
Evaluating Community Sex Offender Treatment Programs: A 12-Year Follow-Up of 724 Offenders

R. KARL HANSON, Public Safety and Emergency Preparedness Canada
IAN BROOM, Carleton University
MARYLEE STEPHENSON, CS/RESORS Consulting Ltd.

Abstract

Although some studies suggest positive effects of treatment for sexual offenders, most studies have been hampered by the unknown influence of selective attrition (e.g., volunteers and drop-outs). In the 1980s, the Correctional Service of Canada began to require weekly community treatment sessions for all sex offenders released in the Pacific Region. This policy change provided a unique opportunity for comparing an unselected cohort of treated sex offenders ($n = 403$) to an untreated cohort ($n = 321$) released in earlier years. After an average 12-year follow-up period, no differences were observed in the rates of sexual (21.1% vs 21.8%), violent (42.9% vs. 44.5%) or general (any) recidivism (56.6% vs 60.4%) for treated and untreated groups, respectively. The outcome remained comparable after controlling for length of follow-up, year of release, age, and seven static risk factors coded from official criminal history records. Retrospective ratings of the treatment quality also showed no relationship to observed recidivism rates. The static risk factors coded in the current study accounted for considerable variance in recidivism and could easily be used to improve statistical controls in future evaluations.

Résumé

Même si certaines études laissent croire à des effets positifs du traitement des délinquants sexuels, la plupart des études ont été retardées par l'influence inconnue de l'attrition sélective (p. ex., les volontaires et les décrocheurs). Au cours des années 1980 le Service correctionnel du Canada a commencé à exiger des séances de traitement communautaire pour tous les délinquants sexuels mis en liberté dans la région du Pacifique. Ce changement de politique nous a donné une occasion unique de comparer une cohorte non sélectionnée de délinquants sexuels traités ($n = 403$) avec une cohorte non traitée ($n = 321$) libérée au cours des années antérieures. Après une moyenne de suivi sur une période de 12 ans, aucune différence n'a été observée dans les taux de récidive sexuelle (21,1 % par opposition à 21,8 %), de récidive avec violence (42,9 % par opposition à 44,5 %) ou de récidivisme

général (de n'importe quelle nature) (56,6 % par opposition à 60,4 %) pour les groupes traités et non traités respectivement. Le résultat est demeuré comparable après avoir contrôlé la longueur de suivi, l'année de libération, l'âge et sept facteurs de risque statiques codés à partir des dossiers criminels officiels. Une évaluation rétrospective de la qualité du traitement n'a également montré aucun rapport avec les taux de récidivisme observés. Les facteurs de risque statique codés dans l'étude actuelle justifiaient l'écart considérable dans le taux de récidivisme et pourraient facilement être utilisés pour améliorer les contrôles statistiques dans les évaluations futures.

Providing treatment for sexual offenders is controversial. Narrative reviews all stress the weakness of existing studies (Furby, Weinrott, & Blackshaw, 1989; General Accounting Office, 1996; MacKenzie & Hickman, 1998), precluding strong conclusions regarding program effectiveness. The central problem concerns potential differences between the treatment and comparison groups. The standard method for minimizing differences is to randomly assign offenders to treatment and no-treatment groups; such designs, however, are difficult to implement and sustain in criminal justice settings. On average, the observed sexual recidivism rate for untreated sex offenders is approximately 15% after 5 years and 20% after 10 years (Hanson & Bussière, 1998; Hanson & Thornton, 2000). When long follow-up periods are required, there is ample opportunity for the research design to be corrupted (e.g., "untreated" offenders receive treatment, administrative support collapses). Consequently, most sex offender treatment outcome studies were not initially designed as such; instead, they have taken advantage of "natural experiments."

The following examples illustrate the challenges faced when comparison groups are not randomly assigned. One of the early influential studies was Marshall and Barbaree's (1988) evaluation of their

community treatment program for child molesters. All of the offenders were assessed at Marshall and Barbaree's clinic and all initially expressed interest in receiving treatment. The comparison group included those who subsequently decided against attending the program (e.g., lived too far away, incarcerated, no longer interested). Based on official criminal records as well as informal reports from police and child welfare agencies, the sexual recidivism rate was 13.2% for the treatment group (9/68) compared to 34.5% for the comparison group (20/58) after an average 3.5 years follow-up period. Because the offenders themselves made the decision whether to attend treatment or not, critics could argue that the differences in recidivism rates were due to pre-existing differences between the groups (e.g., lifestyle instability).

