Psychology, Crime & Law, Vol. 5, pp. 251–278
Reprints available directly from the publisher
Photocopying permitted by license only

THE INFLUENCE OF TREATMENT PROGRAMMES ON THE RECIDIVISM OF JUVENILE AND ADULT OFFENDERS: AN EUROPEAN META-ANALYTIC REVIEW

SANTIAGO REDONDO[a,*], JULIO SÁNCHEZ-MECA[b] and VICENTE GARRIDO[c,1]

[a] *Centre of Legal Studies of the Autonomous Catalan Government C/Roger de Flor, 196, 08013 Barcelona, Spain;*
[b] *Departamento de Psicología Básica y Metodología, Facultad de Psicología, Universidad de Murcia (Campus Universitario de Espinardo), Apartado 4021, 30080 Murcia, Spain;*
[c] *Universidad de Valencia, Facultad de Psicología y Pedagogía, 46010 Valencia, Spain*

Meta-analyses have examined the effectiveness of different treatment techniques on delinquent and criminal behaviour in America and Europe. In one of the last meta-analyses, that integrated the results of programmes applied in European countries (Redondo, Garrido, and Sánchez-Meca, 1997), the global effect size obtained for all treatment and effectiveness measures (Psychological factors, education, etc.) was $d = 0.3039$ ($r = 0.15$). In general, it could be interpreted that treatment groups surpassed controls by 15%. In this work, the results of a meta-analysis to determine which treatments are more effective in reducing on recidivism rates are presented. A total of 32 European studies that evaluated recidivism during an average follow-up period of two years obtained a global effect size of $d = 0.243$ ($r = 0.12$), equivalent to a 12% reduction in recidivism. Regarding the differential influence of treatment models, behavioural and cognitive-behavioural techniques were most beneficial in reducing recidivism.

Key words: meta-analysis; correctional treatment evaluation.

1. INTRODUCTION

In the 1960s and 1970s, many criminal justice researchers and practitioners displayed considerable confidence in the 'rehabilitative ideal' and its possibilities for reducing criminal recidivism. During the 1975–1981 period, this optimism was replaced by pessimism and the idea that "nothing works". However, by 1988–1989 several literature

* Corresponding author.

reviews and meta-analyses published during the 1980s helped to reverse this perspective: there were effective programmes, although not many, but they definitively reduce recidivism (Palmer, 1995). This area of research has continued through the 1990s. In fact, concerning crime correction, a series of meta-analytic reviews has analysed over the last decade the effectiveness of different treatment techniques on criminal behaviour. These analyses have included both juvenile and adult offenders, and institutional as well as community programmes. These meta-analyses have assessed, first in America and afterwards in Europe, a lot of treatment programmes that were applied to thousands of delinquents and offenders (Andrews, Zinger, Hoge, Bonta, Gendreau, and Cullen, 1990; Antonowicz and Ross, 1994; Cleland, Pearson, and Lipton, 1996; Garrett, 1985; Gensheimer, Mayer, Gottschalk, and Davidson II, 1986; Gottschalk, Davidson II, Gensheimer, and Mayer, 1987; Gottschalk, Davidson II, Mayer, and Gensheimer, 1987; Hall, 1995; Izzo and Ross, 1990; Lab and Whitehead, 1988; Lipsey, 1992; Lösel and Köferl, 1989; Quinsey, Harris, Rice, and Lalumière, 1993; Redondo, 1994; Redondo, Garrido, and Sánchez-Meca, 1997; Wells-Parker, Bangert-Drowns, McMillen, and Williams, 1995; Whitehead and Lab, 1989). The average effectiveness of the programmes analysed varied between 5% and 18%. For example, the European meta-analysis of Redondo *et al.* (1997) integrated the results of 57 programmes applied to youthful and adult offenders, obtaining a global effect size of $r = 0.150$ for all treatment and all outcome measures (psychological adjustment, school adjustment, recidivism, etc.). In this case, it could be interpreted that treatment groups surpassed controls by 15%.

Meta-analysis is a technique for synthesising the empirical findings of multiple studies that research the same issue. Its employment in Criminology has not been limited to the issue of offender rehabilitation, but has covered related areas of interest such as drink/drive offenders (Wells-Parker, Bangert-Drowns, McMillen and Williams, 1995), the predictors of adult offender recidivism (Gendreau, Little and Goggin, 1996), and alternative education programmes which prevent school maladjustment and delinquency (Cox, Davidson and Bynum, 1995), to name a few examples. In all the cases, the purpose is to develop generalisations about research findings across different studies. This is very different from the traditional (narrative) reviews, which tend to suffer from: (a) selective inclusion of studies, (b) differential subjective weighting of studies in the interpretation of a set of findings, (c) misleading interpretations of study findings, (d) failure to examine other study characteristics as potential explanations for consistent results across studies, and (e) failure to examine the effect of moderator variables in relationship to the outcome variable (Wolf, 1986). It is true that

recently there have been criticisms of the capability of meta-analysis to build knowledge in psychology (Eysenck, 1995; Sohn, 1995), but Lipsey and Wilson (1995, p. 114) have made clear that this methodology "assumes that the contact with nature made by the primary researcher has been dutifully communicated and that the validity of the empirical findings so communicated does not degrade just because it is incorporated into a larger analysis that examines the generalisability of such findings across settings, subjects, methodological variations, and the like" (see also Cooper, 1989; Glass, McGaw, and Smith, 1981; Gómez, 1987; Rosenthal, 1991; Sánchez-Meca and Ato, 1989). In order to achieve this aim, the results of different studies on a given topic are translated into such effect size indexes as the correlation coefficient, standardised mean difference, or phi coefficient. Effect size represents the effect magnitude of the results and, in terms of the Rosenthal's (1991) Binomial Effect Size Display, BESD, could be directly interpreted as the improvement percentage of treatment groups vis-à-vis controls. Applied to the programme evaluation context, meta-analytic techniques achieve three main objectives: (a) to obtain a global index of programme effectiveness; (b) to determine the homogeneity of the results related to this global index; and (c) if homogeneity is not met, to search for study characteristics that may explain the variability in effect size.

In the evaluation of delinquency and crime treatment programmes different outcome measures have been used as effectiveness criteria. If we consider the above-mentioned meta-analyses, the more commonly used outcome measures were those related to criminal behaviour and recidivism, followed in order of frequency, psychological adjustment, school attendance, academic achievement, vocational improvement, social and interpersonal adjustment, institutional adjustment, and programme implication. The meta-analyses which used a greater number of outcome indexes were those of Garrett (1985), Gottschalk, Davidson II, Gensheimer, and Mayer (1987), Lipsey (1992), and Redondo (1994; see also Redondo *et al.*, 1997).

The use of the above measures, in addition to recidivism, in order to weigh the efficacy of programmes is often justified. Criminological literature has shown that recidivism is just one final product of intermediate stages in which school failure, unemployment, psychological disorders and violent behaviour are important antecedents. Due to this, most researchers have included some of these variables. In addition, all correctional meta-analyses have used some recidivism criteria. It has been evident for both primary researchers and reviewers that, in matters of delinquency and crime treatment, it is necessary to evaluate, one way or another, whether treated subjects relapse or not into crime after participation in a treatment programme. Recidivism is necessarily the bottom

line criterion for criminal justice efficacy. Society, public opinion, and authorities hope that correctional services will reduce the frequency and seriousness of future criminal conduct of treated offenders. Otherwise it could not be concluded that the programmes had been useful.

The main objective of the present research has been to analyse the effectiveness on recidivism of a set of crime treatment programmes recently applied in Europe. In fact, this meta-analysis introduces us to one of the questions that seems to be central at present in criminal justice systems: Do treatment programmes bring reductions in recidivism rates? Another critical issue is to determine whether different theoretical models produce a differential effectiveness in preventing recidivism, as well as the possible influence of methodological, subject, setting, and extrinsic variables on the study results.

2. METHOD

Literature Search

Literature was collected using several information channels. First, computer databases including the *Criminal Justice Periodical Index*, *Pascal*, and *PsycLIT* were consulted using the key-words *delinquen**, *offender*, *inmate*, *probation*, *treatment*, *rehabilitation*, *intervention*, *parole*, *and therap**. Second, a manual search of 55 (mostly European) specialised journals was carried out. Third, letters were written requesting studies about the topic, and they were sent to a total of 118 key researchers and to 82 European institutions linked to the field. All references cited in the selected studies were reviewed for inclusion as well. The literature search covered the years 1980 to 1991 and included both published and unpublished literature.

