



Home Office

BUILDING A SAFE, JUST
AND TOLERANT SOCIETY

A Feasibility Study into Using a Randomised Controlled Trial to Evaluate Treatment Pilots at HMP Whitemoor

David P. Farrington
Darrick Jolliffe

Home Office Online Report 14/02

Acknowledgements

We are very grateful to many people who have helped us with our research, most of whom are mentioned explicitly in this report. However, we are particularly grateful to Siobhan Campbell and Ricky Taylor for helpful comments, and to Gareth Hancock for a great deal of help in the beginning. We are also indebted to Maureen Brown for her speedy and efficient word processing skills.

David Farrington

Darrick Jolliffe

Executive summary

The aim of this report was to assess the feasibility of evaluating the Whitemoor intervention pilot for individuals assessed as DSPD (Dangerous and Severe Personality Disorder) using a randomised controlled trial (RCT).

Background

DSPD individuals are those who pose a significant risk of serious harm to others as a result of their severe personality disorder. It has been estimated that there are about 1,400 DSPD prisoners at any time. The operational definition of DSPD currently involves an assessment of: dangerousness (based on five risk assessment instruments); personality disorder (based on two instruments); and the link between them (based on file data and one assessment inventory).

The main advantage of an RCT is that it equates persons in a treated group with persons in a control group on all possible (measured and unmeasured) variables that influence key outcomes. In this way it is possible to disentangle the effects of the treatment from pre-existing differences between the groups. However, very few previous RCTs have been carried out in British prisons.

Current practice

Currently, potential DSPD prisoners are identified in the eight prisons in the high security estate, are screened, and asked if they are willing to come to the Whitemoor DSPD unit. Selection for the DSPD unit is voluntary, but many prisoners are keen to receive treatment that will reduce their perceived risk and hence (in their eyes) increase their probability of being released. In April 2002, 60 prisoners were either waiting to be screened or were screened and willing and waiting to be admitted to the DSPD unit.

The Whitemoor DSPD unit contains 90 beds, of which 36 are currently assigned for assessment and 54 for treatment. Dialectical Behaviour Therapy (DBT) has been chosen as the basis for the treatment. About 50 prisoners per year are currently being assessed. Up to March 2002, 63 prisoners had been assessed, of whom about half were considered suitable for DBT. Only four of those to complete the assessment were not assessed as DSPD.

The main aims of the DBT-based treatment are to improve personal and social functioning, to improve emotional regulation (mood, anger, irritability), to reduce the risk of harm to self and others, to reduce the person's distress, and to improve the person's amenability to treatment for sexual and violent offending.

Outcome measures

Because few of the DSPD prisoners will be released within the next three years and so will not have the opportunity to reoffend, the main outcome measures in any evaluation would have to be based on psychometric tests, ratings, behaviour and prison disciplinary offences. Research is needed to develop instruments that could measure change in these outcomes, and especially in the risk of committing serious violent or sexual crimes. The report recommends that a sample of prisoners in the high security estate should complete a variety of instruments, and the results correlated with past serious violent and sexual offences, to develop a set of dynamic items that can be used to measure changing risk. Research is also needed on the validity of self-reporting in these measures by DSPD prisoners.

Numbers

Statistical power analysis was carried out to establish the minimum sample size that would be needed to detect the likely effect of the treatment. Assuming that 50 prisoners were in the treated group, it was concluded that at least 300 prisoners would be needed in the control group. This would apply to both experimental and quasi-experimental evaluations. If the number of treated prisoners could be increased from 50 to 80 or 100, this would have disproportionate benefits for statistical power.

Ethical issues

The main ethical issues centre on the denial of treatment to prisoners in the control group and the need for prisoners to freely consent to participate in the research without any inducements.

These issues can be countered by, first, the fact that because the number of eligible prisoners greatly exceeds the number of available treatment beds, arguably, random allocation is the fairest way to select prisoners for

assessment and treatment in the Whitemoor DSPD unit. Second, prisoners who are controls at one time could be eligible to be selected for treatment at a later time. Third, prisoners can freely consent (or not) to participate in the research.

While prisoners may hope that their chances of release on parole would be increased by the treatment, in fact the probability of DSPD prisoners on determinate sentences being released early is likely to be low irrespective of the treatments they have received: no single treatment is likely to have a crucial effect. Therefore, the report is cautiously optimistic that an RCT to evaluate DBT-based treatment in Whitemoor could be approved by the South East Multi-Centre Research Ethics Committee (MREC).

Conclusion

The main problems threatening the feasibility of an RCT to evaluate the Whitemoor DSPD unit are the small number of prisoners who will receive the treatment; the heterogeneity of DSPD prisoners; the likely length of the treatment (2 to 3 years); the assessment case flow (50 per year); the occupation of treatment beds by previously assessed prisoners; and the possibility of dropouts from the treatment. The report concludes that an RCT would be feasible if:

- potential subjects were assessed for DSPD in the eight feeder prisons
- several hundred DSPD prisoners were identified and were willing to participate in the RCT
- the case flow for DSPD assessment in Whitemoor was increased
- the length of the DBT-based treatment was shortened
- DSPD prisoners deemed unsuitable for DBT were given some other treatment.

If an RCT is not feasible, the treatment should be evaluated by using matched treated and control groups, by comparing before and after outcomes in each group, or by statistical adjustment (e.g. in a regression equation) for pre-existing differences between groups. All of these methods require research to devise dynamic risk measures that are related to serious sexual and violent offending. The key dynamic items could be combined into a risk score that was measured before and after the treatment.

Our conclusion is that efforts should continue to create the conditions that would make it feasible to evaluate the Whitemoor DSPD unit using an RCT, because this is the “gold standard” design.

Introduction

The main aim of this report is to assess the feasibility of evaluating the Whitemoor pilot treatment unit for DSPD (Dangerous and Severe Personality Disorder) prisoners using a randomised controlled trial (RCT). The White Paper on *Reforming the Mental Health Act* (Department of Health/Home Office, 2000, Part II, paragraph 6.53) noted that “there is a need to define and refine current treatment goals for offenders with personality disorder and to undertake long term randomised trials with long term follow up”.

The main advantage of an RCT is that it equates persons in the treated group with persons in the control group on all possible (measured and unmeasured) variables that influence key outcomes – providing that there are a sufficiently large number of people (e.g. at least 50) in each group. This equating makes it possible to disentangle the effects of the treatment from the effects of other variables that influence key outcomes (pre-existing differences between the groups). Technically, an RCT has higher internal validity than alternative quasi-experimental evaluation methods that involve matched treated and control groups, the comparison of predicted and actual outcomes in each group, or statistical adjustment for pre-existing differences between groups (Cook and Campbell, 1979; Shadish *et al.*, 2002).

In assessing the feasibility of an RCT, the authors visited Whitemoor, Rampton Hospital, the Home Office and the Prison Service Headquarters and benefited from information supplied by:

Gill Attrill:	Principal psychologist, head of high intensity programmes in the prison service, developing the psychopathy treatment programme
Roberta Babb:	Psychologist at Whitemoor
Jane Bell:	Lead psychologist, high secure services project, DSPD programme
Jamie Bennett:	Head of DSPD Units, HMP Whitemoor
Jeremy Coid:	Professor of Forensic Psychiatry at St. Bartholomew's Hospital
Jason Davies:	Lead psychologist on DBT (Dialectical Behaviour Therapy) at Rampton
Jo Dobry:	Member of the Parole Board
Caroline Doyle:	Psychiatric nurse concerned with selection for the Whitemoor DSPD Unit (she is now at HMP Frankland)
Sue Evershed:	Psychologist at Rampton who has evaluated their DBT-based programme
Gareth Hancock:	Psychiatric Nurse, HMP Whitemoor DSPD Intervention Unit (he has since left)
Val Hawes:	Psychiatrist at Whitemoor concerned with DSPD Unit
Todd Hogue:	Psychologist at Rampton, consultant psychologist to DSPD programme
Ian Keitch:	Lead psychiatrist concerned with DSPD treatment in Rampton and Whitemoor, and a member of the Parole Board
Jane Martin:	Secretary of the South East Multi-Centre Research Ethics Committee (MREC)
Mary Piper:	Public health consultant at prison health policy unit, Department of Health
Deborah Rutter:	Research assistant for Peter Tyrer
Yvonne Schell:	Lead psychologist for assessment in Whitemoor DSPD Unit (she has since left)
Peter Tyrer:	Professor of Forensic Psychiatry, Imperial College Faculty of Medicine, responsible for the evaluation of the DSPD assessment procedure at Whitemoor and Rampton
Mollie Weatheritt:	Member of the Parole Board
Pam Wilson:	Head of psychology in the high security prison estate.

Previous RCTs in British prisons

Randomised controlled trials in criminal justice settings were reviewed by Farrington (1983). Four RCTs in British prisons, borstals or training schools were completed in the 1960s and 1970s, but very few or no large scale ones seem to have been carried out in these settings in the UK since.

Mark Williams (1970, 1975) randomly allocated 610 male borstal inmates to one of three open borstals, differing in their regimes (traditional, group counselling or case-work). The experimental analysis did not begin until these borstals were full of randomly allocated inmates. The inmates treated in the case-work borstal were significantly less likely to be reconvicted in a two-year follow-up period (51%, compared with 63% in the other two conditions, and a 63% rate for all borstal boys at that time). Contrary to the psychologists' predictions, the most disturbed boys did best and the least disturbed boys did worst in the traditional borstal.

