

Does sexual offender treatment work? A systematic review of outcome evaluations

Martin Schmucker and Friedrich Lösel*

University of Erlangen-Nuremberg (Germany) and * University of Cambridge (UK)

The article reports a systematic review of controlled outcome evaluations of psychosocial and organic sexual offender treatment. A comprehensive search of the literature in five languages revealed 80 independent comparisons between treated and untreated groups of sexual offenders ($N= 22,181$). The majority of studies confirmed a positive treatment effect. Overall, 11.1% of treated offenders and 17.5% of controls showed sexual recidivism (37% difference). Findings for violent and general recidivism were similar. Studies on surgical castration showed the strongest effect; however, this was confounded with methodological and offender characteristics. Hormonal medication, cognitive-behavioural, and behavioural approaches also revealed a positive effect. Non-behavioural treatments did not show a significant impact. Other moderators such as small sample size, authors' affiliation with the program, program completion versus dropout, or type of outcome measure had a significant impact. Methodological study characteristics explained the largest proportion of effect size variance. Overall, findings are promising but more differentiated evaluations of high quality are needed.

¿Es efectivo el tratamiento de los delincuentes sexuales? Una revisión sistemática de los resultados evaluados. Este artículo presenta una revisión sistemática de los resultados de evaluaciones controladas acerca del tratamiento psicosocial y orgánico de los delincuentes sexuales. Un estudio exhaustivo de la literatura destacó 80 comparaciones independientes entre grupos de delincuentes sexuales tratados y no tratados ($N= 22.181$). La mayoría de los estudios mostraron un efecto positivo. En conjunto, los delincuentes tratados reincidieron el 11,1%, mientras que los no tratados llegaron al 17,5% (es decir, un 37% de diferencia). Los hallazgos en reincidencia violenta y no violenta fueron parecidos. Los efectos más grandes los obtuvo la castración quirúrgica, aunque ciertas variables metodológicas y de los sujetos contaminan ese dato. Los tratamientos no conductuales no mostraron ser efectivos. Otras variables moderadoras como el tamaño pequeño de la muestra, la relación de los autores de los informes con el programa, acabar o no el mismo, o el tipo de resultado analizado tuvieron un efecto significativo. Las características metodológicas de los estudios explicaron la mayor parte de varianza de los resultados. En conjunto, los resultados son prometedores, aunque se necesitan evaluaciones diferentes de alta calidad.

Sexual offending is a kind of crime about which the general public is particularly concerned. Discussions in the media and in the political arena are emotionally laden and even legislation often seems to be driven by serious single cases of sexual offending. Accordingly, in the last decade many Western countries have revised their criminal justice reactions to sexual offending towards harsher punishment and incarceration. Because most sexual offenses are not so serious to justify lifetime sentences there are also increased attempts to prevent reoffending through correctional treatment. In Germany, for example, a penal law reform in 1998 introduced mandatory treatment for sexual

offenders who received a prison sentence of more than two years. However, the changes in policy have been primarily based on good intentions and not on sound knowledge of the effectiveness of sexual offender treatment (Lösel, 2002). As Garrido, Farrington & Walsh (2006) pointed out in the previous Psicothema section devoted to crime prevention, policy and evaluation are not always walking the same path. Accordingly, in many countries there is an ongoing discussion of the question: What works in the treatment of sexual offenders?

Early reviews of this issue revealed considerable inconsistency. For example Marshall et al. (1991) proposed the effectiveness of treatment, whereas Quinsey et al. (1993) were critically opposed to their point of view. They mainly based their sceptical stance on the methodological shortcomings of the available evaluations. This is in accordance with what Furby, Weinrott and Blackshaw (1989) concluded in a first systematic meta-analysis of findings on the effects of sexual offender treatment. Some years later, various meta-analyses came to more positive conclusions (e.g., Alexander, 1999; Gallagher, Wilson, & MacKenzie, 2000; Hall, 1995a;

Hanson et al., 2002; Lösel, 2002). Such changes may have been due to a substantial increase of evaluation research in this field. While Furby et al. (1989) could find only 11 controlled evaluations of sexual offender treatment, Hall (1995a) integrated 12 controlled studies that have been issued since Furby's publication. Gallagher et al. (2000) drew upon 26 comparisons and Hanson et al. (2002) found 43 studies that applied minimum quality standards in evaluation design. Although these and other meta-analyses found significant effects, there are substantial differences in the size of the mean effect reported (Lösel & Schmucker, 2003). For example, Hanson et al. (2002) reported a relatively small mean effect of $d = .12$ for sexual reoffending, whereas Gallagher et al. (2000) found a much larger overall effect of $d = .47$. Such differences seem to be related to specific selection criteria regarding the methodological quality of the studies and the kinds of treatment that have been analyzed. In addition, previous meta-analyses only considered the English-language literature.

Against this background, we conducted a comprehensive meta-analysis of the international literature on sexual offender treatment that contained both psychosocial and biological modes of treatment and studies that have been published in various languages. The study was carried out within the framework of the Campbell Collaboration Crime and Justice Group. In the following, we present selected data on the characteristics of the evaluation studies, their general findings on outcome, and on moderators that account for differences in effectiveness.

Method

Eligibility criteria. In order to use all relevant evaluations on sexual offender treatment we did not set any historical or regional limits regarding the treatments evaluated. However, studies had to meet minimal methodological standards to be eligible:

- a) *Treatment of sexual offenders.* Treatment included any measures that used therapeutic interventions and aimed at reducing reoffending. The nature of therapy could both be psychosocial or organic, i.e. pharmacological or surgical. Purely deterrent measures or managerial practices that did not include therapeutic elements were not eligible.
- b) *Evaluation by means of recidivism outcomes.* Recidivism was defined broadly and included new convictions, incarcerations, arrests or charges as well as lapse behavior. Measures of personality, e.g. empathy, or therapists' ratings of treatment success were not sufficient. The results had to be presented in a way that allowed a reasonably accurate estimation of the treatment effect.
- c) *Inclusion of a comparison group that did not receive the treatment in question.* This would either be no treatment, «treatment as usual» or other less intensive or less specific treatment. Treatment dropouts were not counted as an appropriate comparison group.

For economic reasons we restricted our selection to studies in English, German, Dutch, French, and Swedish. However, there seemed to be not much more controlled research published in other languages. Minimum sample size was set to $N = 10$. This is a compromise between the small samples of many studies and the problem that very small samples may lead to excessive effect sizes based on minor differences in absolute numbers.

Literature search. We used a wide variety of sources to identify relevant literature. First, the references of existing reviews and meta-analyses formed a basic study pool. In addition we performed searches of relevant databases (e.g. PsycInfo, MedLine, Dissertation Abstracts International). Handsearches of journals pertinent to the topic were performed and the references of identified studies were scanned for further trials. To be sensitive to unpublished evaluations of sexual offender treatment an internet search was conducted. Finally, researchers in the field of sexual offender treatment were personally contacted and asked for further studies.

In total, over 2,000 citations were identified. Of these, 66 reports included 69 studies that met the inclusion criteria. When studies reported results for more specific subgroups (e.g. according to offense type) and reported characteristics of the subgroups in as much detail as for the whole group, we chose these subgroups as units of analysis. This allowed a higher differentiation in moderator analysis while the samples remained independent. Following this principle 80 comparisons were available for the integration.