To be included in the meta-analysis, the study had to (a) be applied to subjects under the control of the criminal justice system (i.e., youthful or adult offenders), (b) use a treatment strategy for some time, (c) include a methodological design permitting a comparison between control/treatment groups or pre/post measures, and finally (d) include recidivism outcome. A total of 29 research papers with 32 studies meeting the inclusion criteria were identified in the search and used in this study. The total sample in these studies was 5,715 participants (see Table 4).

Coding of Variables

In order to explain the heterogeneity showed in the study results on correctional treatment, potential moderator variables of the studies

were defined and coded for subsequent analysis. Variables were classified in treatment, subject, setting, methodological, and extrinsic clusters (Lipsey, 1994; Sánchez-Meca, 1997). The cluster of *treatment variables* included such characteristics of the programmes as the underlying theoretical model (nonbehavioural, educational/informational, behavioural, cognitive-behavioural, deterrence theory, therapeutic community, and diversion), the duration of the programme (in months), the intensity of the programme (in hours per week/subject), and the magnitude of the programme (in total hours/subject). The cluster of *subject variables* included the subject age (adolescents, juveniles, mixed, and adults), the mean age of the sample (in years), the most common offence type in the sample (property, violent, sexual, drugs traffic, alcohol, and mixed), and the sample gender (percent male). The cluster of *setting variables* was composed of the setting in which the programme was implemented (juvenile reform centre, juvenile prison, adult prison, community, and other), the regime of the participants (closed, semi-open, open, and other), and the country. In the cluster of *methodological variables* the following characteristics were coded: the design type (between- versus within-group designs), the subjects' assignment to groups (random versus nonrandom assignment), the sample size, the selection criteria of the offenders (all subjects, sentence duration, specific typologies, violent subjects, and specific needs), the percentage of attrition in the treated group, the design quality[2] (scoring from 0, low quality, to 7, high quality), and the follow-up period (in months). Finally, the cluster of *extrinsic variables* included the year of the paper and the publication source (published versus unpublished).

All coding of the variables was completed independently by the first and third authors of this paper achieving, on average (86.2%), a satisfactory level of interrater reliability (Orwin, 1994). The discrepancies between the ratters were resolved by consensus.

Statistical Procedures

The standardised mean difference, d, was chosen as the effect size index, using a correction for small sample size (Hedges and Olkin, 1985, p. 81). In between-group designs, the d statistic was defined as the difference between the experimental and control means at follow-up, divided by the within-group standard deviation; in within-group designs, the difference between post-treatment and pre-treatment means divided by the within-group standard deviation was the common measure of the results. When the outcome measures were proportions (e.g., recidivism rates), an equivalent d statistic proposed in the program DSTAT (Johnson, 1991) was applied. Positive values of d indicate the amount

of improvement the treated group had beyond the control group or pretest assessment. To facilitate the interpretation of the results, each d statistic was also translated to a Pearson correlation coefficient, r, by means of: $r = d/(d^2+4)^{0.5}$ (Rosenthal, 1991). The studies frequently reported several recidivism measures (e.g., new police contacts, return to prison, new detentions, etc.); in those cases, an overall effect size was calculated, averaging across measures so that each study yielded only one effect to safeguard the assumption of independence among effect sizes.

The most common recidivism measure encountered is that which classified the subjects into recidivists versus nonrecidivists. When the study applied a dichotomous measure of recidivism and included two groups (treated versus control groups), the effect size index was also defined as the difference between the recidivism proportions, d_p, in both treated and control groups (p_E and p_C, respectively), that is, $d_p = p_C - p_E$ (Fleiss, 1994).

All calculations of effect sizes were carried out by the second author of this paper and an independent researcher, obtaining a satisfactory interrater reliability ($r = 0.826$, $p = 0.000$). The discrepancies were resolved reviewing the study together until consensus was achieved. In Table 1 the list of sample sizes, effect sizes, and some moderator characteristics of each study is presented.[3]

The goals of this meta-analysis were to assess the overall effectiveness of correctional treatment on recidivism, to determining whether the studies are homogeneous regarding outcome, and if that is not the case, to determine whether moderator variables explain the variation among the studies. The statistical techniques used are outlined in Hedges and Olkin (1985) and in Cooper and Hedges (1994). The overall effect, d_{++}, is calculated by using a weighted mean that takes into account the differing variances in the individual studies, although other descriptive statistics are also calculated (unweighted mean, median, quartiles, etc.). To determine whether studies share a common population effect size, homogeneity of variance is calculated by the Q_T statistic which is distributed as the chi-square with $k-1$ degrees of freedom, where k is the number of studies. When homogeneity is rejected, moderator variables are tested for their possible contribution in explaining the variability in the effect sizes. For categorical moderator variables, an analogue to analysis of variance (weighted by inverse-variance of effect sizes) partitions the total variance, Q_T, into Q_B the explained variance, and Q_W, the residual variance. The significance of Q_B and Q_W are tested in the same fashion as Q_T, using $p-1$ and $k-p$ degrees of freedom, respectively, where p is the number of categories. For moderator variables that are metric,

Table 1 Effect sizes, sample sizes, and some characteristics of the studies

Study	*Country*	*Treat.*[a]	*Duration*[b]	*Age*[c]	*Follow-up*[d]	n^E	n^C	d	r
Belfrage (1991)	Sweden	PT	—	31	36	188	132	0.454	0.22
Berggren and Svärd (1990)	Sweden	TC	9.5	29	24	280	834[e]	0.259	0.12
Bishop *et al.* (1987)	Sweden	TC	4.5	25	16	42	38	0.099	0.04
Bovens (1987)	Netherlands	ED	0.25	31	12	91	62	0.263	0.13
Brown (1985)	Gr. Britain	BT	8	14	12	8	—[f]	−0.194	−0.11
Collins and Tate (1988)	Gr. Britain	DI	6	31	24	29	19	0.811	0.38
Cook *et al.* (1991), study 1	Gr. Britain	PT	22	37	57	33	11	−0.207	−0.10
Cook *et al.* (1991), study 2	Gr. Britain	PT	6	31	57	11	—[f]	−1.260	−0.54
Cooke (1989)	Gr. Britain	TC	41	31	24	12	—[f]	0.718	0.34
Cooke (1991)	Gr. Britain	PT	3	39	28	120	120[g]	0.603	0.29
Day (1988)	Gr. Britain	BT	17.7	21	39	20	—[f]	0.310	0.16
Dem. Unit (1986)	Gr. Britain	PT	—	31	24	109	—[f]	0.039	0.02
Dünkel (1982)	Germany	PT	14.7	33	54	323	889	0.480	0.23
Kruissink (1990)	Netherlands	DI	—	15	12	124	68	0.421	0.20
Kury (1989), study 1	Germany	CBT	2	18	24	64	106	0.144	0.07
Kury (1989), study 2	Germany	PT	2	18	—	32	106[h]	0.112	0.07
Legaz *et al.* (1990)	Spain	DI	7.7	12	1	10	6	0.281	0.14
McMurran and Boyle (1990), study 1	Gr. Britain	ED	0.25	18	15	15	13	0.000	0.00
McMurran and Boyle (1990), study 2	Gr. Britain	ED	0.25	18	15	15	13[h]	−0.247	−0.12
Petterson *et al.* (1986)	Sweden	TC	12	26	24	70	61	0.658	0.31
Redondo *et al.* (1990)	Spain	BT	3.5	18	36	288	—[f]	0.503	0.24
Robertson and Gunn (1987)	Gr. Britain	TC	—	32	120	61	61	−0.168	−0.08
Rosner (1988)	Germany	PT	1	31	30	420	47	0.405	0.19
Scholte and Smit (1989)	Netherlands	DI	—	14	6	71	71	0.217	0.10
Singer (1991)	Gr. Britain	ED	2	20	12	152	—[f]	−0.031	−0.01

Table 1 (*Continued*)

Study	*Country*	*Treat.*[a]	*Duration*[b]	*Age*[c]	*Follow-up*[d]	n^E	n^C	*d*	*r*
Slot (1983)	Netherlands	BT	—	17	6	9	17	0.419	0.21
Slot (1984)	Netherlands	BT	12.5	16	6	6	—[f]	0.447	0.22
Slot and Bartels (1983)	Netherlands	CBT	—	17	7	29	29	1.237	0.53
Slot and Heiner (1986)	Netherlands	BT	—	16	8	22	—[f]	0.547	0.27
Thornton (1987)	Gr. Britain	DI	—	31	12	1,000	—[f]	−0.012	−0.01
Van Dalen (1989)	Netherlands	ED	0.1	31	12	250	250[g]	0.281	0.13
Weaver and Fox (1984)	Gr. Britain	CBT	15	31	65	38	—[f]	0.647	0.31

[a]Treatment typology: PT = Psychotherapy (non behavioural): psychodynamic individual or group therapy. ED = Education (school, educational materials), BT = Behavioural Therapy (operant conditioning): token economy, environmental design based on contingencies. CBT = Cognitive-behavioural therapy (social skills training, psycho-social competence programmes). ID = Improvement/Discipline (classical penal theory, deterrence, 'shock incarceration', increasing levels of institutional control). TC = Therapeutic community (relationships between inmates/staff similar to patient/nurses, decreasing levels of institutional control). DI = Diversion.
[b]Programme duration in months.
[c]Average age of the sample in years.
[d]Duration of follow-up period in months.
[e]Control group is obtained from an external source.
[f]Within-group design.
[g]n^C is not reported; we assumed $n^C = n^E$.
[h]Same control group as in study 1.

regression analyses, by weighted least squares, were used to determine the association with effect sizes; in this case, the total variability, Q_T, is partitioned into Q_R, the explained variability, and Q_E, the unexplained variability. The significance of both Q_R and Q_E is tested as above with one and $k-2$ degrees of freedom, respectively. Finally, in order to examine the relationship between effect size and moderator variables partialling the influence of other clusters of variables, multivariate analyses were carried out applying multiple regression by weighted least squares (Hedges, 1994; Hedges and Olkin, 1985).