Margaret Shaw (1974) randomly allocated 176 male prisoners to receive special case-work assistance from welfare officers during the last six months of their sentences or to a control condition in two prisons (Ashwell and Gartree). The experimental prisoners were significantly less likely to be reconvicted in a two-year follow-up period (57% versus 76%). However, in a replication of this experiment by Tony Fowles (1978) in one prison (Liverpool) with 304 male short-term prisoners, the experimentals were not significantly less likely to be reconvicted in a one-year follow-up period (39% versus 43%). These researchers justified these experiments ethically by pointing out that control prisoners could seek welfare help and were not refused, whereas the welfare officers actively sought out the experimental prisoners to try to help them.

In the fourth experiment, Derek Cornish and Ron Clarke (1975) randomly allocated 173 Kingswood training school boys to one of two houses, respectively operating a therapeutic community and a traditional regime. Reconviction rates in a two-year follow-up period did not vary according to the regime (70% therapeutic community, 69% traditional). Possibly because of these negative results, Clarke and Cornish (1972) became disillusioned with the RCT method and concluded that "it is particularly unlikely that its widespread use [i.e. the widespread use of RCTs] at present would significantly advance our knowledge about institutional treatment in ways that could not be otherwise achieved" (p.21).

Undoubtedly, one of the reasons for the death of RCTs after the 1970s was because of the conventional wisdom at that time that "nothing works" (see e.g. Brody, 1976), which suggested that treatment experiments were unnecessary. However, the climate has now changed, and Home Office-funded RCTs are now being encouraged, as in the above quote from the White Paper (Department of Health/Home Office, 2000, Part II, paragraph 6.53).

There have been some small scale RCTs carried out in British prisons since 1975. For example, Simon Shepherd (1991) randomly assigned 37 young offenders at Swinfen Hall either to an intervention programme or to a control group. The intervention programme used cognitive-behavioural techniques designed to reduce symptoms of anxiety and depression. Before and after self-report measures of anxiety and depression showed that the treatment was effective (although the statistical power of evaluations with such numbers should be borne in mind). The authors have also heard about an RCT carried out in Winchester prison by Cliff Howells on drug treatment but have not managed to obtain any report from it.

Finally, there is an RCT currently underway in Britain involving one prison (HMP Bullingdon). This trial is part of a wider RCT-based study testing the effects of restorative justice at different stages of the criminal justice system and within four different agencies: Metropolitan Police, Northumbria Police, Thames Valley Area of the National Probation Service, and the Prison Service. This study is still in the middle of its work, however, and there are not any results yet available.

HMP Whitemoor

Selection for the Whitemoor DSPD unit

Potential DSPD prisoners for assessment in the Whitemoor DSPD unit are referred by psychologists from the eight prisons in the high security estate (Belmarsh, Durham, Frankland, Full Sutton, Long Lartin, Wakefield, Whitemoor and Woodhill). The DSPD unit has been publicised widely. For example, a “road show” has been held in all eight prisons, encouraging staff to refer prisoners. An “open day” was held at Whitemoor, attended by prison staff, psychologists and psychiatrists from the eight prisons. There was an article about the Whitemoor unit in the prisoners’ newspaper *Inside Times*. However, the eight prisons have referred rather different numbers of prisoners. The DSPD unit has so far received prisoners mainly from Wakefield, Frankland and Whitemoor. This may be at least partly because of the willingness of the staff at these prisons to refer: all prisons received the same amount of information. However, it may also relate to the fact that staff in Whitemoor (which has a DSPD unit) and Frankland (which will have a DSPD unit) are particularly aware of DSPD prisoners.

Possibly eligible prisoners are screened by a psychiatric nurse in Whitemoor’s DSPD unit. The DIA (Dispersion Induction Assessment) is consulted where it is available, giving assessments for dyslexia, basic literacy and numeracy skills, the CNI (Criminological Needs Interview), the MCMI (Millon Clinical Multiaxial Inventory), the WAIS (Wechsler Adult Intelligence Scale), drug tests and prison officer reports. It used to include the PCL-R (the Psychopathy Checklist-Revised) but this has generally been suspended. However, the PCL-R can be completed as part of the DIA after a recommendation by a psychologist. According to the Prison Service some prisoners are reluctant to have PCL-R assessments because of the possible negative consequences for them of having high PCL-R scores.

The screening consists of reading the prisoner’s files (including medical and psychiatric records, probation reports and histories), interviewing staff, and compiling evidence about the prisoner’s personality disorder and dangerousness (based on their criminal history). Each prisoner is then interviewed to tell them about HMP Whitemoor and his rights; to ensure that he is not mentally ill; and to see if he is willing to come to the DSPD unit. This information is then fed back to the multidisciplinary DSPD team at HMP Whitemoor who then select prisoners who seem likely to be DSPD to come for the full assessment at HMP Whitemoor. In general, disruptive prisoners, and those considered likely to move out of the DSPD unit, are screened out.

Selection for the DSPD unit is voluntary. However, it is believed in Whitemoor that many prisoners are keen to receive treatment that will reduce their risk and hence (in their eyes) increase their probability of being released. On 26 April, 2002, there were 21 prisoners who had been screened, selected and were waiting for admission to Whitemoor, and a further 39 were waiting to be screened.

The definition of DSPD

The concept of “Dangerous and Severe Personality Disorder” (DSPD) was introduced in the July 1999 Green Paper on *Managing Dangerous People with Severe Personality Disorder* (Home Office/Department of Health, 1999). In the December 2000 White Paper on *Reforming the Mental Health Act*, DSPD individuals are defined as those “who pose a significant risk of serious harm to others as a result of their severe personality disorder” (Department of Health/Home Office, 2000, Part II, paragraph 1.5).

Of course, the definitions of what is “significant” risk, “serious” harm and “severe” personality disorder can be rather subjective. Severe personality disorder overlaps with “psychopathy” but these are two different concepts. The 1999 Green Paper stated that the overwhelming majority of DSPD persons would have committed offences such as murder, manslaughter, grievous bodily harm, arson or serious sex crimes (Home Office/Department of Health, 1999, p.9). An estimate of the number of DSPD prisoners was obtained from the Office for National Statistics (1998) survey of psychiatric morbidity in prisoners. DSPD cases were defined as those with antisocial personality disorder and six out of ten common risk factors (not specified) for recidivism. On this basis, it was estimated that there were about 1,420 adult male sentenced prisoners in this extreme category (Home Office/Department of Health, 1999, p.34).

The DSPD assessment process was developed by a group chaired by David Thornton (Department of Health/Home Office, 2000, Part II, paragraph 6.38). The operational definition of DSPD is as follows (Thornton and Hogue, 2001):

(a) *Dangerousness*: A significant risk of committing a serious violent or sexual offence after release (within the next 15 years). This is assessed using a number of risk assessment tools:

(i) HCR-20 (Historical, Clinical, Risk Management Items; Webster *et al.*, 1997)

(ii) VRAG (Violence Risk Appraisal Guide; Quinsey *et al.*, 1998)

- (iii) Static-99 (Hanson and Thornton, 1999)
- (iv) SVR-20 (Sexual Violence Risk; Boer *et al.*, 1997)
- (v) Risk Matrix 2000 (Thornton *et al.*, 2001)

There is some repetition of the items in these instruments and they are likely to be highly intercorrelated. (This is currently being investigated by the DSPD assessment evaluation.) Generally, the risk must be greater than 50 per cent according to at least two of the instruments for an individual to be assessed as dangerous.

(b) *Personality disorder*: This is measured using the PCL-R (Hare, 1991) and IPDE (International Personality Disorder Examination). Prisoners must have either:

- (i) a PCL-R score of 30 or more; or
- (ii) a PCL-R score of 25 to 29 plus at least one personality disorder according to ICD-10 or DSM-4 criteria (excluding antisocial); or
- (iii) two or more personality disorders according to ICD-10 or DSM-4 criteria (one of which can be antisocial).

(c) *Link*: There is a demonstrable link between personality disorder and dangerous behaviour in at least two past offences and/or a link between personality disorder and offence-like behaviour in prison. This is measured using file data and the IPRF (Individual Psychological Risk Factor) inventory (Hogue and Thornton, unpublished).

In order to be assessed as DSPD, a prisoner must have (a) risk, (b) personality disorder, and (c) the link. This has been described as:

“The operational definitions of risk and severe PD being used in the pilot project were arrived at somewhat pragmatically as no such definitions existed especially in the recognised clinical literature. With regard to risk we have tried to identify only those who are at greatest risk of serious violent and/or sexual reoffending. Unfortunately many of the risk assessment tools do not give the specificity we would have hoped for... With regard to severe PD we again have tried to identify only the most severely affected individuals, hence the need for people to meet quite stringent criteria.”