Coding. Every study underwent a detailed coding procedure. Variables included general study features, methodological characteristics, treatment variables and characteristics of the treated offenders. A detailed manual was used in order to ensure objectivity. The first author coded all studies on the basis of a detailed manual. To test interrater agreement, ten studies were coded by another experienced rater. On average the interrater agreement was at 91 % with no variable below 60 %. Core variables such as design or principal treatment strategy were at 100 %.

To evaluate the overall methodological quality we adapted the Maryland Scale of Scientific Rigor (Sherman et al., 1997) for our purposes. This is a 5-point rating that integrates various methodological study features regarding the validity of causal interpretations. Level 1 of the Maryland Scale refers to uncontrolled studies. These were not eligible for our meta-analysis. In our adaptation the eligible studies were assigned to the different levels according to the following definitions: (2) non-equivalent control group (e.g. demonstrated differences in psychosocial characteristics, treatment refusers); (3) group assignment is incidental but equivalence can be assumed, i.e. although assignment is not actively controlled by the researchers the groups seem largely comparable (e.g. demonstrated equivalence on important variables, assignment followed principles unrelated to risk of recidivism); (4) systematic procedures to ensure comparability of treatment and control group (e.g. pairwise matching, statistical control); (5) uncompromised random designs. If random assignment procedures were used but flawed because of dropout or other reasons, studies were downgraded.

Study sample characteristics. Although we did not set historical limits most of the studies were very recent. About three quarters appeared since the 1990s and one third even in the new millennium. The studies are mainly North-American (USA: 31; Canada: 17). However, we also located eight studies from German-speaking countries, eight from Great Britain, and five from other countries. Unpublished evaluations comprised 36 % of our study pool.

The most common treatment approach was cognitive-behavioral therapy (46 %). Other approaches each comprised less than 15 % of the comparisons. This includes six comparisons on hormonal medication and eight comparisons on surgical castration. One third of the cases refer to outpatient treatment only. Residential treatment was somewhat more frequent (prison: $k=25$; hospital: $k=14$). Ten comparisons referred to a mixture of outpatient and residential treatment settings. Seven comparisons exclusively addressed adolescent offenders while the majority ($k=45$) referred to adults only.

The design quality of the evaluations was generally poor. There were only six uncompromised random designs on Level 5 of the Maryland Scale. Seven comparisons corresponded to Level 4 and 19 to Level 3. The vast majority applied obviously non-equivalent control groups (Level 2; $k=48$). As far as risk variables were reported, the majority of Level 2 studies contained treated groups at higher risk ($k=14$ vs. $k=5$). For 29 comparisons the direction was unclear or no data on relevant variables were reported. The follow-up period ranged from 1 to 10 years and averaged 5.22 years ($SD=3.46$). Small samples were common. 25 comparisons referred to samples of 50 offenders or less. However, in one comparison total sample size was as large as $N=2,557$. The median was at 118. Overall, the analysis refers to 22,181 offenders of which 9,512 had been in the treatment groups.

Effect size computation and statistical integration. Official recidivism rate was the typical outcome criterion. As recommended for dichotomous data (Fleiss, 1994; Lipsey & Wilson, 2001) we used Odds Ratios (*OR*) as effect size measure and conducted the statistical analyses on the natural log of the Odds Ratio (*LOR*). Only some studies either reported the data in other formats or presented more sophisticated analyses. In the latter case we used the more sophisticated data if possible.

The unit of analysis was an individual study. If a study contained more than one dependent (sub)sample we chose the comparison that showed the best internal validity. Some studies reported results for different independent subgroups (e.g. according to offense type, age groups, or risk classification) but did not qualify as independent comparisons as defined above. In regard of the comparability of such subgroups we separately calculated effect sizes for each subgroup and then used the weighted average to obtain a study effect size.

We considered and integrated different domains of recidivism (e.g. sex offenses, non-sexual violent offenses, any offenses)

separately. If more than one indicator of recidivism was reported for an offense type (e.g. arrests and convictions) we averaged the resulting effect sizes to one study effect size. This seemed to be appropriate because comparisons between the different indices of recidivism did not reveal systematic differences.

Effect size integration followed the procedure developed by Hedges & Olkin (1985) which weighs the individual study effect sizes according to their standard error. A *Q* test revealed that the effect size distribution was significantly heterogeneous and we thus applied the random effects model. Moderator analyses were carried out using a mixed effects model accordingly (see also Wilson, 2001).

Results

Overall effects

74 of the 80 comparisons reported sexual recidivism outcomes. Data regarding (non-sexual) violent reoffending were provided for 20 comparisons and in 49 studies the authors presented data on overall recidivism. The simplest approach to examine the overall effect is to compare the direction of effects, irrespective of their statistical significance. If there is no treatment effect, one would expect as many comparisons indicating positive outcomes (i.e. lower recidivism rates in the treated group) as negative outcomes (i.e. the treated group recidivated at the same or an even higher rate than the untreated group). Table 1 shows that positive outcomes were significantly more frequent for all domains of recidivism.

Meta-analytic integration allows a more detailed look as it pays attention to the size of the individual effects. The mean odds ratios were quite consistent across the different domains of reoffending ranging from 1.67 for any recidivism to 1.90 for violent reoffenses. For sexual recidivism the average treatment effect was $OR=1.70$ (each $p<.001$; see Table 1). Taking the *n*-weighted average sexual recidivism rate of the treated groups (11.1 %) as a base rate, this average effect translates to a 17.5 % average recidivism rate in the comparison groups. Accordingly, the recidivism rate of the treated offenders was 6.4 percentage points or 37 % lower than in the control groups. Although the base rates vary considerably for different domains of reoffending, in terms of proportions the results are similar (violent recidivism: 44 % lower; any recidivism: 31 % lower).

However, the average effects are only a rough indicator of the treatment success in individual studies. Except for violent

Table 1
Total outcomes and mean effects

Outcome	Balance					Recidivism (%)	
	k	TG : CG ^a	OR	CI _{95%}	Q	TG ^b	CG ^c
Sexual recidivism	74	53 : 18***	1.70***	1.35 - 2.13	237.14***	11.1	17.5
Violent recidivism	20	18 : 2***	1.90***	1.49 - 2.33	19.68	6.6	11.8
Any recidivism	49	36 : 13**	1.67***	1.33 - 2.08	159.80***	22.4	32.5

Note: *k*= number of comparisons; Balance TG : CG = number of comparisons favoring treated vs. untreated group, i.e. positive vs. negative treatment outcomes; *OR*= mean odds ratio; *CI*_{95%}= 95 % confidence interval; *Q*= test of homogeneity (χ^2 , $df= k - 1$); TG= treated group; CG= comparison group.

^a χ^2 tests; ^b *n*-weighted average; ^c estimated recidivism rate

*** $p<.01$; ** $p<.001$

recidivism homogeneity analyses indicate considerably heterogeneous effect size distributions, i.e. the differences between the results of individual evaluations are clearly above what would be expected by chance. We thus conducted moderator analyses in order to isolate variables that might account for these differences. We restricted these analyses to the domain of sexual recidivism for reasons of space and also because this is the main area of interest and provides a larger database.

