Psychology in Spain, 2001, Vol. 5. No 1, 47-62

TREATMENT OF OFFENDERS AND RECIDIVISM: ASSESSMENT OF THE EFFECTIVENESS OF PROGRAMMES APPLIED IN EUROPE

Santiago Redondo Illescas*, Julio Sánchez-Meca** and Vicente Garrido Genovés***
**Universidad de Barcelona. **Universidad de Murcia. ***Universidad de Valencia*

Meta-analyses have examined the effectiveness of different treatment techniques on delinquent and criminal behaviour in America and Europe. In one of the most recent meta-analyses, which integrated the results of programmes applied in European countries (Redondo, Garrido, & Sánchez-Meca, 1997), the global effect size obtained for all treatments and effectiveness measures (psychological factors, education, etc.) was d = 0.3039 (r = .15). In general, it could be concluded that the improvement in treatment groups was 15% greater than that of controls. This work presents the results of a meta-analysis to determine which treatments are most effective in reducing recidivism rates. 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 = .12), equivalent to a 12% reduction in recidivism. With regard to the differential influence of treatment models, behavioural and cognitive-behavioural techniques were found to be most beneficial in reducing recidivism.

Diversos meta-análisis han analizado en América y Europa la efectividad que tienen diferentes técnicas de tratamiento sobre la conducta delictiva. En uno de los últimos meta-análisis, que integró los resultados de programas aplicados en países europeos (Redondo, Garrido y Sánchez-Meca, 1997), el tamaño del efecto global obtenido para el conjunto de los tratamientos y de las medidas de efectividad (factores psicológicos, educación, etc.) fue de d = 0.3039 (r = 0.15). Este resultado permite concluir que los grupos de tratamiento obtuvieron una mejora del 15% sobre los resultados obtenidos por los grupos de control. En el presente trabajo, se presentan los resultados de un nuevo meta-análisis que hemos realizado para determinar qué tratamientos, de los aplicados en Europa, resultan más efectivos para reducir las tasas de reincidencia de los delincuentes. Para ello hemos analizado un total de 32 programas Europeos que evaluaban la reincidencia durante un período de seguimiento medio de dos años. Estos programas, globalmente considerados, obtuvieron un tamaño del efecto de d = 0.243 (r = 0.12), lo que implica una reducción del 12% de la reincidencia. En lo que respecta al efecto diferencial logrado por distintos modelos de tratamiento, se observó que las terapias conductuales y cognitivo-conductuales produjeron las tasas más bajas de reincidencia.

In the history of criminal justice, rehabilitation philosophy has fluctuated in pendular form between periods in which rehabilitation of offenders was considered feasible and other periods in which it was not. Up to the 1970s, researchers and those working in the field of justice placed a great deal of confidence in the so-called "rehabilitation ideal" which held that offender recidivism could be reduced by means of treatment. However, the 1980s marked the beginning of a more pessimistic period in which the predominant belief was that "nothing worked" as far as offender rehabilitation was concerned. Nonetheless, a certain optimism re-emerged in 1988/89 with respect to the treatment of offenders. The evaluative research published in the 1980s concluded that some programmes had been moderately effective in terms of reducing criminal recidivism (Palmer, 1995).

This line of investigation has continued to develop in the 1990s. Throughout the decade, a number of meta-analytic studies both in North America and, later, in Europe, were carried out in order to assess the degree of effectiveness of treatment techniques applied to 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;

The original Spanish version of this paper has been previously published in *Anuario de Psicología Jurídica*, 1999, 11-37

Correspondence concerning this article should be addressed to Santiago Redondo Illescas. Dpto, Personalidad, Evaluación y Tratamientos Psicológicos. Facultad de Psicología. Universidad de Bercelona. Paseo del Valle de Hebrón, 171, Edificio Ponent. 08035 Barcelona. E-mail: sredondo@psi.ub.es

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). Taken as a whole, the average effectiveness of programmes applied to offenders ranged from 5% to 18%. In the European context the meta-analysis by Redondo *et al.* (1997) integrated the results of 57 programmes carried out in different countries with young and adult offenders and obtained an overall effect size of r=0.150 for the set of treatments applied and results evaluated (which included different variables, such as the improvement of the institutional environment, school participation, the reduction of recidivism, etc.). This result demonstrates that treatment groups did 15% better than control groups. That is, in terms of the recidivism variable, there was a decrease of 15% with respect to what might have been expected had the subjects not received treatment.

Meta-analysis is a statistical technique (of second-order or secondary analysis) that allows the results of various primary studies (i.e., empirical studies) to be summarised. In other words the meta-analyst does not work directly with samples of individuals (offenders who have undergone treatment in the case of our study). Instead, the sample to be evaluated consists of studies carried out previously. In our area of interest, the use of meta-analysis was not limited to the study of the effectiveness of treating offenders, but also covered other areas of criminological interest such as the analysis of risk or damage offences committed by drivers under the effects of alcohol (Wells-Parker, Bargert-Drowns, McMillen and Williams, 1995), predictors of recidivism among adult offenders (Gendreau, Little and Goggin, 1996), or alternative education programmes for the prevention of educational maladjustment and juvenile delinquency (Cox, Davidson and Bynum, 1995). In all of these cases the objective of the meta-analysis was to reach general conclusions in a specific research field on the basis of previous studies carried out in the same field.