Another approach to examining treatment effects is to compare offenders released before and after the implementation of a treatment program. For example, Proctor (1996) used a cohort design that compared all the sex offenders (rapists, exhibitionists, child molesters) who started a community treatment program between 1989 and 1992 with a matched group from the same jurisdiction released between 1986 and 1989 when no treatment was available. After a fixed 5-year follow-up period (the same for both groups), the sexual reconviction rate was 5.6% for the treatment group (3/54) compared to 13.0% (7/54) for the comparison group. The difference was not statistically significant.

Proctor's (1996) study is among the better studies of treatment outcome with sexual offenders. A common problem with his study, and many similar studies, is that little information was provided concerning the screening processes by which the treatment group was selected from all potential candidates. Consequently, the "unscreened" comparison group would be expected to contain offenders who would be unable to receive treatment due to factors such as language difficulties, major mental illness, cognitive impairment, or lack of motivation. Another problem with cohort designs is that there can be systematic differences in the criminal justice response to sex offenders during different years (Friendship & Thornton, 2001).

Researchers attempt to address the threat of pre-existing group differences through matching or post-hoc statistical controls. Exact matching on more than two variables is difficult, and often fails in practice (Hanson & Nicholaichuk, 2000; Proctor, 1996). Post-hoc statistical controls are useful, but only provide limited protection because there could always be unmeasured variables that systematically vary across groups. Statistical controls are most persuasive when

they consider a wide range of variables to be related to recidivism, which is rarely observed in the existing recidivism studies.

In the absence of strong evidence, narrative reviewers have come to contradictory conclusions concerning the effectiveness of sex offender treatment effectiveness (e.g., Furby et al., 1989; Harris, Rice, & Quinsey, 1998; Polizzi, MacKenzie, & Hickman, 1999). Meta-analytic reviews (Gallagher, Wilson, Hirshfield, Coggeshall, & MacKenzie, 1999; Hall, 1995; Hanson et al., 2002) offer a number of advantages over the earlier qualitative, narrative reviews. By explicitly defining the criteria upon which the studies are aggregated, reviewers can transform their subjective impressions into hypotheses open to empirical scrutiny. As well, given that single studies typically lack the statistical power to detect small effects, many apparent inconsistencies can often be attributed to the chance fluctuations expected among studies with varying sample sizes (Rosenthal & DiMatteo, 2000).

Hall's (1995) meta-analysis included 12 studies that appeared after Furby et al.'s (1989) narrative review. In contrast to earlier reviews, Hall only considered studies that compared a treatment group to a comparison group. Hall reported a small, positive treatment effect, with the most effective treatment being cognitive-behavioural and hormonal. One major limitation of this study was that the observed treatment effect was mainly derived from comparisons between treatment completers and noncompleters (Hall). Given that men who fail to complete treatment are younger, less educated, and more antisocial than treatment completers (Wierzbicki & Pekarik, 1993), critics have interpreted Hall's study as further evidence of the stability of individual differences in recidivism potential (Harris et al., 1998).

In the next major meta-analysis, Gallagher et al. (1999) examined 22 studies (25 treatment comparisons). Gallagher et al., like Hall (1995), found a significant treatment effect for cognitive-behavioural treatment, but did not find an effect for hormonal treatments. Gallagher et al. focused on the best available studies, but some of these studies had significant threats to validity. For example, some of the studies involved comparisons between treatment completers and treatment drop-outs, and others were preliminary reports contradicted by later versions.