Moderator analyses for sexual recidivism outcomes

Content of treatment. One set of comparisons clearly stood out of the study pool, both in terms of treatment and in terms of effect sizes. The eight comparisons on surgical castration showed an average odds ratio of 15.03 ($z=9.03, p<.001$). Moreover, the effect size distribution is very homogenous with $Q(7)=1.76, p=.97$. The remaining 66 comparisons' average effect is clearly lower but remains significant ($OR=1.38, z=3.16, p<.01$). For various reasons we decided to exclude the comparisons on surgical castration from further analysis: First, this approach is rarely used in contemporary practice. Second, it contains serious legal and ethical problems. Third, because of the legal conditions for surgical castration the respective studies do not contain equivalent control groups. Fourth, the extremely high and homogenous effects would have distorted any further analysis considerably.

Results of the moderator analyses of the more common treatment programs are shown in Table 2. Even after the exclusion of the surgical castration studies, the effect size distribution remains highly heterogeneous, $Q(65)=163.92, p<.001$, and the main treatment approach used still exerts a significant influence on the evaluation results. Hormonal medication shows the highest mean effect. Of the psychosocial approaches only cognitive-behavioral as well as classic behavioral approaches indicate significant treatment effects. The odds ratios of the other psychosocial approaches center around 1 and indicate no difference in recidivism rates between treated and untreated offenders.

The coding of general treatment approaches is not very subtle because modern treatment programs do not strictly refer to one therapeutic approach only but are more eclectic. For example, hormonal treatment is often accompanied by psychotherapy and psychosocial programs frequently contain various therapeutic modules. Therefore, we additionally rated how far the evaluated programs incorporated elements from different therapeutic approaches. We used a 4-point scale ranging from 0 (not at all) to 3 (mainly). A multiple regression analysis revealed 21 % of variance explained by these combinations ($p=.05$). Overall, the initial analyses on the treatment approaches were confirmed in that only the inclusion of cognitive and behavioral treatment elements as well as hormonal medication revealed significant β -weights (each $p<.05$).

Other treatment characteristics. Only outpatient treatment showed a significant effect. The mean odds ratios for treatment in institutional settings were considerably lower and not significant. For mixed settings there was an intermediate effect. While these differences did not reach significance, an ordering of the setting variable from institutional to outpatient treatment revealed a significant correlation of $r=.27 (p=.02)$. However, this variable is somewhat confounded with the treatment approach. If one

controls the latter by only analyzing comparisons on cognitive-behavioral treatments the relationship is much weaker ($r=.12, p=.45$).

What does make a difference, however, is whether the treatment program was specifically designed for sexual offenders or a general offender program that also incorporated sexual offenders. This difference was also found in outcomes of general recidivism ($p=.05$). Another finding related to the implementation of the programs. Evaluations in which the author was in some way involved in the program delivery showed clearly significant effects but programs that were evaluated by independent researchers did not. That model projects reveal better outcomes than programs that were implemented in everyday routine was a related finding. However, the difference is not as clear-cut as for author affiliation and both kinds of implementation revealed significantly positive mean effects.

Offender characteristics. Unfortunately, the description of offender characteristics was often scarce. For example, only some studies differentiated the results by the type of sexual offense. In these, there were significant reductions in recidivism for any type of offense except for incest offenders. The latter finding is mainly due to the low recidivism base rate of incest offenders who are often taken out of the respective family. The rather high effect reported for rapists seems to contradict the common notion that these offenders are particularly difficult to treat. However, this finding is based on only five comparisons and should be regarded with caution. Furthermore, rapists may be less sexually disordered than, for example, child molesters with a preference for male children.

Although programs that exclusively addressed juvenile offenders show somewhat stronger effects than program for adult offenders, this was not a significant difference. A related finding refers to the age homogeneity of the treated offenders. It showed a tendency of higher effects in more homogeneous age groups ($r=.23, p=.10$).

As an indicator of therapy motivation, we compared programs with voluntarily participating offenders with programs that involved a more or less coerced treatment. Only the voluntary treatment showed a significant mean effect. However, there was considerable heterogeneity within any of the categories (each $p<.02$) and the between differences did not reach statistical significance.

In contrast, whether offenders regularly terminated a program or dropped out of treatment had a strong impact. Because the control groups for these comparisons were partially dependent, this difference could not be tested statistically but was rather obvious. The odds to not recidivate for offenders who dropped out of treatment was only half the odds for offenders who were not treated at all. In contrast, for offenders who terminated treatment according to the treatment plan the mean odds ratio was clearly in the positive range.

Methodological characteristics. Although the design quality indicates how much one can trust a result, it did not have a systematic impact on the effect sizes. At $p=.06$, comparisons rated at Level 3 or higher on the Maryland Scale actually fared somewhat better ($OR=1.69, CI_{95\%}: 1.26 - 2.28$) than comparisons on Level 2 ($OR=1.16$; see Table 2). However, there was no linear trend in the relation between effect size and design quality ($r=.11$,

Table 2
Effects of selected moderator variables

Variables	<i>k</i>	<i>OR</i>	<i>CI</i> _{95%} <i>lower - upper</i>	<i>Q</i> _{bet}	<i>p</i>
Total effect	66	1.38**	1.13 - 1.69	—	
Moderator analyses					
Treatment approach				12.60*	.027
Cognitive-behavioral	35	1.46**	1.12 - 1.89		
Classical behavioral	7	2.18*	1.20 - 3.97		
Insight oriented	5	1.01	0.52 - 1.96		
Therapeutic community	8	0.87	0.54 - 1.39		
Other psychosocial, unclear	5	0.94	0.52 - 1.68		
Hormonal medication	6	3.11**	1.39 - 6.95		
Specific treatment for sex offenders				4.70*	.030
Yes	56	1.56***	1.27 - 1.93		
No	5	0.76	0.41 - 1.41		
Setting of treatment				5.10	.165
Prison	21	1.16	0.84 - 1.60		
Hospital	8	1.10	0.62 - 1.94		
Outpatient	27	1.93***	1.35 - 2.77		
Mixed	10	1.37	0.78 - 2.41		
Status of treatment program				2.63	.105
Model project	9	2.41**	1.20 - 4.86		
Routine practice	57	1.32**	1.07 - 1.63		
Author affiliation to treatment program				10.95***	.001
Yes	32	1.92***	1.44 - 2.56		
No	30	0.99	0.76 - 1.29		
Age group				1.19	.275
Adolescents only	7	2.35*	1.01 - 5.43		
Adults only	36	1.43*	1.08 - 1.90		
Offense type				9.04*	.029
Rape	5	4.91**	1.64 - 14.68		
Child molestation (not in family)	9	2.15*	1.11 - 4.16		
Child molestation (incl. incest)	10	1.02	0.58 - 1.80		
Exhibitionism	4	3.72*	1.27 - 10.93		
Treatment participation				2.22	.329
Voluntary	28	1.45*	1.08 - 1.93		
Non-voluntary	15	1.05	0.70 - 1.58		
Mixed	7	1.01	0.57 - 1.77		
Treatment termination				— ^a	
Treatment completed regularly	44	1.58***	1.23 - 2.05		
Dropped out of treatment	14	0.51***	0.39 - 0.67		
Sample size				31.43***	.000
10 to 50	18	4.03***	2.50 - 6.50		
51-100	10	1.32	0.76 - 2.27		
101-200	16	1.65**	1.13 - 2.41		
201-500	12	1.00	0.72 - 1.38		
More than 500	10	0.88	0.64 - 1.21		
Design quality: Maryland Scale				6.13	.106
Level 2 (non-equivalent)	37	1.16	0.90 - 1.50		
Level 3 (equivalence assumed)	17	2.08***	1.40 - 3.08		
Level 4 (matching, statistical control)	6	1.19	0.67 - 2.12		
Level 5 (randomization)	6	1.48	0.74 - 2.96		
Control group formation				1.64	.200
Treatment refused	11	1.96**	1.20 - 3.20		
Other	47	1.37*	1.07 - 1.75		
Source of recidivism data				4.56*	.033
Criminal records only	57	1.28*	1.04 - 1.57		
Also self report	5	3.32**	1.42 - 7.78		
Publication status				2.91	.088
Published	40	1.62***	1.24 - 2.12		
Unpublished	26	1.14	0.84 - 1.54		