There is a great deal of research, mainly carried out in Canada, and to a lesser extent in Sweden, on the predictive accuracy of the HCR-20, VRAG and PCL-R. For example, Rice and Harris (1995) found that the VRAG predicted violent recidivism among offender populations, although results in a sex offender population were less impressive (Rice and Harris, 1997). Douglas *et al.* (1999) reported that the HCR-20 predicted later violence among a psychiatric population, and that the H (Historical) scale was the best predictor. Grann *et al.* (2000) showed that the VRAG and the H scale predicted violent recidivism among mentally disordered offenders in Sweden. Hemphill *et al.* (1998) carried out a meta-analysis of studies investigating the ability of the PCL-R to predict violent and sexual recidivism. However, no studies seem to have examined correlations between changes in PCL-R scores over time and changes in offending. Hanson and Thornton (2000) found that the Static-99 predicted violent and sexual recidivism in Canadian and British samples.

There have also been studies comparing different risk assessment instruments. For example, Douglas and Webster (1999) reported that the H scale of the HCR-20 correlated 0.62 with the VRAG and 0.50 with the PCL-R. Sjostedt and Langstrom (2002) compared the VRAG, the PCL-R, the SVR-20 and the Rapid Risk Assessment for Sex Offence Recidivism (a component of the Static-99) in predicting recidivism among personality-disordered rapists in Sweden. Only the RRASOR predicted sexual recidivism, but the other three instruments predicted violent non-sexual recidivism. Kroner and Mills (2001) in Canada compared the ability of five instruments (including the VRAG, HCR-20 and PCL-R) to predict recidivism among violent offenders, and found that the HCR-20 was the best predictor of future violence, but the VRAG was the best predictor of reconvictions in general. More research of this kind needs to be carried out in Great Britain.

Assessment in the Whitemoor DSPD unit

The assessments detailed above are completed for each prisoner in the DSPD unit. In addition, prison staff complete the DBRS (Daily Behaviour Rating Scale) and prison files are consulted. A 22-page assessment report is then completed by the psychology team for each prisoner. The authors wonder whether it is necessary to have such a detailed assessment procedure.¹ If the assessment period were shorter, an RCT would be more feasible

¹ The reason for the length of the assessment process relates not only to the time it takes to conduct the structured risk assessments but also allows time for detailed behaviour observation, case formulation and coming to a judgement about the link between risk and personality disorder.

because more prisoners could be treated (see the section, “Can a feasible RCT be designed?”). There is then an intervention case conference for each prisoner that yields a clinical judgement about whether he is suitable for the proposed intervention based on DBT (Dialectical Behaviour Therapy).

The Whitemoor DSPD unit, D-wing, contains “Red Spur”, “Blue Spur” and “Green Spur”, with a total of 90 beds. For some, conditions in the DSPD unit can be more pleasant than in their feeder prisons. The prison staff are specially selected, highly motivated and committed. Prisoners are given normal employment; a huge step forward for some who are typically unable to hold down a job in a conventional prison environment. Through this opportunity, some relatively good money can be made by prisoners – £15 a week compared to £2.50 when doing no work.

The DSPD assessment procedure and the DSPD treatment programme are quite separate and will be evaluated separately. One of the aims of the DSPD assessment procedure is to compare the existing assessment instruments with the Personality Assessment Schedule – Severe Personality Disorder version (Tyrer, 2000). Another aim is to assess the effect of the existing assessment procedure on violent episodes and costs, using an RCT. Finally, it also aims to assess the reliability and validity of all instruments, and their predictive accuracy in relation to the DSPD diagnosis. There is also a qualitative element to the project.

The assessment is completed in “Red Spur”, which has a maximum of 36 beds. It is planned that two prisoners will be admitted every two weeks and that the assessment process will take 17 weeks. Assessments began in September 2000 and the DSPD unit was closed for building work between April and July 2001. The original assessment procedure included some group work treatment which the prisoners liked. This was dropped because it was considered that treatment should not be included in the assessment. Of the first 31 prisoners admitted to the DSPD unit before April 2001, 22 completed the assessment. These prisoners had heterogeneous offending histories. Their current offences were murder or manslaughter in six cases, rape and murder in five cases, rape in two cases, armed robbery in four cases, sexual offences against children in two cases, plus one case each of attempted murder using a firearm, GBH and false imprisonment. Eleven were in the “Blue Spur” treatment unit in February 2002.

Of the 63 prisoners admitted to the DSPD unit up to March 2002, 32 had completed the assessment. Of these, 13 were considered suitable for the intervention based on DBT, 13 were considered unsuitable (2 of whom were classified as not DSPD), and six were unclear. Hence, it seems likely that roughly half of the assessed prisoners will be considered suitable for the intervention based on DBT. Of the 16 still being assessed (1 of whom was not DSPD) five were considered suitable for DBT. Fifteen prisoners did not complete the assessment: eight withdrew and seven were removed (1 inappropriately placed near to the victim, 1 inappropriate sex behaviour, 1 intimidation, 1 assault, 1 refused drug test, 1 mental illness, 1 not DSPD). Only four of the 63 admissions were classified as not DSPD.

The DSPD assessment procedure may be explicitly specified in future legislation. The White Paper on *Reforming the Mental Health Act* (Department of Health/Home Office, 2000, Part II) proposed new compulsory powers of detention and treatment. Prisoners would be required to undergo a four-week screening assessment to establish whether there was sufficient evidence of DSPD to justify a longer-term intensive assessment (paragraph 6.39). After the case was considered by a Mental Health Tribunal, a full specialist DSPD assessment could be required, lasting 12 weeks (paragraph 6.42). If and when these proposals become law, the DSPD assessment process would not be voluntary. Compulsory assessment would pose problems; in particular, prisoners may be even more likely to tell lies in interviews in an attempt to avoid the consequences of being labelled as DSPD.

Treatment in the Whitemoor DSPD unit

Dialectical Behaviour Therapy (DBT) within a structured therapeutic setting has been chosen as the basis of the treatment to be used in “Blue Spur” and “Green Spur” in Whitemoor. This was mainly because of the encouraging results obtained with DBT in Rampton and partly because many of the potentially DSPD prisoners who were first assessed in “Red Spur” showed borderline traits that DBT could address. However, later intakes have been more likely to show solely psychopathic, rather than borderline, traits. There will also be motivational work to get prisoners ready for DBT, and criminogenic needs work after DBT. The primary aim of the DBT-based treatment is to treat the personality disorder rather than to reduce the risk of violent or sexual offending. However, the precise nature of the intervention has not yet been finalised.

“Blue Spur” and “Green Spur” each have 27 beds, giving a total of 54 treatment beds. The DBT-based treatment is scheduled to begin in April 2003. In the meantime, there is a “supportive regime”, and in April 2002, an activity centre opened, providing classroom activities such as art, design and information technology.

DBT is a manualised therapeutic approach developed by Linehan (1993a, 1993b). It was originally developed as a treatment for self-harming women with Borderline Personality Disorder. It is based on a model of Borderline Personality Disorder as consisting of impairments in interpersonal abilities, emotion regulation, tolerance of distress and self-control. DBT aims to improve these abilities using weekly individual behavioural psychotherapy combined with group sessions of skills training and telephone consultation with a therapist as and when required.

A detailed systematic review of the literature on the effectiveness of treatment for DSPD (Warren *et al.*, 2001), concluded that "there is no convincing evidence that psychopathic or antisocial personality disorder can or cannot be treated" (p.164). In regard to DBT:

"There is little from the studies of CBT [Cognitive-Behavioural Treatment] that suggests that any one approach is more promising than any other in terms of research evidence. DBT is marginally ahead, although the authors of the study conducted with patients in high security suggest that DBT (or CBT in general) may not be suitable for patients with limited cognitive capacity... All the evidence reviewed in this chapter regarding DBT pertains to the treatment of women. This treatment is, as yet, untested in male patients" (p.90).

In the Personality Disorder unit at Rampton Hospital, it was found that most of the violent impulsive men had Borderline Personality Disorder or Borderline traits. About five years ago, a modified version of DBT began to be used. The aim of the treatment is to provide individuals with skills to manage their impulsivity and emotional dysregulation (mood swings, anger, aggression, irritability and self-harm) and thereby reduce the risk of violence to themselves or others. An evaluation by Sue Evershed (2002), based on psychometric measures before, during and after treatment, showed that the treated men benefited from the intervention, with better coping skills, lower impulsiveness and a lower violence propensity. This evaluation was based on small numbers (8 treated and 9 comparison forensic patients with similar personality disorders). A DBT-based approach is also being used in Colorado by Robin McCann, with similar benefits (see McCann *et al.*, 2000).

DBT treats the PD that interferes with conventional treatments for sexual and violent offending. It is therefore often combined with other treatments (e.g. for deviant sexual arousal). Essentially, DBT is a vehicle to motivate and engage patients and provide them with skills so that they can benefit from other treatments to decrease their PD and dangerousness. The treatment goals and targets are specifically identified for each patient and therefore, although the DBT skills groups are manualised, the individual focus of the work is not. It requires a high level of clinical experience and is likely to change over time in light of experiences with particular types of patients. At Rampton, modifications have been made to the standard DBT programme to work with an offending population. However, it would be possible to manualise this specific version in a relatively short time.