The meta-analysis procedure is quite different from that of traditional research review processes, which consist in the qualitative analysis of a specific scientific field. Researchers, on studying the scientific literature of interest, identify those conclusions they consider, according to their personal criteria, to be most relevant. However, these qualitative or narrative reviews tend to present the following problems: (a) a considerable bias in the selection of studies to be analysed, (b) a subjective examination and interpretation of the results of each study when drawing general conclusions about the field under analysis, and (c), this methodology does not permit the valid exploration of the most relevant moderator and explanatory variables that account for the results obtained in these studies (Wolf, 1986). It is also true that some authors have recently criticised the meta-analytic technique, casting doubt on its capacity to improve our knowledge above and beyond the direct results of the primary studies themselves (Eysenck, 1995; Sohn, 1995). Nevertheless, Lipsey and Wilson (1995, p. 114) have pointed out that meta-analytic methodology does not aim to replace the primary studies that form the basis of scientific discoveries. The aim of meta-analysis is the consolidation of this isolated knowledge "within a wider analysis examining the extent to which such knowledge can be generalised across diverse contexts, subjects, methodologies, etc." (see also Cooper, 1989; Glass, McGaw, and Smith, 1981; Gómez, 1987; Rosenthal, 1991; Sánchez-Meca and Ato, 1989). In order to achieve this objective, meta-analysis converts the primary study results into effect size indices such as correlation coefficient, typified mean difference or the *phi* coefficient.

The effect size represents the magnitude of the results. According to the BESD (Rosenthal's binomial effect size display), the effect size reflects the percentage of improvement of the treatment group with respect to the control group. Applied to the context of programme assessment, meta-analytic techniques permit three fundamental objectives to be fulfilled: (a) to obtain an overall index for programme effectiveness; (b) to determine whether the results of different programmes are homogenous with respect to this overall index; and (c) if they are not homogenous, to find those programme characteristics that might serve to explain this variability in results (i.e., the factors that appear to account for the effects observed).

In the evaluation of programmes with offenders, different measures have been used as effectiveness criteria. In the above-mentioned meta-analyses, the most common measures were those related to criminal behaviour and recidivism, psychological adjustment of the subjects, participation and performance in academic activities, improvements at work, interpersonal adjustment, institutional adjustment and subject's involvement in the treatment programme. The meta-analyses using the highest number of result indices were those of Garrett (1985), Gottschalk, Davidson II, Gensheimer and Mayer (1987), Lipsey (1992) and Redondo (1994); see also Redondo *et al.*, 1997).

In order to assess the effectiveness of treatment programmes, it is often necessary to use a variety of measures over and above the evaluation of recidivism. Recidivism into crime is a behavioural and juridical product linked to such factors as academic failure, unemployment, certain psychological disorders, previous criminal behaviour, consumption of certain drugs, and so on. Consequently, most researchers have identified diverse variables related to these problems as programme effectiveness criteria. Nonetheless, all meta-analyses of treatment programmes have used some form of recidivism criterion. Subsequent to the application of a treatment programme for offenders it is obviously necessary to assess, by one means or another, whether the subjects treated relapse into crime or not. Recidivism is, of necessity, the ultimate criterion of the efficiency of criminal justice. Society, public opinion and the authorities expect treatment programmes applied to offenders to reduce the frequency and gravity of their future criminal behaviour. Otherwise it would be impossible to conclude that the programmes were useful.

In accordance with the arguments outlined above, the main objective of this study was to evaluate the capacity of European treatment programmes to reduce recidivism among offenders. A second objective was to decide whether some treatment models were more effective than others. Finally, we analysed the relationship between the effectiveness of a programme and the characteristics of the subjects treated, the context in which the programmes were applied and the assessment methodology applied in the primary studies that served as the basis for our research.

METHOD

Our search for treatment programmes

We used a variety of information channels in order to gain access to studies of European treatment programmes. First of all, the computerised data bases *Criminal Justice Periodical Index, Pascal* and *PsycLIT* were consulted. To this end the key words: *delinquen*, offender, inmate, probation, treatment, rehabilitation, intervention, parole,* and *therap** were used. Secondly, a manual search of 55 specialised (mainly European) journals was conducted. We also wrote letters requesting studies on treatment programmes involving offenders. These letters were sent to a total of 118 specialised researchers and 82 European institutions working in this field. In addition, all the bibliographical references contained in the previously-selected studies were examined by means of an *ancestry approach (cursna)*, with a view to finding other relevant studies. The literature search covered the period from 1980 to 1991, and included both published and unpublished material.

In order to be included in the meta-analyses, each study had to meet the following requirements: (a) it must have been applied to subjects under the control of the criminal justice system (e.g., young or adult offenders), (b) it must have used some treatment strategy for a certain period of time, (c) it must have employed a methodological design allowing comparison between control and treatment groups or pre-test/post-test measures, and finally (d) it must have included some recidivism measure. A total of 29 studies referring to 32 programmes were selected, all of these having met the requirements outlined above. The total sample of subjects involved in these programmes was 5,715 (see Table 4).

Coding procedure

With a view to explaining the heterogeneity of the treatment programme results, the potential moderator variables were defined and codified for subsequent analysis. Variables were classified into the following treatment clusters: subject, context, method and extrinsic (Lipsey, 1994; Sánchez-Meca, 1997). The *treatment variables* cluster included programme characteristics such as the underlying theoretical model, (non-behavioural therapy, educational/informative, behavioural therapy, cognitive behavioural therapy, classical penal theory, therapeutic community and diversion), duration of the programme (in months), intensity of the programme (in hours per week per subject) and size of the programme (in total hours per subject). The *subject variables* cluster included age of the subjects treated (adolescents, youths, mixed age and adults), average age of the sample (in years), most common type of crime in the sample (property-related, crimes against persons, sexual crimes, drug trafficking, alcohol-related crimes and mixed offences) and gender of the sample (defined by the percentage of males). The *context variables* cluster was composed of the place in which the programme was applied (juvenile reform centre, prison for young offenders, adult prison, the community and others), security conditions of the centre (closed, semi-open, open and others), and country. The following characteristics were codified in the *methodological variables* cluster: design type (between-group versus within-group designs), assignment of subjects to groups (random versus non-random assignment), sample size, offender selection criteria (all subjects from one institution, according to sentence length, belonging to specific typologies, violent

subjects and subjects with specific needs), attrition within the treatment group, design quality[i] (scoring from 0, low quality, to 7, high quality) and follow-up period (in months). Finally, the *extrinsic variables* cluster included the year in which the study was carried out and its publication source (published versus unpublished).

