

The effectiveness of drug treatment programs in reducing criminal behavior: A meta-analysis

Katy R. Holloway, Trevor H. Bennett and David P. Farrington*
University of Glamorgan and * University of Cambridge

The main aim of this paper is to assess the effectiveness of drug treatment programs in reducing criminal behavior. The research is based on a systematic review and meta-analysis of evaluations of drug treatment programs. The research investigated various drug treatment programs including methadone treatment, therapeutic communities, post-release supervision for drug-misusing offenders, and drug courts, and focused on drug treatment programs that were primarily concerned with heroin, crack and cocaine misuse. The main outcome variable analyzed was reoffending following treatment as measured by self-reports or official records. The study was based on a meta-analysis of 28 evaluations of drug treatment programs. The meta-analysis involved calculating individual effect sizes for each study and weighted mean effect sizes for groups of studies. The mean odds of offending following treatment were significantly lower among clients of the drug treatment programs than among the comparison groups. However, the results varied by type of program, type of evaluation methods used, and characteristics of clients.

La efectividad de los programas de tratamiento de la drogadicción en la reducción de la delincuencia: un metaanálisis. El principal objetivo de este artículo es evaluar la efectividad de los programas de tratamiento aplicados a drogodependientes para reducir la conducta delictiva. Para ello se realizó una revisión sistemática y un metaanálisis en diferentes tipos de tratamiento, como el que se basa en la metadona, los que emplean la comunidad terapéutica, los que supervisan a los delincuentes adictos en la comunidad y los llamados «tribunales de la droga». Los programas tomaban en consideración principalmente el abuso de crack, la heroína y la cocaína. La medida de resultado elegida fue la reincidencia posterior al tratamiento, de acuerdo a registros autoinformados y oficiales. La investigación que se presenta aquí se basó en un metaanálisis de 28 evaluaciones de programas de tratamiento de drogodependientes, e implicó calcular los tamaños del efecto de cada estudio y luego ponderar los tamaños del efecto medios para los grupos de estudios. Se observa que los grupos de comparación, en promedio, reinciden menos que los grupos de comparación, aunque esto habría que matizarlo en función del tipo de programa, tipo de método de evaluación usado y las características de los sujetos.

There have been many reviews of the literature on the effectiveness of drug treatment programs (e.g., Chanhasilpa et al, 2000; Hall, 1996; Vaughn and Howard, 2004). However, most of these have investigated the effects of programs on drug misuse. It is much less common for reviews to study the effects of programs on other problem behaviors, such as criminal behavior. This neglect of criminal behavior is important as research has shown that the majority of drug misusers presenting to drug treatment programs are self-reported offenders (Gossop et al, 2003) and a notable proportion of these commit crimes at a high rate (Strang et al, 2000). It has also been shown that drug-misusing offenders often continue to offend both during and after drug treatment (Hutchinson et al, 2000).

There have been some prior reviews of the literature that have included criminal behavior as an outcome measure and have used meta-analytic techniques. In total, we found five systematic reviews that used meta-analysis to investigate the effect of drug treatment programs on criminal behavior. All were conducted in the USA. Two of the five reviews were based on single treatment programs (Marsch, 1998; Wilson et al, unpublished manuscript) and three were based on multiple programs (Mitchell et al, 2005; Pearson and Lipton, 1998; Prendergast et al, 2002).

The two reviews of single treatment programs both showed modest desirable effect sizes for the program. Marsch (1998) investigated the effects of methadone maintenance programs on various problem behaviors. Seventeen of the 24 studies providing results on criminal behavior showed a desirable and significant effect size ranging from $r=0.01$ to $r=0.76$, with a weighted mean $r=0.16$ for all studies. The mean effect size was greater for studies that examined drug-related crime ($r=0.67$) than those that examined drug- and property-related crime combined ($r=0.14$). Wilson et al (unpublished manuscript) examined the results of 38 evaluations of drug courts that placed drug-misusing offenders in

Methods

treatment programs. The mean odds-ratio for all offense types was 1.79. The effect size was greater in relation to drug offenses (1.68) compared with non-drug offenses (1.29), and for juveniles (2.11) compared with adults (1.69).

The three reviews based on multiple treatment modalities also indicated significant differences between the experimental and comparison groups. Prendergast et al (2002) conducted a systematic review and meta-analysis of 25 studies that investigated the effects of drug treatment on crime. The review included five treatment modalities: methadone maintenance programs, therapeutic communities, outpatient drug-free programs, detoxification programs, and private sector treatment. Studies were eligible for inclusion if they were conducted in the United States, published between 1965 and 1996, and were based on adult drug abusers. Overall, the mean effect size for crime outcomes for all treatments combined was $r = 0.13$. The authors concluded that drug treatment was effective in reducing criminal behavior. However, there were no significant differences in effect sizes across treatment modalities.

Mitchell et al (2006) conducted a meta-analysis of 26 evaluations of incarceration-based drug treatment programs. The study included evaluations of therapeutic communities, group counselling, boot camps, and methadone maintenance. The overall mean odds ratio for all programs combined was 1.25, which represented a statistically significant reduction in post-treatment offending. However, there were some important differences in outcomes by program type. Only therapeutic communities ($OR = 1.47$) and group counselling programs ($OR = 1.25$) were associated with lower rates of offending. There was no difference in post-treatment offending between participants and non-participants in boot camps, and those in methadone maintenance programs were significantly more likely than the comparison groups to offend following treatment. Pearson and Lipton (1998) also conducted a meta-analysis of incarceration-based drug treatment programs. They investigated 6 studies of boot camps and 7 studies of drug-focused group counselling and concluded that neither was effective in reducing criminal behavior. However, their analysis of 7 studies of therapeutic communities concluded that these were effective ($r = 0.13$).

