Protocol for a Systematic Review of the Effects of Sexual Offender Treatment

Friedrich Lösel and Martin Schmucker

Institute of Psychology
University of Erlangen-Nuremberg
Germany

Submission to the Campbell Collaboration
Crime and Justice Group

March 2003

1. COVER SHEET

Title

Effects of Sexual Offender Treatment

Reviewers

Friedrich Lösel and Martin Schmucker
Institute of Psychology
University of Erlangen-Nuremberg
Bismarckstr. 1
91054 Erlangen
Germany
Tel: +49-9131-852-2330 or -2333
Fax: +49-9131-852-2646
Email: fhloesel@phil.uni-erlangen.de
mnschmuc@phil.uni-erlangen.de

Sources of Support

Social Science Research Center of the University of Erlangen-Nuremberg
German Federal Government (Applied)

2. BACKGROUND

The effectiveness of sexual offender treatment is currently a very urgent topic in many countries. Media reports on extremely serious cases of child abuse and related murder raised strong public concern about this issue. The respective crime policies contain both measures of incapacitation and offender treatment as means of protecting potential victims and preventing recidivism. However, the efficacy of both approaches are questionable. Incarceration alone only protects as long as the offenders can be locked up according to legal rules of proportional sanctioning. Measures of sex offender treatment are particularly questioned by the general public because of extreme single cases of serious recidivism of treated sex offenders.

However, in the scientific community the effectiveness of sexual offender treatment has been also a matter of controversy. In an early systematic review of sexual offender recidivism, Furby, Weinrott and Blackshaw (1989) found a lack of methodologically sound treatment evaluations. As a consequence, the authors concluded that there was no evidence that sexual offender treatment reduces recidivism. Later narrative reviews took a more positive view (e.g., Marshall, Jones, Ward, Johnston, & Barbaree, 1991). However, this was questioned by other authors for methodological reasons, i.e. the lack of randomized designs (Quinsey, Harris, Rice, & Lalumière, 1993). When a systematic review from the Cochrane Collaboration applied the latter criterion, only three studies on very heterogeneous modes of treatment were included (White, Bradley, Ferriter, & Hatzipetrou, 1998). Thus, the authors could not make any generalizations for policymaking and practice except to state a serious lack of internally valid evaluations. In contrast to this rather restrictive approach, other authors emphasized that it is rarely possible to implement *and* maintain a randomized design with no threats to internal validity in a field such as the treatment of sexual offenders (Barbaree, 1997; Lösel, 2001; Marshall, 2002). Particularly in cases of serious sex offenders, legal and practical reasons make it difficult to form randomized untreated control groups. Therefore, policymaking and practice cannot wait until an unrealistic gold standard evaluation study but must be informed about the results of sub-optimal, quasi-experimental studies. This is in line with the notion that methodological quality is not a unidimensional and all-or-none category (Cook & Shadish, 2001; Cronbach et al., 1980; Lösel & Köferl, 1989).

Accordingly, a number of meta-analyses on quasi-experimental evaluations of sex offender treatment have been published since the mid-1990s (Alexander, 1999; Gallagher, Wilson & MacKenzie, 2000, Grossman, Martis, & Fichtner, 1999; Hall, 1995, Hanson, et al., 2001; Lösel, 2000; Polizzi, MacKenzie & Hickman, 1999). All meta-analyses found a lower average rate of sexual recidivism in the treated groups than in the control groups. However, a systematic comparison of the various syntheses revealed not only substantial differences in the overall effect sizes but also in the type of treatment included, prevailing design quality, type of outcome measures, mean follow-up periods, categorization of programs, treatment settings, publication language, meta-analytical techniques, and other features (Lösel & Schmucker, 2001). Although the more current cognitive-behavioral/relapse prevention programs and hormonal treatment showed encouraging results (e.g., Hall, 1995; Hanson et al., 2001), the pattern of general and moderator effects was only partially consistent. Insofar, an actualized systematic review is needed. In comparison to the recent and relatively sound meta-analyses, it should

(a) be more comprehensive with respect to modes of treatment, in particular by addressing both psychosocial and pharmacological treatment;

(b) include studies that are not only published in English language;

(c) be more restrictive with respect to the methodological quality of primary studies and focus on the best empirical evidence;

(d) include also the most studies that were previously not yet available.