The first report of the Collaborative Data Project (Hanson et al., 2002) examined all the relevant studies in the Hall (1995) and Gallagher et al. (1999) reviews, as well as many recent studies (42 usable studies; 20 published, 22 unpublished). Hanson et al. focused on two questions: 1) To what extent could

the different research designs (e.g., random assignment, incidental assignment, drop-outs versus completers) inform the evaluation of treatment effectiveness? and 2) What was the evidence for treatment effectiveness given the best available research designs and treatment programs that meet current standards? In response to these questions, Hanson et al. found that studies comparing treatment drop-outs to completers consistently found higher recidivism rates among the treatment drop-outs – regardless of the type of treatment received. Such a finding confirmed previous concerns that the results of drop-out studies are largely determined by selective attrition of high-risk cases rather than by differences in the amount of treatment received. Surprisingly, Hanson et al. did not find any difference in the sexual recidivism rates between offenders who explicitly refused treatment and those who volunteered.

There were too few random assignment studies ($k = 3$) to make meaningful conclusions about this method. Consequently, most of the results of the Hanson et al. (2002) meta-analysis were based on matching/incidental assignment studies. In incidental assignment studies ($k = 17$), the comparison groups were selected from offenders in which there were no a priori reasons to expect differences from the treatment group. Based on the random and incidental assignment studies, Hanson et al. found a significant difference between the older (pre-1980) forms of treatment and the current treatment programs for sexual offenders. Current treatments (cognitive-behavioural and systemic) were associated with significant reductions in sexual (17% to 10%) and general recidivism (51% to 32%) after 4-5 years of follow-up. No treatment effects were found for older treatments (purely behavioural, unspecified psychotherapy).

The Hanson et al. (2002) study said very little about which current forms of treatment are most effective. Sex offender treatment has evolved during the last decades based partly on theory (e.g., Laws, 1989), and partly on progress made in the “what works” literature for general offenders (e.g., Andrews et al., 1990; Lösel, 1995). For general offenders, programs that target criminogenic needs are skills based, and are delivered in a manner consistent with the learning styles of the offenders, are most effective (Andrews et al.; Lösel). There has been insufficient research to know whether the features important for treating general offenders are also important for treating sexual offenders. Recent treatment approaches for sexual offenders find stronger treatment effects than the earlier studies, but such changes could also be attributed to general changes in the offenders, the

victims, or the criminal justice system during the past 20 years.

Reviews are only as good as the studies that go into them. Almost all of the studies in the sex offender treatment reviews were secondary analyses of natural experiments. Given the difficulties associated with random assignment studies, it is likely that our information about sexual offender treatment will be based on such natural experiments for many years to come. Consequently, it is important that researchers using “incidental” assignment studies carefully address potential threats to validity in order to contribute to knowledge of sexual offender treatment.

During the late 1980s, the Correctional Service of Canada (CSC) implemented a policy of mandatory treatment for all sexual offenders released in the Pacific Region (British Columbia). The systematic introduction of the Community Sex Offender Program (CSOP) provided a unique opportunity to observe the potential effects of treatment with minimal concerns about selection biases (CS/RESORS, 1991). A limited amount of treatment had been offered to sex offenders in the Abbotsford Regional Treatment Centre since 1972. It was not until 1983 that sex offender specific group treatment was offered in the community (CS/RESORS, 1991). After the local CSOP start date, all sex offenders received weekly sex offender treatment until the end of their sentences (no attrition without re-incarceration). The initial comparison between offenders released before, and released after, the program suggested little overall differences in the recidivism rates (CS/RESORS, 1991). The CSOP treatment providers varied, however, in orientation. The most promising results were associated with well-managed, cognitive-behavioural programs that rigorously targeted sex offence specific issues (Stephenson, 1991).

The current research built on the original evaluation of the CSC community treatment program operated in the 1980s. The original CSOP evaluation was promising, but the follow-up period for the original study was too short (less than 4 years) to justify strong conclusions. In the present study, the follow-up period was extended to 12 years, with sufficient increases in the recidivism base rates/statistical power to detect moderate to small treatment effects (see discussion by Barbaree, 1997). The current study also improved upon the previous CSOP evaluation (and most other sex offender treatment evaluations) by controlling for several relevant risk factors, including time-at-risk, year of release, treatment quality, age, and seven static risk factors derived from official criminal records.