Based on experience in the Whitemoor DSPD Unit, about 25 of 50 prisoners assessed in a year might be considered suitable for DBT. The main criterion for suitability is whether the prisoner shows emotional dysregulation. In general, disruptive and psychopathic prisoners are excluded. Men with high PCL-R scores are not excluded if they have emotional dysregulation and it is considered that they would benefit from DBT. The "unsuitable" prisoners are returned to prisons in the high security estate. One possibility in the future would be for these prisoners to receive the psychopathy treatment programme that is currently being developed and is due to commence piloting in March 2004. This is designed for prisoners scoring 25 or above on the PCL-R. It is a cognitive behavioural treatment targeting risk factors such as egocentricity, impulsivity and hostile attributions and aiming to change antisocial thinking and socio-cognitive skills deficits (Attrill, 2001).

Outcome measures

Since few of the DSPD prisoners will be released within the next three years and thus not in a position to reoffend, the main outcome measures used in any evaluation would have to be based on psychometric tests, ratings, behaviour and prison disciplinary offences. The main aims of the DBT-based treatment are to improve personal and social functioning, to improve emotion regulation, to reduce the risk of harm to self and others, to reduce the person's distress, and to improve the person's amenability to treatment for sexual and violent offending. Therefore, the tests and observations should be chosen to assess how far these aims have been achieved.

Items on many tests can be divided into static versus dynamic ones. Static items can never change (e.g. age at first violent incident) whereas dynamic items can change (e.g. impulsivity). For example, the first ten items on the HCR-20 (Webster *et al.*, 1997) are static, while the second ten items are dynamic. The instruments currently used to assess risk or dangerousness contain mainly static items. The PCL-R contains both static and dynamic items. Dynamic items are obviously more relevant for assessing change over time. For example, all HCR-20 items may be assessed in a pre-test (to assess risk for violence) but only the ten dynamic items in a post-test. The IPRF contains 26 dynamic items and so might be used in assessing treatment effects. In regard to the risk for sexual violence, we understand that David Thornton has developed a "Structured Risk Assessment" for use in the Sex Offender Treatment Programme. This covers four risk domains: sexual interests, distorted attitudes, socio-affective functioning and self-management.

Any outcome measures would have to be usable not only in Whitemoor (for experimentals) but also in many other settings (for controls). This is one reason why complex or highly specialised measures such as the PPG (penile plethysmograph) would not be feasible. The Daily Behaviour Rating Scale currently being used in Whitemoor to observe behaviour seems potentially useful and usable in other institutional settings, although a better developed instrument such as the Prison Behaviour Rating Scale (Cooke, 1998) could be used. Useful psychometric tests that could detect change and that could be used widely include the Personality Assessment Schedule (Tyrer 2000), the Custodial Adjustment Questionnaire (Thornton, 1987), the Emotion Control Questionnaire (Roger and Najarian, 1989), the Psychological Inventory of Criminal Thinking Styles (Walters, 1995a, 1995b, 1996), the Level of Service

Inventory, Revised (Andrews and Bonta, 1995), the General Health Questionnaire (Goldberg, 1978), the Beck Depression Inventory (Beck and Steer, 1987), the Barratt Impulsiveness Scale (Barratt, 1994), the Novaco Anger Scale (Novaco, 1980) and the Sex Offender Need Assessment Rating (Hanson and Harris, 2000). However, it is unclear how far DSPD prisoners would give valid responses on self-report questionnaires. Research on validity is essential. If DSPD prisoners do not give valid self-reports, outcome measures would have to be based on ratings, prison records and behavioural observations.

It would be useful to carry out a research project in which a sample of prisoners in the high security estate completed all these instruments. The main aim of this project would be to determine which dynamic items were correlated with past serious violent and sexual offending. These dynamic items could then be used to develop a dynamic risk scale for estimating the effect of DBT, or any other treatment. In the interests of drawing independent conclusions, this research should be carried out by persons who were not involved in constructing these instruments and have no personal stake in them.

Randomised Controlled Trials

General problems of an RCT

Farrington (1983) strongly advocated the use of RCTs and discussed both their advantages and problems. Problems associated with the intervention included: How far is the treatment based on theory? What was the exact nature of the treatment implemented? Did the treatment change over time? Was there a Hawthorne effect contingent upon the extra attention given to the experimental group? How far did all the experimental subjects receive treatment? Were experimental subjects who co-operated unrepresentative? How far did all the control group not receive treatment? Was there differential attrition from experimental and control conditions? What was the “active ingredient” of the treatment that caused the effect? How far was there an interaction between types of subjects and types of treatments?

Problems associated with outcomes included: How far were outcome measures guided by theory? How far were outcome measures reliable and valid? How can problems associated with reconviction be overcome (e.g. separation of rehabilitation from individual deterrence, gaps between reoffending and reconviction, high reconviction probabilities, underestimation of true offending)? How far are follow-up results biased by the knowledge of who are experimentals or controls? What should be the length of the follow-up period (e.g. depending on the causal lag between treatment and effect)? It is desirable to have pre-test and post-test measures of offending and measures of effect size as well as statistical significance; issues of statistical power need to be considered in the design (see later).

Practical problems include how to get co-operation from policy makers and programme staff, how to avoid case flow problems (which always seem to occur with RCTs), and how to maintain the randomisation in the face of practical issues. For example, in evaluating the impact of Colchester YOI, Farrington *et al.* (2000, 2002) began randomly allocating young offenders to Colchester or to control institutions, but the randomisation could not be maintained because of: the need to allow young offenders to complete educational courses e.g. for A levels; and the problem that family members could not visit if young offenders were allocated to a YOI hundreds of miles from their homes. Another practical problem is how to achieve a “double-blind” trial where neither the subjects nor the evaluators know who is in the experimental condition.

Statistical power analysis

Statistical power analysis can be used to specify what sample size is needed for given values of the effect size, statistical significance and statistical power. Statistical significance, conventionally set at $p=0.05$, is the probability of falsely rejecting the null hypothesis of no effect when it is true. Statistical power, set at a minimum of 0.70, is the probability of correctly rejecting the null hypothesis when it is false.

What is the likely effect size of the DBT-based treatment? This is unknown. Perhaps the most useful information is provided in the meta-analysis by Redondo *et al.* (1999) of the influence of European treatment programmes on recidivism. The average effect size (d) over all subjects and all programmes was 0.24, with a 95% confidence interval between 0.20 and 0.29. This 0.24 effect size corresponds to a change in the percentage reconviction rate from 50 per cent to 38 per cent (Lipsey and Wilson, 1998). Behavioural ($d = 0.48$) and cognitive-behavioural ($d = 0.46$) programmes were the most effective, and the effectiveness of programmes with juveniles ($d = 0.36$) was greater than with adults ($d = 0.20$). Since it is proposed to use a cognitive-behavioural treatment with adults, it might be predicted that the most likely effect size will be between $d = 0.20$ (a 10% reduction in recidivism) and $d = 0.40$ (a 20% reduction in recidivism). An effect size of 0.10 (a 5% reduction in recidivism) seems almost too small to be worth detecting.

Given that the DBT-based treatment is predicted to improve functioning and decrease recidivism (a directional prediction), it would be defensible to use a one-tailed test of statistical significance. However, Table 1 shows sample sizes required for both one-tailed and two-tailed tests. In Table 1 (obtained using the “Power and Precision” statistical package), power is set at 0.70. For example, the table shows that with 50 in the treated group, a control group of at least 172 is required to detect an effect of $d = 0.35$ at $p = 0.05$ (one-tailed). The problem is that the sample size in the treated group is non-linearly related to statistical power. The required size of the control group decreases dramatically as the number of treated prisoners increases from 50 to 80 or 100.

Table 1 shows that, if the effect size is 0.30, 60 treated prisoners could be compared with 306 controls in order to achieve statistical significance at $p = 0.05$ (one-tailed). If the effect size is 0.25, 100 treated prisoners would have to be compared with 381 controls to achieve significance in a one-tailed test.

An alternative approach is to ignore statistical power and merely investigate the sample sizes required to achieve a statistically significant result for different values of d .

These are shown in Table 2. With $d = 0.25$ and 50 treated prisoners, 320 controls would be required to reach statistical significance at $p = 0.05$ (one-tailed). If $d = 0.20$, 80 treated prisoners and 435 controls would be required to reach statistical significance at $p = 0.05$ (one-tailed).

Table 1: Sample sizes required to detect different effect sizes with statistical power = 0.70

Treated N	Minimum Control N				
	Effect Size (d)				
	.20	.25	.30	.35	.40
One-Tailed					
50	x	x	x	172	62
60	x	x	306	110	52
80	x	x	135	76	43
100	x	381	101	64	39
125	1386	217	84	57	36
150	487	168	76	53	34
Two-Tailed					
50	x	x	x	x	133
60	x	x	x	338	92
80	x	x	376	141	67
100	x	x	194	104	57
125	x	622	140	86	52
150	x	340	118	78	48

x = Over 2000.