3. OBJECTIVES

This review will survey the existing empirical evidence of the effectiveness of treatment and management programs for sexual offenders on recidivism. Sex offenders are persons who have been previously arrested, convicted or incarcerated for a sexual offense such as child molesting, rape, or exhibitionism. Although the focus is on sexual offenses, the respective delinquents could have committed also other crimes (no delict specialization). The main target of outcome is sexual recidivism. However, because many sex offenders also engage in non-sexual violence, property, and other crimes, these categories will also be regarded if available. Sex offender treatment programs are broadly defined as all explicit, and systematic interventions that intend to reduce sexual reoffending. This includes psychosocial

interventions such as psychotherapy and relapse prevention programs as well as biological treatments such as hormonal drug therapy or castration. Because hormonal treatment is often combined with psycho-social measures, separated meta-analyses on both areas would not reveal the complex picture of intervention in practice. The review will address three questions:

1. What are the general effects of sexual offender treatment programs on reoffending?
2. Are there differential effects that relate to
(a) general study characteristics (e.g., recency of programs, country of origin),
(b) methodological characteristics (e.g., internal validity of the design, length of follow-up),
(c) treatment characteristics (e.g., type of program, treatment intensity),
(d) offender characteristics (e.g., offense type, age).
3. Where are the main knowledge gaps and research deficits according to the descriptive characteristics of the existing evaluations?

4. METHODS

Criteria for inclusion and exclusion of studies

Sex offender treatment. The studies must address some sort of intervention on convicted sex offenders as defined above. Treatment of sexual disorders in general or of individuals who are at risk of sexual offending will not be included. There is no restriction with respect to intervention type. Thus, all forms of psychotherapy, behavior modification, education, and organic treatment will be considered as long as they aim to have impact on sexual offending.

Publication. Eligible studies may be published or unpublished. There will be no limit with respect to the time of appearance. Unpublished studies that are already accepted for publication will later be coded as published studies.

Country of origin. Studies may be conducted in any country and be published in any language. However, due to limited financial resources, only studies published in English, French or German language will be fully analyzed in a first round of the review.

Methodological quality. Only studies using a control group design will be eligible. The treatment and control groups must be randomly or nonrandomly assigned. Nonrandom assignment must either contain a matching procedure or a post-hoc comparison that provides evidence for initial equivalence of both groups on type of sex offending (e.g., rape), prior offense history (e.g., number of convictions for sex and other offenses), relevant demographic variables, (e.g., age), and/or clinical offender characteristics (e.g., paraphilia). A score of 3 or above on an adaptation of the Maryland Scale of Methodological Rigor (Sherman et al., 1997) must be reached for eligibility. We adapted the Maryland Scale for our purposes. Our adaptation (see coding section) is slightly stricter but a little more differentiated at the upper end of the scale that is of special interest for the synthesis of methodologically sound studies. We use the following categories of the Maryland Scale:

(1) No control or comparison group;
(2) Non-equivalent control or comparison group: Differences on relevant variables effecting recidivism are reported or are to be expected (e.g., subjects who refuse treatment);
(3) Control or comparison group with no reason for assuming subject differences related to recidivism between groups or sound statistical control of such differences;
(4) Systematic strategy to reach equivalence of control or comparison group (e.g., individual matching procedure basing on relevant variables) or comprised random assignment;
(5) Random assignment of treated and untreated subjects without obvious differences between groups.

Comparison subjects may have been drawn from groups that received no treatment, placebo, treatment "as usual," or measures of offender rehabilitation that are unspecific for sexual offenders. Waiting-list control groups will also be included if the design allows for testing a program effect (see outcome measures).

Outcome measure. Studies have to report a follow up regarding recidivism as an outcome measure. Studies that only report outcome measures of institutional adaptation, behavior ratings of therapists, and so forth will not be included. Recidivism measures must use the

same sources and definition for both the treatment and comparison group as well as a comparable follow-up period.

Search Strategy for Identification of Relevant Studies

Basing on previous literature searches (Lösel, 2000; Lösel & Schmucker, 2001) we already have gathered 18 studies that meet our inclusion criteria. Although more than 200 studies on sex offender treatment have to be scanned, we expect not more than 25-30 studies that will fulfill our eligibility criteria. The strategy of our more expanded literature search will use the following strategies:

1. Literature databases. Although there is normally much overlap between databases, we will analyse the following systems:

- Cochrane Controlled Trials Register
- MEDLINE
- DARE (Database of Abstracts of Reviews of Effectiveness; medicine)
- ERIC
- PsychoInfo/PsychLit
- Psyndex (a German database of psychological publications)
- Dissertation Abstracts International/Online
- SPECTR 2
- SCI, SSCI
- NCJRS (The National Criminal Justice Reference Service)
- NCCAN (National Clearinghouse on Child Abuse and Neglect)
- PAVNET (database on US federally funded violence-related research)
- UK National Health Service National Research Register (NRR)
- Center for Sex Offender Management (US) documents database
- Social Services Abstracts
- Social Work Abstracts
- JURIS (a German database on research in the legal system)

Because controlled evaluations of sex offender treatment is a relatively new topic, there will be no restrictions with respect to the beginning of the time frame; 2002 will be the last date.