Note: *k*= number of comparisons; *OR*= odds ratio; *CI*_{95%}= 95 % confidence interval; *CG*= comparison group; *Q*_{bet}= test of between group differences (χ^2 -distributed with *df*= number of categories - 1).

^a Comparisons are based on identical *CG* in part; therefore, between-group differences could not be tested.

* *p*<.05; *** *p*<.01; **** *p*<.001

$p = .36$). In addition, randomized trials did not differ significantly from non-randomized designs; $Q(1, k = 66) = 0.07, p = .79$. Whether the comparison group consisted of offenders who had refused treatment or not had a small but insignificant influence. Similar to design quality, the length of follow-up time did not correlate with effect size ($r = .00$).

However, other methodological variables had a clear relation to effect size. This was particularly the case for sample size ($r = -.26; p = .03$). This correlation resulted mainly from large effects in trials with very small samples ($N \leq 50$; see Table 2). The recidivism base rate was also significantly related to effect size ($r = .30, p = .01$). Furthermore, studies using self reported recidivism alongside criminal records showed higher effects than studies that relied on official data only.

Another set of variables that can be regarded as methodological in a broader sense refers to descriptive validity (see Lösel & Köferl, 1989), i.e. the quality of study reporting. In particular, a lack of reporting details on the treatment concept and on outcome statistics correlated significantly with effect size ($r = -.33, p < .001$ and $r = -.24, p = .03$).

Although there was no significant difference between published and unpublished studies, a significant mean effect appeared in published studies only.

Moderator effects of treatment features after controlling for confounding variables

Many of the moderator variables presented above are confounded with each other. In order to test whether treatment-specific variables remain significant after controlling for non-specific variables, we conducted a hierarchical regression analysis. In four steps we first entered unspecific and methodological variables, then offender variables, general treatment characteristics, and finally, variables referring to treatment content. In a sample of only 66 studies it was not reasonable to include all moderator variables in the model. Therefore, on each hierarchical step we firstly included those variables that seemed to be important from a theoretical point of view or had proven to be empirically related in the bivariate analyses ($r \geq .20$). Following a stepwise procedure, we then excluded any variable that did not significantly contribute to the

explanation of variance ($p > .10$). Only for the treatment content cluster we added all types of concept and excluded them stepwise as described above. In order to retain the full sample, missing values were plugged by the sample mean (see Cohen & Cohen, 1983).

For the full model there was not much heterogeneity; $Q(55, k = 66) = 65.40, p = .16$. Therefore, the further calculations followed the assumptions of a fixed effects model (see Lipsey & Wilson, 2001). Table 3 shows a summary of the analysis. With 60 %, the full model explains a large proportion of effect size variance ($p < .001$). The largest part of this was due to unspecific features which accounted for almost half of the variance. Nevertheless, on any consecutive step the added variables provided for a significant increase in explained variance. Most notably, even after controlling for an array of possible confounding variables, a cognitive orientation of the treatment program still added significantly to the explanation of effect size variance. Although the increase in explained variance at this step is not large, one must take the conservative nature of this estimate into account: Because of confounding variables the shared variance needs always to be regarded as originating from the lower step variable. This caveat holds similarly for other treatment contents. Rather than suggesting that other treatment methods do not work in sexual offender treatment the present regression analysis gives further evidence that the adoption of a cognitive framework serves to enhance treatment effects independently of other factors.

Discussion

To our knowledge the present study is the most comprehensive meta-analysis on the effects of sexual offender treatment. This is in part due to our inclusion of the non-English literature as well as unpublished trials. Another reason is the recent increase in the evaluation of sexual offender treatment. Almost one third of the integrated studies have been published since 2000. Although Alexander's (1999) study pool is larger than ours, our analysis is restricted to controlled trials. Including a control group is a minimum standard in order to come to valid conclusions regarding the effectiveness of an intervention. However, we did not restrict the analysis to randomized trials that can be considered as the «gold standard» in program evaluation. In fact, most studies are

Table 3
Hierarchical regression with effect size as dependent variable

Cluster of predictor variables	ΔR^2	ΔQ	p
<i>Methodological characteristics</i> Variables: Quality of outcome reporting, Quality of treatment description, Small sample ($N \leq 50$), Treatment refusers as CG, No dropouts in TG	.45	73.21	.000
<i>Offender characteristics</i> Variables: Age homogeneity of TG	.03	4.33	.037
<i>General treatment characteristics</i> Variables: Involvement of authors, Group format, Specific for sexual offenders	.10	15.71	.001
<i>Content of treatment</i> Variables: Cognitive orientation	.03	5.27	.022
<i>Total</i>	.60	98.52	.000

Note: Variable changes in index direction correspond to higher effect sizes.
 ΔQ = Change in the sum of squares with each hierarchical step (χ^2 distributed with df = number of added variables)

less than perfect. 60 % used clearly non-equivalent control groups. Only six studies applied a sound randomization procedure. There are several reasons why we also included non-randomized studies. First of all, there are too few randomized trials on sexual offender treatment with too heterogeneous modes of treatment as to carry out a differentiated analysis. Secondly, not even randomization can fully rule out relevant group differences; see Marques, Wiederanders, Day, Nelson, & von Ommeren (2005) as an example. Thirdly, the effect size of the randomized trials did not differ significantly from the non-randomized evaluations. The non-equivalent comparisons even indicated weaker effects. Therefore, the overall effect seems not to be positively biased by the inclusion of studies with poorer designs. Nevertheless, we should remain critical about the overall positive effect. In order to reach a more definitive answer on the question «Does sexual offender treatment work?» we need more high quality studies.