All codification of the variables was carried out independently by the first and third authors of this study and a satisfactory level of intercoder reliability (Orwin, 1994), an average of 86.2%, was achieved. Discrepancies between coders were resolved by consensus.

Statistical procedures

Typified mean difference, *d*, was chosen as a measure of effect size, using the correction for small samples (Hedges and Olkin, 1985, p. 81). In the between-group designs, the *d* statistic was defined as the difference between the means of the experimental and control groups for the measures made during the follow-up divided by the inter group standard deviation. In the within-group designs the common measure of the results was the difference between the post-treatment and pre-treatment means divided by the within-group standard deviation. When the measures of results were proportions (e.g., reci-

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)	Holland	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 y Tate (1988)	Gr. Britain	DE	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	DE	—	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	DE	7.7	12	1	10	6	0.281	0.14
McMurran and Boyle (1990),study 1	Gr. Britain	ED	0.25	18	15	13	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)	Holland	DE	—	14	6	71	71	0.217	0.10
Singer (1991)	Gr. Britain	ED	2	20	12	152	—[f]	-0.031	-0.01
Slot (1983)	Holland	BT	—	17	6	9	17	0.419	0.21
Slot (1984)	Holland	BT	12.5	16	6	6	—[f]	0.447	0.22
Slot and Bartels (1983)	Holland	CBT	—	17	7	29	29	1.237	0.53
Slot and Heiner (1986)	Holland	BT	—	16	8	22	—[f]	0.547	0.27
Thornton (1987)	Gr. Britain	DE	—	31	12	1,000	—[f]	-0.012	-0.01
Van Dalen (1989)	Holland	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): individual or group psychodynamic therapy. ED = Education (school, educational materials). BT = Behavioural therapy (operant conditioning): token economy, environmental contingency design. CBT= Cognitive-behavioural therapy (social skills training, psychosocial competence programmes). DI = Dissuasion/Discipline (classic penal theory, dissuasion, 'shock prison', increase in institutional control). TC = Therapeutic Community (relationships between inmates/personal relationships similar to those of patients/nurses, relaxation of institutional control). DE = Diversion.

[b] Duration of the programme in months.

[c] Average sample age in years.

[d] Duration of follow-up period in months.

[e] Control group obtained from external source.

[f] Within-group design.

[g] n^C is unknown; we assume $n^C = n^E$.

[h] Identical control group to that of Study 1.

divism rates), an statistic equivalent to *d* proposed in the DSTAT programme (Johnson, 1991) was applied. Positive *d* values indicated the amount of the improvement in the treated group with respect to the control group or the pre-test evaluation. In order to facilitate an appropriate interpretation of results, each *d* statistic was also converted into Pearson's correlation coefficient *r*, using: $r = d/(d^2 + 4)^{0.5}$ (Rosenthal, 1991). It is quite common for studies on treatment programmes to include diverse recidivism measures (e.g., new arrests, new convictions, return to prison, etc.). In these cases, an overall effect size was calculated by taking an average of all the recidivism measures in each study, with a view to providing a single measure for each of them, thus safeguarding the assumption of independence of the effect sizes.

The most common recidivism measure classifies subjects into recidivists and non-recidivists. Where a particular study used a dichotomous recidivism measure and included two groups (treated group and control group), the difference between the proportions of recidivism, d_p, in the treated and control groups (p_E and p_C, respectively), i.e., $d_p = p_C - p_E$ (Fleiss, 1994) was also defined as an effect size.

All effect size calculations were carried out by the second author of this study and an independent researcher, and satisfactory intercoder reliability ($r = 0.826$, $p = 0.000$) was obtained. Discrepancies in the calculations were resolved by means of a combined review of the study until consensus was reached. Table 1 shows a list of sample sizes, effect sizes and some of the moderator characteristics of each study[ii].

As we have already mentioned, the objectives of this meta-analysis were to evaluate overall effectiveness of treatment of offenders with respect to subsequent recidivism, determining whether the programme results were homogenous in their results, or identifying which moderator variables which may serve to explain any variability found in the effects[iii].

RESULTS AND DISCUSSION

Characteristics of the studies

Tables 2 and 3 show the descriptive characteristics of the 32 studies included in the meta-analysis. A great hetero-

Table 2
Descriptive characteristics of categorical variables

Cluster characteristics	*Frequency*	*%*
Treatment cluster		
Theoretical model (*k* = 32):		
Non-behavioural	8	25
Educational	5	15.6
Behavioural	6	18.8
Cognitive-behavioural	3	9.4
Dissuasion theory	1	3.1
Therapeutic community	5	15.6
Diversion	4	12.5
Subjects cluster		
Sample age (*k* = 32):		
Adolescents (<16)	7	21.9
Youths (16-21)	6	18.7
Mixed ages	4	46.9
Adults (>21)	15	12.5
Offence typology (*k* = 27):		
Property	6	22.2
Persons	1	3.7
Sexual	3	11.1
Drug-dealing	2	7.4
Drunk driving	7	25.9
Mixed	8	29.7
Context cluster		
Place of programme (*k* = 29):		
Juvenile reform centre	1	3.5
Youth prison	5	17.2
Adult prison	7	24.1
Community	12	41.4
Others	4	13.8
Conditions (*k* = 29):		
Closed	11	37.9
Semi-open	2	6.9
Open	10	34.5
Others	6	20.7
Country (*k* = 32):		
Germany	4	12.5
Great Britain	14	43.8
Holland	8	25.0
Spain	2	6.2
Sweden	4	12.5
Methodological cluster		
Design type (*k* = 32):		
Between-groups	22	68.8
Without groups	10	31.2
Subject Assignment (*k* = 22):		
Random	3	13.6
Non-random	19	86.4
Subject selection criterion (*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 sources (*k* = 32):		
Published	25	78.1
Unpublished	7	21.9

[i] The design quality of the studies was assessed by means of a scale containing seven items related to methodological issues, such as: type of design, attrition, sample size, assignment of subjects to groups and type of dependent measures. The questionnaire can be requested from the second author.