The following terms will be used to search these databases: (sex* or paraphil* or rape or rapist or molest* or exhibitionis* or voyeur* or pedophil* or incest* or fetish* or necrophil* or frotteur*) and (offen* or crim* or delinquen* or perpetrator* or prison*) and (treat* or therapy or psychotherapy or intervention or training or correction* or rehabilitation or prevention or management) and (evaluation or evaluate or evaluated or outcome or outcomes or effect or effects or effectiveness or impact or success* or recidivism or re-offen* or reoffen* or recurrence or follow-up or followup or relapse). The terms will have to be adapted to fit the peculiarities of individual databases. E.g. searches in the PAVNET database do not allow for such complex combinations or the search in Psyndex, a German database, will naturally afford the use of German terms.

2. Existing research reviews. The reference sections of retrieved reviews dealing with sexual offender treatment will be scanned for further studies (e.g. Furby, Weinrott, & Blackshaw, 1989; Gallagher, Wilson, & MacKenzie, 2000; Hall, 1995; Hanson et al., 2002; Lösel, 2000; Polizzi, MacKenzie, & Hickman, 1999; White et al., 1999).

3. Primary studies. The snowball system will be applied to find further studies cited in already retrieved evaluations. For reasons of time and resources this will only be done for those studies that meet the inclusion criteria for our review.

4. Internet search. An internet search will be performed to retrieve unpublished studies and materials in progress. To our knowledge there are no systematic guidelines on how to perform a most efficient internet search. However, according to our experiences, the Google search engine already revealed relevant materials. We will locate specific sites and use links to find sites that deal with sexual offender treatment. A focus will be on institutional sites that promote correctional treatment (e.g. the Correctional Service of Canada, U.S. State Departments for Corrections, UK Home Office etc.) and sites that specifically deal with sexual offending (e.g. Center for Sex Offender Management). Unfortunately, the WWW is

not a systematically built up database but rather a collection of more or less systematically linked information. Thus, we cannot rule out missing spots that might be of interest.

5. Personal contacts. To compensate for eventual deficits of all other search strategies, we will contact leading researchers in the field of sexual offender treatment. This is expected to be especially useful to tap studies from Non-English-speaking countries that may not show up in international databases. In addition studies in progress with useful outcome data may be retrieved by this strategy.

Whether located studies will be included depends on a three step procedure of decision making: First the title will be evaluated. If it is obvious that the study does deal with another topic it will be discarded. In a second step, abstracts will be screened to determine if the study reports evaluative data on sex offender treatment and whether it contains a comparison group. The remaining studies will be fully analyzed and the methodological quality be rated using the Maryland Scale.

The search procedure will be carried out by the second author. He will apply a rating scale that contains three categories: (1) Clearly fitting the inclusion criteria; (2) Ambiguous; (3) Clearly not fitting the inclusion criteria. In cases of the second category, the first author will independently analyze the study using the three categories. If this leads to a rating of 3, the study will be excluded. In all other cases, the studies will be discussed by both authors until agreement is reached.

Typical Methods used in the Primary Studies

Most evaluation studies regarding sex offender treatment use quasi-experimental designs. Although randomized trials do exist (e.g. Borduin et al., 1990; Borduin & Schaeffer, 2001; Marques, 1999; Romero & Williams, 1983), they remain rare in this field of research. Usually treatment groups are compared to incidental comparison groups from other facilities, periods before the implementation of a sex offender program or national samples that are matched for crime-related characteristics. Outcomes are mostly presented as sexual reoffending or reconviction rates according to national registers. In some studies, additional information in

further offending is reported. Follow-up periods vary considerably in and between studies. Studies on specific sexual offense behaviors are rare and differential outcome data for different types of sex offenders are a minority. These and other specific issues of sex offender treatment evaluation have to be taken into account when planning both the analysis of general effects and of potential different outcomes. A typical research example may illustrate the situation:

Worling and Curwents (2000) evaluated a community-based program for adolescent sex offenders. Participants in the treatment program were counted as members of the treatment group as long as they did not drop-out before twelve months of treatment. The comparison group consisted of 46 adolescents who were referred for assessment only, 27 wh odropped out of the program before completion of twelve months and 17 who refused treatment. Of the 90 comparison subjects, 67% received some form of treatment elsewhere. Subjects were not randomly assigned to treatment or comparison conditions. A pretreatment assessment on measures linked to sexual and non-sexual recidivism revealed no significant differences between the treatment group and the different comparison groups. Criminal charges regarding sexual, violent non-sexual, and nonviolent offenses based on a national registry were used as outcome measures with a mean follow-up time of approximately six years. Recidivism rates were separately reported for the whole comparison group and the various subgroups. In addition, a survival analysis was conducted, however, χ^2-statistics were reported only for the comparison groups as a whole.

This is a study at the lower end of our eligibility criterion of methodological quality (approximately point 3 on the adapted Maryland Scale). The final inclusion will probably become ambiguous. If this study will finally be included in our meta-analyses, we will only use the data of the 46 assessed subjects who did not receive the treatment as a comparison. The drop-outs will be regarded as treated participants of the program. Refusers will be completely rejected from analyses. The data from the survival analysis will be used for computing effect sizes for different times of follow-up (if sufficient comparable data from other studies are available).

Criteria for Determination of Independent Findings

As a general principle, our analysis intends as much independence between the various data as possible. In order to be regarded as an independent outcome, treated subjects have to be unique to one primary study. Thus, an overlap of subjects between various studies will lead to the exclusion of all but one study. In such cases, the report with the largest sample, the most typical outcome measure, and the longest follow-up period will be selected for our analysis.

Some studies will report various outcome measures. If a study reports outcome measures on sexual recidivism as well as violent, non-sexual or other recidivism, the respective effect sizes will be analyzed separately. However, only the measure on sexual recidivism will be used for the estimation of general treatment efficacy. If there are various indicators of sexual recidivism such as rearrest or reconviction, the respective effect sizes will be coded separately to allow for differential analyses. For the estimation of the general treatment effect, however, only one study effect size will be used. If the various outcomes do not differ significantly, we will compute an average effect size as study effect. If the various outcomes are heterogeneous, we will select the outcome measure that is most frequently represented in the other studies of our data base. In case of multiple follow-up periods reported or study updates the data referring to the longest follow up period will be used to determine the respective study effect size. In cases of strong differences in attrition rates or other factors that might invalidate the results, a previous time of follow-up may be chosen. However, the data on the longest period will be regarded in differential analyses on the impact of the time of follow up on effect size.

Details of Study Coding Categories

Studies will be coded by two reviewers and interrater agreement will be analyzed for the whole study sample. As indicators of reliability/objectivity we will compute both percentages of interrater agreement and more conservative Kappa coefficients. In cases of continuous quantitative categories, we will further use correlation coefficients and – for reasons of comparability – percentages based on classes of scores.

Although the content of treatment is a particularly important issue in offender treatment, it is often only vaguely described (Lösel & Köferl, 1989). Therefore, if necessary, we will try to get additional information directly from the authors. Similarly, there is often not enough information on the conditions in the control groups, although these are as relevant for effect size as the treatment conditions (Lösel, 2002). Accordingly, we will focus also on this issue and will code some categories of descriptive validity as important study characteristics.

Our coding scheme contains the following categories:

General study characteristics:

- Publication type (book, chapter, journal article, dissertation, thesis, unpublished report, conference paper, internet presentation, other)
- Year of publication
- Country of origin (US, Canada, UK, German speaking country, Other)
- Profession of senior author
- Confounding/Dependency of evaluation (Had the authors carried out the treatment themselves? yes, no)

Sample

- Original sample size (treatment, control)
- Drop-outs (treatment, control)
- Age (mean; treatment, control)
- Age group (adult, adolescent, mixed)
- Homogeneity of age (large, medium, low)
- Offense types (separately yes-no-coded: rape, child molestation, incest, exhibitionism, other)
- Offender risk (low, high, medium/mixed/unknown; rated as described by the authors or [if not available] roughly according core items of sex offender risk assessment instruments such as the SVR-20 or RASOR)
- Referral to treatment (voluntary, non-voluntary, both)
- Source of referral (court, other institution, both, none)