The positive overall effect corresponds to a reduction in recidivism of about a third. Results are similar for general and violent reoffending. Many sexual offenders are not «specialized» in sexual offending but have a broader range of offenses (Hanson & Morton-Bourgon, 2005). Therefore, the finding that treatment also reduces general and violent recidivism is of some importance. The mean odds ratios for different areas of recidivism correspond to a *d*-value of approximately 0.30 which is in the typical range of the effects found for general offender treatment (Lösel, 1995; J. McGuire, 2002). The mean effect reported by Hanson et al. (2002) for sexual reoffending is lower than ours. One reason for this is that Hanson and colleagues did not include organic treatment which showed the highest effects in our analyses. This is particularly true for surgical castration and it should be noted that after removal of these studies the mean effect for sexual reoffending decreases (*OR*= 1.38, equalling a 24 % reduction in recidivism).

The interpretation of the very high effects of surgical castration is not straightforward. On the one hand, the operated upon offenders are hardly equivalent to the control groups. At least, they have to be regarded as a highly selected and motivated group. This sets them at a lower risk compared to the non-operated upon offenders who often had refused the surgery. Quite probably then, the reported effect is an overestimation. On the other hand, the effects are very homogeneous across the studies and the recidivism rate of the operated upon offenders is quite consistently at about 5 % after relatively long follow-up periods of ten years or more. Apart from the question of effectiveness, ethical considerations have to be taken into account. Without going into details, arguments can be raised for both pro and contra surgical castration (Weinberger, Sreenivasan, Garrick, & Osran, 2005). As for now, one can only say that abandoning this approach without closer inspection may not be in the best interest of the society nor the most serious sex offenders. However, it is necessary to collect more solid knowledge on circumstances and modes that may prove such a treatment to be reasonable.

This is also true for hormonal medication or the so-called «chemical castration» by lowering the testosterone level. This approach did show encouraging results in the present analysis. One advantage over surgical castration is its reversible nature. However, this is a disadvantage at the same time. In the light of possibly serious side effects, non-compliance and treatment dropout are common problems for hormonal treatments and after cessation of the medication, not only do testosterone levels reach

initial levels but also recidivism rates show a marked increase (e.g. Berlin & Meinecke, 1981; Meyer, Cole, & Emory, 1992). One should also consider that hormonal treatment is usually accompanied by psychotherapeutic measures. More over, in our analysis, hormonal treatment was highly confounded with other variables and did not show a significant impact in the hierarchical regression.

Only a cognitive-behavioral treatment orientation consistently demonstrated a positive impact. This effect is based on a relatively solid number of 35 comparisons. For all other treatment approaches the available number of independent findings was much smaller. Still, the pattern of our findings on different modes of treatment fits in well with what is known from general offender therapy. The usually less clearly structured insight-oriented and milieu-therapeutic approaches seem to be of little benefit while highly structured cognitive-behavioral treatment shows good effects (Lösel, 1995; J. McGuire, 2002). However, even in the category of cognitive-behavioral treatment, there is considerable variance in effects. For example, albeit applying this approach and being one of the soundest studies so far, the recently finalized California Sexual Offender Treatment and Evaluation Program (Marques et al., 2005) did not show a positive overall effect.

Type of treatment is only one feature that moderated outcomes. In particular, we found methodological factors to be strongly related to effect size. This cluster of variables explained almost half of the effect size variance. Saying this, a cautionary note seems to be necessary. Applying a hierarchical model that introduced methodological factors first, one may overestimate their influence. In addition, the meaning of methodological variables needs also to be regarded from a treatment perspective. For example, descriptive validity was negatively related to effect size. This may indicate that the respective evaluation referred to a less well-elaborated treatment or to insufficient monitoring of implementation. The latter may also relate to sample size. In larger samples it may be more difficult to control and maintain a thorough treatment implementation. Usually, smaller effects in larger samples are discussed in terms of publication bias (e.g. Light & Pillemer, 1984). However, we found just the same relationship between sample size and effect size in the unpublished studies. Therefore, this relation may be linked to treatment integrity and this to author involvement in implementation. If evaluators actively drive the treatment under study they probably care that it is implemented properly (Petrosino & Soydan, 2005). Actually, our data do indicate that those studies in which measures to ensure treatment integrity have been undertaken show better effects than studies in which implementation problems were obvious. Unfortunately this aspect could be rated only for a few primary studies and we cannot rule out that the difference is due to sampling error. In the literature on general offender treatment implementation issues have been raised repeatedly (Bernfeld, Farrington, & Leschied, 2001; Lösel & Wittmann, 1989; Weisburd, Lum, & Petrosino, 2001). Accordingly, we need more information on treatment implementation and its relationship to effect size in the field of sexual offender treatment. However, the fact that treatments in bigger samples and without author involvement show small or no effects does not allow us to generalize the positive overall results to large-scale implementations in routine practice.

Also other moderator analyses have to be interpreted cautiously, both regarding the moderators we found as well as those we failed to identify. There are several reasons for this:

Firstly, from a statistical point of view, the random model is less sensitive in revealing moderator effects (Overton, 1998). On the other hand, multiple significance testing contains the danger of inflated α -error. Secondly, variables are confounded and it is virtually impossible to disentangle the multiple relationships with regard to their individual causal impacts on effect size (Lipsey, 2003). Thirdly, sometimes the necessary information is not provided in primary studies and analyses cannot be conducted or are based on only a few studies. For example, offender characteristics are probably more influential than our analyses can demonstrate. But the lack of detailed information hampers more differentiated analyses (Lösel, 2001). A clearly offender related factor that shows worse outcomes is treatment dropout. However, one should not reduce dropping out of treatment to a stable offender characteristic but perceive it as a mismatch between offender and therapy or a lack of motivational treatment elements (Beyko & Wong, 2005; McMurrin, 2002). This also seems to be an important point with respect to whether offenders enter treatment voluntarily or not. Effects tend to be better in voluntarily participating offenders. But the difference to more or less coerced treatment was not significant and effect sizes within the categories were heterogeneous. This means that neither voluntary participation is a sufficient condition for a positive development nor does an external treatment enforcement preclude effective

therapy. Obviously, readiness for treatment is a more complex construct and should be aimed at under both conditions (Ward, Day, Howells, & Birgden, 2004).

In summary, our results indicate that sexual offender treatment can significantly reduce recidivism rates. The size of the effect is small to moderate but it is in accord with what we know from the larger research literature on general offender treatment evaluation. However, the evidence is based on studies that mostly apply a weak methodological standard. Restricting the analysis to a few randomized trials shows a comparable mean effect but it does not render it statistically significant. As for now, our results indicate that cognitive-behavioral treatment is promising. Also, hormonal medication seems to improve outcomes. Although promising findings have been reported on other pharmacological treatments (Hill, Briken, Kraus, Strohm, & Berner, 2003), there are as yet no controlled evaluations that show its usefulness with regard to reoffending. Obviously, we need more high quality evaluations on the whole range of sexual offender treatments to come to unequivocal conclusions. This is also true with regard to the descriptive validity of studies. More differentiated reports of the treatment, its implementation, offender characteristics, and the respective outcomes would clearly improve our ability to give a more detailed answer to the question of «What works for whom and under what conditions?».