3. RESULTS AND DISCUSSION

Study Characteristics

Tables 2 and 3 show the descriptive characteristics of the 32 studies included in the meta-analysis. Great heterogeneity in the programmes implemented with respect to the theoretical model can be observed, with a median duration of 6 months, a median intensity of 32 hours week per subject, and a median programmes magnitude of 840 hours/subject. The distribution of the magnitude of interventions also showed great variability, from programmes that had a minimal magnitude of 1.5 hours per subject to others whose magnitude was greater than 9,000 hours per subject (see Table 3). Nevertheless, concerning magnitude it is necessary to make a distinction between *ambulatory* and *residential* programmes. The intervention magnitude of ambulatory programmes ranged from 1.5 to 180 hours, and the median was 30.6 hours. In this set of ambulatory programmes the following treatment categories were included: psychotherapy, education, behavioural interventions, and cognitive-behavioural treatments. In all of these cases, magnitude represents the real time of treatments application. On the other hand, in the case of residential programmes, where discipline, therapeutic community, and environmental contingency treatments were considered, magnitude was artificially calculated starting from when treated/control offenders entered the units where the above programmes operated. Obviously, the apparent magnitude of residential programmes was very high in comparison with ambulatory programmes. It ranged from 840 to 9,840 hours, and had a median magnitude of 2,508 hours.

The characteristics of participants in the studies were also heterogeneous; the median age of subjects was 25.5 years and the most common offences were property related offences, mixed, and alcohol abuse. However, with the exception of one study (Bishop, Sundin-Osborne,

Table 2 Descriptive characteristics of categorical variables

Cluster characteristic	*Frequency*	*%*
Treatment cluster		
Theoretical model ($k = 32$)		
Nonbehavioural	8	25
Educational	5	15.6
Behavioural	6	18.8
Cognitive-behavioural	3	9.4
Deterrence theory	1	3.1
Therapeutic community	5	15.6
Diversion	4	12.5
Subject cluster		
Sample age ($k = 32$)		
Adolescents (<16)	7	21.9
Juveniles (16–21)	6	18.7
Mixed	4	46.9
Adults (>21)	15	12.5
Offence type ($k = 27$)		
Property	6	22.2
People	1	3.7
Sexual	3	11.1
Drugs traffic	2	7.4
Alcohol	7	25.9
Mixed	8	29.7
Setting cluster		
Programme setting ($k = 29$)		
Juvenile reform center	1	3.5
Juvenile prison	5	17.2
Adult prison	7	24.1
Community	12	41.4
Other	4	13.8
Regime ($k = 29$)		
Closed	11	37.9
Semi-open	2	6.9
Open	10	34.5
Other	6	20.7
Country ($k = 32$)		
Germany	4	12.5
Great Britain	14	43.8
Netherlands	8	25.0
Spain	2	6.2
Sweden	4	12.5
Methodologic cluster		
Design type ($k = 32$)		
Between-group	22	68.8
Within-group	10	31.2
Subject assignment ($k = 22$)		
Random	3	13.6
Nonrandom	19	86.4

Table 2 (*Continued*)

Cluster characteristic	*Frequency*	*%*
Subject selection criteria ($k = 32$)		
All subjects	6	18.8
Sentence duration	2	6.2
Specific typologies	7	21.9
Violent subjects	1	3.1
Specific needs	16	50.0
Extrinsic cluster		
Publication source ($k = 32$)		
Published	25	78.1
Unpublished	7	21.9

Table 3 Descriptive characteristics of quantitative variables

Cluster Characteristics	*k*	*Min.*	*Max.*	*Mean*	*S.D.*	*Median*
Treatment cluster						
Programme duration (months)	23	0.1	41	8.3	9.5	6
Programme intensity (hours week)	24	0.7	56	29.6	27	32
Programme magnitude (hours/subject)	23	1.5	9,840	1,323.4	2,203.1	840
Subject cluster						
Average sample age (years)	32	12.5	39.3	24.7	7.8	25.5
Sample gender (% of men)	28	0	100	91.8	21.5	100
Methodological cluster						
Sample size	32	6	1,212	179	267	114
Attrition (in treated group)	21	0	82.1	27.9	24.4	23.1
Design quality	32	1	6	3.4	1.2	3
Follow-up (months)	31	1	120	26.4	24.2	24
Extrinsic cluster						
Date (year)	32	1982	1991	1988	2.6	1988

Min.: Minimum value. Max.: Maximum value. S.D.: Standard deviation.

and Petterson, 1987), all of them were applied to male populations (71.4%) or to a mixture of male and female populations (25%).

Great Britain was the country most represented in the sample of studies (43.8%). The programmes were applied mostly to subjects in closed (37.9%) or open (34.5%) regimes, and the most frequent programme setting was the community (41.4%).

With respect to methodological characteristics, 22 studies applied between-group and 10 within-group designs, with very heterogeneous

sample sizes ranging from 6 to 1,212 subjects (median: 114 subjects). The median percentage of attrition in treated groups was 23.1% and median design quality was of 3 points (on a scale from 0 to 7).

One of the main features of recidivism research concerns follow-up duration. For the programmes included in this meta-analysis the follow-up was an average of 26.4 months (median: 24 months). Follow-up periods of two years may seem insufficient to evaluate the recidivism of a sample of offenders. In fact, in many studies researchers evaluate recidivism after three, five, or more years of release from prison. However, although the follow-up period here is not as long as would be suitable, it can be considered that in general the obtained data are relevant, since the greater percentage of recidivism occurs during the period that immediately follows the offender's release. For example, Redondo, Funes, and Luque (1994) assessed recidivism in a sample of 485 Spanish ex-prisoners during almost four years and obtained an average recidivism rate of 37.9%. From this total recidivism rate, 58% of the recidivism took place in the first year after release, and in the second year 80% of the recidivism already had occurred. European recidivism rates referring to different countries were analysed in Tournier and Barre (1990), and they found that between 49.5% and 80% of recidivism, depending on the study, took place during the first and second follow-up year, while between 20% and 50.5% took place from the third year on. On average, 67.7% of the recidivism occurred in the two first years of follow-up, while the remaining 32.3% happened from the third year onwards. These results coincide with the meta-analytic review presented in Sánchez-Meca, Marín, and Redondo (1996) that compared recidivism rates of European and American countries. According to that, one can suppose that having information on recidivism for two follow-up years, means that we have the lion's share of information on the recidivism of subjects that were treated in the analysed programmes.

Overall Effect Size

For the objectives of the present meta-analysis, recidivism was defined as any measure related to new delinquent or criminal behaviour, contacts with the police, arrests, return to correctional institutions, and so on. Consequently, our analyses included programmes that had used such measures of recidivism as: parole and probation revocation, self-informed offences, new offences and new specific offences (serious, sexual, drugs traffic), new sentences and new serious sentences, return to prison by new offences, vandalism, police contacts, and return to juvenile institutions. Table 4 presents the descriptive statistics of the

standardised mean differences (d index). Most of the programmes (75%) showed lower recidivism rates in treatment than in control groups, or lower recidivism in pre-test than in post-test measures. Only one study (McMurran and Boyle, 1990, study 1) obtained a null effect size and seven studies evidenced counteractive results for the treatment. These studies showing worse results in the treated subjects included two nonbehavioural programmes (Cook *et al.*, 1991, studies 1 and 2), two educational programmes (McMurran and Boyle, 1990, study 2; Singer, 1991), one therapeutic community programme (Robertson and Gunn, 1987), one contingency environmental programme (Brown, 1985), and one deterrence programme (Thornton, 1987).