Study design

- Unit of assignment (individual, facility, regional, other)
- Type of control group (volunteered for treatment, treatment not offered at that time or at that place/region, not suitable for treatment, refused treatment, other)
- Assignment to treatment (random without matching or stratification, random after individual matching or stratification, non-random with individual post hoc matching, available sample with equivalence in relevant characteristics)
- Group differences (negligible differences, some relatively unimportant differences, some differences judged to be important, not tested)
- Maryland Scale Rating of general methodological quality (see above)

Treatment

- Treatment elements used (each judged 0 = no, 1 = few, 2 = some among others, 3 = mainly)
  - cognitive-behavioral
  - behavioral
  - relapse-prevention
  - insight oriented
  - systemic
  - therapeutic community (low structured)
  - therapeutic community (structured)
  - psycho-educational
  - hormonal or otherwise medicinal
  - surgical castration
- Overall treatment approach (which of the above provides the main framework for the treatment and is the main treatment approach used)
- Specificity of sex offender treatment (yes, no)
- Setting (prison, hospital in-patient, out-patient)
- Provider (state/public, private)
- Treatment length (average)

- Frequency of sessions (per week)
- Treatment intensity (5-point scale integrating length and frequency)
- Format of treatment (individual only, group only, mainly individual, mainly group, both)
- Aftercare (obligatory, optional, not offered)
- Intensity of aftercare (5-point scale)
- Combination with community surveillance measures (5-point scale)
- Time of treatment (First year of treatment)
- "Treatment" of comparison group (non, judged as ineffective/Placebo, unspecific, psycho-educational, psychotherapeutic, hormonal/surgical, other)
- Treatment surplus of treatment group in relation to comparison group (5-point scale)
- Integrity of treatment implementation (5-point scale)
- Treatment providers/therapists (medical doctors, psychologists, social workers, other psychosocial professionals, prison officers, nurses, others; more than one category)
- Researcher-monitored model project or routine practice (yes or no)

Follow up

- Treatment group (average in months)
- Control group (average in months)
- Start of follow-up period (beginning of treatment, end of treatment, time at risk)
- Drop-outs (reported separately, counted as part of treated group, counted as part of comparison group, partly counted as treated partly as untreated, drop-outs not included in analysis, no drop-out subjects)

Outcome data

- Source (official registries, additional sources)
- Type of recidivism (sexual, violent, any sexual or violent, non-sexual, neither sexual nor violent, any recidivism, including parole violations)
- Definition of recidivism (rearrest, reconviction, new charge, inappropriate behavior/lapse)
- Sample sizes for ES computation (treatment, control)

- Effect size measure
  - ES calculation based on (proportions/frequencies, M & SD, test statistics, p-values, other)
  - Confidence in ES calculation (highly estimated, moderate, some, slight, no)

Descriptive validity

- Description of treatment concept (3-point scale; Lösel & Köferl, 1989)
- Assessment of treatment realization (3-point scale)
- Assessment of treatment goals (3-point scale)
- Data on integrity (3-point scale)
- Data on elementary statistics (3-point scale)
- Overall transparency of report (5-point scale)

Statistical Procedures and Conventions

Drop-out subjects will be counted as part of the treatment group. Although this is a very conservative strategy in cases that only got a short amount of intervention, the aim of a sex offender treatment should not only be to prevent recidivism for those who stay in a program but to motivate those who started treatment to complete it. However, if detailed outcome data for drop-outs are reported, we will perform additional comparisons for both dropouts and regular program completers. These analyses will not be used for the general estimation of treatment effectiveness.

Usually, recidivism measures are reported as dichotomous data (i.e. recidivism: yes or no). We will therefore use odds ratios as effect size estimates (Fleiss, 1994). To avoid the problem of zero cell frequencies in contingency tables, 0.5 will be added to all cell frequencies in such cases. For statistical reasons analyses will be conducted on the natural log of the odds ratios (see Lipsey & Wilson, 2001).

In case of continuous data (e.g. number of offenses) the data will be dichotomized (recidivism: yes or no). If the data reported to not allow dichotomization (e.g. descriptive

statistics or test of significance reported only) and simple recidivism data is not provided, Cohen's d will be estimated and the respective value transformed to a logged odds ratio. This conservative strategy will yield somewhat lower effect size estimates but is expected for very rare cases only. Results will be reported as original odds ratios. Other effect size measures will be presented as well because these may be more familiar for practitioners and policy makers.