The combined results of these five meta-analyses are far from conclusive. One of the three reviews of methadone treatment programs concluded that it was associated with reduced offending (Marsch, 1998), one concluded that it resulted in higher rates of offending (Mitchell et al, 2006), and one found that its effect was positive, but no different from that of any other treatment modality (Prendergast et al, 2002). All three reviews that investigated therapeutic communities concluded that they were effective in reducing offending (Mitchell et al, 2006; Pearson and Lipton, 1998; Prendergast et al, 2002), although Prendergast et al (2002) reported that they were no more effective than other drug treatment methods. One of the two reviews that included group counselling concluded that it was effective in reducing recidivism (Mitchell et al, 2006) and the other reported that it was not effective (Pearson and Lipton, 1998).

The main aim of the current review is to investigate the effectiveness of drug treatment programs in reducing criminal behavior. This adds to the work of previous meta-analyses by including drug treatment implemented in the UK and Europe, as well as the US, programs initiated by the criminal justice system as well as through conventional routes, and more recent research covering modern types of drug treatment.

This research is a systematic review of the literature on the effects of different kinds of intervention for problematic drug use on criminal behavior. Systematic reviews use rigorous methods for locating, analyzing, and collating evidence from a number of studies. They have explicit objectives and criteria for including or excluding studies and are based on extensive searches of the literature for eligible evaluations. They are also based on careful extraction and coding of key features of the studies and are sufficiently detailed to allow replication. Information about the methods of systematic reviews can be found in a number of recent publications (Farrington and Petrosino, 2000; Farrington and Welsh, 2002; Welsh and Farrington, 2002).

Criteria for inclusion

In selecting evaluations for inclusion in this review, three main criteria were used, concerning the type of intervention, the type of evaluation method used, and the type of outcome measures.

The main criteria for the type of intervention were that the evaluation should be based on either treatment programs that aimed to reduce drug use (e.g. methadone maintenance, detoxification, or self-help programs) or criminal justice programs that aimed to reduce both drug use and drug-related crime (e.g. drug courts or Drug Treatment and Testing Orders [DTTOs]). Treatment programs were included in the review if they were primarily concerned with heroin, crack, or cocaine users. Programs that aimed mainly to reduce other kinds of substance misuse (such as alcohol or amphetamine use) were excluded. Criminal justice programs were included if they had the specific aim of reducing drug misuse.

The criteria for the type of method used were that the evaluation should use methods of sufficient quality to provide interpretable results. The current research broadly follows the methodological quality criteria adopted by Sherman et al (1997) in their version of the Scientific Methods Scale (SMS). The SMS is based on a five-point scale that ranks studies according to their ability to establish causality and to minimize threats to internal validity. Levels 1 and 2 are the lowest levels and include studies that seek to determine either a simple correlation at one point in time or differences between before and after measures over time without comparable control conditions. Levels 3 to 5 designs provide more robust findings and include studies that compare before and after measures for experimental and control conditions and evaluations based on random assignment to program or control conditions. Evaluations are deemed eligible for inclusion in this review if they were at least Level 3 on the SMS scale (see also Farrington, 2003; Farrington et al, 2002).

The main criterion for outcome measures was that the study must include a measure of criminal behavior. The review includes evaluations that used a measure of crime and evaluations that used measures of both drug use and crime. Studies that evaluated the effect of an intervention on drug use alone were excluded from the review. This was because the primary objective of the research was to investigate the effects of drug treatment programs on crime.

Other selection criteria were that the evaluation was published in the English language and that the study was available during the period of the research. Studies were only included if they were published during the period between January 1980 and March 2004 when the selection component of the research was

completed. The evaluation could be presented in any form and included reports, journal publications and other manuscripts.

Search methods

Evaluations were obtained mainly by searching online databases, reviewing citations in eligible studies, and contacting key researchers in the field. The databases included: Criminal Justice Abstracts, IBSS (International Bibliography of the Social Sciences), C2-SPECTR, Psychological Abstracts, and the Home Office Research Development and Statistics (RDS) publications database.

Each database was investigated using a list of predetermined search terms. Each search term yielded a list of titles and abstracts that were carefully reviewed. Studies that were clearly not evaluations of drug treatment programs were removed from the list. Obtained studies were screened for eligibility using the inclusion criteria described above and relevant data from eligible studies were entered into the research database.

Attrition rates

The searches of the five online databases resulted in a total of over 9,700 'hits': Criminal Justice Abstracts (n= 3,550), IBSS (n= 2,585), C2-SPECTR (n= 286), PsycINFO (n= 3,273), and the Home Office RDS data base (n=6). The titles and abstracts of these studies were then checked for relevance. Studies that were not *prima facie* evaluations of drug treatment programs were excluded at this point. This resulted in 597 studies selected from the searched databases. In addition, we already had in our possession, or selected from bibliographies, 80 further studies of possible relevance, making a total of 677 studies initially selected. Of these, 504 were obtained during the study period. The main reason for not obtaining publications was that the inter-library loan system was unable to locate them. The obtained studies were then checked for eligibility using the criteria mentioned above. This resulted in 55 eligible studies. The main reasons for exclusion were that the study was not an evaluation of a treatment program (n= 224) or the SMS methods score was below Level 3 (n= 183). Twenty-eight of the 55 eligible studies presented sufficient information in their results to enable raw data to be extracted for the purpose of the meta-analysis. Hence, the current paper is based on the results of these 28 evaluations of programs for drug misusers.

Eligible studies

The details of the 28 studies included in the meta-analysis are shown in table 1. Most studies (24) were conducted in the USA. The others were conducted in the UK (3) and Switzerland (1). Two of the studies were published in the 1980s, 15 in the 1990s and 11 in the 2000s. Most of the evaluations (14) were based on sample sizes of less than 100, whereas 11 had between 100 and 1,000 and two studies used samples of over 1,000.

The majority of evaluations were based on a single treatment type and a single comparison group. When studies included multiple treatments, a random-selection procedure was used to select the experimental program. The comparison condition was usually no treatment. When there was no obvious no treatment condition, then a comparison group was also selected using a

random-selection procedure. Using this method, the programs included in the review were therapeutic communities (n= 7), followed by drug courts and drug-testing (n= 5) maintenance prescribing (n= 5), probation and parole (n= 3), and other treatment (n= 3). In most cases, only one data source was used (either self reports or official records). When both data sources were used, self report measures were chosen over official records on the grounds that they had the potential to provide fuller and more recent evidence of offending. Overall, the majority of studies used self-report data (17) and the remainder (11) used only official records.