TABLE 1
Group Demographics

Variable	Treatment (N = 403)		Comparison (N = 321)		p
	M	SD	M	SD	
Release date (years)	1987.6	2.0	1986.1	2.3	<.001
Follow-up in years	11.7	1.9	13.2	2.7	<.001
Age at release	37.7	11.2	37.0	11.2	.37
Static Risk Score (1-10)	1.7	1.5	1.4	1.4	.001
Prior sex offences	0.6	0.9	0.5	0.9	.17
	P	N	P	N	p
Any prior sex offences	29.3	118	20.6	66	.007
Age greater than 25	11.2	45	13.1	42	.43
Any conviction for non-contact sex offence	2.5	10	1.9	6	.58
Index conviction for nonsexual violence	19.8	80	14.3	46	.05
Any conviction for prior nonsexual violence	34.0	137	28.4	91	.104
Four or more prior sentencing dates	42.7	172	35.2	113	.041
Any breach of conditional release	46.4	187	38.3	123	.029
Less than four years at liberty prior to index conviction	65.5	264	55.1	177	.005

Method

Participants

Participants were the complete sample of male sexual offenders released between 1980 and 1992 onto community supervision in the Pacific Region of Correctional Service of Canada. Of the 870 cases in the original CSOP roster, valid follow-up information was obtained for 724 unique cases. Duplicate cases were excluded by randomly selecting one release date as the index offence. Also excluded were cases with release dates prior to 1980, cases with no identifiable sexual or violent offence linked to index date, and cases in which no criminal history records could be matched to the identifying information. All the men had received a sentence of two years or more for an explicitly sexual offence (contact offence against adults or children) or for an offence with a sexual motivation (e.g., sexually motivated assault). Participants were divided into two groups: treated ($n = 403$) and comparison ($n = 321$). Groups were assigned based on the year of implementation of the CSOP treatment programs. The comparison group included offenders whose release date occurred prior to the CSOP program implemented in their specific location (1980-1991). The treated group comprised offenders released after the implementation of the program (1981-1992). Treatment in this context refers only to the CSOP program; some offenders received treatment during, or prior to, incarceration, although the amount of other treatment was unknown.

As dictated by the research design, the comparison offenders were released, on average, about 1.5 years earlier than the treated offenders (see Table 1). The average age of the two groups was similar (37-38

years old). Unexpectedly, the treated offenders were at higher risk to re-offend than the comparison offenders, as indicated by the number of prior sexual offences.

Procedure

The CSOP treatment was provided by nine different therapists (psychologists) contracted by the Correctional Service of Canada. Table 2 provides an overview of these programs according to geographic region. The type of treatment delivered by each therapist was consistent within region, with the exception of Program D, in which the therapists adopted divergent approaches (psychodynamic versus cognitive-behavioural).

Treatment quality. Ratings of treatment quality (better or worse) were based on detailed information collected during the original evaluation conducted in 1990/1991 (CS/RESORS, 1991), as well as telephone interviews conducted in 2002 with five of the nine service providers. The original evaluations included reviews of program descriptions along with interviews of both therapists and offenders. The recently conducted telephone interviews, lasting approximately one hour, closely followed CSC Accreditation Criteria (CSC, 1998). These criteria are intended to identify programs that are likely to have an effect on reducing criminal recidivism:

- the treatment must be based on an explicit, empirically based model of change;
- the problems addressed in the program must be related to criminal behaviour (criminogenic needs);

TABLE 2
Program Descriptions

Program	Region	N	Orientation	Conformity to Current Standards
A	Metropolitan & outskirts	299	General skill-building group psychotherapeutic model paired with individual sessions targeted at specific issues to provide support for daily living.	Worse
B	Northern	45	Applied behavioural analysis and behavioural and individual therapy.	Better
C	Small town/rural	41	Cognitive behavioural group and individual therapy.	Better
D	Urban	9	1. The first contractor provided psychodynamic-oriented individual psychotherapy.	Worse
		6	2. The second contractor provided individual cognitive-behavioural treatment.	Better

- c) the amount of treatment received must be sufficient;
- d) the program must be delivered in a manner consistent with the learning styles of the offenders (responsivity);
- e) the methods used for delivering treatment must be proven to be effective with offenders;
- f) offenders should learn new skills;
- g) there should be a continuity of care such that offenders have opportunity for follow-up practice and booster session; and
- h) there must be mechanisms in place to monitor program integrity, protect against therapist burn-out, and prevent program drift.