In terms of the standardised mean difference, the effectiveness of correctional treatment on recidivism achieves a positive mean value of $d_+ = 0.242$ (95% confidence interval: 0.196 and 0.287) and, in terms of correlation coefficient, it supposes an average value of $r_+ = 0.120$. With respect to the difference in recidivism proportions between treated and control groups, of 32 studies 17 of them applied dichotomous measures of recidivism and included two groups (Table 4). The weighted mean in $d_p = 0.15$ reflects a differential recidivism rate of 15% in favour of treated groups compared to the recidivism rate of controls (95% confidence interval: 11.9% and 18.1%). Therefore, our results are very similar to those obtained in other meta-analyses on recidivism measures, such

Table 4 Summary statistics of ESs distribution (d and d_p indexes)

Statistic	*d index*	*d_p index*
k	32	17
Number of treated subjects	3,964	1,772
Number of control subjects	1,751	2,503
Total number of subjects	5,715	4,275
Minimum	−1.2596	−0.545
Maximum	1.2370	0.449
Unweighted mean[a]	0.2572	0.086
Weighted mean[b]	0.2418	0.150
Weighted mean correlation[c]	0.120	—
Q_1[d]	0.019	0.000
Median	0.2809	0.123
Q_3[d]	0.491	0.207
Normal-based SD[e]	0.354	0.155
Confidence interval (95%)	0.196; 0.287	0.119; 0.181
Proportion of positive *ESs*	24/32 = 0.75	12/17 = 0.70
Homogeneity test (DF), p	124.070 (31), $p = 0.0000$	59.374(16), $p = 0.0000$

[a] Unweighted mean = $\Sigma d_i/k$.
[b] Weighted mean = $\Sigma w_i d_i/\Sigma w_i$, being w_i the inverse variance of each d_i.
[c] To the exception of this statistic, all calculations were obtained on d values.
[d] Q_1 and Q_3 are quartiles 1 and 3, respectively.
[e] Normal-based $SD = 0.7413(Q_3 - Q_1)$.

as Lösel (1987) and Whitehead and Lab (1989) both with $r=0.12$, Pearson *et al.* (November, 1995) with $r=0.157$, Andrews *et al.* (1990) with $r=0.10$, and the lower effectiveness found in the meta-analyses of Garrett (1985), with $r=0.065$, and Lipsey (1992), with $r=0.05$.

Although the results showed that, in general, the treatment of offenders is effective, the effect size distribution was very heterogeneous [d index: $Q_T(31)=124.070$, $p=0.0000$; d_p index: $Q_T(16)=59.374$, $p=0.0000$]. The remaining analyses test moderator variables in order to explain the differences among the results of the studies, being all of them based on the d index.

Analysis of Moderator Variables

The first of these analyses concerns the theoretical model of the treatment. As shown in Table 5 (see also Figure 1), this characteristic was statistically related to effect sizes [$Q_B(6)=59.565$, $p<0.01$], with 48% of the explained variance. Behavioural and cognitive-behavioural programmes were most successful at recidivism reducing ($r_+=0.231$ and $r_+=0.226$, respectively), obtaining double the effectiveness of the average of all programmes ($r_+=0.120$). On the contrary, educational programmes were less effective than the average ($r_+=0.08$) and, dramatically, the only deterrence programme here considered (Thornton, 1987) produced more recidivism than no intervention ($r_+=-0.006$). The contribution of this study to the average effect magnitude was very high because of its large sample size ($n=1{,}000$ subjects). Thus, if this study is deleted from the database, the average effect size increases from $r_+=0.120$ to $r_+=0.165$, and the heterogeneity statistic decreases from $Q_T=124.070$ to $Q_T=80.203$.

Other treatments characteristics positively related to effect sizes were programme duration (in months) [$Q_R=5.388$, $p<0.05$] and programme magnitude (in hours/subject) [$Q_R=3.899$, $p<0.05$], but their percentages of explained variance were very low (see Table 6).

In the cluster of subject characteristics, a negative relationship was found between the sample age and effect sizes [$Q_B=8.620$, $p<0.05$]. Although some effectiveness was evident at all ages, better results were obtained with adolescents ($r_+=0.206$) and juveniles ($r_+=0.179$) than with adult offenders ($r_+=0.101$).

Related to crime typology (Table 5), the only programme specifically applied to violent offenders (Cooke, 1989) obtained the highest effectiveness ($r_+=0.338$), followed by the mixed category ($r_+=0.184$), and the lowest effectiveness programmes occurred in sex offenders ($r_+=0.068$). However, the differences among the offence types were not very pronounced, not achieving statistical significance [$Q_B=7.042$, $p>0.05$].

Table 5 Results of the analyses of variance for categorical variables

Cluster characteristic	k_j	d_{+j}	95% C.I.	r_{+j}	Q_{wj}	Q_B	R^2
Treatment cluster						59.565**	0.480
Theoretical model ($k=32$)							
Nonbehavioural	8	0.390	+0.302; 0.478	0.191	28.828**		
Educational	5	0.162	0.037; 0.286	0.080	6.281		
Behavioural	6	0.476	0.327; 0.625	0.231	2.266		
Cognitive-behavioural	3	0.464	0.229; 0.698	0.226	11.952**		
Deterrence theory	1	−0.012	−0.099; 0.076	−0.006	—		
Therapeutic community	5	0.255	0.141; 0.369	0.126	12.154*		
Diversion	4	0.382	0.179; 0.586	0.188	3.025		
Subject cluster						8.620**	0.069
Sample age ($k=32$)							
Adolescents (<16)	7	0.420	0.231; 0.610	0.206	11.203		
Juveniles (16–21)	6	0.363	0.233; 0.493	0.179	9.987		
Mixed	4	0.241	0.139; 0.343	0.119	9.177		
Adults (>21)	15	0.202	0.145; 0.259	0.101	85.082**		
Offence type ($k=27$)						7.042	0.065
Property	6	0.210	0.147; 0.272	0.104	62.372**		
People	1	0.718	−0.107; 1.544	0.338	—		
Sexual	3	0.136	−0.217; 0.489	0.068	14.621**		
Drugs traffic	2	0.245	0.116; 0.375	0.122	0.465		
Alcohol	7	0.218	0.105; 0.331	0.108	12.278		
Mixed	8	0.374	0.248; 0.499	0.184	11.610		
Setting cluster						17.384**	0.150
Programme setting ($k=29$)							
Juvenile reform center	1	0.419	−0.397; 1.235	0.205	—		
Juvenile prison	5	0.354	0.221; 0.487	0.174	9.609		
Adult prison	7	0.164	0.103; 0.225	0.082	46.006**		
Community	12	0.244	0.146; 0.342	0.121	41.830**		
Other	4	0.500	0.327; 0.674	0.243	1.305		

Table 5 *(Continued)*

Cluster characteristic	k_j	d_{+j}	95% C.I.	r_{+j}	Q_{+j}	Q_B	R^2
Regime ($k = 29$)						12.484**	0.105
Closed	11	0.209	0.153; 0.265	0.104	57.800**		
Semi-open	2	−0.074	−0.400; 0.251	−0.37	1.672		
Open	10	0.229	0.124; 0.335	0.114	17.514		
Other	6	0.484	0.314; 0.654	0.235	29.469**		
Country ($k = 32$)						52.908**	0.426
Germany	4	0.404	0.298; 0.511	0.198	6.108		
Great Britain	14	0.051	−0.018; 0.121	0.026	46.500**		
Netherlands	8	0.351	0.232; 0.470	0.173	11.789		
Spain	2	0.497	0.334; 0.661	0.241	0.179		
Sweden	4	0.331	0.224; 0.438	0.163	6.587		
Methodological cluster							
Design type ($k = 32$)						25.823**	0.208
Between-groups	22	0.345	0.285; 0.405	0.170	56.912**		
Within-group	10	0.109	0.040; 0.177	0.054	41.336**		
Subject assignment ($k = 22$)						1.854	0.032
Random	3	0.037	−0.410; 0.484	0.018	1.410		
Nonrandom	19	0.351	0.290; 0.411	0.172	53.648**		
Subject selection criteria ($k = 32$)						27.307**	0.220
All subjects	6	0.134	0.069; 0.204	0.068	42.949**		
Sentence duration	2	0.531	0.381; 0.681	0.257	0.604		
Specific typologies	7	0.235	0.120; 0.351	0.117	18.523**		
Violent subjects	1	0.718	−0.107; 1.544	0.338	—		
Specific needs	16	0.307	0.226; 0.389	0.152	34.687**		
Extrinsic cluster							
Publication source ($k = 32$)						2.462	0.020
Published	25	0.225	0.176; 0.275	0.112	115.26**		
Unpublished	7	0.322	0.212; 0.431	0.159	6.342		

*$p < 0.05$, **$p < 0.01$.

