stated that information relating to the DRR court review process within their courts was being collated. This included information on number of reviews, the consistency of reviewer and review outcomes. By contrast, around half the respondents reported that there were some systems in place for monitoring the extent to which DRR cases were listed together (11), the training offered to the judiciary and court staff (10) and levels of inter-agency working and information sharing (11).

It is not clear from the responses given to what extent this level of monitoring – and the manner in which this information is being collated and stored -- has wider utility for assessing the impact of the DDC approach. However, there would seem to be little scope for accurately defining and measuring most of the key features associated with the DDC model (i.e. inter-agency working, information sharing and training) in other areas using existing systems and data sources. On a more promising note, almost all (19) the jurisdictions without a DDC reported that, on average, they routinely collate data on two-thirds (30) of the 44 variables being recorded by the six dedicated sites. That said, the researchers were unable to assess the quality and consistency of these data collection processes.

---

32 Excludes two probation respondents who replied 'don't know' to all these questions.

## Appendix 3: Costs

### RCT

Total RCT costs are estimated to be between £1 million and £1.5 million. The estimated cost for each stage is given below.

**Table A3 RCT costs**

Stage	£	Researcher days		
		senior	middle	junior
Day rate		£750-£1000	£500-£600	£300-£450
Design	60,000-90,000	60-80	10-15	15-20
Set up	170,000-250,000	90-120	125-165	130-175
Fieldwork	500,000-780,000	15-20	80-110	90-125
Data matching 1	50,000-75,000	5-10	15-20	90-125
Analysis (early outcomes)	80,000-130,000	30-40	50-70	90-125
Data matching 2	30,000-45,000	5-10	10-15	55-75
Analysis (later outcomes)	90,000-140,000	35-45	60-80	90-125
<b>TOTAL</b>	<b>£1 - £1.5 million</b>	<b>245-330</b>	<b>350-475</b>	<b>570-775</b>

The main elements of costs and assumptions are as follows.

#### Design

Substantial senior input is required for designing the randomisation system, including investigation of how this should ensure operation across all those involved in referring cases to the DDCs.

#### Set up

This includes substantial education work within the DDCs to gain the co-operation and assistance of magistrates and court staff, plus police and CPS staff.

Random assignment protocols to detail the process of assigning each individual to the DDC or not would be developed in consultation with relevant stakeholders (including piloting). Training would be given to relevant stakeholders in the operationalisation of the random assignment.

#### Fieldwork

This covers costs of operating the random assignment, monitoring its operation (monitoring the characteristics of offenders assigned to control and DDC groups to check for signs of bias), providing further education to stakeholders where problems are identified and where staff changes make this necessary.

The costs include the cost of a junior administrator to be available at all times to make random allocation of offenders.

Monitoring and random allocation has to take place throughout the four to five years of fieldwork, whilst further education will be sporadic across this period.

### **Data matching 1**

This covers negotiating access to datasets and gaining ethical approval (including NHS ethical approval) for HOPNC, DIR, TOP, probation case management records, IAPS and Enforcement Tracker and for matching these datasets to the sample. Gaining permission for access (including ethical approval) can be time-consuming.

### **Analysis 1**

This covers analysis of the RCT for outcomes at six months (recidivism, drug-misuse and sentence compliance). It also allows for other analysis of the data (using matching techniques, if appropriate) and reporting.

### **Data matching 2**

This covers accessing relevant datasets (HOPNC, DIR, TOP, probation case management records, IAPS and Enforcement Tracker) and matching outcomes after one or two years to the dataset produced at Data matching 1.

### **Analysis 2**

This covers analysis of the RCT for outcomes. It includes an allowance for some statistical analysis to adjust for any systematic differences between the treatment and the control groups. This may not be necessary, but there may be issues with the randomisation process (e.g. inconsistent implementation at one or more courts) which lead to the groups having different characteristics. This may result in errors in measuring outcomes. Statistical matching can offset these errors. It also covers reporting.

### **Between-area comparison costs**

Total between-area comparison costs are estimated to be between £260,000 and £325,000. The estimated cost for each stage is given below.

**Table A4 Between-area comparison costs**

Stage	£	Researcher days		
		senior	middle	junior
Day rate		£750-£1000	£500-£600	£300-£450
Design	23,000-27,000	10-15	15-20	20-25
Additional design, model evaluation	20,000-24,000	5-10	5-10	30-40
Data collation and matching 1	74,000-87,000	10-15	20-30	135-170
Analysis (early outcomes)	70,000-82,000	25-30	40-50	85-105
Data collation and matching 2	44,000-51,000	5-10	10-51	85-105
Analysis (later outcomes)	49,000-58,000	20-25	35-45	45-60
<b>TOTAL: pilot evaluation</b>	<b>260,000-300,000</b>	<b>70-80</b>	<b>133-155</b>	<b>405-470</b>
<b>TOTAL: model evaluation</b>	<b>280,000-325,000</b>	<b>80-95</b>	<b>140-165</b>	<b>435-505</b>

These main elements and assumptions are as follows.