Table 2: Sample sizes required for statistical significance at $p = 0.05$

Treated N	Minimum Control N				
	Effect Size (d)				
	.20	.25	.30	.35	.40
One-Tailed					
50	x	320	73	37	23
60	x	153	58	33	21
80	435	92	46	28	
100	207	74	41		
125	145	64	37		
150	121	59			
Two-Tailed					
50	x	x	288	81	43
60	x	x	145	62	37
80	x	262	88	48	
100	x	156	71		
125	411	118	61		
150	264	101			

x = Over 2000

Given that the effect size seems unlikely to exceed 0.30 (corresponding to a 15% decrease in recidivism), the main implications of these analyses are that an increase in the treated group above 50 is highly desirable and the control group should contain at least 300 prisoners. For illustrative purposes, this report assumes that a total of 400 prisoners are randomly allocated to treated (50) or control (350) groups. These numbers would apply in both experimental and quasi-experimental evaluations.

Ethical issues

The ethical issues obviously depend on the particular design of the RCT. The authors assume that potentially DSPD prisoners will be identified in the eight prisons in the high security estate and will be asked if they are willing to participate in an RCT to evaluate the treatment in the Whitemoor DSPD unit. It is also assumed that about 400 will be identified as willing and will be randomly allocated to experimental (50) or control (350) groups.

The first ethical issue arises from the denial of the DBT-based treatment to those in the control group. All the clinicians interviewed were concerned about this, and generally were opposed to an RCT because they thought that everyone should be treated. However, the problem is that the number of DSPD treatment beds is not sufficient to treat all prisoners who are likely to be defined as DSPD. The Office for National Statistics (ONS) estimate of 1,420 DSPD prisoners (Home Office/Department of Health, 1999, p.34) is vastly in excess of the 50 treatment beds in the Whitemoor DSPD unit, and also in excess of the total 300 DSPD high security beds that are expected to be available by 2004. Where the number of persons who want treatment greatly exceeds the number who can be treated, random allocation may be the fairest way to select prisoners for treatment. It is also possible to select 50 treated prisoners every year (or every two years) out of the eligible population, so that control prisoners at one time would have the chance of going to the DSPD unit at a later time. Also, of course, it is not known whether the DBT-based treatment is effective, and hence whether control prisoners would have derived any benefit from it (or indeed any harm).

For the report, the secretary of the South East Multi-Centre Research Ethics Committee; Jeremy Coid, Professor of Forensic Psychiatry at St. Bartholomew's Hospital; and a public health consultant at the Prison Health Policy Unit, Department of Health were consulted about ethical issues. In addition, Sue Evershed, whose evaluation project in Rampton was approved by an ethics committee was also interviewed. Sue Evershed said that her ethics committee was concerned that no treatment was offered to the control group. Also, the controls were less co-operative than the experimentals in completing assessments, because they thought that there was nothing in it for them. The Prison Service's psychopathy programme is currently carrying out research on motivational factors in psychopaths.

A key issue for MREC is that prisoners should freely consent to participate in research without any inducements, and should be able to refuse or withdraw at any time.

Currently, prisoners are keen to be treated in the DSPD unit because they hope that it will decrease their perceived risk and hence improve their chances of release on parole. The authors consulted representatives of the parole board to discuss parole issues. They thought that a prisoner might benefit if he had received an effective treatment, although in practice prisoners (including those in the control group) may receive many different programmes, any single programme may not be crucial, and the effectiveness of the DBT-based treatment is unclear. The Parole Board receives a great deal of information about each prisoner, and the actual programmes attended are less important than the conclusions drawn by report writers. The representatives of the parole board were much more concerned about the DSPD label, which they felt could stigmatise prisoners in the eyes of the Parole Board. Being labelled as DSPD seemed to be the main risk for prisoners in agreeing to participate in an experiment at present. (As mentioned, new legislation in the future might make the DSPD assessment compulsory, although there is no guarantee that the term 'DSPD' would remain.) The probability of DSPD prisoners on determinate sentences being released early would be very low irrespective of what treatments they had received. Prisoners who were allocated to the control group would have few risks or benefits.

An important ethical argument is that the benefits of an RCT (in proving most convincingly whether or not a treatment is effective) might outweigh its costs (e.g. of denial of treatment to control prisoners). In light of the potential risk to the community of DSPD persons, it is important that they receive treatment in prison that has been proved to be effective and cost-effective. The authors are cautiously optimistic that an RCT to evaluate DBT-based treatment in the Whitemoor DSPD unit could be approved by MREC (under present legislation).

Specific problems of an RCT to evaluate the DSPD unit

The main problems threatening the feasibility of an RCT to evaluate the Whitemoor DSPD unit are as follows:

- the current case flow for assessment of 50 prisoners per year
- the small number of DSPD prisoners who will receive DBT. If only 25 prisoners were in the treated group, this number would be too small to achieve the main benefit of randomisation of equating treated and control groups on all possible (measured and unmeasured) extraneous variables that influence outcomes. Also, the number would be too small to detect the effects of the treatment reliably (see the power analysis above)
- the heterogeneity of the DSPD prisoners. This is especially a problem when combined with the small number. The clinicians, especially, emphasised the variability of DSPD prisoners
- the projected length of the DBT-based treatment: 2 to 3 years. This again compounds with the small numbers resulting in a very small sample of treated DSPD prisoners
- the fact that the DBT-based treatment is not constant for every prisoner (since it involves individual one-to-one work) and is likely to change over time, in the light of experience
- the possibility of dropouts from the treatment. DSPD prisoners are not very stable and keep moving between different settings (e.g. to the segregation unit, to other prisons). This is a problem at present because the treatment is voluntary

- the fact that ratings of prisoners are not blind to condition (treated versus control) may bias them
- the possibility that DSPD prisoners may not give truthful answers in interviews.

Many of these problems are interconnected, and many also would be problems in a quasi-experimental evaluation (not using an RCT). The uncertainty of the new legislation poses a further problem, but this report will discuss these issues in the context of the present legislation.

Prisoners could be randomly allocated to treated or control groups either before or after reception to the Whitemoor DSPD unit. However, there are many problems of allocation after reception. First, few prisoners are currently assessed: about 50 per year. Second, prisoners come to the unit expecting treatment. Third, only about 25 out of each 50 prisoners who are assessed are likely to be considered suitable for the DBT-based treatment. There would be little point in carrying out an RCT based on a total N of 25 (e.g. 12 treated and 13 control), because this number would not ensure the equivalence of the experimental and control groups.

An RCT is likely to be more appropriate and feasible if prisoners are assessed in the eight feeder prisons and then randomly allocated to treated or control groups. As mentioned, out of 400 eligible prisoners, 50 could be randomly allocated to the Whitemoor DSPD unit in the first year and 350 could be randomly allocated to a control group. However, if only 25 out of 50 prisoners who are randomly allocated to the DSPD unit receive the DBT-based treatment, this is a problem. The random allocation ensures comparability of the 50 treated to the 350 controls, but it does not ensure comparability of the 25 who receive the DBT-based treatment to the 350 controls.

In analysing the results of the experiment, the 50 treated prisoners would have to be compared to the 350 controls. However, if only half of the treated prisoners received DBT, this would not be a good test of the effects of the DBT-based treatment. This problem would be alleviated somewhat if the other 25 received some other kind of treatment (e.g. for psychopathy). Then, the success of the overall treatment strategy (either DBT or psychopathy) could be determined by the RCT, and the success of each individual type of treatment could be determined quasi-experimentally by comparing pre-test and post-test assessments.

If the DBT-based treatment lasts at least two years, this would limit the number of treated subjects in an RCT to about 25 in the time period April 2003 – April 2005 (assuming that the treatment begins in April 2003). Arguably, the success of a treatment should only be evaluated when it has been completed. On some present assumptions, 50 prisoners would not have completed their treatment in the Whitemoor DSPD unit until April 2007.

Can a feasible RCT be designed?

It is clear from the report so far that an RCT would be desirable and feasible in some circumstances but not in others. It would be most feasible if:

- potential subjects were assessed for DSPD in the eight feeder prisons
- several hundred DSPD prisoners were identified by the assessment and were willing to be treated in the DSPD unit in Whitemoor
- the case flow for DSPD assessment in Whitemoor was increased
- the length of the DBT-based treatment was shortened to one year
- DSPD prisoners deemed unsuitable for DBT were allocated to some other treatment programme
- an outcome measure was developed consisting of dynamic items that predicted serious sexual and violent offending.

There are various other circumstances that are not considered in this report. For example, if the operational definition of DSPD was made less restrictive, more DSPD prisoners would be identified in the high security estate. If the operational definition of DSPD was simplified to include fewer instruments, the DSPD assessment would take less time. Generally, the current assessment system is taken as given (it is being separately evaluated by Professor Peter Tyrer for the DSPD programme) and the report does not question whether the DBT-based treatment is the best choice.