Results

In order to conduct a meta-analysis of program effectiveness in reducing criminal behavior it was necessary to calculate a comparable effect size for each study. The most appropriate measure for data based on proportions of respondent (e.g. percentage offending) is the odds ratio (Lipsey and Wilson, 2001, p. 52).

Overall effectiveness

The individual effect sizes for each study and the mean effect size for all studies combined are shown in table 2. The odds ratio (OR) was greater than 1 (suggesting that the treatment group was associated with lower offending than the control group) for 19 of the 28 studies and statistically significant ($p < 0.05$) for 11 of the 19. The OR was less than 1 (suggesting that the control group was associated with lower offending than the treatment group) for 9 of the 28 studies and statistically significant for only 1 of the 9 (Hubbard et al, 1997). This single negative result might in part be explained by the fact that the comparison condition was another treatment program rather than no treatment. Overall, just under half of the studies showed that the evaluated treatment was significantly more effective than the comparison in reducing criminal behavior.

One of the main advantages of meta-analysis is that a mean effect size can be calculated for groups of studies. There are two common ways of calculating a mean effect size. The first is the fixed effects (FE) model in which each effect size is weighted by the inverse of its variance ($1/\text{VAR}$). In this model, it is assumed that each effect size is drawn from a random (normal) distribution of effect sizes. The second is the random effects (RE) model in which each effect size is weighted by the inverse of its variance, plus an additional factor. In this case, it is assumed that the variance associated with each effect size is based on sampling error and a second component that reflects differences between the studies, such as in procedures or settings.

If the ESs are not significantly heterogeneous (measured by the Q statistic), then it can be assumed that they are all randomly drawn from the same population. In this case, the best way of estimating the mean of the distribution is the FE model. However, if the heterogeneity of the ESs is significant, then it is possible that they are not all drawn from the same distribution. In this case, there are at least two possibilities: (1) calculate the mean ES for more homogeneous subcategories (e.g. types of program or crimes), or (2) use the RE model which adds a constant to the variance of each ES so that they are no longer significantly heterogeneous. Each method has advantages and disadvantages. In the case of (1), this method would lose the overall findings for all

studies combined and would be limited only to homogeneous sub-groups. In the case of (2), small-sample studies would be given much the same weighting as large-sample studies. Given that FE and RE models have advantages and disadvantages, it is usually considered good practice to report results obtained with both. In the following, we include findings for all studies combined and various sub-groups and present the results using both the FE and RE models.

Table 2 shows that the weighted mean OR for the 28 studies combined was 1.41 using the FE model and 1.56 using the RE model. Both were statistically significant. Hence, it can be concluded that, taken together, the studies show that drug treatment is effective in reducing criminal behavior. These means show that the odds of offending for the treatment groups were reduced by 29 per cent (because $1/1.41 = 0.71$) according to the FE model and by 36 per cent (because $1/1.56 = 0.64$) according to the RE model.

Variations by research design

Previous research has shown that measured effects of treatment can be influenced by features of the research design. It has been argued, for example, that quasi-experimental designs might produce larger effect sizes than random allocation designs (Weisburd et al, 2001; Wilson et al, unpublished manuscript). One possible reason for this is that there might be a selection effect in quasi-experimental designs in which the most promising clients are selected for the treatment program. The findings shown in Table 3 offer some support for this view. The mean effect size for random allocation and quasi-experimental studies was 1.28 and 1.60 respectively for the FE model and 1.33 and 2.09 for the RE model. Hence, quasi-experimental designs tended to produce higher ORs than random allocation designs. However, both methods yield the same overall conclusion that drug treatment was associated with a reduction in criminal behavior.

Table 1
Description of studies included in the meta-analysis

Author	Date	Location	Design	Treatment group sample size	Treatment type [1]	Data source [2]
Bale et al	1980	USA	Random	59	Methadone treatment	Self report
Simpson and Sells	1982	USA	Quasi	83	Methadone treatment	Self report
Britt et al	1992	USA	Random	Unknown	Drug testing	Records
Latessa and Moon	1992	USA	Random	182	Other treatment [3]	Records
Hoffmann and Norman	1992	USA	Quasi	4,541	Other treatment [3]	Self report
Martin and Scarpitti	1993	USA	Random	130	Probation	Self report
McBride and Inciardi	1993	USA	Random	531	Drug testing	Self report
Magura et al	1993	USA	Quasi	195	Methadone treatment	Self report
Deschenes et al (a)	1995	USA	Random	76	Probation	Records
Deschenes et al (b)	1995	USA	Random	46	Probation	Records
Inciardi et al	1997	USA	Random	43	TC	Self report
Hubbard et al	1997	USA	Quasi	1,203	Methadone treatment	Self report
Simpson et al	1997	USA	Quasi	342	TC	Self report
Perneger et al	1998	Switz.	Quasi	27	Heroin treatment	Self report
Nemes et al	1999	USA	Random	194	TC	Records
Turner et al	1999	USA	Random	363	Drug testing	Records
Wexler et al	1999	USA	Random	425	TC	Records
Farrell	2000	USA	Random	41	TC	Self report
Gordon et al	2000	USA	Random	254	TC	Records
Dynia and Sung	2000	USA	Quasi	184	TC	Records
Hutchinson et al	2000	UK	Quasi	50	Methadone treatment	Self report
Strang et al	2000	UK	Quasi	18	Methadone treatment	Self report
Brown et al	2001	USA	Random	94	Supervision	Self report
Farabee et al	2001	USA	Quasi	681	Probation	Self report
Hser et al	2001	USA	Quasi	457	TC	Self report
Haapanen and Britton	2002	USA	Random	172	Drug testing	Records
Ghodse et al	2002	UK	Quasi	22	Supervision	Self report
Gottfredson et al	2003	USA	Random	139	Drug courts	Records

Total n= 28

Notes: 'TC'= Therapeutic Community. 'Random'= Random allocation studies. 'Quasi'= Quasi-experimental designs.