References

- (Studies included in the meta-analysis are indicated with an asterisk).
- Alexander, M.A. (1999). Sexual offender treatment efficacy revisited. *Sexual Abuse: A Journal of Research and Treatment*, 11(2), 101-116.
- * Allam, J. (1998). *Effective practice in work with sex offenders: A reconviction study comparing treated and untreated offenders*. Birmingham: West Midlands Probation Service Sex Offender Unit.
- * Bakker, L., Hudson, S.M., Wales, D.S., & Riley, D. (1998). «And there was Light»: Evaluating the Kia Marama Treatment Programme for New Zealand sex offenders against children. Christchurch, NZ: Psychological Service, Department of Corrections.
- * Barbaree, H.E., & Seto, M.C. (1998). *The ongoing follow-up of sex offenders treated at the Warkworth Sexual Behaviour Clinic*. Toronto: Centre for Addiction and Mental Health.
- * Barnes, J.M. (2000). *Recidivism in sex offenders: A follow-up comparison of treated and untreated sex offenders released to the community in Kentucky*. Unpublished Dissertation, University of Louisville, Louisville, KY.
- Berlin, F.S., & Meinecke, C.F. (1981). Treatment of sex offenders with antiandrogenic medication: Conceptualization, review of treatment modalities, and preliminary findings. *American Journal of Psychiatry*, 138, 601-607.
- * Berner, W., & Karlick-Bolten, E. (1986). *Verlaufsformen der Sexualkriminalität. 5-Jahres-Katamnesen bei 326 Sexualdelinquenten unter Berücksichtigung von Frühsozialisation, vorausgegangener Delinquenz, psychiatrisch-psychologischer Diagnostik und Therapie*. Stuttgart: Enke.
- Bernfeld, G.A., Farrington, D.P., & Leschied, A.W. (Eds.). (2001). *Offender rehabilitation in practice. Implementing and evaluating effective programs*. Chichester: Wiley.
- Beyko, M.J., & Wong, S.C.P. (2005). Predictors of treatment attrition as indicators for program improvement not offender shortcomings: A study of sex offender treatment attrition. *Sexual Abuse*, 17(4), 375-389.
- * Bluglass, R. (1980). Indecent exposure in the West Midlands. In D. West (Ed.): *Sex offenders in the criminal justice system* (pp. 171-180). Cambridge: Cambridge Institute of Criminology.
- * Borduin, C.M., Henggeler, S.W., Blaske, D.M., & Stein, R.J. (1990). Multisystemic treatment of adolescent sexual offenders. *International Journal of Offender Therapy and Comparative Criminology*, 34, 105-113.
- * Borduin, C.M., & Schaeffer, C.M. (2001). Multisystemic treatment of juvenile sexual offenders: A progress report. *Journal of Psychology and Human Sexuality*, 13(3/4), 25-42.
- * Byrne, S.M. (1999). *Treatment efficacy of a juvenile sexual offender treatment program*. Unpublished Dissertation, Memorial University of Newfoundland, St. Johns, Newfoundland, Canada.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- * Cornu, F. (1973). *Katamnesen bei kastrierten Sittlichkeitsdelinquenten aus forensisch-psychiatrischer Sicht*. Basel: Karger.
- * Craissati, J., Falla, S., McClurg, G., & Beech, A. (2002). Risk, reconviction rates and pro-offending attitudes for child molesters in a complete geographical area of London. *Journal of Sexual Aggression*, 8(1), 22-38.
- * Di Fazio, R., Abracen, J., & Looman, J. (2001). Group versus individual treatment of sex offenders: A comparison. *Forum on Corrections Research*, 13(1), 56-59.
- * Dünkel, F., & Geng, B. (1994). Rückfall und Bewährung von Karrieretätern nach Entlassung aus dem sozialtherapeutischen Behandlungsvollzug und dem Regelvollzug. In M. Steller, K.-P. Dahle & M. Basqué (Eds.): *Straftäterbehandlung: Argumente für eine Revitalisierung in Forschung und Praxis* (pp. 35-74). Pfaffenweiler: Centaurus.
- * Fedoroff, J.P., Wisner-Carlson, R., Dean, S., & Berlin, F.S. (1992). Medroxy-progesterone acetate in the treatment of paraphilic sexual disorders: Rate of relapse in paraphilic men treated in long-term group psychotherapy with or without medroxy-progesterone acetate. *Journal of Offender Rehabilitation*, 18, 109-123.
- Fleiss, J.L. (1994). Measures of effect size for categorical data. In H. Cooper & L.V. Hedges (Eds.): *The handbook of research synthesis* (pp. 245-260). New York: Russell Sage Foundation.
- * Friendship, C., Mann, R.E., & Beech, A.R. (2003). Evaluation of a national prison-based treatment program for sexual offenders in England and Wales. *Journal of Interpersonal Violence*, 18(7), 744-759.