The CSC Accreditation Criteria were closely modeled after those developed by Her Majesty's Prison Service (England and Wales), which were based on meta-analytic reviews of "what works" in correctional treatment for general offenders (Lipton, Thornton, McGuire, Porporino, & Hollin, 2000). In addition to the ratings based on the accreditation criteria, ratings of program quality were also provided in the original CSOP evaluation (CS/RESORS, 1991). Given that the international accreditation criteria were not available in 1991, the criteria used in the original evaluation were as follows: a) the treatment orientation must be predominantly cognitive-behavioural, but other techniques should be included; b) treatment should also address factors relating specifically to sexual offending such as deviant sexual drive, crime cycle, pathway to the offence, relapse-prevention, and awareness of factors increasing risk to re-offend, and c) the program must focus on countering denial. Both the original and the accreditation approaches sorted the providers into two groups ("better" and "worse") with relatively little difference within the groups. Comparison between the original and current ratings found that the group placement was identical for eight of the nine providers ($Kappa = .78$).

It is interesting to note, however, that none of the programs would have met current CSC Accreditation Criteria. The most common weaknesses concerned a lack of mechanisms for controlling program drift, and an insufficient dose of treatment (treatment time was confined to the time remaining in the offenders' sentences).

Recidivism

Recidivism information was based on official police records obtained in October 1999, from the Royal Canadian Mounted Police (RCMP) Canadian Police Information Centre (CPIC). The RCMP records contain all indictable convictions in Canada, and may contain summary convictions as well as charges that did not result in convictions. Three types of recidivism were examined: sexual, violent, and general (any) recidivism. Sexual recidivism was defined as any charge or conviction for a sexual offence committed after the index offence. Violent recidivism included any charge or conviction for nonsexually violent or sexual offences. Sexual and nonsexual violent offences are commonly combined in sexual offender research because it gives a policy-relevant measure of "serious" recidivism (e.g., Dempster & Hart, 2002). General criminal recidivism included a charge or conviction for a sexual, violent, or nonviolent offence. Parole revocations were not included unless they were accompanied by new charges.

Control Variables

Static risk instrument. Given observed differences in prior criminal history between the treatment and comparison groups (see Table 1), some method was required to control for pre-existing differences on static risk variables. Although it would be desirable to use an established risk instrument, the only information that was consistently available were the offenders' ages and criminal records. Consequently, a risk

TABLE 3
Group Comparisons in Recidivism Rates

Analysis	Treatment		Comparison		Effect Size	
	<i>P</i>	<i>n</i>	<i>P</i>	<i>n</i>	Odds Ratio	95% C. I.
Sexual Recidivism						
Overall	21.1	85	21.8	70	0.96	0.67-1.37
7-year follow-up	15.1	61	14.3	46	1.07	0.70-1.61
Violent Recidivism						
Overall	42.9	173	44.5	143	0.94	0.70-1.26
7-year follow-up	33.7	136	30.8	99	1.14	0.83-1.56
Any Recidivism						
Overall	56.6	228	60.4	194	0.85	0.63-1.15
7-year follow-up	49.4	199	49.8	160	0.98	0.73-1.32

Note. C. I. refers to 95% confidence interval.

TABLE 4
Effect of Treatment Controlling for Risk and Date of Release

Outcome Criterion	Static Risk Score (0-10)		Release Dates (Year)		Treatment Group (0 = No, 1 = Yes)	
	Odds Ratio	95% C. I.	Odds Ratio	95% C. I.	Odds Ratio	95% C. I.
Sexual Recidivism	1.23	1.14-1.32	0.95	0.89-1.02	1.00	0.71-1.40
Violent Recidivism	1.29	1.23-1.36	0.97	0.93-1.02	0.99	0.78-1.26
Any Recidivism	1.31	1.26-1.37	0.96	0.92-0.97	0.90	0.73-1.12

Note. C. I. refers to 95% confidence interval.

scale was created that included age and criminal history items from Static-99 (Hanson & Thornton, 2000) and Static-2002 (Hanson & Thornton, 2003): age greater than 25, any conviction for noncontact sex offence, index conviction for nonsexual violence, any conviction for prior nonsexual violence, four or more prior sentencing dates, prior sex offence (using Static-99 coding), any breach of conditional release, and less than 4 years at liberty prior to index conviction. All items were scored as "0" or "1" except for the number of prior sex offences (0-3). The maximum total score of the instrument was 10 ($\alpha = .62$). This risk instrument was significantly related to sexual recidivism (AUC of .63), violent recidivism (AUC of .70), and general (any) recidivism (AUC of .74).