.....

[ii] A table with all the moderator variables of each study included in the meta-analysis can be requested from the first author.

geneity can be observed in the programmes applied with respect to the theoretical model on which they were based. Mean duration of the programmes (from when they started to when they finished) was 6 months. Mean intensity (number of treatment hours per week) was 32 hours per subject per week, and mean magnitude (treatment hours) was 840 hours per subject. Distribution of treatment magnitudes also showed great variability, from programmes with a minimum magnitude of 1.5 hours per subject to those with a magnitude of more than 9,000 hours per subject (see Table 3)[iv].

The characteristics of the participants on the programmes were also heterogeneous. Mean age was 25.5 years and the most frequent offences were "crimes against property", "mixed crimes" and "crimes related to alcohol abuse". On the other hand, with the exception of one case, (Bishop, Sundin-Osborne, and Petterson, 1987), all the programmes were applied to male population (71.4%) or to mixed groups of male and female populations (25%).

Great Britain was the country with the highest representation (43.8%) in the sample of studies (or treatment programmes). Most of the programmes were applied in closed (37.9%) or open (34.5%) conditions. The most frequent context in which the programmes were applied was that of the community (41.4%).

With regard to methodological characteristics, 22 studies applied between-group designs and 10 used within-group designs. Their sample sizes were extremely heterogeneous, and ranged from 6 to 1,212 subjects (median: 114 subjects). Mean percentage of attrition for the treated groups was 23.1%, and median of design quality score was 3 points (on a scale of 0 to 7).

One of the main moderator factors of a study on criminal recidivism is the duration of the follow-up period used to evaluate it. The recidivism follow-up period for the programmes included in this meta-analysis was 26.4 months on average (median: 24 months). A follow-up period of two years might, at first sight, seem insufficient to evaluate recidivism in a sample of offenders. In accordance with this, researchers in many studies used follow-up periods of three, five or more years subsequent to the release of offenders. However, despite the apparent brevity, the recidivism rates produced in the first two years of follow-up are significantly representative of the absolute rates, since recidivism normally occurs in the period immediately following the release of the offenders. For example, Redondo, Funes and Luque (1994) evaluated recidivism in a sample of 485 released Spanish prisoners for almost four years, and obtained a mean recidivism rate of 37.9%. Of this total recidivism rate, 58% occurred in the first year after release, and 80% of all recidivism had taken place by the end of the second year. Tournier and Barre (1990) analysed the recidivism rates in different European countries and found that between 49.5% and 80% of recidivism, depending on the studies, took place in the first or second year of follow-up, while between 20% and 50.5% occurred after the end of the second year. An average of 67.7% of recidivism took place in the first two

[iii] The statistical techniques used are presented in Hedges and Olkin (1985) and in Cooper and Hedges (1994). The overall effect, d_{++}, was calculated using a weighted mean that took into account the different variances of the individual studies, although other descriptive statistics were also calculated (unweighted mean, median, quartiles, etc.). In order to determine whether the studies shared a common population effect size, the homogeneity of the variance was calculated by means of the statistic Q_T, which is distributed according to chi-square with k - 1 degrees of freedom, k being the number of studies. When homogeneity is rejected, the contribution of moderator variables to the explanation of variability of effect sizes is put to the test. For the moderator variable categories, an analogue to the variance analysis (weighted by the inverse of the variance of the effect sizes) breaks down 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 is tested in the same way as Q_T, using p - 1 and k - p degrees of freedom, respectively, where p is the number of categories. For the moderator variables that are metric, we used regression analyses, by weighted least squares, in order to determine their relationship with the effect sizes; in this case, the total variability, Q_T, breaks down into Q_R, the explained variability, and Q_E, the non-explained variability. The significance of Q_R and Q_E is tested in the same way as before, with one and k - 2 degrees of freedom, respectively. Finally, with the aim of examining the relationship between effect size and the moderator variables while controlling the influence of other variable clusters, multi-varied analyses were carried out, applying multiple regression analysis by weighted least squares (Hedges, 1994; Hedges and Olkin, 1985).

.....

[iv] Nonetheless, with respect to magnitude, it is necessary to distinguish between residential and non-residential programmes. The treatment magnitude of the non-residential programmes ranged from 1.5 to 180 hours, with a mean of 30.6 hours. In this combination of non-residential programmes, the following treatment categories were included: psychotherapy, education, behavioural therapy, and cognitive behavioural therapy. In all these cases the magnitude is of real significance and represents the real treatment application time. With respect to residential programmes (which included dissuasive programmes, therapeutic community programmes and environmental contingency systems), the magnitude was calculated artificially based on the time the treated and control subjects stayed in the programme application centres. Obviously, the apparent magnitude of the residential programmes was very high in comparison to the non-residential programmes. It ranged from 840 to 9,840 hours, with a mean of 2,508 hours.

years of follow-up, whereas the remaining 32.3% took place from the third year on. These results coincide with the meta-analysis presented by Sánchez-Meca, Marín and Redondo (1996), which compared recidivism rates in European and American countries. Therefore, it would not be unreasonable to suppose that on having information on recidivism for the first two years of follow-up, we have substantial and significant (even though incomplete) information on the overall recidivism of the samples of subjects evaluated.