[1] In the case of evaluations of multiple treatments, one program was chosen using a random selection procedure (see text for details).

[2] In the case of multiple outcome measures, one measure was selected using a priority system (see text for details).

[3] Includes acupuncture and abstinence.

It is also possible that the type of comparison used in the study might also influence the measured outcome. Common sense suggests that an evaluation is more likely to produce a desirable effect if the treatment group is compared with a group receiving no treatment than if it is compared with a group receiving another form of treatment. One possible reason for this is that the comparison treatment might also be effective in changing the outcome variable. The studies used in the current review used three different kinds of comparison: (1) treatment with no treatment (T1 versus T0), (2) treatment with another kind of treatment (T1 versus T2), and treatment at one level of intensity with the same treatment at another level of intensity (T1a versus

T1b). According to the above, there could be measurable differences in the mean effect sizes from these three comparison methods.

The results shown in table 3 indicate that this indeed is the case. The larger effect sizes were obtained with research designs that compared one type of treatment with no treatment (FEOR= 1.68 and REOR= 1.76, where FEOR= Fixed Effects OR and REOR= Random Effects OR) and a strong version of a treatment with a weak version of the same treatment (FEOR= 1.50 and REOR= 1.84) (significant for the FE model, but not significant for the RE model). The smaller effect sizes (which generated a non-significant result overall) were obtained from studies that compared one type

Table 2
Individual effect sizes

Author	Date	Outcome measure [1]	Treatment type [2]	Treatment comparison[3]	OR	CI	p
Bale et al	1980	Arrest	Methadone treatment	1	1.24	0.70-2.20	ns
Simpson and Sells	1982	Arrest	Methadone treatment	1	2.02	1.09-3.75	0.0257
Britt et al	1992	Arrest	Drug testing	1	0.93	0.52-1.69	ns
Latessa and Moon	1992	Arrest	Other treatment [4]	1	0.88	0.38-2.06	ns
Hoffmann and Norman	1992	Arrest	Other treatment [4]	2	0.80	0.53-1.20	ns
Martin and Scarpitti	1993	Imprisonment	Probation	1	1.19	0.58-2.45	ns
McBride and Inciardi	1993	Arrest	Drug testing	2	0.92	0.58-1.47	ns
Magura et al	1993	Illegal income	Methadone treatment	1	1.57	0.57-4.29	ns
Deschenes et al (a)	1995	Arrest	Probation	3	0.54	0.23-1.25	ns
Deschenes et al (b)	1995	Arrest	Probation	1	1.54	0.70-3.35	ns
Inciardi et al	1997	Arrest	TC	1	3.86	1.79-8.29	0.0006
Hubbard et al	1997	Offended	Methadone treatment	2	0.69	0.50-0.97	0.0278
Simpson et al	1997	Arrest	TC	2	1.44	0.83-2.51	ns
Perneger et al	1998	Charged	Heroin treatment	2	27.02	1.64-445.98	0.0214
Nemes et al	1999	Arrest	TC	3	1.73	1.07-2.79	0.0257
Turner et al	1999	Arrest	Drug testing	2	0.61	0.33-1.13	ns
Wexler et al	1999	Imprisonment	TC	1	10.97	5.14-23.44	0.0001
Farrell	2000	Recidivism	TC	2	1.02	0.41-2.52	ns
Gordon et al.	2000	Recidivism	TC	1	1.70	1.17-2.48	0.0054
Dynia and Sung	2000	Arrest	TC	1	2.16	1.10-4.23	0.0143
Hutchinson et al	2000	Drug offenses	Methadone treatment	3	3.07	0.45-20.82	ns
Strang et al	2000	Offended	Methadone treatment	3	2.90	0.34-24.94	ns
Brown et al	2001	Offended	Supervision	1	0.82	0.31-2.19	ns
Farabee et al	2001	Arrest	Probation	1	3.74	2.41-5.80	0.0001
Hser et al	2001	Arrest	TC	2	3.77	2.53-5.62	0.0001
Haapanen and Britton	2002	Arrest	Drug testing	1	0.74	0.51-1.09	ns
Ghodse et al	2002	Offended	Supervision	3	13.13	1.59-108.32	0.0168
Gottfredson et al	2003	Arrest	Drug courts	1	2.21	1.19-4.12	0.0121
Total n= 28							
Fixed effects					1.41	1.26-1.58	0.0001
Random Effects					1.56	1.18-2.07	0.0018
Notes: 'TC'= Therapeutic Community. OR= Weighted Mean Odds Ratio. CI= Confidence Interval. p= Probability of Z. ns= Not significant at p<.05.							
[1] In the case of multiple outcome measures, one measure was selected using a priority system (see text for details).							
[2] In the case of evaluations of multiple treatments, one program was chosen using a random selection procedure (see text for details).							
[3] '1'='T1 v T0' (one kind of treatment versus no treatment); '2'='T1 v T2' (one kind of treatment versus another kind of treatment); and '3'='T1a v T1b' (one level of intensity of treatment versus another level of intensity of the same treatment).							
[4] Includes acupuncture and abstinence.							

of treatment with another type of treatment (FEOR= 1.11 and REOR= 1.20). However, there was no statistically significant difference between the ORs of the different types of comparison.

Methods effects of these kinds could have implications for the conclusions drawn from the research. First, they suggest that the main treatment effect shown for all studies combined was mainly driven by the larger effect sizes in studies that compared treatment with no treatment. Second, there are no clear effects of drug treatment on criminal behavior among studies that compare one type of treatment with another type. This could be taken to mean that all programs have similar effects on offending. Third, according to the FE model, intensive programs are more effective than non-intensive programs in reducing criminal behavior. The measures of program intensity in these studies covered high or low dosages of prescribed drugs, long or short-term programs, continuous versus interrupted programs, and enhanced versus standard versions of an intervention.

Variations by type of treatment program

One of the most important research questions is whether there are any variations in the effectiveness of different types of treatment program. In order to avoid the possible distorting effects of treatment intensity (comparing one level of treatment with another level of the same treatment), the mean effect sizes were calculated only for those studies that compared treatment with no treatment, and treatment with an alternative treatment (see table 4).