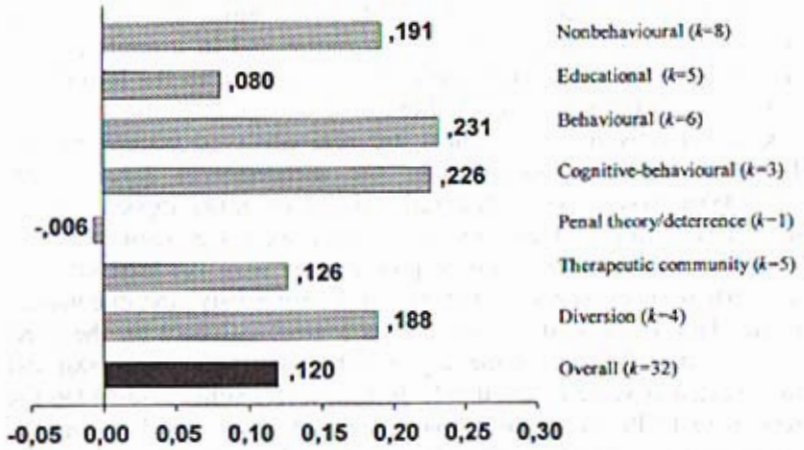

Figure 1 Treatment cluster: theoretical model of the treatment effectiveness (r).

Table 6 Results of the simple regression analyses for quantitative variables

Cluster Characteristic	k	B	Q_R	Q_E	R^2	R^2_{adj}
Treatment cluster						
Programme duration (in months)	23	0.011	5.388*	49.882**	0.097	0.054
Programme intensity (hours week)	24	−0.001	0.711	85.621**	0.008	0.000
Programme magnitude (hours/subject)	23	5.310^{-5}	3.899*	68.527**	0.054	0.009
Subject cluster						
Average sample age (years)	32	−0.004	1.166	122.774**	0.009	0.000
Sample gender (% of male)	28	−0.002	1.447	118.944**	0.012	0.000
Methodological cluster						
Sample size	32	-1.110^{-4}	4.467*	119.473**	0.036	0.004
Attrition (in treated group)	21	0.004	3.451	47.442**	0.068	0.019
Design quality	32	0.100	24.364**	99.576**	0.196	0.170
Follow-up (months)	31	0.004	9.676**	113.846**	0.078	0.046
Extrinsic cluster						
Date (year)	32	−0.014	2.815	121.125**	0.023	0.000

*$p < 0.05$, **$p < 0.01$.

Concerning the context cluster we found that programme's place was related to effectiveness (Table 5), something which is connected to the previous association with subjects. Lower rates of recidivism were obtained in the category 'other settings' (e.g., psychiatric units), with

$r_+ = 0.243$. On the other hand, double effectiveness was achieved in the only program applied in a youth centre and in juvenile prisons ($r_+ = 0.205$, and $r = 0.174$, respectively) in contrast to the lowest one obtained in adult prisons ($r_+ = 0.082$).

Clear effectiveness differences are also observed concerning the European countries included in the meta-analysis [$Q_B = 52.908$, $p < 0.0000$]: from the evaluated studies the programmes applied in Spain ($r_+ = 0.241$) and in Germany ($r_+ = 0.198$) were the most effective. The fact that the Great Britain programmes were the least effective ($r_+ = 0.026$), is explained by the studies' heterogeneity, and due mainly to the Thornton's study. This one has a great influence on the average size effect, because of the large number of subjects ($n = 1{,}000$) and the negative d value ($d = -0.012$). In fact, if this study is excluded the remaining 13 British studies show a mean of $r_+ = 0.080$, lowering the whole heterogeneity from 52.908 to 14.477, although still maintaining statistical significance ($p < 0.01$).

We also analysed the relationship between effectiveness and some methodological factors. With respect to the criteria used for inclusion of subjects in the programmes (Table 5), three studies corresponding to samples selected by violence and sentence duration criteria achieved the greatest effect sizes ($r_+ = 0.338$ and $r_+ = 0.257$, respectively). Conversely, studies that included 'all subjects' of the institution were less effective ($r_+ = 0.068$).

An important methodological aspect refers to the influence of design type on study results. Although it is common that one group pre-test/post-test designs have greater effect sizes than between-group ones, in our meta-analysis a contrary result has been obtained (Table 5): one-group studies achieved an average effect size ($r_+ = 0.054$) that was significantly lower than between-group studies ($r_+ = 0.170$). However, again the great weight assigned to Thornton's (1987) study can explain this anomalous result; in fact, if this study is deleted from the analysis, the remaining nine pre-test/post-test studies reach an average effect size of $r_+ = 0.147$, and no longer statistically differ from between-group designs [$Q_B = 0.578$, $p > 0.05$]. In this respect, Gottschalk, Davidson II, Gensheimer, and Mayer (1987) obtained a similar result. On the other hand, the 22 between-group studies were classified as a function of assignment type of the subjects into random versus nonrandom assignment. As is common (e.g., Cleland, Pearson, and Lipton, 1996; Hall, 1995; Lipsey, 1992; Whitehead and Lab, 1989), studies with random assignment obtained a lower average effect size ($r_+ = 0.018$) than those with nonrandom assignment ($r_+ = 0.172$), but the difference was not statistically significant [$Q_B = 1.854$, $p > 0.05$]. Nevertheless, because only three studies with random assignment were included in our analysis, this finding has to be interpreted cautiously.

Another important variable in the explanation of the heterogeneity found in the studies results is the duration of the follow-up period. As shown in Table 6, the weighted regression analysis of follow-up time on d values achieved a statistically significant result [$Q_R = 9.676$, $p < 0.01$]; this relationship is positive, and is a somewhat unusual finding (Hall, 1995).

In each study design quality was assessed through a scale that took into account such aspects as sample size, random assignment, attrition, utilisation of normalised dependent variables, presence of a control group, equivalence between reported outcome measures in pre-test and post-test, existence of some pre-test measure, and so forth. The relationship between design quality and effect size also obtained an unusual result. In particular, the studies with high quality were associated with higher effect sizes [$Q_R = 24.360$, $p < 0.01$], with a 17% of explained variance. Nevertheless, other meta-analyses have obtained similar findings (Lipsey and Wilson, 1993). Finally, sample size and attrition in the treated group barely reached a significant relationship with effect size, with a very low percent of explained variance.

With respect to extrinsic variables, the date of the study and the publication source were analysed. The date of the study did not hold a significant relationship with effect size (Table 6), although a weak decrease of effect size was observed with time, similar to results obtained in Gottschalk, Davidson II, Gensheimer, and Mayer (1987). In order to examine whether publication bias may threaten the representativeness of our meta-analytic database, the studies were classified as published versus unpublished (Table 5). The comparison did not achieve significant differences between both mean effect sizes [$Q_R = 2.462$, $p < 0.05$]; furthermore, unpublished studies reached a higher average effect size ($r_+ = 0.159$) than published ones ($r_+ = 0.112$). Thus, publication bias can be discarded as a threat to validity in our meta-analysis.

Multivariate Analysis

The variables' inter-relationships can shadow the true associations between effect size and each moderator variable. Specifically, the differences found in types of treatment can show the influence of the studies' methodologies in effectiveness (for example, differences in design type, in the design quality, etc.), as well as the influence of subjects' characteristics (for example, the mean age of the samples). In order to control these inter-correlations we have used techniques of weighted multiple regression analysis. It is possible in this way to answer such critical questions as the following: With these variables controlled, is the type of treatment still a moderator factor in the effectiveness results? What explanatory power do the different moderator variable clusters hold if the remaining clusters are controlled?

Because of the small number of studies meta-analysed ($k=32$), it was necessary to select a small number of moderator variables: specifically, those that are conceptually and statistically more relevant. From the cluster of the treatment variables the most important for our objectives was selected: the type of treatment. For the codification of treatment categories five dichotomous variables were initially defined: Nonbehavioural (0: no; 1: yes), educational programmes (0: no; 1: yes), behavioural and cognitive-behavioural (0: no; 1: yes), therapeutic community (0: no; 1: yes), and diversion (0: no; 1: yes). Due to the similar level of effectiveness found both in behavioural and behavioural-cognitive programmes, they were joined in the same category. From the subjects cluster we selected the justice system, separating juvenile justice (0) and adult justice (1). For that purpose the variable age of the sample was dichotomised, giving a 0 value to 'adolescents' and 'juveniles', and a value of 1 to 'adults' and 'mixed'. Finally, from the methodological cluster we included the quality of design (0: low; 7: high) and the follow-up time in months (this last variable had a missing score; it was decided to use instead the median of follow-up, or six months).