The feasibility of an RCT would be increased if potential subjects were assessed for DSPD in the eight feeder prisons. At present, prisoners in the five dispersal prisons (but not the three core locals) receive the DIA, which takes about eight to ten days per inmate. As explained above, the DIA includes the CNI, MCMI, WAIS and other basic assessments, although this is currently being audited by the Home Office's Research, Development and Statistics Directorate (RDS) to confirm. The current DSPD assessment in Whitemoor is largely based on eight instruments: the HCR-20, VRAG, Static-99, SVR-20, Risk Matrix 2000, PCL-R, IPDE and IPRF. Assuming that each instrument takes two days to complete, it does not seem impossible for this assessment to be completed in

about 16 days per inmate. In principle, therefore, (given sufficient resources), sentenced prisoners in the eight feeder prisons could be assessed for DSPD in four weeks.²

The feasibility of an RCT would be increased if several hundred DSPD prisoners were identified in feeder prisons. This could be achieved most efficiently by first screening prisoners on the basis of their current (or past?) offences, using the Inmate Information System (IIS). The identification of possibly DSPD prisoners could be initially on the basis that they have committed an offence falling within the Crime (Sentences) Act 1997 or CSA: murder, attempted murder, conspiracy to murder, manslaughter, wounding with intent to do GBH, robbery involving the use of real or imitation firearms, possession of a firearm with intent to endanger life, rape, attempted rape or unlawful sexual intercourse with a girl under 13. The CSA specifies that an offender aged 18 or over committing a second offence of this type should receive an automatic life sentence unless there are exceptional circumstances.

Table 3 shows that there were 2,840 inmates in the eight feeder prisons on 28 February, 2002 whose current offence fell within the CSA. If about half of these prisoners were severely personality disordered, it might be estimated that there would be about 1,420 DSPD prisoners, surprisingly agreeing with the ONS estimate for the prevalence of DSPD prisoners. The ONS (1998) also estimated that about 70 per cent of all prisoners had an identifiable personality disorder. However, the main point is that there would be hundreds of eligible DSPD prisoners at any given point in time (especially if past as well as current offences were considered).

Table 3: Number of male prisoners in the high security estate who have committed CSA crimes

Crime	No.	Prison	No.
Murder	938	Belmarsh	372
Attempted Murder	123	Durham	150
Conspiracy to Murder	26	Frankland	509
Manslaughter	68	Full Sutton	478
Wounding/GBH	256	Long Lartin	317
Robbery with intent	495	Wakefield	493
Possess Firearm with Intent	124	Whitemoor	296
Rape/Attempted Rape	804	Woodhill	225
Unlawful sex under 13	6		
Total	2,840	Total	2,840

Note: CSA = Crime Sentences Act (1997)
Data from 28 February, 2002.

Table 4 divides up the 2,433 sentenced prisoners out of the 2,840 according to whether they were serving life (1,301, or 53%) or determinate sentences (1,132, or 47%). About one-third had already passed their EDR or earliest date of release (9%) or would reach it within three years (26%). The remaining 12% would not reach their EDR within the next three years. It should be noted that DSPD prisoners with a relatively short time to serve will still be assessed and treated. It is planned that they will receive individualised treatment that prepares them for release.

Table 4: Time to serve of male prisoners in the High Security Estate who have committed CSA crimes

Prison	Serving Life	Passed EDR	EDR within 3 years	EDR over 3 years	Total
Belmarsh	31	13	69	29	142
Durham	18	6	38	3	65
Frankland	224	54	157	74	509
Full Sutton	254	53	104	66	477
Long Lartin	203	26	56	32	317
Wakefield	360	27	68	38	493
Whitemoor	174	26	57	38	295
Woodhill	37	13	72	13	135
Total	1,301	218	621	293	2,433

Note: CSA = Crime Sentences Act (1997)
Data from 28 February, 2002

EDR = Expected Date of Release (based on automatic release date, conditional release date or parole eligibility date, depending on which one was relevant)

² However, again it must be noted that the structured risk assessments are only one part of the assessment process at HMP Whitemoor, with other factors such as behavioural observation, case formulation and professional judgement also requiring substantial time.

It seems likely that most DSPD prisoners identified in the feeder prisons would be willing to come to the Whitemoor DSPD unit if they thought that they would receive treatment that would reduce their risk and improve them in other ways.

The feasibility of an RCT would be increased if the Whitemoor assessment period was shortened. If the assessment for DSPD was made in the feeder prisons, it may be feasible to determine suitability for the DBT-based treatment in four weeks. This might mean that each assessment bed could be used for (say) ten prisoners per year. If there were 15 assessment beds, this might allow 150 prisoners to be assessed per year, producing (say) 75 prisoners deemed suitable for the DBT-based treatment. This is a hypothetical example, but the general point is that reducing the Whitemoor assessment period would allow more prisoners to be assessed and treated. Given the total of 90 beds in D-wing, possibly 75 beds could be used for treatment. Increasing the number of prisoners in the treated group from 50 to 75 would have disproportionate benefits according to the power analysis discussed above.

The feasibility of an RCT would be increased if the length of the DBT-based treatment was shortened to one year. Taking the above example, possibly 75 DSPD prisoners could be treated in the first year, and a new allocation procedure could then occur in each succeeding year.

A possible RCT flow chart

In any RCT, it is important to specify a flow chart of cases, with numbers. Figure 1 specifies a possible flow chart for an RCT to evaluate the DBT-based treatment provided by the Whitemoor DSPD unit. Initially, N1 CSA prisoners could be identified in the high security estate, using the IIS, as explained above. A certain number (N2) of these N1 prisoners would then be assessed in the high security estate to determine if they were DSPD, and conversely (N1 – N2) would not be assessed for various reasons (e.g. because they refuse).

N3 prisoners would then be identified as DSPD, and conversely (N2 – N3) prisoners would be identified as not DSPD. The expectation is that N3 would be at least many hundreds at any given time. Taking the ONS estimate of 1,420 DSPD prisoners, possibly only half of the prisoners in the high security estate (chosen at random) would need to be assessed to yield sufficient numbers for an experimental evaluation. Possibly, the sample of prisoners to be assessed could be stratified according to time left to serve, so that a specified fraction of those assessed would be released within three years. This could make it possible (in principle, depending on numbers) to evaluate the DBT-based treatment according to reoffending.

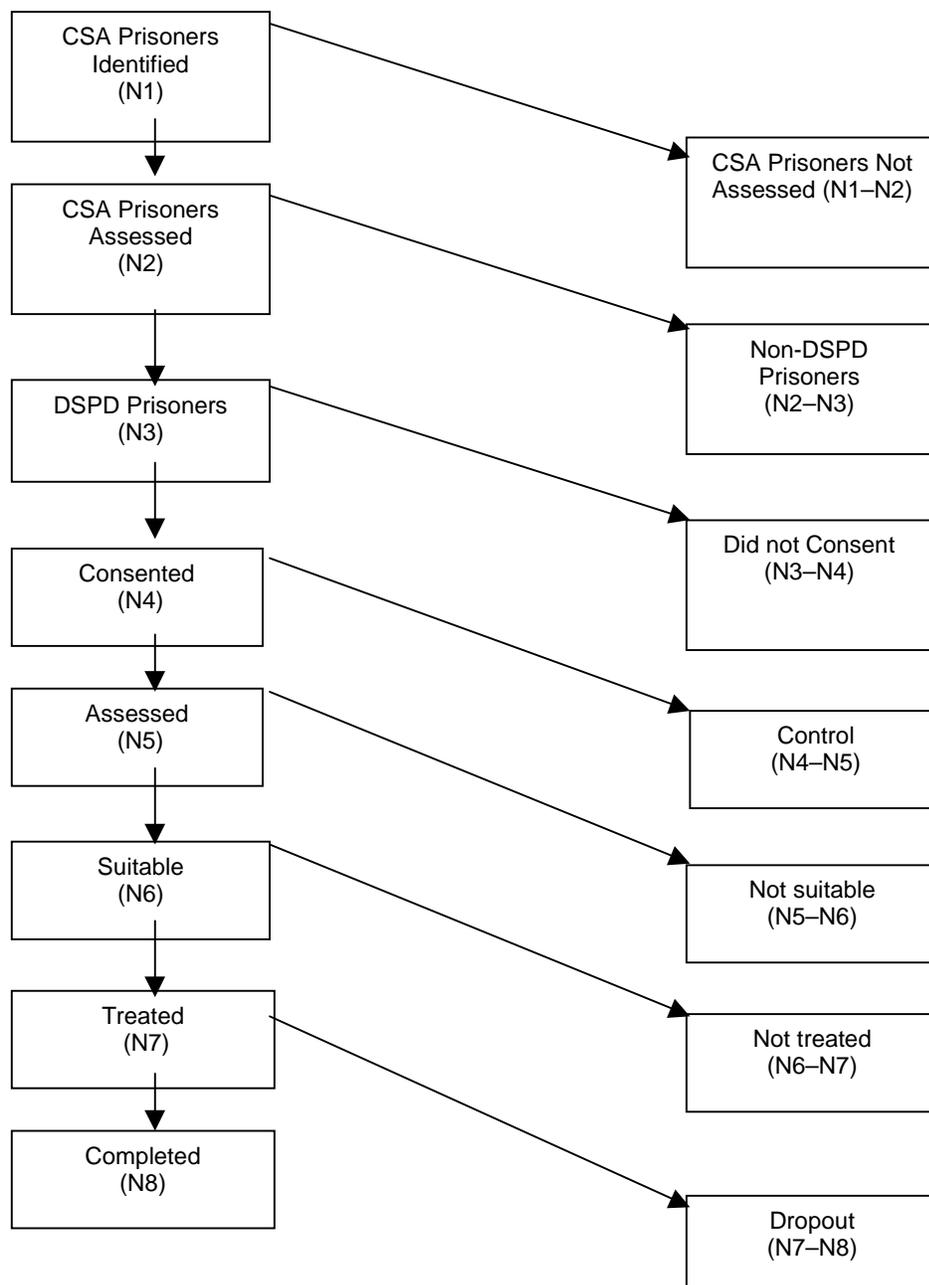
Of the N3 DSPD prisoners, N4 would consent to participate in the Whitemoor assessment and treatment programme if selected, and conversely (N3-N4) would decline. Alternatively, the (N3 – N4) could include some prisoners who were screened out because they were considered likely to be unsuitable or too mobile for treatment. Of the N4 who consented, N5 would be randomly allocated to be assessed in the Whitemoor DSPD unit, and conversely (N4 – N5) would be randomly allocated to the control group. Under current plans, N5 would be about 50 per year, and (N4 – N5) would be of the order of several hundreds. By reducing the assessment period in Whitemoor, as mentioned above, it might be possible to increase N5 to 100 or even 150 per year.