- * Frisbie, L.V. (1969). *Another look at sex offenders in California* (Mental Health Research Monograph No. 12). Sacramento, CA: State of California, Department of Mental Hygiene.
- Furby, L., Weinrott, M.R., & Blackshaw, L. (1989). Sex offender recidivism: A review. *Psychological Bulletin*, *105*(1), 3-30.
- Gallagher, C.A., Wilson, D.B., & MacKenzie, D.L. (2000). A meta-analysis of the effectiveness of sex offender treatment programs. Retrieved 5/23/2001, from <http://www.wam.umd.edu/~wilsondb/papers/sexoffender.pdf>
- Garrido, V., Farrington, D.P., & Welsh, B.C. (2006). The importance of an evidence-based approach in the current Spanish policy for crime prevention. *Psicothema*, *18*, 591-595.
- * Groth, N.A. (1983). Treatment of the sexual offender in a correctional institution. In J.G. Greer & I.R. Stewart (Eds.): *The sexual aggressor* (pp. 160-176). New York: Van Nostrand Reinhold.
- * Guarino-Ghezzi, S., & Kimball, L.M. (1998). Juvenile sex offenders in treatment. *Corrections Management Quarterly*, *2*(1), 45-54.
- Hall, G.C.N. (1995a). Sexual offender recidivism revisited: A meta-analysis of recent treatment studies. *Journal of Consulting and Clinical Psychology*, *63*(5), 802-809.
- * Hall, G.C.N. (1995b). The preliminary development of a theory-based community treatment for sexual offenders. *Professional Psychology: Research and Practice*, *26*(5), 478-483.
- * Hanson, R.K., Broom, I., & Stephenson, M. (2004). Evaluating community sex offender treatment programs: A 12-year follow-up of 724 offenders. *Canadian Journal of Behavioural Sciences*, *36*(2), 87-96.
- Hanson, R.K., Gordon, A., Harris, A.J.R., Marqués, J.K., Murphy, W.D., Quinsey, V.L., et al. (2000). First report of the collaborative outcome data project on the effectiveness of psychological treatment for sex offenders. *Sexual Abuse: A Journal of Research and Treatment*, *14*(2), 169-194.
- Hanson, R.K., & Morton-Bourgon, K.E. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology*, *73*(6), 1154-1163.
- * Hanson, R.K., & Nicholaichuk, T.P. (2000). A cautionary note regarding Nicholaichuk et al. (2000). *Sexual Abuse: A Journal of Research and Treatment*, *12*(4), 289-293.
- * Hanson, R.K., Steffy, R.A., & Gauthier, R. (1992). Long-term follow-up of child molesters: Risk predictors and treatment outcome.
- * Hedderman, C., & Sugg, D. (1996). *Does treating sex offenders reduce offending?* (Research Findings No. 45). London: Home Office Research and Statistics Directorate.
- Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando: Academic Press.
- Hill, A., Briken, P., Kraus, C., Strohm, K., & Berner, W. (2003). Differential pharmacological treatment of paraphilias and sex offenders. *International Journal of Offender Therapy and Comparative Criminology*, *47*(4), 407-421.
- * Huot, S.J. (2002, October). *Recidivism, recidivism, recidivism! An update of several Minnesota recidivism studies*. Paper presented at the 21st Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers on 'Best Practices: Clinical and Research Collaborations', Montréal, Québec, Canada.
- * Janner, J. (1959). Über die Bedeutung psychiatrischer Betreuung verurteilter Sexualdelinquenten. *Schweizerische Zeitschrift für Strafrecht*, *74*, 310-319.
- * Kansas Department of Corrections. (2000). *Offender programs evaluation: Volume IV*. Topeka: Kansas Department of Corrections.
- * Kaul, J., Huot, S.J., Epperson, D., & Dornfeld, M. (1994). *Sex offenders released in 1988* (Unpublished report). St. Paul, MN: Minnesota Department of Corrections.
- * Lab, S.P., Shields, G., & Schondel, C. (1993). Research note: An evaluation of juvenile sexual offender treatment. *Crime and Delinquency*, *39*(4), 543-553.
- * Langevin, R., Paitich, D., Hucker, S.J., Newman, S., Ramsay, G., Pope, S. et al. (1979). The effect of assertiveness training, Provera, and sex of therapist in the treatment of genital exhibitionism. *Journal of Behavior Therapy and Experimental Psychiatry*, *10*, 275-282.
- Light, R.J., & Pillemer, D.B. (1984). *Summing up: The science of reviewing research*. Cambridge: Harvard University Press.
- Lipsey, M.W. (2003). Those confounded moderators in meta-analysis: Good, bad and ugly. *Annals of the American Academy of Political and Social Science*, *587*(May), 69-81.
- Lipsey, M.W., & Wilson, D.B. (2001). *Practical meta-analysis*. Thousand Oaks: Sage.
- * Looman, J., Abracens, J., & Nicholaichuk, T.P. (2000). Recidivism among treated sexual offenders and matched controls: Data from the Regional Treatment Centre (Ontario). *Journal of Interpersonal Violence*, *15*(3), 279-290.
- Lösel, F. (1995). The efficacy of correctional treatment: A review and synthesis of meta-evaluations. In J. McGuire (Ed.): *What works: Reducing reoffending. Guidelines from practice and research* (pp. 79-111). Chichester: Wiley.
- Lösel, F. (2001). Evaluating the effectiveness of correctional programs: Bridging the gap between research and practice. In G.A. Bernfeld, D.P. Farrington & A.W. Leschied (Eds.): *Offender rehabilitation in practice. Implementing and evaluating effective programs* (pp. 67-92). Chichester: Wiley.
- Lösel, F. (2002). ¿Sirve el tratamiento para reducir la reincidencia de los delincuentes sexuales? (Englisches Manuskript: Evaluation of sex offender treatment). In S. Redondo (Ed.): *Delincuencia sexual y sociedad* (pp. 361-396). Barcelona: Ariel.
- Lösel, F., & Köferl, P. (1989). Evaluation research on correctional treatment in West Germany: A meta-analysis. In H. Wegener, F. Lösel & J. Haisch (Eds.): *Criminal behavior and the justice system* (pp. 334-355). New York: Springer.
- Lösel, F., & Schmucker, M. (2003). *The efficacy of sex offender treatment: A brief synthesis of meta-analyses*. Preliminary Report for the Campbell Crime and Justice Group (Updated). University of Erlangen-Nuremberg: Institute of Psychology.
- Lösel, F., & Wittmann, W.W. (1989). The relationship of treatment integrity and intensity to outcome criteria. *New Directions for Program Evaluation*, *42*, 97-108.
- * Lowden, K., Hetz, N., Harrison, L., Patrick, D., English, K., & Pasini-Hill, D. (2003). *Evaluation of Colorado's prison therapeutic community for sex offenders: A report of findings*. Denver, CO: Office of Research and Statistics, Division of Criminal Justice.
- * Maletzky, B.M. (1991). The use of medroxyprogesterone acetate to assist in the treatment of sexual offenders. *Annals of Sex Research*, *4*(2), 117-129.
- * Maletzky, B.M., & Field, G. (2003). The biological treatment of dangerous sexual offenders. A review and preliminary report of the Oregon pilot depo-Provera program. *Aggression and Violent Behavior*, *8*(4), 391-412.
- * Marqués, J. K. (1999). How to answer the question «Does sex offender treatment work?». *Journal of Interpersonal Violence*, *14*(4), 437-451.
- Marqués, J.K., Wiederanders, M., Day, D.M., Nelson, C., & von Ommeren, A. (2005). Effects of a relapse prevention program on sexual recidivism: Final results from California's Sex Offender Treatment and Evaluation Project (SOTEP). *Sexual Abuse: A Journal of Research and Treatment*, *17*, 79-107.
- * Marshall, W.L., & Barbaree, H.E. (1988). The long-term evaluation of a behavioral treatment program for child molesters. *Behaviour Research and Therapy*, *26*, 499-511.
- * Marshall, W.L., Eccles, A., & Barbaree, H.E. (1991). The treatment of exhibitionists: A focus on sexual deviance versus cognitive and relationship features. *Behaviour Research and Therapy*, *26*, 129-135.
- Marshall, W.L., Jones, R.L., Ward, T., Johnston, P., & Barbaree, H.E. (1991). Treatment outcomes with sex offenders. *Clinical Psychology Review*, *11*, 465-485.
- * McConaghy, N., Blaszczyński, A., & Kidson, W. (1988). Treatment of sex offenders with imaginal desensitization and/or medroxyprogesterone. *Acta Psychiatrica Scandinavica*, *77*, 199-206.
- * McGrath, R.J., Cumming, G.F., Livingston, J.A., & Hoke, S.E. (2003). Outcome of a treatment program for adult sex offenders. From prison to community. *Journal of Interpersonal Violence*, *18*(1), 3-17.
- * McGrath, R.J., Hoke, S.E., & Vojtisek, J.E. (1998). Cognitive-behavioral treatment of sex offenders. A treatment comparison and long-term follow-up study. *Criminal Justice and Behavior*, *25*, 203-225.
- McGuire, J. (2002). Integrating findings from research reviews. In J. McGuire (Ed.): *Offender rehabilitation and treatment: Effective programmes and policies to reduce re-offending* (pp. 3-38). Chichester, UK: Wiley.
- * McGuire, T.J. (2000). Correctional institution based sex offender treatment: A lapse behavior study. *Behavioral Sciences and the Law*, *18*(1), 57-71.