Overall effect size

For the purposes of the present meta-analysis, recidivism is defined as any measurement related to the committing of new offences. In consequence, our analyses included programmes which had used recidivism measures such as: revocation of probation or parole, confessing to new crimes, the committing of specific crimes (serious crimes, sex crimes or drug trafficking), new sentences and serious new sentences, return to prison, juvenile vandalism, new arrests and re-admittance to juvenile institutions. Table 4 shows the descriptive statistics and the typified mean differences (*d* index). The majority of the programmes (75%) showed lower recidivism in treatment groups than in control groups, or lower recidivism in post-test as compared to pre-test measures. Only one study (McMurran and Boyle, 1990, Study 1) obtained a null effect size (i.e., neither positive nor negative), and seven studies showed unfavourable results for the treatment groups (i.e., treated groups obtained poorer results than non-treated groups). The studies with negative results included two non-behavioural programmes (Cook *et al.*, 1991, Studies 1 and 2), two educational programmes (McMurran and Boyle, 1990, Study 2; Singer, 1991), a therapeutic community programme (Robertson and Gunn, 1987), an environmental contingency programme (Brown, 1985) and a dissuasion programme (Thornton, 1987).

With respect to typified mean difference, treatment effectiveness in relation to recidivism reached a positive mean value of $d_+ = 0.242$ (confidence interval at 95%: 0.196 and 0.287), and in terms of correlation coefficient,

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

Statistic	*d index*	*dp index*
k	32	17
Number of subjects treated	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
Correlation weighted mean[c]	0.120	-
Q_1[d]	0.019	0.000
Median	0.2809	0.123
Q_3[d]	0.491	0.207
Normalised 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 = $\sum d_i/k$.

[b] Weighted mean = $\sum w_i d_i / \sum w_i$, w_i being the inverse variance of each d_i.

[c] With the exception of this statistic, all the calculations were obtained in *d* values.

[d] Q_1 and Q_3 are quartiles 1 and 3, respectively.

[e] Normalised standard deviation SD = $0.75(Q_3 - Q_1)$.

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			
Programme intensity (hours/week)	24	0.7	56	8.3	9.5	6
Programme size (hours/subj.)	23	1.5	9,840	29.6	27	32
				1,323.4	2,203.1	840
Subjects cluster:						
Average sample age (years)	32	12.5	39.3			
Sample gender (% of men)	28	0	100	24.7	7.8	25.5
				91.8	21.5	100
Methodological cluster:						
Sample size	32	6	1,212			
Attrition (in treated group)	21	0	82.1	179	267	114
Design quality	32	1	6	27.9	24.4	23.1
Duration of follow-up (months)	31	1	120	3.4	1.2	3
				26.4	24.2	24
Extrinsic cluster:						
Publication date (year)	32	1982	1991	1988	2.6	1988

Max.: Maximum value. S.D.: Standard deviation.Min.: Minimun value. K:Number of studies.

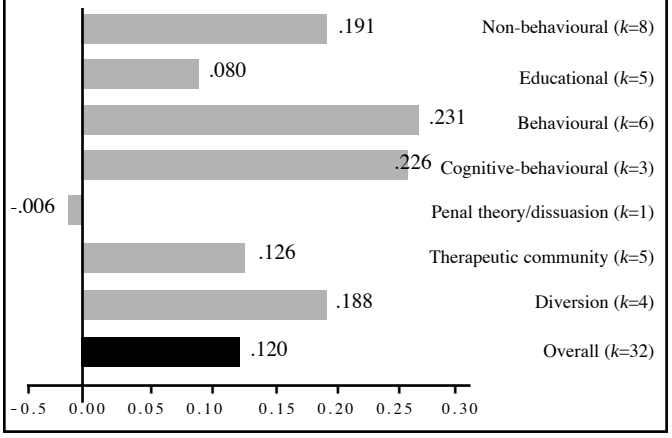
this represents a mean value of $r_+ = 0.120$. With regard to differences in the proportions of recidivism between treated and control groups in the 32 studies assessed, 17 applied dichotomous recidivism measurements and included two groups (Table 4). The weighted mean d_p = 0.15 reflects a differential rate of recidivism of 15% in favour of the treated groups with respect to the recidivism rate of the control groups (confidence interval at 95%: 11.9% and 18.1%). Thus, our results are quite similar to those obtained in other meta-analyses on recidivism measures, such as those of 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 to those of lower effectiveness in the meta-analyses by Garrett (1985), with $r = 0.065$, and by Lipsey (1992), with $r = 0.05$.

Although the results show that in general the treatment of offenders is effective, the effect size distributions were strongly heterogeneous [index *d*: Q_T (31) = 124.070, $p = 0.0000$; index d_p: Q_T (16) = 59.374, p = 0.0000]. The analyses presented below (all based on the *d* index) were aimed at identifying the moderator variables that might serve to explain the differences between the results of the different studies.

Analysis of moderator variables

The first of these analyses refers to theoretical model of treatment. As can be seen from Table 5 (see also Figure 1), the theoretical model of treatment bears a close and statistically significant relationship to the effect sizes [$Q_B(6) = 59.565$, $p < 0.01$], with an explained variance of 48%. The behavioural and cognitive behavioural programmes showed the greatest reductions in recidivism ($r_+ = 0.231$ and $r_+ = 0.226$, respectively), obtaining twice as much effectiveness as the average of all the programmes ($r_+ = 0.120$). On the other hand, the educational programmes were less effective than the mean (r_+ = 0.08), and the only dissuasive programme included in the meta-analysis (Thornton, 1987) produced more recidivism than non-intervention ($r_+ = -0.006$). The contribution of this last-named study to the mean effect magnitude was very high due to its large sample size (n = 1,000 subjects). Therefore, if we eliminate this study from the database, the average effect size increases from $r_+ = 0.120$ to $r_+ = 0.165$, and the heterogeneity statistic drops from $Q_T = 124.070$ to $Q_T = 80.203$.

Figure 1
Treatment cluster: theoretical models and effectiveness (r)



Within the programme application context cluster, it was found that the place of application had a close relationship to its effectiveness (Table 5), a relationship in which the age of subjects variable played a decisive role. The lowest recidivism rates were to be found in the category "other contexts" (e.g., psychiatric units), with r_+ = 0.243. Double the effectiveness was obtained in the only programme to be applied in a juvenile reform centre and in those applied in prisons for young offenders (r_+ = 0.205, y $r_+ = 0.174$, respectively). This contrasted with the figure for adult prisons, in which the lowest effectiveness was obtained ($r_+ = 0.082$).