The results of the analysis showed that the most effective interventions in reducing criminal behavior were therapeutic communities (FEOR= 2.49 and REOR= 2.61), post-release supervision (FEOR= 2.46 and REOR= 1.99), and maintenance prescribing (FEOR= 1.64 and REOR= 1.75). In the case of therapeutic communities, the mean effect size was positive and significant using both the FE and RE methods. This can be interpreted to mean that the odds of offending among those receiving therapeutic community

Table 3
Study design characteristics

Group	N	Fixed effects				Random effects			
		OR	CI	Zp	Qp	OR	CI	Zp	Qp
All studies	28	1.41	1.26-1.58	0.0001	<0.0001	1.56	1.18-2.07	0.0018	ns
Research design									
Random allocation	16	1.28	1.10-1.48	0.0011	<0.0001	1.33	0.96-1.84	ns	ns
Quasi-experimental	12	1.60	1.36-1.89	0.0001	<0.0001	2.09	1.25-3.50	0.0048	ns
Comparison type									
T1 v T0 studies	15	1.68	1.44-1.96	0.0001	<0.0001	1.76	1.23-2.51	0.0019	ns
T1 v T2 studies	8	1.11	0.94-1.33	ns	<0.0001	1.20	0.70-2.06	ns	ns
T1a v T1b studies	5	1.50	1.01-2.22	0.0444	0.027	1.84	0.75-4.52	ns	ns

Notes: OR= Weighted Mean Odds Ratio. 'CI'= Confidence Interval. 'Zp'= Probability of Z. 'Qp'= Probability of Q. 'ns'= Not significant at p<.05. 'T1 v T0'= One kind of treatment versus no treatment. 'T1 v T2'= One kind of treatment versus another kind of treatment. 'T1a v T1b'=One level of intensity of treatment versus another level of intensity of the same treatment. None of the differences in ORs was statistically significant.

Table 4
Type of treatment program

Group	N	Fixed effects				Random effects			
		OR	CI	Zp	Qp	OR	CI	Zp	Qp
Maintenance prescribing	4	1.64	1.12-2.41	0.0114	0.1571	1.75	0.99-3.11	0.0549	0.2509
Therapeutic communities	7	2.49	2.03-3.06	<0.0001	<0.001	2.61	1.58-4.33	0.0002	0.2689
Post-release supervision	3	2.46	1.75-3.45	<0.0001	0.0126	1.99	0.92-4.31	0.0801	0.4382
Drug courts and drug testing	5	0.91	0.73-1.14	0.4179	0.0321	0.95	0.65-1.39	0.7949	0.3263
Other treatment	3	0.81	0.58-1.15	0.242	0.9753	0.91	0.56-1.47	0.7039	1.0000
Total	22	1.55	1.37-1.75	<0.0001	<0.0001	1.61	1.18-2.18	0.0023	0.3156

Notes: 'Maintenance prescribing' includes heroin and methadone treatment. 'Drug courts and drug testing' include drug courts, drug testing, and DTTOs. 'Other treatment' includes supervision and alternative treatments (e.g. acupuncture). OR= Weighted Mean Odds Ratio. 'CI'= Confidence Interval. 'Zp'= Probability of Z. 'Qp'= Probability of Q. 'ns'= Not significant at p<.05. The table shows results for treatment versus no treatment (T1 v T0) and treatment versus another treatment (T1 v T2) studies only. One study that investigated maintenance prescribing (Hubbard et al. 1997) has been excluded from the meta-analysis because the control condition selected was therapeutic community treatment (a program shown from our results to be particularly effective in reducing criminal behavior). This resulted in a strong negative effect size for the individual study and would have resulted in an artificially lower mean effect size for the 'maintenance prescribing' group as a whole. The mean ORs for this group with the study included were FEOR= 1.00 and REOR= 1.41 (neither was statistically significant). None of the differences in ORs was statistically significant.

treatment were reduced by 60 per cent (1/2.49) according to the FE method and 62 per cent (1/2.61) according to the RE method.

In the case of post-release supervision and maintenance prescribing the mean OR was significant using the FE method, but just failed to reach significance using the RE method. The remaining two intervention types, drug courts and drug testing (combined) and other treatment programs (e.g. acupuncture), were not effective using either FE or RE methods.

Variations by the characteristics of clients

It is possible that some interventions have a greater effect on some types of drug misuser than on others. The most common demographic breakdowns discussed in the treatment literature are gender, age, and ethnic background. Unfortunately, it was not common for studies to provide separate results for different subgroups in a form suitable for meta-analysis. None of the studies included an analysis of different ethnic groups. However, five studies provided results disaggregated by the gender of the participant in a suitable form (4 for males and 1 for females) and 26 studies provided results according to the age of the participant; 22 studies were based on samples with a mean age of 17 years or above (coded as 'adults') and 4 studies were based on samples with a mean age of less than 17 years (coded as 'juveniles'). Hence, it was possible to provide some disaggregation of results in terms of the characteristics of clients.

The breakdown by gender (table 5) shows that the mean odds ratio was FEOR= 2.05 and REOR= 2.51 for males compared with FEOR= 1.02 and REOR= 1.02 for females. The mean effect size using both models was significant for males and non-significant for females. Hence, the results suggest that interventions are effective in reducing drug-related crime among males, but not females. However, the number of studies included in the comparison is small and there was no significant difference in the ORs of the two groups. The mean odds ratios for studies based on juveniles and adults were both statistically significant. However, the effect size was noticeably higher for juveniles than for adults (FEOR= 1.95 and REOR= 2.05 compared with FEOR= 1.29 and REOR= 1.57). This finding is based on a larger overall number of studies and provides a stronger indication of variations in the effectiveness of interventions by client type.