### Design

Design includes sampling design and a literature review to identify sample matching characteristics. For the model evaluation, desk-based and qualitative research would be required to identify comparator jurisdictions which were similar to the DDCs other than in drugs court characteristics.

### Data collation and matching 1

This covers negotiating access to datasets and gaining ethical approval (including NHS ethical approval) for HOPNC, DIR, TOP, probation case management records, IAPS and Enforcement Tracker and for matching these datasets. Gaining permission for access (including ethical approval) can be time-consuming.

### Analysis 1

This covers analysis (using matched comparison techniques) to identify impact on recidivism, drug misuse and sentence compliance over six months and reporting.

### Data matching 2

This covers accessing relevant datasets (HOPNC, DIR, TOP, probation case management records, IAPS and Enforcement Tracker) and matching outcomes (recidivism and drug-misuse) after one or two years to the dataset produced at Data matching 1.

### Analysis 2

This covers analysis (using matched comparison techniques) to identify impact on recidivism, drug-misuse and sentence compliance over one or two years and reporting.



## Appendix 4: Propensity score matching

In a random assignment experiment the underlying assumption is that any observed difference between the outcomes of the treatment group and the control group can be attributed to the experiment. This is based on the premise that the characteristics that might influence the outcome (for instance, personal circumstances, motivation, educational background) are similarly distributed across both samples. Thus, with random assignment, the average impact of treatment on the treated group is the same as the average impact of the treatment overall.

Where a non-experimental approach is used, either with a non-randomly selected comparison group, or with other observational methods such as the use of administrative data or comparison with the wider population, it is likely that these characteristics which influence outcomes are not distributed in the same way across both the treatment group and the comparators. This means that the observed difference in outcomes between the treatment group and the comparator group is likely to be biased. This bias may operate in either direction. The treatment group might have characteristics that make them more likely to have a positive outcome (perhaps because they are motivated volunteers). In this case, the impact of treatment on the treated group is larger than the average impact of the treatment overall. Conversely the treatment group might also have characteristics that make them less likely to have a positive outcome (because they have difficult personal circumstances). In this case, the impact of treatment on the treated is less than the average overall impact of the treatment.

Rosenbaum and Rubin (1983) developed propensity score matching. This is a method of taking account of the probable impact of different personal characteristics and local circumstances on the outcomes of both the treatment group and the comparator group and combining them into a single score with a value between zero and one (the propensity score). The propensity score represents the probability that an individual has a set of circumstances and characteristics that would lead them to volunteer for treatment (whether or not they have done so). The score is estimated (typically using a logit or probit model) across the combined treatment and control groups using variables which are thought likely to influence the probability that someone will take part. They could be selected on the basis of previous research evidence, or in the absence of such evidence, the views of researchers based on theory, or the views of practitioners based on experience, or some combination of the two.

Instead of the comparison of outcomes between the treatment group and the comparison group being based on the outcomes for the entire group, the adjusted outcomes are based on the outcomes for a group who have received the treatment and a subgroup of those who have not received treatment who have a similar distribution of propensity scores to the treatment group. Sometimes members of the comparator group are matched with more than

one member of the treatment group. This has been shown to improve the accuracy of the estimate where some characteristics are more common in the treatment group than in the comparison group (Dehejia and Wahba, 1999). Similarly, sometimes several members of the comparison group can be matched to an individual member of the treatment group.

There are several varieties of matching that can be used, depending on the estimation method used for the outcome measurement and on the closeness of the match between the treatment and comparison samples. The most commonly used is nearest neighbour matching using identical or very close propensity scores. Alternatives are caliper matching, based on rules about the degree of difference allowed or kernel matching, which is usually used when propensity score matching is combined with estimating the difference in the changes observed between the two samples (difference in differences estimation) using panel data with repeated observations of the same people. Strata matching is a variant where the outcomes are estimated for matched subgroups based on their propensity scores. Propensity score matching can also be combined with other forms of matching.

An important validation of whether the use of propensity score matching does reduce bias is to compare the mean values of the variables used in the matching equation before and after matching, for both the treatment and the comparison groups. The differences between the two groups should generally be markedly lower after matching than they were before (in fact ideally there should be no difference after matching).

Successful propensity score matching generally requires relatively large samples, a substantial degree of overlap in terms of characteristics between the two groups, and accurate measurement of the variables used in the matching process.