Clear differences in effectiveness were also found between the different European countries included in the meta-analysis [$Q_B = 52.908$, $p < 0.0000$], with the programmes applied in Spain ($r_+ = 0.241$) and in Germany ($r_+ = 0.198$) showing the highest reductions in recidivism. Although the lowest effectiveness was to be found in the British programmes ($r_+ = 0.026$), it must be pointed out that the great heterogeneity observed in these studies was mainly due to Thornton's study (1987). This study, with its very high sample size (n = 1,000) and a negative d value (d = -0.012), exercises a very strong influence over the mean effect size. In fact, if we eliminate this work from the analysis, the remaining 13 British studies achieve an average of $r_+ = 0.080$, and the overall heterogeneity between the countries falls dramatically from 52.908 to 14.477, though it is still statistically significant ($p < 0.01$).

We also analysed the relationship between effectiveness and some methodological factors. With respect to the criteria used for the inclusion of the subjects on the programmes (Table 5), the greatest effect sizes were obtained by three studies in which the subjects were selected using criteria of violence or sentence duration ($r_+ = 0.338$ y $r_+ = 0.257$, respectively). In contrast, studies that included "all the subjects" of the institution were less effective ($r_+ = 0.068$).

Table 5

Results of the analyses of variance for categorical variables

Cluster characteristics	k_j	d_{+j}	95% C.I.	r_{+j}	Q_{wj}	Q_B	R^2
Treatment cluster							
Theoretical Model (*k* = 32):						59.565**	0.480
Non-behavioural	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**		
Dissuasion 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		
Subjects cluster							
Sample age (*k* = 32):						8.620**	0.069
Adolescents (<16)	7	0.420	0.231; 0.610	0.206	11.203		
Youths (16-21)	6	0.363	0.233; 0.493	0.179	9.987		
Mixed ages	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 typology (*k* = 27):						7.042	0.065
Property	6	0.210	0.147; 0.272	0.104	62.372**		
Persons	1	0.718	-0.107; 1.544	0.338	—		
Sexual	3	0.136	-0.217; 0.489	0.068	14.621**		
Drug dealing	2	0.245	0.116; 0.375	0.122	0.465		
Drunk driving	7	0.218	0.105; 0.331	0.108	12.278		
Mixed	8	0.374	0.248; 0.499	0.184	11.610		
Context cluster							
Place of programme (*k* = 29):						17.384**	0.150
Juvenile reform centre	1	0.419	-0.397; 1.235	0.205	—		
Youth 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**		
Others	4	0.500	0.327; 0.674	0.243	1.305		
Conditions (*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		
Others	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**		
Holland	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-group	22	0.345	0.285; 0.405	0.170	56.912**		
Within-group	10	0.109	0.040; 0.177	0.054	41.336**		
Assignment of subjects (*k* = 22):						1.854	0.032
Random	3	0.037	-0.410; 0.484	0.018	1.410		
Non-random	19	0.351	0.290; 0.411	0.172	53.648**		
Subject selection criterion (*k* = 32)						27.307**	0.220
All subjects in an institution	6	0.134	0.069; 0.204	0.068	42.949**		
According to sentence duration	2	0.531	0.381; 0.681	0.257	0.604		
According to specific typology	7	0.235	0.120; 0.351	0.117	18.523**		
Violent subjects	1	0.718	-0.107; 1.544	0.338	-		
According to treatment 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		

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

An important methodological aspect is the influence of type of design on the results of the studies. Although pre-test/post-test designs with only one group commonly produce greater effect sizes than between-group designs, in our meta-analysis we obtained the opposite result (Table 5): one-group studies obtained an mean effect size ($r_+ = 0.054$) significantly smaller than that of between-group studies ($r_+ = 0.170$). However, this anomalous result can once again be explained by the great weight assigned to the study by Thornton (1987). In fact, if we eliminate this study from the analysis, the nine remaining pre-test/post-test studies obtain a mean effect size of $r_+ = 0.147$, and there are no longer statistically significant differences between them and the between-group designs [$Q_B = 0.578$, $p > 0.05$]. Gottschalk, Davidson II, Gensheimer and Mayer (1987) obtained a similar result.

The 22 between-group studies were classified in two categories according to the assignment of subjects to groups: random and non-random. As is customary (e.g., Cleland, Pearson, and Lipton, 1996; Hall, 1995; Lipsey, 1992; Whitehead and Lab, 1989), studies with a random assignment obtained a smaller mean effect size ($r_+ = 0.018$) than those with non-random assignment ($r_+ = 0.172$), but this difference was not statistically significant [$Q_B = 1{,}854$, $p > 0.05$]. However, this result must be interpreted with caution, given that only three studies with random assignment of subjects to groups were included in the meta-analysis.

The follow-up period during which the recidivism data of the samples is registered is another critical variable in the explanation of heterogeneity of results between the studies. As can be seen from Table 6, the weighted regression analysis of follow-up period on the *d* values produced a statistically significant result [$Q_R = 9.676$, $p < 0.01$]. In contrast to what might normally be expected, this relationship was positive, although in some previous meta-analyses this anomalous result has also been found (Hall, 1995).

For each study the design quality was assessed by means of a scale that took into account aspects such as: sample size, random assignment, attrition, use of normalised dependent variables, presence of a control group, equivalence between dependent measures reported in the pre-test and the post-test, the existence of some measurements in the pre-test, etc. The relationship between quality of design and effect size was also unusual. Specifically, it was found that high quality studies were associated with the greatest effect sizes [$Q_R = 24.360$, $p < 0.01$], with an explained variance of 17%. Nevertheless, other meta-analyses have obtained similar results (Lipsey and Wilson, 1993). Finally, the sample size and attrition in the treated group barely attained a significant relationship with effect size, with a very low percentage of explained variance.