Year of study

One possible source of variation in outcome is the year of publication of the study. Studies published in the 2000s tended to have moderate to large effect sizes (FEOR= 1.92 and REOR= 1.98), whereas studies published in the 1980s and 1990s tended to have smaller effect sizes (FEOR= 1.15 and REOR= 1.35). Overall, the results show that the mean ORs of studies published in the 1980s and 1990s were not statistically significant, whereas the ORs of studies published in the 2000s were significant. This might

Table 5
Variations by age and gender

Group	N	Fixed effects				Random effects			
		OR	CI	Zp	Qp	OR	CI	Zp	Qp
Gender									
Males	4	2.05	1.58-2.67	0.0000	<0.0001	2.51	1.16-5.42	0.0193	ns
Females	1	1.02	0.41-2.52	ns	ns	1.02	0.41-2.52	ns	ns
Age									
Adults [1]	22	1.29	1.12-1.48	0.0004	<0.0001	1.57	1.15-2.13	0.0042	ns
Juveniles	4	1.95	1.60-2.38	0.0000	<0.0001	2.05	0.96-4.37	ns	ns

OR= Weighted Mean Odds Ratio. 'CI'= Confidence Interval. 'Zp'= Probability of Z. 'Qp'= Probability of Q. 'ns'= Not significant at p<.05.
 [1] Samples were coded as 'Adult' if the mean age of the group was aged 17 years or over and as 'Juvenile' if the mean age was less than 17 years.
 The table includes only those studies that provided information on gender or age. None of the differences in ORs was statistically significant.

Table 6
Year of study and country of origin

Group	N	Fixed effects				Random effects			
		OR	CI	Zp	Qp	OR	CI	Zp	Qp
Year of study									
1980s & 1990s	17	1.15	1.00-1.32	ns	<0.0001	1.35	0.97-1.88	ns	ns
2000s	11	1.92	1.61-2.28	0.0000	<0.0001	1.98	1.25-3.13	0.0034	ns
Country of origin									
USA	24	1.39	1.24-1.55	0.0000	<0.0001	1.45	1.09-1.93	0.0099	ns
Europe	4	6.20	2.08-18.45	0.001	ns	6.03	2.23-16.33	0.0004	ns

OR= Weighted Mean Odds Ratio. 'CI'= Confidence Interval. 'Zp'= Probability of Z. 'Qp'= Probability of Q. 'ns'= Not significant at p<.05. None of the differences in ORs was statistically significant.

be because treatment programs have become more effective over time in reducing crime or because more recent studies use different methods that might more easily demonstrate success. The previous analysis showed that quasi-experimental designs produced higher ORs than random allocation methods and that 'treatment-versus-no-treatment' comparisons produced higher ORs than 'treatment-versus-another-treatment' designs. Can the higher ORs in the 2000s be explained by a higher proportion of quasi-experimental designs or a higher proportion of 'treatment-versus-no-treatment' comparisons among these studies? Overall, there was a slightly higher proportion of quasi-experimental designs in the more recent period (55% in the 2000s versus 33% in the 1980s and 1990s) and slightly more 'treatment-versus-no-treatment' comparisons (55% compared with 47%). However, neither of these differences was statistically significant (Fisher's Exact Test). Hence, it is possible that the higher ORs of studies published in the 2000s is a result of treatment programs becoming more effective over time.

Country of origin

The majority of studies included in this review (24 of 28) were published in the USA and the remainder were published in Europe (4 of 28). The results shown in table 6 indicated much higher ORs among studies published in Europe (FEOR= 6.20 and REOR= 6.03) than studies published in the USA (FEOR= 1.39 and REOR= 1.45). The same explanations discussed in the previous section might apply here. The difference might be a result of method differences or differences in the effectiveness of the interventions. In relation to methods, all four European studies used quasi-experimental designs (shown previously to be associated with higher ORs). Conversely, all of the random allocation studies reviewed were conducted in the USA. Hence, the fact that European studies were based wholly on quasi-experimental designs might explain some of the difference. While the number of cases is too small to arrive at any strong conclusions, it is possible that the difference between American and European studies can be explained by differences in research methods.

Discussion

Overall findings

The results of a meta-analysis of 28 studies showed that the mean odds of offending following treatment for all studies combined were between 1.41 and 1.56 times greater for the comparison groups than the treatment groups. This means that the odds of offending for the treatment group were between 29 per cent and 36 per cent lower following treatment. Hence, the analysis suggests that drug treatment is effective in reducing offending. However, these overall findings mask some important variations in the results.

The study showed that the results varied according to the type of methods used in the original evaluations. Quasi-experimental designs produced larger mean effect sizes than random allocation designs and 'treatment-versus-non-treatment' comparisons produced larger effect sizes than 'treatment-versus-other-treatment' comparisons. There was also some variation relating to the type of program. There was some evidence that therapeutic communities, post-release supervision for drug-misusing offenders, and maintenance prescribing were effective

in reducing offending. However, there was no evidence that drug courts and drug testing (combined) and alternative treatment programs were effective. This is not to say that these programs might not be effective under some conditions and in relation to some clients. There was also some variation in outcome according to the characteristics of the participants. Studies based on males produced larger mean effect sizes than those based on females and evaluations based on juveniles produced larger effects than those based on adults. This review also has shown that high-intensity programs (in terms of dosage levels, duration, and continuity) were more effective in reducing criminal behavior than low-intensity programs.

Limitations of the research

There are a number of reasons why the above findings should be viewed as tentative. First, it was not possible to search all known sources during the time period allocated to the research. However, we obtained initial information on over 9,000 citations and found over 50 eligible evaluations. Second, the relatively small number of studies suitable for meta-analysis (n= 28) limited the ability of the review to disaggregate the findings by features of the program or the participants. This was particularly problematic in investigating variations by gender and in treatment effect by type of program. It also meant that the statistical power of the tests used to determine the significance of differences in ORs was limited. Third, the selection process focused on published articles rather than unpublished manuscripts. It was mentioned earlier that this might result in 'publication bias' in that evaluations producing desirable results might be more likely to be published than evaluations producing undesirable or non-significant results. However, there has been at least one study (Wilson et al, unpublished manuscript) that has investigated differences between the results of published and unpublished materials in relation to evaluations of drug courts and it concluded that there was no significant difference between the two in measured effectiveness.