In total, eight moderator variables were included in the regression model. As shown in Table 7, the multiple regression analysis applied on the full model reached high statistical significance with an explained variance percentage of 55.6%. The explanatory power of the model is remarkable, compared with that obtained by Lipsey (1992) (47%).

The regression analyses for each cluster show that the methods cluster explains 21.9% of variance, very similar to Lipsey's of 25% However, when the effects of the treatment and subject clusters were partialled out, the methodological variables explanatory power lowered to 4.2, which was not significant ($p=0.074$). This result indicates that there is a strong association between the method and treatment variables. Another interesting result of this analysis was the reverse relationship between effect size and follow-up time: in the simple regression analysis the relationship was positive; controlling the other variables turned it negative (see Table 8).

Table 7 Multiple regression analyses for each moderator cluster adjusted and unadjusted for the remaining moderators

Cluster	$Q_R(DF)$	R^2	$Q_E(DF)$	$Q_{R_{part}}(DF)$	R^2_{part}
Method	27.176(2)**	0.219	96.764(29)**	5.197(2)	0.042
Subject	8.127(1)**	0.065	115.813(30)**	7.649(1)**	0.062
Treatment	59.747(5)**	0.482	64.193(26)**	41.602(5)**	0.336
Full model	68.968(8)	0.556	54.972(23)**		

Table 8 Partialised unstandardised regression coefficients obtained in the multiple regression analysis for each moderator variable

Moderator variable	B_i
Design quality (0: low; 7: High)	0.062
Follow-up time (in months)	−0.004
Justice system (0: Juvenile; 1: Adult)	0.407
Nonbehavioural programmes (0: No; 1: Yes)	0.434
Educational programmes (0: No; 1: Yes)	0.087
Behavioural and cognitive-behavioural programmes (0: No; 1: Yes)	0.785
Therapeutic community (0: No; 1: Yes)	0.331
Diversion programmes (0; No; 1: Yes)	0.587

The subjects cluster, composed of one dichotomised variable (juvenile versus adult justice system), resulted in 6.5% of explained variance, and maintained 6.2% in this analysis. In spite of this, the direction of the relationship with the effect size also was reversed compared with the univariate regression analysis.

Other variation occurred with respect to the sample age. Table 5 showed better results with young offenders than adult offenders. Now adult criminals obtain better outcomes in the programmes (Table 8). The explanation for this finding lies in the different types of treatment applied: seven of the thirteen juvenile programmes employed the most effective techniques (behavioural and cognitive-behavioural). On the contrary, only two of the adult programme came from this category: seven studies described nonbehavioural techniques and five therapeutic communities. This imbalance explains why the analysis of variance (in Table 5) supported greater effectiveness for juvenile offenders. Now, with the influence of the type of treatment controlled, the better outcomes correspond to adult offenders.

Finally, the greater level of explained variance (48%) corresponded to the treatment cluster, and decreased only slightly in the multivariate regression analysis (to 33.6%), which indicates the robustness of the differential effectiveness of the employed treatments, irrespective of method or subject variables. In Table 8 one can see that the differential effectiveness of the treatment models is preserved: first are the behavioural and cognitive-behavioural programmes, and then the diversion and non-behavioural programmes.

It must be said that the contrasts made in order to specify the multiple regression analysis (with the Q_E test) were positive, which tell us that some important variables (such as programme characteristics or context of implementation) were lacking in this analysis. The reason for this deficiency is the small number of studies meta-analysed. So, we have to take this model as a exploratory one.

4. CONCLUSIONS

The present research has analysed the contingent effectiveness of correctional programmes applied in European countries over a decade. In terms of correlation coefficient, the average effect size obtained from the 32 studies was $r_+ = 0.120$, that is, treated subjects showed a lower recidivism rate than controls. This result had a moderate size that, in average, was of 12% ($r_+ = 0.120$). The practical importance of this result can be interpreted composing the Binomial Effect Size Display,[4] *BESD*, as shown in Table 9 (Rosenthal, 1991). Thus, assuming 100 subjects in a treatment group and another 100 subjects in a control group, the recidivism rate in the treated group would be 44%, whereas in the control group, the recidivism rate would be 56%. In other words, our results showed a differential recidivism rate of 12% in favour of the treatment.

Considering the heterogeneity of effectiveness observed in the set of analysed programmes, the influence of different factors on effectiveness has been analysed. In this way, recidivism reductions appear related to the following factors:

1. The most effective programmes are based on behavioural and cognitive-behavioural theoretical models. Similar conclusions have been obtained by Gendreau and Ross (1979), Ross and Fabiano 1985), Ross *et al.* (1990), Andrews *et al.* (1990), Palmer (1992), McGuire (1992), Lösel (1995a,b, 1996) and Redondo *et al.* (1997).
2. Juvenile offenders are more successfully rehabilitated than adult offenders, because they are treated with the most successful techniques (behavioural and cognitive-behavioural). Nevertheless, when the influence of the type of treatment is controlled, adult offenders show lower level of recidivism. This clearly supports the importance of also using the most effective techniques with adult offenders. The idiosyncrasy of the juvenile and adult justice systems can explain why practitioners are more likely to use more defined and structured techniques with juveniles, and using more "open and personal decision" type of treatment with adult offenders. But as the outcomes show, adult offenders need structured programmes too. That juvenile offenders are more hard to deal with is not a surprise: usually they are at the height of their criminal careers, and the important point is that at least we have well prepared programmes for them; otherwise, many years of antisocial reinforcement would not be easily changed.
3. Regarding crime typology, programmes applied to violent offenders (not sex offenders) show the highest effectiveness. At this point, despite the fact that the reduced number of studies by category hinders reaching general conclusions, the obtained results endorse one of the quality criterion of intervention effectiveness proposed by Andrews *et al.* (1990): the risk principle. The risk principle suggests that treatment will be more effective for higher risk cases than for those whose risk of reoffending is judged to be low, and that more intensive services should therefore be reserved for high-risk offenders. According to this, and considering that violent offenders have more criminological needs and are "high risk" compared to property offenders, in general one could expect greater implicit quality and effectiveness of programmes directed to high risk criminals, that receive the best correctional services, compared to those applied to low risk offenders, that receive only ordinary services.
4. Finally, with respect to the setting of interventions the greatest effectiveness is achieved in correctional institutions for youths, either centres for delinquents or juvenile prisons.

Table 9 Binomial effect size display for $r_+ = 0.120$

Group	*Recidivists*	*Nonrecidivists*	*Total*
Treatment	44	56	100
Control	56	44	100
Total	100	100	200

The results of this European meta-analysis are similar to the main results of previous reviews on correctional treatment. Most of the reviewers (Andrews *et al.*, 1990; Garrett, 1985; Gensheimer *et al.*, 1986; Gottschalk, Davidson II, Gensheimer, and Mayer, 1987; Gottschalk, Davidson II, Mayer, and Gensheimer, 1987; Lipsey, 1992; Lösel, 1996; Lösel and Köferl, 1989; Palmer, 1990; Redondo, 1994; Redondo *et al.*, 1997; Taylor, 1992) reach a general conclusion: Recent correctional meta-analyses contradict the "nothing works" affirmation of initial researchers; many treatment programmes of delinquency and crime are effective with broad groups of offenders.

Our research also coincides with previous reviews in terms of the greater effectiveness of cognitive and behavioural methods. For example, Palmer (1995) examined 23 literature reviews and 9 meta-analysis conducted up to 1989, and concluded that "the following five were the most successful or promising: behavioural, cognitive-behavioural or cognitive, life-skills or skill orientated, multimodal and family intervention" (p. 101).

However, "it is important to recognise that the meta-analytic process necessarily decontextualises in all but the broadest terms the

situation in which treatment occurs or the way decisions about treatment are made (...) In the process of working back from broad meta-analytic trends to the circumstances of individual settings, it is therefore necessary for practitioners to consider what aspects of the local environment might influence the application of these principles – in other words, what are the contingent effects of unmeasured contextual factors. What this means in practical terms is that the application of generally derived knowledge is far from straightforward, requiring a careful consideration of local environmental contingencies" (Brown, 1996).

From our point of view, in upcoming years there has to be a focus upon researching sets of factors that are effective in well-specified settings (see Palmer, 1995). That "local environmental contingencies" are very difficult – if not impossible – to measure in a meta-analytic procedure is true, but individual empirical studies must help in this task. With more focused studies, perhaps we will be able to know more about the future of the rehabilitation of offenders in Europe.