With respect to extrinsic variables, the date of the study and the source of publication were analysed. The date of the study did not show a significant relationship with effect size (Table 6), although a slight reduction in effect size over time could be observed. This was similar to the result obtained in Gottschalk, Davidson II, Gensheimer and Mayer (1987). For the purposes of checking whether the publication bias might affect the representative nature of our meta-analytic database, the studies were

Table 6
Results of the simple regression analyses for quantitative variables

Cluster characteristics	*k*	*B*	Q_R	Q_E	R^2	R_{adj}^2
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	$5.3\ 10^{-5}$	3.899*	68.527**	0.054	0.009
Subjects cluster						
Average sample age (years)	32	-0.004	1.166	122.774**	0.009	0.000
Sample gender (% of males)	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
Duration of follow-up (in months)	31	0.004	9.676**	113.846**	0.078	0.046
Extrinsic cluster						
Publication date (year)	32	-0.014	2.815	121.125**	0.023	0.000

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

categorised into "published" and "unpublished" (Table 5). The comparison did not show significant differences between the two mean effect sizes [Q_R = 2.462, p > 0.05]. Moreover, the unpublished studies obtained a greater mean effect size (r_+ = 0.159) than the published studies (r_+ = 0.112). Thus, publication bias can be discarded as a threat to the validity of the results of our meta-analysis.

Multivariate analysis

The inter-relationships between the variables of the studies may mask the real associations between the effect size and each individual moderator variable. More specifically, the differences in effectiveness found between the studies might be influenced by differences in the methodology applied (e.g., differences in type of design, quality of the study, etc.), as well as by differences in subject characteristics (e.g., average age of the samples). In order to control these inter-relationships we applied weighted multiple regression analysis techniques. Thus, answers can be provided to several relevant questions, such as: once the other variables are controlled, is treatment type still a moderator variable of the results? What is the explanatory value of the different moderator variable clusters once the other clusters have been controlled?

Given the small size of our meta-analytic database (k = 32 studies), we had to choose a small group of moderator variables, specifically, those which were most important from a conceptual and statistical point of view. From the cluster of treatment variables we chose the most relevant for our purposes: treatment type. In order to codify the treatment categories, five dichotomous variables were initially established: Non-behavioural therapy (0: no; 1: yes), educational programmes (0: no; 1: yes), behavioural and cognitive behavioural therapies (0: no; 1: yes), therapeutic community (0: no; 1: yes), and diversion programmes (0: no; 1: yes). Owing to the similar level of effectiveness obtained by the behavioural and cognitive-behavioural programmes, the two were placed in the same category. From the subject cluster we selected the judicial system, distinguishing between the juvenile system (0) and the adult system (1). For this purpose we dichotomised the age variable of the sample, assigning the value 0 (juvenile system) to the categories 'adolescent' and 'youth', and the value 1 to the categories 'adult' and 'mixed' (mixed ages). Finally, from the methodological cluster we included design quality (0: low; 7: high) and the follow-up period (in months). There was data missing from this latter variable, which was replaced by the median of the follow-up period (6 months).

In total, eight moderator variables were included in the regression model. As can be seen in Table 7, the multiple regression analysis applied to the complete model produced high statistical significance with an explained variance of 55.6%. The explanatory potential of the model is by no means insignificant when compared to that obtained by the most relevant meta-analysis in this field, Lipsey (1992), which obtained an explained variance of 47%.

The regression analyses performed separately for each cluster indicate that the method cluster alone accounts for 21.9% of effect size variance, a result similar to the 25% obtained by Lipsey (1992). However, when the effect of the other two clusters of variables (treatment and subject) is taken out, the explanatory potential of the methodological variables is reduced to 4.2%, and its contribution is non-significant (p = 0.074). This result indicates that there is a strong inter-relationship between the method and treatment variables. Another interesting result of the separated regression analysis was the inversion of the relationship between effect size and the follow-up period. While a positive relationship was obtained in the simple regression analysis, the relationship became negative when the other variables in the model were controlled (see Table 8).

The subject cluster, composed of just one dichotomous variable (juvenile versus adult justice) obtained an explained variance of 6.5%, and this remained at 6.2% when the influence of the other two clusters was removed. However, its relationship with effect size was also reversed with respect to the univariate analysis carried out previously.

Table 7
Multiple regression analyses for each moderator cluster, adjusted and non-adjusted for the remaining moderator clusters

Cluster	$Q_R(DF)$	R^2	$Q_E(DF)$	$Q_{Rpart}(DF)$	R_{part}^2
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
Complete model	68.968(8)	0.556	54.972(23) **		

Table 8
Non-standardised regression coefficients obtained in the multiple regression analysis for each moderator variable

Moderator variable	B_j
Design quality (0: low; 7: high)	0.062
Follow-up period (in months)	-0.004
Judicial system (0: Young people; 1: Adults)	0.407
Non-behavioural programmes (0: No; 1: Yes)	0.434
Educational programmes (0: No; 1: Yes)	0.087
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

Another variation was produced with respect to the subjects' age variable. Table 5 shows that better results were achieved with young offenders than with adults. However, in the multiple regression analysis adults achieved better results than young offenders (Table 8). The explanation for this contradictory finding resides in the different treatments utilised: seven of the thirteen programmes applied to young people used the techniques found to be most effective by our analysis (behavioural and cognitive-behavioural). Only two adult programmes applied these techniques, as against seven that employed non-behavioural methods and five that used therapeutic communities. This imbalance in the types of programme applied to adults and young people explains why the variance analysis (in Table 5) shows greater effectiveness for young offenders. In contrast, when the influence of treatment type was controlled, the better results corresponded to adult offenders.

Finally, the treatment cluster obtained the highest percentage of explained variance (48.2%), which was only slightly lower (33.6%) when the effect of the other clusters was taken out. This result shows that the different effectiveness of the programmes is irrefutable, since the explained variance is maintained even when the subject and method variables are controlled. Moreover, as can be seen in Table 8, the differential effectiveness of the different treatment models is similar to that found in the univariate analysis: the behavioural and cognitive-behavioural programmes achieve the greatest effectiveness, followed by diversion programmes and non-behavioural therapy.