Implications for future research

There are a number of implications for future evaluation research on the effectiveness of drug treatment on criminal behavior. First, this review has drawn attention to the relatively small number of high-quality evaluations. Most of the studies reviewed were conducted in the USA, whereas only a few were conducted in Europe or the UK. Second, it has shown that there are problems with the quality of research evaluations. The majority of studies originally selected were eventually excluded on the grounds that they did not meet the minimum standards of methodological quality. The most common reason for rejection was that no comparison group was included. Third, there was a strong reliance among evaluations on quasi-experimental designs; all of the European studies were based on these. It is possible that quasi-experimental designs are prone to selection bias in relation to the treatment samples, whereby the most promising clients are allocated to the experimental treatment. It would improve the overall quality of results if evaluations were based on random allocation designs.

Fourth, the results of evaluations can be affected by the type of comparison group used. Treatment-versus-no-treatment comparisons provide the strongest and most encouraging results. However, many evaluations are based on comparisons of treatment-versus-other-

treatment. These can underestimate the effectiveness of a program because the comparison intervention might also be effective. Hence, future research should try to use comparison groups that do not receive treatment. Fifth, many studies were excluded from the meta-analysis because the results were not presented in a way that would allow an effect size to be calculated. It would be helpful if published evaluations included raw data, cell sizes and other relevant information in order to facilitate future meta-analyses or if their data were deposited in data archives. Finally, very few evaluations disaggregated the findings in a way that would show differential effects for subgroups. As some programs might be effective for some clients rather than others, it is important that this information is included in a research report.

Implications for policy

Government drugs policy in the UK is not particularly specific about the type of treatment that should be used to reduce crime. The most recent policy documents (Home Office, 2002; Home Office, 2004) give support to a wide range of measures, from drug testing to heroin prescription. However, the current review suggests that some programs work better than others. In particular, there is evidence that therapeutic communities, post-release supervision, and maintenance prescribing reduce criminal

behavior. Hence, there might be something to be gained from prioritizing certain kinds of interventions over others. The results of the current research also show that more intensive interventions are more successful than less intensive programs. This applies to dosage levels, whether the program is continuous or interrupted, time in treatment, and whether the participant completes or terminates the program. Hence, government policy might also aim to promote more intensive treatment programs.

There is also some evidence that more favorable results are obtained with males compared with females and with young compared with old clients. These results highlight the need for government policy to ensure that treatment is better suited to meet the needs of women and older drug users in order to obtain more successful outcomes. While the current research has not investigated the interaction between type of program and type of participant, the evidence that has been provided at least suggests that this is an area that might be worth investigating further.

Acknowledgements

The findings reported in this paper derive from research funded by the UK Home Office. The opinions expressed in the paper are not necessarily the official view of the Home Office and should not necessarily be considered an indication of Home Office policy.

References

- Bale, R.N., Van Stone, W.W., Kuldau, J.M., Engelsing, T.M.J., Elashoff, R.M. and Zarcone, V.P. (1980). Therapeutic communities versus methadone maintenance. *Archives of General Psychiatry*, 37, 179-193.
- Britt, C.L., Gottfredson, M.R. and Goldkamp, J.S. (1992). Drug testing and pretrial misconduct: an experiment on the specific deterrent effects of drug monitoring dependents on pretrial release. *Journal of Research in Crime and Delinquency*, 29, 62-78.
- Brown, B.S., O'Grady, K.E., Battjes, R.J., Farrell, E.V., Smith, N.P. and Nurco, D.N. (2001). Effectiveness of a stand-alone aftercare program for drug-involved offenders. *Journal of Substance Abuse Treatment*, 21, 185-192.
- Chanhatisilpa, C., MacKenzie, D. and Hickman, L. (2000). The effectiveness of community-based programs for chemically dependent offenders: a review and assessment of the research. *Journal of Substance Abuse Treatment*, 19, 383-393.
- Deschenes, E.P., Turner, S. and Petersilia, J. (1995). A dual experiment in intensive community supervision: Minnesota's prison diversion and enhanced supervised release programs. *Prison Journal*, 75, 330-356.
- Dynia, P. and Hung-En, S. (2000). The safety and effectiveness of diverting felony drug offenders to residential treatment as measured by recidivism. *Criminal Justice Policy Review*, 11, 299-311.
- Farabee, D., Shen, H., Hser, Y., Grella, C. and Anglin, M. (2001). The effect of drug treatment on criminal behavior among adolescents in DATOS-A. *Journal of Adolescent Research*, 16, 679-696.
- Farrell, A. (2000). Women, crime and drugs: testing the effect of therapeutic communities. *Women and Criminal Justice*, 11, 21-48.
- Farrington, D.P. (2003). Methodological quality standards for evaluation research. *Annals of the American Academy of Political and Social Science*, 587, 49-68.
- Farrington, D.P., Gottfredson, D.C., Sherman, L.W. and Welsh, B.C. (2002). The Maryland Scientific Methods Scale. In L.W. Sherman, D.P. Farrington, B.C. Welsh and D.L. MacKenzie (eds.): *Evidence-based crime prevention*. (pp. 13-21). London: Routledge.
- Farrington, D.P. and Petrosino, A. (2000). Systematic reviews of criminological interventions: the Campbell Collaboration Crime and Justice Group. *International Annals of Criminology*, 38, 49-66.
- Farrington, D.P. and Welsh, B. (2002). *Effects of improved street lighting on crime: a systematic review* (Home Office Research Study No. 251). London: Home Office.
- Ghodse, H., Reynolds, M., Baldacchino, A., Dunmore, E., Byrne, S., Oyefeso, A., Clancy, C. and Crawford, V. (2002). Treating an opiate-dependent inpatient population: a one-year follow-up study of treatment completers and noncompleters. *Addictive Behaviors*, 27, 765-778.
- Gordon, J., Moriarty, L. and Grant, P. (2000). The impact of a juvenile residential treatment center on minority offenders. *Journal of Criminal Justice*, 16, 194-208.
- 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, 291-303.
- Gottfredson, D. and Exum, M. (2002). The Baltimore city drug treatment court: one-year results from a randomized study. *Journal of Research in Crime and Delinquency*, 39, 337-356.
- Haapanen, R. and Britton, L. (2002). Drug testing for youthful offenders on parole: an experimental evaluation. *Criminology and Public Policy*, 1, 217-244.
- Hall, W. (1996). Methadone maintenance treatment as a crime control measure. *Crime and Justice Bulletin*, 29, 1-12.
- Hoffmann, N. and Norman, S. (1992). Treatment outcomes for abstinence-based programs. *Psychiatric Annals*, 22, 402-408.
- Home Office (2002). *Updated drug strategy 2002*. London: Home Office.
- Home Office (2004). *Tackling drugs, changing lives*. London: Home Office.
- Hser, Y.I., Grella, C.E., Hubbard, R.L., Hsieh, S.C., Fletcher, B.W., Brown, B.S. and Anglin, M.D. (2001). An evaluation of drug treatments for adolescents in 4 US cities. *Archives of General Psychiatry*, 58, 689-695.
- Hubbard, R.L., Marsden, M.E., Rachal, J.V., Harwood, H.J., Cavanaugh, E.R. and Ginzburg, H.M. (1989). *Drug abuse treatment: a national study of effectiveness*. London: University of North Carolina Press.
- Hutchinson, S.J., Taylor, A., Gruer, L., Barr, C., Mills, C., Elliott, L., Goldberg, D.J., Scott, R. and Gilchrist, G. (2000). One-year follow-up of opiate injectors treated with oral methadone in a GP-centered program. *Addiction*, 95, 1.055-1.068.