Notes

1. We acknowledge the participation of Dr. Fulgencio Marín-Martínez in the reliability study of effect size calculations and the assistance of Dr. Rosemary Barberet and Antonio Marchal in the review of the English version of this article.
2. The design quality of the studies was assessed by means of a scale with seven items related to methodological issues such as design type, attrition, sample size, assignment of subjects to the groups, and type of outcome measures. The questionnaire is available from the second author.
3. A table with all moderator variables for each study included in the meta-analysis is available from the first author.
4. The *BESD* is obtained by computing the treatment nonrecidivism rate as $0.5+r/2$ and the control nonrecidivism rate as $0.5-r/2$, with $r=0.12$ (cf. Rosenthal, 1991).

References

(References preceded by an asterisk (*) were included in the meta-analysis.)

Andrews, D., Zinger, I., Hoge, R., Bonta, J., Gendreau, P., and Cullen, F. (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology*, **28**, 369–404.

Antonowicz, D. H., and Ross, R. R. (1994). Essential components of successful rehabilitation programs of offenders. *International Journal of Offender Therapy and Comparative Criminology*, **38**, 97–104.

*Belfrage, H. (1991). The crime preventive effect of psychiatric treatment on mentally disordered offenders in Sweden. *International Journal of Law and Psychiatry*, **14**, 237–243.

*Berggren, O., and Svärd, H. (1990). *The Österaker Project: A further follow-up of the drug misuser treatment programme at Österaker prison*. Kriminalvarden. Forsknings-gruppen, I, Swedish Prison and Probation Administration (19 pp.).

*Bishop, N., Sundin-Osborne, A., and Pettersson, T. N. (1987). *The drug free programme at the Hinseberg prison for women*. National Prison and Probation Administration (Report 1987: 4) (27 pp.).

*Bovens, R. (1987). The alcohol program: An educational program for drunken drivers in prison. In M. Brand-Koolen (Ed.), *Juvenile Delinquency in the Netherlands* (pp. 151–157). Kugler Publications bv.

*Brown, B. (1985). An application of social learning methods in a residential programme for young offenders. *Journal of Adolescence*, **8**, 321–331.

Brown, M. (1996). Refining the risk concept: Decision context as a factor mediating the relation between risk and program effectiveness. *Crime and Delinquency*, **41**(2), 435–455.

Cleland, C. M., Pearson, F. S., and Lipton, D. S. (November 1996). *A meta-analytic approach to the link between needs-targeted treatment and reductions in criminal offending*. Paper presented at the Annual Meeting of the American Society of Criminology, Chicago, Illinois.

*Collins, S. A., and Tate, D. H. (1988). Alcohol related offenders and a voluntary organisation in a Scottish community. *The Howard Journal*, **27**, 44–57.

*Cook, D. A., Fox, C. A., Weaver, C. M., and Rooth, F. G. (1991). The Berkeley group: Ten years' experience of a group for non-violent sex offenders. *British Journal of Psychiatry*, **158**, 238–243.

*Cooke, D. J. (1989). Containing violent prisoners: An analysis of the Barlinnie Special unit. *British Journal of Criminology*, **29**, 129–143.

*Cooke, D. J. (1991). Psychological treatment as an alternative to prosecution: a form of primary diversion. *The Howard Journal*, **30**, 53–65.

Cooper, H. M. (1989). *Integrating Research: A Guide for Literature Reviews* (2nd ed.). Newbury Park, CA: Sage.

Cooper, H., and Hedges, L. V. (Eds.) (1994). *The Handbook of Research Synthesis*. New York: Sage.

Cox, S. M., Davidson, W. S., and Bynum, T. S. (1995). A meta-analytic assessment of delinquency related outcomes of alternative education programs. *Crime and Delinquency*, **41**(2), 219–234.

*Day, K. (1988). A hospital-based treatment programme for male mentally handicapped offenders. *British Journal of Psychiatry*, **153**, 635–644.

*Demonstration Unit 1981–84, Members of the (1986). Increasing the use of probation. *Probation Journal, March*, 87–90.

*Dünkel, F. (1982). Selection and recidivism after different modes of imprisonment in West-Berlin. In: *Research in Criminal Justice. Stock-taking of Criminological Research at the Max-Planck-Institute for Foreign and International Penal Law after a Decade* (pp. 452–471). Freiburg: Max-Planck-Institute.

Eysenck, H. J. (1995). Meta-Analysis squared: Does it make sense? *American Psychologist*, **50**, 110–111.

Fleiss, J. L. (1994). Measures of effect size for categorical data. In H. Cooper, and L. V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 245–260). New York: Sage.

Garrett, C. J. (1985). Effects of residential treatment on adjudicated delinquents: A meta-analysis. *Journal of Research in Crime and Delinquency*, **22**, 287–308.

Gendreau, P., Little, T., and Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology*, **34**(4), 575–607.

Gendreau, P., and Ross, R. R. (1979). Effective correctional treatment: Bibliotherapy for cynics. *Crime and Delinquency*, **25**, 463–489.

Gensheimer, L. K., Mayer, J. P., Gottschalk, R., and Davidson II, W. S. (1986). Diverting youth from the juvenile justice system: A meta-analysis of intervention efficacy. In S. Apter and A. Goldstein (Eds.), *Youth Violence: Program and Prospects*. New York: Pergamon Press.

Glass, G. V., McGaw, B., and Smith, M. L. (1981). *Meta-analysis in Social Research*. Beverly Hills, CA: Sage.

Gómez, J. (1987). *Meta-análisis* [Meta-analysis]. Barcelona: PPU.

Gottschalk, R., Davidson II, W. S., Gensheimer, L. K., and Mayer, J. P. (1987). Community-based interventions. In H. C. Quay (Ed.), *Handbook of Juvenile Delinquency* (pp. 266–289). New York: Wiley.

Gottschalk, R., Davidson II, W. S., Mayer, J. P., and Gensheimer, L. K. (1987). Behavioral approaches with juvenile offenders: A meta-analysis of long-term treatment efficacy. In E. K. Morris, and C. K. Braukmann (Eds.), *Behavioral Approaches to Crime and Delinquency* (pp. 399–422). New York: Plenum Press.

Hall, G. C. N. (1995). Sexual offender recidivism revisited: A meta-analysis of recent treatment studies. *Journal of Consulting and Clinical Psychology*, **63**, 802–809.

Hedges, L. V. (1994). Fixed effects models. In H. Cooper, and L. V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 285–299). New York: Sage.

Hedges, L. V., and Olkin, I. (1985). *Statistical Methods for Meta-analysis*. Orlando, FL: Academic Press.

Izzo, R. L., and Ross, R. R. (1990). Meta-analysis of rehabilitation programs for juvenile delinquents: A brief report. *Criminal Justice and Behavior*, **17**, 134–142.

Johnson, B. T. (1991). *Software for the Meta-analytic review of Research Literatures* (2nd ed.). Hillsdale, NJ: Erlbaum.

*Kruissink, M. (1990). The Halt program: Diversion of juvenile vandals. Dutch Penal Law and policy. *Notes on Criminological Research from the Research and Documentation Centre*, **1**, 1–8.

*Kury, H. (1989). Treatment of young remand prisoners: Problems and results of a research project. In H. Wegener, F. Lösel, and J. Haisch (Eds.), *Criminal Behavior and the Justice system: Psychological Perspectives* (pp. 356–381). New York: Springer-Verlag.

Lab, S. P., and Whitehead, J. T. (1988). An analysis of juvenile correctional treatment. *Crime and Delinquency*, **34**, 60–83.

*Legaz, F., López-Pina, J. A., Sánchez-Meca, J., and Velandrino, A. P. (1990). *Evaluación del Programa Experimental de Intervención en Medio Abierto*. Unpublished technical report, Instituto de Servicios Sociales, Murcia (Spain).

Lipsey, M. W. (1992). Juvenile delinquency treatment: A meta-analytic inquiry into the variability of effects. In T. D. Cook, H. Cooper, D. S. Cordray, H. Hartman, L. V. Hedges, R. J. Light, T. A. Louis, and F. Mosteller (Eds.), *Meta-analysis for Explanation: A casebook* (pp. 83–127). New York: Sage.

Lipsey, M. W. (1994). Identifying potentially interesting variables and analysis opportunities. In H. Cooper, and L. V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 111–123). New York: Sage.

Lipsey, M. W., and Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, **48**, 1181–1209.

Lipsey, M. W., and Wilson, D. R. (1995). Reply to comments on Lipsey and Wilson (1993). *American Psychologist*, **50**, 113–115.

Lösel, F. (1995a). Increasing consensus in the evaluation of offender rehabilitation? Lessons from recent research syntheses. *Psychology, Crime, and law*, **2**, 19–14.