The misspecification testing for the multiple regression model using Q_E were significant, indicating that relevant variables are missing from the model. In fact, the small size of our meta-analytic database obliged us to omit conceptually important moderator variables from the model (e.g., the characteristics of the programmes and their application context). Consequently, the model proposed should be considered only as a tentative one.

Table 9
Binomial Effect Size Display for $r_+ = .120$

Group	Recidivists	Non-recidivists	Total
Treatment	44	56	100
Control	56	44	100
Total	100	100	200

CONCLUSIONS

The present study explored the effectiveness of a set of treatment programmes for offenders applied in European countries in the course of a decade. In terms of correlation coefficient, mean effect size for the 32 programmes assessed was $r_+ = 0.120$. This result means that moderate effectiveness was obtained, on average 12% ($r_+ = 0.120$). The practical importance of this can be interpreted on the basis of the Binomial Effect Size Display, *BESD*, as shown in Table 9 (Rosenthal, 1991). Thus, assuming that there are 100 subjects in the treatment group and another 100 in the control group, the recidivism rate would be 44% in the treated group and 56% in the control group. In other words, treated subjects would relapse 12% less than subjects that had not been treated.

Bearing in mind the heterogeneity of programme effectiveness, the influence of different factors on effectiveness was analysed. On the basis of this analysis, recidivism rates appeared to be linked to the following factors:

1. The most effective programmes are those based on behavioural and cognitive-behavioural theoretical models. Similar conclusions have been reached in 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. In our first analysis, juvenile offenders appeared to be more responsive to rehabilitation, due to the fact that they were generally treated by means of the most effective techniques (behavioural and cognitive-behavioural). Nevertheless, in a second analysis in which the influence of treatment type was controlled, the greater effectiveness was obtained with adult offenders. This data indicates that positive results can be achieved with both juvenile and adult offenders. Looking beyond the age of the subjects, which is without doubt a conditioning factor, the most relevant factor is the application of programmes based on plausible theoretical models. Research has repeatedly shown that behavioural and cognitive-behavioural programmes produce the best results.

3. With respect to offence typology, programmes applied to offenders against persons (excluding sex offenders) show the greatest effectiveness. On this point, despite the fact that the small number of studies in the crime categories restricts the drawing of general conclu-

[v] The *BESD* is obtained by calculating the non-recidivism rate of the treated group by means of 0.5 + *r*/2 and the non-recidivism of the control group by means of 0.5 - *r*/2, with *r* = 0 .12 (cf. Rosenthal, 1991).

sions, the results obtained are consistent with the criteria of quality of interventions proposed by Andrews *et al.* (1990), based on the so-called "risk principle". The risk principle suggests that treatment is more effective with high-risk offenders (violent offenders, recurrent offenders, etc.) than with lower-risk offenders, since more intensive and higher quality services and treatments are generally used with the former group. Thus, and taking into account that violent offenders are "higher risk" (than offenders against property) and have a greater need for treatment, it is reasonable to expect that programmes of higher implicit quality will have been carried out with them, and hence the greater effectiveness.

4. Finally, with respect to treatment context, the greatest effectiveness was produced in institutions for young people, both in reform centres and in prisons for young offenders.

The results of this European meta-analysis are similar to the principal results of previous reviews on the treatment of offenders that we have mentioned. Most 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) agree on one general conclusion: recent meta-analyses on the treatment of offenders contradict the idea that "nothing works", which was the main conclusion drawn by early research in the field. Many juvenile and adult treatment programmes are effective with wide-ranging groups of offenders. As Dilulio puts it, "the notion that "nothing works" can be thrown into the dustbin of correctional evaluation history" (quoted in Taylor, 1992, p. 133).

Our study is also consistent with those of previous reviewers in its conclusion that greater effectiveness is achieved with behavioural and cognitive-behavioural methods. For example, Palmer (1995) examined 23 qualitative reviews and 9 meta-analyses, all carried out before 1989, and concluded that the most effective and promising models in the treatment of offenders were "behavioural, cognitive-behavioural or cognitive, life skills, multi-modal and family" (p 101).

Nevertheless, "it is important to acknowledge that the meta-analytic process de-contextualises the situation in which the treatment and/or the decisions taken in order to carry it out take place (...). On inverting the analysis process, from the meta-analysis to the concrete circumstances in which the programmes were applied, it is essential to bear in mind those aspects of the specific environment that may have an influence on the application of these principles – in other words, the influence of contextual factors that are difficult to quantify. In practical terms, this means that the application of knowledge obtained from general analyses (in this case, meta-analyses) is not so simple, and requires the careful consideration of specific environmental contingencies" (Brown, 1996).

In our opinion, future research should consider it a priority to study those factors that positively influence the effectiveness of treatment programmes in particular contexts (see Palmer, 1995). The "specific environmental contingencies" are difficult to assess in meta-analytic studies, but they can be appropriately studied through direct primary research, which would undoubtedly help to improve the effectiveness of offender rehabilitation.

ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to Dr. Fulgencio Marín Martínez for his participation in the study of the reliability of the calculation of the effect sizes, and to Silvia Sebaté García for her help with the final preparation of the text for printing.

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. Forskningsgruppen, 1, 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 educa-

tional 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. 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. 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 of correctional 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. Unpublished doctoral thesis, Universidad de Barcelona.
Redondo, S. Funes, J., and Luque, E. (1994). Justicia Penal y Reincidencia. Barcelona: Jaume Callís.
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. Delincuencia / 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. Delincuencia/Delinquency, 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. 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. In J. Funes, E. Luque, A. Ruíz, and J. Sánchez-Meca (Eds.), Reincidència: En la justícia de menors/Avaluació internacional (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: Démographie 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.