- Inciardi, J., Martin, S., Butzin, C., Hooper, R. and Harrison, L. (1997). An effective model of prison-based treatment for drug-involved offenders. *Journal of Drug Issues*, 27, 261-278.
- Latessa, E.J. and Moon, M.M. (1992). The effectiveness of acupuncture in an outpatient drug treatment program. *Journal of Contemporary Criminal Justice*, 8(4), 317-331.
- Lipsey, M.W. and Wilson, D.B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Magura, S., Rosenblum, A., Lewis, C. and Joseph, H. (1993). The effectiveness of in-jail methadone maintenance. *Journal of Drug Issues*, 23, 75-99.
- Marsch, L. (1998). The efficacy of methadone maintenance interventions in reducing illicit opiate use, HIV risk behavior and criminality: a meta-analysis. *Addiction*, 93, 515-532.
- Martin, S. and Scarpitti, F. (1993). An intensive case management approach for paroled IV drug users. *Journal of Drug Issues*, 23, 43-59.
- McBride, D.C. and Inciardi, J.A. (1993). The focused offender disposition program: philosophy, procedures and preliminary findings. *Journal of Drug Issues*, 23, 143-160.
- Mitchell, O., MacKenzie, D.L. and Wilson, D.B. (2006) Incarceration-based drug treatment. In B.C. Welsh and D.P. Farrington (eds.): *Preventing crime: what works for children, offenders, victims and places* (pp. 103-116). Dordrecht, Netherlands: Springer.
- Nemes, S., Wish, E. and Messina, N. (1999). Comparing the impact of standard and abbreviated treatment in a therapeutic community. *Journal of Substance Abuse Treatment*, 17, 339-347.
- Pearson, F. and Lipton, D. (1999). A meta-analytic review of the effectiveness of corrections-based treatments for drug abuse. *Prison Journal*, 79, 384-410.
- Perneger, T.V., Giner, F., del Río M. and Mino A. (1998). Randomized trial of heroin maintenance program for addicts who fail in conventional drug treatments. *British Medical Journal*, 317, 13-18.
- Prendergast, M., Podus, D., Chang, E. and Urada, D. (2002). The effectiveness of drug abuse treatment: a meta-analysis of comparison group studies. *Drug and Alcohol Dependence*, 67, 53-72.
- Sherman, L.W., Gottfredson, D., MacKenzie, D.L., Reuter, P., Eck, J. and Bushway, S. (1997). *Preventing crime: what works, what doesn't, what's promising*. Washington, DC: U.S. Department of Justice.
- Simpson, D.D. and Sells, S.B. (1982). Effectiveness of treatment for drug abuse: an overview of the DARP research program. *Advances in Alcohol and Substance Abuse*, 2, 7-29.
- Simpson, D., Joe, G.W. and Brown, B.S. (1997). Treatment retention and follow-up outcomes in the drug abuse treatment outcome study (DATOS). *Psychology of Addictive Behaviors*, 11, 294-307.
- Strang, J., Marsden, J., Cummins, M., Farrell, M., Finch, E., Gossop, M., Stewart, D. and Welch, S. (2000). Randomized trial of supervised injectable versus oral methadone maintenance: report of feasibility and 6-month outcome. *Addiction*, 95, 1.631-1.645.
- Turner, S., Greenwood, P., Fain, T. and Deschenes, E. (1999). National Drug Court Institute review: perceptions of drug court: How offenders view ease of program completion, strengths and weaknesses, and the impact on their lives. *National Drug Court Review*, 2, 61-86.
- Vaughn, M.G. and Howard, M.O. (2004). Adolescent substance abuse treatment: a synthesis of controlled evaluations. *Research on Social Work Practice*, 14, 325-335.
- Weisburd, D., Lum, C.A. and Petrosino, A. (2001). Does research design affect study outcomes in criminal justice? *Annals of the American Academy of Political and Social Science*, 578, 50-70.
- Welsh, B.C. and Farrington, D.P. (2002). *Crime prevention effects of closed circuit television: a systematic review* (Home Office Research Study No. 252.) London: Home Office.
- Wexler, H., De Leon, G., Thomas, G., Kressel, D. and Peters, J. (1999). The Amity Prison TC evaluation. *Criminal Justice and Behavior*, 26, 147-167.
- Wilson, D.B., Mitchell, O. and MacKenzie, D.L. (unpublished manuscript). *A systematic review of drug court effects on recidivism*. Fairfax, VA: George Mason University.