- McMurran, M. (Ed.). (2002). *Motivating offenders to change: A guide to enhancing engagement in therapy*. Chichester, UK: Wiley.
- * Meyer, W.J., Cole, C., & Emory, E. (1992). Depo provera treatment for sex offending behavior: An evaluation of outcome. *Bulletin of the American Academy of Psychiatry and the Law*, 20(3), 249-259.
- * Mulloy, R., & Smiley, W.C. (1996, August). *Recidivism and treated sexual offenders*. Paper presented at the International Congress of Psychology, August 16-21, Montreal, Canada.
- * Nicholaichuk, T.P. (1996). Sex offender treatment priority: An illustration of the risk/need principle. *Forum on Corrections Research*, 8(2), 38-41.
- * Nutbrown, V., & Stasiak, E. (1987). Research monograph: A retrospective analysis of O.C.I. cost effectiveness 1977-1981. *Ontario Correctional Institute*, 2, 1-16.
- * Ohio Department of Rehabilitation and Correction. (1996). *Five year recidivism follow-up of sex offender releases* (Unpublished Report). Columbus, OH: Ohio Department of Rehabilitation and Correction.
- * Oregon Department of Corrections. (1997). Outcome evaluation of the Jackson County Sex Offender Supervision and Treatment Program. Retrieved 5/29/01, from <http://www.doc.state.or.us/research/jackrpt2.pdf>.
- * Ortmann, R. (2002). *Sozialtherapie im Strafvollzug: Eine experimentelle Längsschnittstudie zu den Wirkungen von Strafvollzugsmaßnahmen auf Legal- und Sozialbewährung*. Freiburg: Edition iuscrim.
- Overton, R.C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods*, 3(3), 354-379.
- * Perkins, D. (1987). A psychological treatment programme for sex offenders. In B.J. MacGurk & D.M. Thornton (Eds.): *Applying psychology to imprisonment: Theory and practice* (pp. 191-217). London: Her Majesty's Stationery Office Books.
- * Peters, J.J., Pedigo, J., Steg, J., & McKenna Jr., J. (1968). Group psychotherapy of the sex offender. *Federal Probation*, 32, 41-45.
- Petrosino, A., & Soydan, H. (2005). The impact of program developers as evaluators on criminal recidivism: Results from meta-analyses of experimental and quasi-experimental research. *Journal of Experimental Criminology*, 1(4), 435-450.
- * Procter, E. (1996). A five-year outcome evaluation of a community-based treatment program for convicted sexual offenders run by the probation service. *Journal of Sexual Aggression*, 2, 3-16.
- Quinsey, V.L., Harris, G.T., Rice, M.E., & Lalumière, M.L. (1993). Assessing treatment efficacy in outcome studies of sex offenders. *Journal of Interpersonal Violence*, 8, 512-523.
- * Rasmussen, L.A. (1995). *Factors related to recidivism among juvenile sexual offenders*. Unpublished Dissertation, The University of Utah.
- * Rattenbury, F.R. (1985). *The outcomes of hospitalized and incarcerated sex offenders: A study of offender types, recidivism rates, and identifying characteristics of the repeat offender*. Unpublished Dissertation, Loyola University of Chicago, Chicago, IL.
- * Rice, M.E., Harris, G.T., & Quinsey, V.L. (1993). Evaluating treatment programs for child molesters. In J. Hudson & J.V. Roberts (Eds.): *Evaluating justice: Canadian policies and programs* (pp. 189-203). Toronto: Thompson.
- * Robinson, D. (1995). *The impact of cognitive skills training on post-release recidivism among Canadian federal offenders* (Research Report No. R-41). Ottawa, Ontario, Canada: Correctional Service of Canada.
- * Romero, J.J., & Williams, L.M. (1983). Group psychotherapy and intensive probation supervision with sex offenders: A comparative study. *Federal Probation*, 47(4), 36-42.
- * Ruddijs, F., & Timmerman, H. (2000). The Stichting ambulante preventie projecten method: A comparative study of recidivism in first offenders in a Dutch outpatient setting. *International Journal of Offender Therapy and Comparative Criminology*, 44(6), 725-739.
- * Schmid, P. (1988). *Was geschieht mit den Sexualstraftätern in der Psychiatrie? Darstellung und Bewertung der psychiatrischen Behandlung von Sexualstraftätern im Psychiatrischen Landeskrankenhaus Bad Schussenried in den Jahren 1978-1987*. Unpublished Dissertation, Universität Tübingen, Tübingen.
- Sherman, L., Gottfredson, D., MacKenzie, D., Eck, J., Reuter, P., & Bushway, S. (1997). *Preventing crime: What works, what doesn't, what's promising. Report to the U.S. Congress*. Washington, DC.
- * Smith, M.A. (2000). *The impact of prison-based drug treatment on the timing and risk of reincarceration of sexual offenders*. Unpublished Dissertation, Florida State University, Tallahassee, FL.
- * Song, L., & Lieb, R. (1995). *Washington State sex offenders: Overview of recidivism studies* (No. 95-02-1101). Olympia, WA: Washington State Institute for Public Policy.
- * Stalans, L.J., Seng, M., Yarnold, P., Lavery, T., & Swartz, J. (2001). *Process and initial impact evaluation of the Cook County Adult Probation Department's Sex Offender Program: Final and summary report for the period of June, 1997 to June, 2000*. Chicago, IL: Illinois Department of Corrections.
- * Stürup, G.K. (1953). Sexual offenders and their treatment in Denmark and other Scandinavian countries. *International Review of Criminal Policy*, 4, 1-19.
- * Stürup, G.K. (1968). Treatment of sexual offenders in Herstedvester Denmark: The rapists. *Acta Psychiatrica Scandinavica Supplement*, 204, 1-62.
- * Taylor, R. (2000). *A seven-year reconviction study of HMP Grendon Therapeutic Community* (Research Findings No. 115). London: Home Office Research and Statistics Directorate.
- Ward, T., Day, A., Howells, K., & Birgden, A. (2004). The multifactor offender readiness model. *Aggression and Violent Behavior*, 9(6), 645-673.
- Weinberger, L.E., Sreenivasan, S., Garrick, T., & Osran, H. (2005). The impact of surgical castration on sexual recidivism risk among sexually violent predatory offenders. *Journal of the American Academy of Psychiatry and the Law*, 33(1), 16-36.
- * Wille, R., & Beier, K.M. (1989). Castration in Germany. *Annals of Sex Research*, 2, 103-133.
- Weisburd, D., Lum, C.M., & Petrosino, A. (2001). Does research design affect study outcomes in criminal justice? *Annals of the American Academy of Political and Social Science*, 578, 50-70.
- Wilson, D.B. (2001). Meta-analytic methods for criminology. *Annals of the American Academy of Political and Social Science*, 578, 71-89.
- * Worling, J.R., & Curwen, T. (2000). Adolescent sexual offender recidivism: Success of specialized treatment and implications for risk prediction. *Child Abuse & Neglect*, 24(7), 965-982.
- * Ziethen, F. (2002). *Rückfallpräventive Effizienz der sozialtherapeutischen Behandlung von Sexualstraftätern. Evaluation der Sozialtherapie in der JVA Berlin-Tegel*. Unpublished Diplomarbeit, Freie Universität Berlin.