Of the N5 who are assessed, N6 (about half, based on experience so far) would be deemed suitable for the DBT-based treatment, and conversely (N5 – N6) would be deemed unsuitable. Of the N6 who were suitable, N7 would be treated, and conversely (N6 – N7) would not be treated. Of the N7 who were treated, N8 would be regarded as “completers” of the treatment in some sense, and conversely (N7 – N8) would be non-completers or treatment dropouts.

Neglecting the problem of prisoners who have already been assessed (who could not be part of an RCT), it can be predicted that about 25 DSPD prisoners will begin treatment in 2003 (N7). From the viewpoint of an RCT, it would be desirable if N7 could be increased to at least 50. Assuming that most suitable prisoners are treated and that half of those who were assessed would be suitable, N5 should be over 100. In turn, based on the statistical power analysis, N4 should be over 400.

The value of N3 is more problematic. If the N4 prisoners are defined as those who consent, it might be expected that the majority of the N3 DSPD prisoners would consent. Hence, N3 could be at least 500. However, if (N3 – N4) includes prisoners who were screened out as probably unsuitable or too mobile for the treatment, N3 might have to be at least 800. In principle, it should be possible to identify at least 1,000 DSPD prisoners in the eight feeder prisons, although in practice this would require considerable resources. (This assessment might eventually be required under the new legislation.)

Figure 1: Participant flow chart for RCT



It is envisaged that the evaluation would focus on six groups who would complete pre-test and post-test instruments:

- (a) N8 treatment completers
- (b) (N7 – N8) treatment dropouts
- (c) (N6 – N7) not treated
- (d) (N5 – N6) not suitable
- (e) (N4 – N5) controls
- (f) (N3 – N4) who did not consent (or were screened out).

Ideally, the pre-test would be given in the feeder prisons and the post-test at the end of the Whitemoor treatment period. As mentioned, those who were controls in one year could be eligible to be treated in a later year.

The main evaluation of the treatment would be on an “intention-to-treat” basis, that is comparing groups (a), (b), (c) and (d) together (N5) with the control group (N4 – N5). Pre-test versus post-test changes in the assessments of group (a) versus group (b) would show quasi-experimentally how far the treatment has an effect; it might be predicted that there would be a dose-response relationship between the amount of treatment received and the amount of improvement observed. Changes in the assessments of group (d) versus groups (a), (b) and (c) might provide some indication of the effects of being selected for the DSPD unit and assessed. Changes in the assessments of group (f) would assist in determining how far the consenters (or those not screened out) were untypical, and hence how far the results of the experiment might be generalisable to all DSPD prisoners.

This flow diagram does not take account of prisoners who have already been assessed in the DSPD unit at Whitemoor. To the extent that they occupy treatment beds, there would be fewer treatment beds available for prisoners participating in the RCT.

Many RCTs experience case flow problems: the number of eligible cases seems large at the beginning but then decreases for one reason or another. Therefore, in the interests of carrying out a successful RCT, it is important to cast the net relatively wide at the beginning and assess a relatively large number of CSA prisoners (N2). This would also help to overcome problems of attrition later on in the flow diagram (e.g. prisoners who do not consent or who drop out of treatment).

An alternative evaluation strategy

As indicated above, there are many circumstances in which using an RCT to evaluate the Whitemoor treatment would not be desirable or feasible. This would be especially true if only 50 DSPD prisoners were assessed and only 25 were treated and if previously assessed prisoners occupied treatment beds from 2003. All clinicians spoken to were against an RCT for a variety of reasons, including the ethics of denying treatment to those allocated to the control group and the heterogeneity of DSPD prisoners (suggesting that the treatment would be different for each prisoner). Prison managers were not opposed to an RCT.

Where an RCT is not desirable or feasible, the treatment can be evaluated by using matched treated and control groups, by comparing before and after outcomes in each group, or by statistical adjustment for pre-existing differences between groups. All of these methods require the research recommended above to devise outcome measures by comparing dynamic items with offending histories. Once the key dynamic items that are related to serious violent and sexual offending have been established, control subjects who match the treated subjects could be selected from the feeder prisons. This matching could be done prospectively (before the Whitemoor assessment) or retrospectively (after the Whitemoor assessment), although prospective matching would be better. Prisoners would have to give informed consent in a quasi-experimental evaluation, and they could not be blind to their condition.

There are problems of choosing a matched comparison group. The DIA does not have an adequate measure of personality disorder that could be used in matching control prisoners to the experimental group. Hence, control prisoners would have to be selected using a DSPD assessment procedure. Also, many of the DIA instruments depend on self-report and hence may be open to manipulation by DSPD prisoners. The PCL-R might possibly be used in selecting control prisoners; the authors understand that the Prison Service has administered the PCL-R to over 200 prisoners as part of their DIA, and these data might possibly be made available to assist in retrospectively selecting a control group.

The key dynamic items could be combined into a dynamic risk score that distinguishes serious violent and sexual offenders from others (or possibly into two risk scores, one for violence and one for sex). Pre-test and post-test

scores could be compared for treated and control subjects. Also, the post-test score (or some key dynamic items measured on the post-test) could be regarded as the outcome, and this could be predicted by pre-test scores (or items) and by treated versus control group in a regression equation. The key question is whether being in the treated or control group predicts key outcomes after controlling for pre-test differences between the groups.

It is also desirable to devise key dynamic outcome measures and to have pre-test and post-test measures of them in case an RCT breaks down and cannot be carried through successfully. In this case, the data would have to be analysed as a quasi-experiment. Also, the pre-test measures could provide a check on the pre-test comparability of prisoners in the treated and control conditions of an RCT.

A possible advantage of a retrospective quasi-experimental evaluation is that the prisoners who have been assessed so far might conceivably be included in it. However, it has to be emphasised that a quasi-experimental evaluation would be far less convincing than an RCT.

Conclusions

An RCT is desirable and feasible in some circumstances but not in others. It is not feasible if about 50 potentially DSPD prisoners are assessed between April 2003 and April 2004, about 25 are considered suitable for the DBT-based treatment, and the other 25 receive no specific treatment (e.g. are returned to the feeder prisons). It is especially not feasible if the DBT-based treatment takes three years, and if treatment beds are occupied by previously assessed prisoners.

An RCT is feasible if several hundred prisoners can be assessed for DSPD in the feeder prisons, if the length of the Whitemoor assessment is shortened (thereby allowing more prisoners to be assessed and treated), if prisoners deemed not suitable for DBT receive some alternative (e.g. psychopathy) treatment, and if the treatment lasts only one year.

Research is urgently needed to devise key risk measures based on dynamic items that correlate with serious violent and sexual offending. Assuming that few DSPD prisoners will be released in the next three years (to allow a reconviction follow-up), these measures are needed whether or not an RCT is carried out. They should be given to treated and control groups before and after any treatment in an RCT or in a quasi-experimental evaluation. If an RCT is not carried out, the treatment can be evaluated by comparing pre-test and post-test risk measures of treated and control groups. If an RCT is carried out, the treatment can be evaluated by comparing post-test measures of treated and control groups. Research is also needed on the validity of self-reporting by DSPD prisoners. It would be desirable to plan for a reconviction follow-up by stratified sampling of prisoners according to their time left to serve.

Ultimately, whether an RCT is desirable and feasible will probably depend on decisions made operationally in the Prison Service and will not be under the control of researchers. For example, the assessment of prisoners for DSPD in feeder prisons might require considerable resources that could not be provided as part of a research project. The authors have made recommendations based on what is best for evaluation research, but of course many other factors have to be taken into account in making decisions about how to evaluate the treatment of DSPD prisoners. Our conclusion is that efforts should continue to create the conditions that would make it feasible to evaluate the Whitemoor DSPD unit using an RCT, because this is the "gold standard" design.