Lösel, F. (1995b). The efficacy of correctional treatment: A review and synthesis of meta-evaluations. In J. McGuire (Ed.), *What Works: Reducing Reoffending* (pp. 79–114). Chichester: Wiley.

Lösel, F. (1996). What recent meta-evaluations tell us about the effectiveness of correctional treatment. In G. Davies, S. Lloyd-Bostock, M. McMurran, and C. Wilson (Eds.), *Psychology, Law, and Criminal Justice* (pp. 537–554). Berlin: De Gruyter.

Lösel, F., and Köferl, P. (1989). Evaluation research on correctional treatment in West Germany: A meta-analysis. In H. Wegener, F. Lösel, and J. Haisch (Ed.), *Criminal Behavior and the Justice System: Psychological Perspectives* (pp. 334–355). New York: Springer-Verlag.

McGuire, J. (1992). Enfocaments psicològics per a la reducció de la conducta delictuosa: Investigació recent i implicacions pràctiques. [Psychological strategies for reduction of offending behaviour: Recent research and practical implications] *Papers d'Estudis i Formació*, **10**, 67–77.

*McMurran, M., and Boyle, M. (1990). Evaluation of a self-help manual for young offenders who drink: A pilot study. *British Journal of Clinical Psychology*, **29**, 117–119.

Orwin, R. G. (1994). Systematic coding for research synthesis. In H. Cooper, and L. V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 125–138). New York: Sage.

Palmer, T. (1992). *The Re-emergence of Correctional Intervention*. Newbury Park, CA: Sage.

Palmer, T. (1995). Programmatic and nonprogrammatic aspects of successful intervention: New directions for research. *Crime and Delinquency*, **41**(1), 100–131.

Pearson, F. S., Lipton, D. S., Cleland, C. M., and O'Kane, J. B. (November, 1995). *Meta-analysis on the effectiveness ofcorrectional treatment: Another approach and extension of the time frame to 1994: A progress report*. Paper presented at the Annual Meeting of the American Society of Criminology, Boston, Massachusetts.

*Petterson, T., Sundin-Osborne, A., and Bishop, N. (1986). *Results of the drug misuser treatment programme at the Österaker prison*. In: National Prison and Probation Administration (Report 1986: 2) (13 pp.).

Quinsey, V. L., Harris, G. T., Rice, M. E., and Lalumière, M. L. (1993). Assessing treatment efficacy in outcome studies of sex offenders. *Journal of Interpersonal Violence*, **8**, 512–523.

Redondo, S. (1994). *El tratamiento de la delincuencia en Europa: Un estudio meta-analítico*. [The treatment of offenders in Europe: A meta-analysis] Unpublished Doctoral Thesis, University of Barcelona.

Redondo, S. Funes, J., and Luque, E. (1994). *Justicia Penal y Reincidencia*. [Criminal Justice and Recidivism] Barcelona: Jaume Callis.

Redondo, S., Garrido, V., and Sánchez-Meca, J. (1997). What works in correctional rehabilitation in Europe: a meta-analytic review. In S. Redondo, V. Garrido, J. Pérez, and R. Barberet (Eds.), *Advances in Psychology and Law: International Contributions* (pp. 499–523). Berlin: De Gruyter.

*Redondo, S., Roca, M., Pérez, E., Sánchez, A., and Deumal, E. (1990). Diseño ambiental de una prisión de jóvenes: Cinco años de evaluación. [Environmental design of a youth prison in Catalonia: Five years of evaluation] *Delincuencial Delinquency*, **2**, 331–357.

*Robertson, G., and Gunn, J. (1987). A one-year follow-up of men discharged from Grendon prison. *British Journal of Psychiatry*, **151**, 674–678.

Rosenthal, R. (1991). *Meta-analytic Procedures for Social Research* (2nd ed.). Newbury Park, CA: Sage.

*Rosner, A. (1988). Evaluation of a drinking-driver rehabilitation program for first offenders. In G. Kaiser, and I. Geissler (Eds.), *Crime and Criminal Justice: Criminological Research in the 2nd Decade at the Max Planck Institute in Freiburg* (pp. 319–336). Freiburg: Max-Planck-Institute.

Ross, R. R., and Fabiano, E. A. (1985). *Time to Think: A Cognitive Model of Delinquency Prevention and Offender Rehabilitation*. Johnson City: Institute of Social Sciences and Arts.

Ross, R. R., Fabiano, E., and Garrido, V. (1990). El pensamiento prosocial. El modelo cognitivo para la prevención y tratamiento de la delincuencia. [Pro-social thinking. The cognitive model for prevention and treatment of delinquency and crime] *DelincuencialDelinquency*, Monographic n. 1, 116 pp.

Sánchez-Meca, J. (1997). Methodological issues in the meta-evaluation of correctional treatment. In S. Redondo, V. Garrido, J. Pérez, and R. Barberet (Eds.), *Advances in Psychology and Law: International Contributions* (pp. 486–498). Berlin: De Gruyter.

Sánchez-Meca, J., and Ato, M. (1989). Meta-análisis: Una alternativa metodológica a las revisiones tradicionales de la investigación. [Meta-analysis: A methodological alternative to traditional reviews of research] In J. Arnau and H. Carpintero (Eds.), *Tratado de Psicología General: Historia, Teoría y Método* (pp. 617–669). Madrid: Alhambra.

Sánchez-Meca, J., Marín, F., and Redondo, S. (1996). Reincidencia: Evaluación internacional. [Recidivism: An international assessment]. In J. Funes, E. Luque, A. Ruiz, and J. Sánchez-Meca (Eds.), *Reincidència: En la justicia de menorsIAvaluació*

internacional. [Recidivism: At the juvenile justice/International assessment] (pp. 93–200). Barcelona: Generalitat de Catalunya.

*Scholte, E. M., and Smit, M. (1987). Early social assistance for juveniles at risk. *International Journal of Offender Therapy and Comparative Criminology*, 209–218.

*Singer, L. R. (1991). A non-punitive paradigm of probation practice: Some sobering thoughts. *British Journal of Social Work*, **21**, 611–621.

*Solt, N. W. (1983). The implementation and evaluation of a residential social skills training programme for youth in trouble. In W. Everaerd, C. B. Hindley, A. Bot, and J. J. van der Werf (Eds.), *Development in Adolescence: Psychological, Social, and Biological Aspects* (pp. 176–205). Amsterdam: Marinus Nijhoff Pub.

*Slot, N. W. (1984). The teaching family model in the Netherlands: First results of a community based program for problem youths. Paper presented at the 14th Conference of the European Association for Behaviour Therapy, Brussels (Belgium).

*Slot, N. W., and Bartels, A. A. J. (1983). Outpatient social skills training for youth in trouble: Theoretical background, practice and outcome. In W. Everaerd, C. B. Hindley, A. Bot, and J. J. van der Werf (Eds.), *Development in Adolescence: Psychological, Social, and Biological Aspects* (pp. 176–191). Amsterdam: Marinus Nijhoff Pub.

*Slot, N. W., and Heiner, J. (1986). *Development of community-based treatment programs for troubled youth in the Netherlands*. Paper presented at the ABA-Conference, Milwaukee (USA).

Sohn, D. (1995). Meta-analysis as a means of discovery. *American Psychologist*, **50**, 108–110.

Taylor, J. M. (1992). Post-Secondary Correctional Education: An Evaluation of Effectiveness and Efficiency. *Journal of Correctional Education*, **43**(3), 132–141.

*Thornton, D. M. (1987). Correctional evaluation of custodial regimes. In B. J. McGurk, D. M. Thornton, and M. Williams (Eds.), *Applying Psychology to Imprisonment* (pp. 467–481). London: Her Majesty's Stationery Office.

Tournier, P., and Barre, M. D. (1990). Enquête sur les systèmes pénitentiaires dans les Etats membres du Conseil de l'Europe: Demographie carcérale comparée. *Bulletin d'Information Pénitentiaire*, **15**, 4–44.

*Van Dalen, W. E. (1989). Education: A successful instrument for reducing drunken driving in the Netherlands. *Prevention and Control/Realities and Aspirations*, **IV**, 717–722.

*Weaver, C., and Fox, C. (1984). The Berkeley sex offenders group: A seven year evaluation. *Probation Journal*, **31**, 143–146.

Wells-Parker, E., Bangert-Drowns, R., McMillen, R., and Williams, M. (1995). Final results from a meta-analysis of remedial interventions with drink/drive offenders. *Addiction*, **90**, 907–926.

Whitehead, J. T., and Lab, S. P. (1989). A meta-analysis of juvenile correctional treatment. *Journal of Research in Crime and Delinquency*, **26**, 276–295.

Wolf, F. M. (1986). *Meta-Analysis: Qualitative methods for research synthesis*. New York: Sage.