Effect sizes will be analyzed for heterogeneity. According to our previous analyses, the random/mixed effects model seems to be more adequate than the fixed model of computation (see, Lipsey & Wilson, 2001). Accordingly, studies will be weighted by the inverse standard error of the logged odds ratio and an additional residual random variance component that remains after controlling for moderator variables. Mean effect sizes and corresponding confidence intervals will be reported. ANOVA as well as regression analyses will be used to identify moderator variables and the impact of specific study features on study effects.

Statistical analyses will be carried out using SPSS (Version 10) and macros written by D. B. Wilson (Lipsey & Wilson, 2001) for meta-analytic procedures.

Treatment of qualitative research

Qualitative studies on sex offender treatment will not be regarded in our review. However, qualitative information in the selected studies may be used for ratings of treatment/control content, program integrity, and descriptive validity (see above).

5. TIME FRAME

Because a part of the search procedure and coding is already done or in progress, the time frame can be relatively narrow. The following table shows our monthly schedule after acceptance of the protocol:

	Months after acceptance of protocol									
	1	2	3	4	5	6	7	8	9	10
Adaptation of coding scheme	X	X								

Test of coding procedure		X	X							
Revision of coding scheme			X							
Search for further studies		X	X	X						
Final decision on inclusion				X	X					
Coding of included studies				X	X	X				
Data analysis						X	X			
Writing of draft report							X	X		
Review by C2 advisers									X	
Final report										X

6. PLANS FOR UPDATING THE REVIEW

The first version of our meta-analysis will include studies that are published in English and German language. Although our current literature search does not suggest that there will be many eligible studies in other languages, we are planning a more comprehensive review 12 months after the first report. This research synthesis will be updated every 3 years thereafter.

7. ACKNOWLEDGEMENTS

Preparation of this protocol was supported by grants from the Social Science Research Center and the Institute of Psychology of the University of Nuremberg. We are currently applying for further financial support for translating studies that are not reported in English and German language (Federal Ministry of Justice).

8. STATEMENT CONCERNING CONFLICT OF INTEREST

Although we have published on general offender treatment and sex offender treatment, we are not engaged in the development or evaluation of any specific sex offender treatment program. Therefore, there is no conflict of interest.

9. REFERENCES

Borduin, C. M., & Schaeffer, C. M. (2001). Multisystemic treatment of juvenile sexual offenders: A progress report. *Journal of Psychology and Human Sexuality, 13*, 25-42.

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.

Furby, L., Weinrott, M.R., & Blackshaw, L. (1989). Sex offender recidivism: A review. *Psychological Bulletin, 105*, 3-30.

Gallagher, C.A., Wilson, D.B., & MacKenzie, D.L. (2000). *A meta-analysis of the effectiveness of sex offender treatment programs*. Paper retrieved from http://www.wam.umd.edu/~wilsondb/papers/sexoffender.pdf

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.

Hanson, R.K., & Bussière, M.T. (1998). Predicting relapse: A meta-analysis of recent offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66*, 348-362.

Hanson, R.K., Gordon, A., Harris, A.J.R., Marques, J.K., Murphy, W., Quinsey, V.L., & Seto, M.C. (2001). *First report of the Collaborative Outcome Project on the Effectiveness of Psychological Treatment for Sex Offenders.* Unpublished paper, June 2001.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Lipsey, M.W., & Wilson, D.B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Looman, J., Abracen, 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*, 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* (pp. 79-111). Chichester: Wiley.

Lösel, F. (2000). The efficacy of sexual offender treatment: A brief review of German and international evaluations. In P. J. van Koppen & N. Roos (Eds.), *Rationality, information and progress in law and psychology* (pp. 145-170). Maastricht, NL: Metajuridica.

Marques, J. K. (1999). How to answer the question "Does sex offender treatment work?" *Journal of Interpersonal Violence, 14,* 437-451.

Marshall, W.L., Jones, R., Ward, T., Johnston, P., & Barbaree, H.E. (1991). Treatment outcome with sex offenders. *Clinical Psychology Review, 11*, 465-485.

Polizzi, D.M., MacKenzie, D.L., & Hickman, L.J. (1999). What works in adult sex offender treatment? A review of prison- and non-prison-based treatment programs. *Journal of Offender Therapy and Comparative Criminology, 43*, 357-374.

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.

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.

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.

Worling, J.R., & Curwen, T. (2000). Adolescent sexual offender recidivism: success of specialized treatment and implications for risk prediction. *Child Abuse & Neglect, 24*, 965-982.