References

- Andrews, D. A. and Bonta, J. L. (1995) *The Level of Service Inventory – Revised*. Toronto: Multi-Health Systems.
- Attrill, G. (2001) *The Psychopathy Programme - Theory Manual*. Unpublished manuscript.
- Barratt, E. S. (1994) Impulsiveness and aggression. In J. Monahan and H. J. Steadman (Eds.) *Violence and Mental Disorder*. Chicago: University of Chicago Press.
- Beck, A. T. and Steer, R. A. (1987) *Manual for Revised Beck Depression Inventory*. San Antonio, Texas: Psychological Corporation.
- Boer, D. P., Hart, S. D., Kropp, P. R. and Webster, C. D. (1997) *Manual for the Sexual Violence Risk-20*. Vancouver: British Columbia Institute against Family Violence.
- Brody, S. R. (1976) *The Effectiveness of Sentencing*. London: HMSO.
- Clarke, R. V. G. and Cornish, D. B. (1972) *The Controlled Trial in Institutional Research*. London: HMSO.
- Cook, T. D. and Campbell, D. T. (1979) *Quasi-Experimentation*. Chicago: Rand McNally.
- Cooke, D. (1998) The development of the Prison Behaviour Rating Scale. *Criminal Justice and Behaviour*, 25, 482-506.
- Cornish, D. B. and Clarke, R. V. G. (1975) *Residential Treatment and its Effects on Delinquency*. London: H.M.S.O.
- Department of Health/Home Office (2000) *Reforming the Mental Health Act*. (Part I: The New Legal Framework; Part II: High Risk Patients). London: The Stationery Office.
- Douglas, K. S., Ogloff, J., Nicholls, T. L. and Grant, I. (1999) Assessing risk for violence among psychiatric patients: The HCR-20 violence risk assessment scheme and the psychopathy checklist: screening version. *Journal of Consulting and Clinical Psychology*, 67, 917-930.
- Douglas, K. S. and Webster, C. D. (1999) The HCR-20 violence risk assessment scheme: Concurrent validity in a sample of incarcerated offenders. *Criminal Justice and Behaviour*, 26, 3-19.
- Evershed, S. (2002) *Evaluation of an Amended Dialectical Behaviour Therapy Intervention with Adult Male Forensic Patients who have a Personality Disorder and who live in a Maximum Secure Setting*. Unpublished doctoral dissertation in clinical psychology.
- Farrington, D. P. (1983) Randomized experiments on crime and justice. In M. Tonry and N. Morris (Eds.) *Crime and Justice*, vol. 4 (pp. 257-308). Chicago: University of Chicago Press.
- Farrington, D. P., Ditchfield, J., Hancock, G., Howard, P., Jolliffe, D., Livingston, M. S. and Painter, K. A. (2002) *Evaluation of Two Intensive Regimes for Young Offenders*. London: Home Office. (Home Office Research Study No. 239.)
- Farrington, D. P., Hancock, G., Livingston, M., Painter, K. A. and Towl, G. (2000) *Evaluation of Intensive Regimes for Young Offenders*. London: Home Office. (Home Office Research Findings No. 121.)
- Fowles, A. J. (1978) *Prison Welfare*. London: H.M.S.O.
- Goldberg, D. (1978) *Manual of the General Health Questionnaire*. Windsor: NFER-Nelson.
- Grann, M., Belfrage, H. and Tengstrom, A. (2000) Actuarial assessment of risk for violence: Predictive validity of the VRAG and the historical part of the HCR-20. *Criminal Justice and Behaviour*, 27, 97-114.
- Hanson, R. K. and Harris, A. (2000) *The Sex Offender Need Assessment Rating (SONAR): A Method for Measuring Change in Risk Levels*. Ottawa: Solicitor General of Canada (User Report 2000-01).
- Hanson, R. K. and Thornton, D. (1999) *Static 99: Improving Actuarial Risk Assessment for Sex Offenders*. Ottawa: Solicitor General of Canada (User Report 1999-02).

- Hanson, K. R. and Thornton, D. (2000) Improving actuarial risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behaviour*, 24, 119-136.
- Hare, R. D. (1991) *The Psychopathy Checklist – Revised*. Toronto: Multi-Health Systems.
- Hemphill, J. F., Templeman, R., Wong, S. and Hare, R. D. (1998) Psychopathy and crime: Recidivism and criminal careers. In D. J. Cooke, A. E. Forth and R. D. Hare (Eds.) *Psychopathy: Theory, Research and Implications for Society* (pp. 375-399). Dordrecht, Netherlands: Kluwer.
- Home Office/Department of Health (1999) *Managing Dangerous People with Severe Personality Disorder: Proposals for Policy Development*. London: Home Office/Department of Health.
- Kroner, D. G. and Mills, J. F. (2001) The accuracy of five risk appraisal instruments in predicting institutional misconduct and new convictions. *Criminal Justice and Behaviour*, 28, 471-489.
- Linehan, M. M. (1993a) *Cognitive-Behavioural Treatment of Borderline Personality Disorder*. New York: Guilford.
- Linehan, M. M. (1993b) *Skills Training Manual for Treating Borderline Personality Disorder*. New York: Guilford.
- Lipsey, M. W. and Wilson, D. B. (1998) Effective intervention with serious juvenile offenders: A synthesis of research. In R. Loeber and D. P. Farrington (Eds.) *Serious and Violent Juvenile Offenders: Risk Factors and Successful Interventions* (pp. 313-345). Thousand Oaks, California: Sage.
- McCann, R., Ball, E. M. and Ivanoff, A. (2000) DBT with an inpatient forensic population: The CMHIP forensic model. *Cognitive and Behavioural Practice*, 7, 447-456.
- Novaco, R. W. (1980) *The Novaco Anger Scale*. Irvine, California: Department of Psychology, University of California.
- Office for National Statistics (1998) *Psychiatric Morbidity in England and Wales*. London: The Stationery Office.
- Quinsey, V. L., Harris, G. T., Rice, M. E. and Cornier, C. A. (1998) *Violent Offenders: Appraising and Managing Risk*. Washington, D.C.: American Psychological Association.
- Redondo, S., Sanchez-Meca, J. and Garrido, V. (1999) The influence of treatment programmes on the recidivism of juvenile and adult offenders: A European meta-analytic review. *Psychology, Crime and Law*, 5, 251-278.
- Rice, M. E. and Harris, G. T. (1995) Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology*, 63, 737-748.
- Rice, M. E. and Harris, G. T. (1997) Cross-validation and extension of the violence risk appraisal guide for child molesters and rapists. *Law and Human Behaviour*, 21, 231-241.
- Roger, D. and Najarian, B. (1989) The construction and validation of a new scale for measuring emotion control. *Personality and Individual Differences*, 8, 845-853.
- Shadish, W. R., Cook, T. D. and Campbell, D. T. (2002) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton-Mifflin.
- Shaw, M. (1974) *Social Work in Prison*. London: HMSO.
- Shepherd, S. (1991) A brief intervention for anxiety and depression. *Proceedings of the Prison Service Psychology Conference*, Scarborough, October 16-18.
- Sjoestedt, G. & Langstroem, N. (2001). Actuarial assessment of sex offender recidivism risk: A cross-validation of the RRASOR and Static-99 in Sweden. *Law and Human Behavior*, 25, 629-645.
- Thornton, D. (1987) Assessing custodial adjustment. In B. J. McGurk, D. M. Thornton and M. Williams (Eds.) *Applying Psychology to Imprisonment* (pp. 445-462). London: HMSO.
- Thornton, D. and Hogue, T. (2001) *Operational Definition of Dangerous and Severely Personality Disordered* (Second draft, July). Unpublished paper.
- Thornton, D., Travers, R., Blud, L., Friendship, C. and Erickson, M. (2001) *Predicting Violence: Relevance of History and Current Functioning*. Unpublished manuscript.

Tyrer, P. (2000, Ed.) *Personality Disorders: Diagnosis, Management and Course* (2nd ed.) Oxford: Butterworth/Heinemann.

Walters, G. D. (1995a) The Psychological Inventory of Criminal Thinking Styles, Part I. Reliability and preliminary validity. *Criminal Justice and Behaviour*, 22, 307-325.

Walters, G. D. (1995b) The Psychological Inventory of Criminal Thinking Styles, Part II. Identifying simulated response sets. *Criminal Justice and Behaviour*, 22, 437-445.

Walters, G. D. (1996) The Psychological Inventory of Criminal Thinking Styles, Part III. Predictive Validity. *International Journal of Offender Therapy and Comparative Criminology*, 40, 105-112.

Warren, F., Preedy, K., McGauley, G., Pickering, A., Norton, K., Geddes, J. R. and Dolan, B. (2001) *Review of Treatments for Dangerous and Severe Personality Disorder*. Unpublished report to the Home Office.

Webster, C. D., Douglas, K. S., Eaves, D. and Hart, S. D. (1997) *HCR-20: Assessing Risk for Violence* (version 2). Burnaby, B.C.: Simon Fraser University.

Williams, M. (1970) *A Study of Some Aspects of Borstal Allocation*. Home Office: Chief Psychologist Report No. 33.

Williams, M. (1975) Aspects of the psychology of imprisonment. In S. McConville (Ed.) *The Use of Imprisonment* (pp. 32-42). London: Routledge and Kegan Paul.

Produced by the Research Development and Statistics Directorate, Home Office

This document is available only in Adobe Portable Document Format (**PDF**) through the RDS website

Home Office
Research, Development and Statistics Directorate
Communication Development Unit
Room 275
50 Queen Anne's Gate
London SW1H 9AT

Tel: 020 7273 2084 (answerphone outside of office hours)

Fax: 020 7222 0211

Email: publications.rds@homeoffice.gsi.gov.uk

ISBN 1 84082 925 7

© Crown copyright 2002