Release dates. Initial analyses searched for systematic differences in recidivism rates based on the year released. Given that the offenders released early were at liberty longer than offenders released later, these analyses used a fixed 7-year follow-up period (the minimum follow-up period). In other words, if an offender was released in 1990 and first re-offended in 1998, he would be considered a nonrecidivist because

the follow-up period was more than 7 years. The year of release showed small negative correlations with recidivism among both the treatment and comparison groups. For the treatment group, the correlations were -.06, -.02, and -.08 with sexual, violent, and general recidivism, respectively. For the comparison group, the correlations with sexual, violent, and general recidivism were -.06, -.03, and -.12, respectively. Only for the comparison group was the correlation between any recidivism and year released statistically significant. Nevertheless, the direction of all correlations indicated that those released early were more likely to be identified as recidivists than those released later (a bias favouring the treatment group). Given that this pattern could be the result of selective attrition of old, inactive records (see Hanson & Nicholaichuk, 2000), year of release was included as a control variable.

Planned Analyses

This study comprised two main sets of comparisons. First, the overall effect of treatment was examined for the treatment and comparison groups. Second, the treated group was divided into those

TABLE 5
Comparison of "Better" and "Worse" Treatments

Analysis	Better (N = 92)		Worse (N = 308)		Effect Size	
	P	n	P	n	Odds Ratio	95% C. I.
Sexual Recidivism	18.5	17	21.8	67	0.82	0.45-1.47
Overall						
7-year follow-up	14.1	13	15.6	48	0.89	0.46-1.72
Violent Recidivism						
Overall	35.9	33	45.1	139	0.68	0.42-1.10
7-year follow-up	31.5	29	34.7	107	0.86	0.52-1.42
Any Recidivism	54.3	50	57.5	177	0.88	0.55-1.41
Overall						
7-year follow-up	47.8	44	50.3	155	0.90	0.57-1.44

Note. C. I. refers to 95% confidence interval.

TABLE 6
Effect of "Better" Treatments Controlling for Risk and Date of Release

Outcome Criterion	Static Risk Score (0-10)		Release Dates (Year)		Treatment Group (0 = No, 1 = Yes)	
	Odds Ratio	95% C. I.	Odds Ratio	95% C. I.	OddsRatio	95% C. I.
Sexual Recidivism	1.27	1.15-1.40	0.90	0.80-1.00	1.23	0.70-2.15
Violent Recidivism	1.33	1.24-1.43	0.94	0.87-1.01	1.03	0.69-1.53
Any Recidivism	1.37	1.29-1.46	0.93	0.87- .99	1.34	0.97-1.86

Note. C. I. refers to 95% confidence interval.

who received "better" and "worse" treatment. Group differences in recidivism rates were identified using proportional hazard survival analyses (Allison, 1984). Survival analysis has the advantage of controlling for time-at-risk, while examining how recidivism rates are influenced by categorical variables (e.g., treated or not) or continuous variables (e.g., risk scores). Proportional hazard analyses were conducted using the Cox Regression procedure from SPSS for Windows Version 10.0 (SPSS, 1999). In this study, offenders were censored if they did not re-offend or if they were lost to follow-up by being continuously incarcerated for another type of offence.

Similar to logistic regression, exponents of the Cox regression weights can be interpreted as odds ratios. For example, given relatively low base rates, an odds ratio of .50 would mean that the observed recidivism rate of the treatment group would be approximately half the recidivism rate of the comparison group (Fleiss, 1994). In our tables, odds ratios less than 1 indicate a positive effect of treatment, while odds ratios more than 1 indicate the treatment group did worse than the comparison group. When the 95% confidence interval contains 1.00, the differences

between the groups should not be considered statistically significant. For continuous variables, the odds ratio represented the relative change in recidivism rates for each unit increase in the predictor variable.

Results

After an average follow-up time of 12.5 years (ranging from 7 to 14 years), the sexual, violent, and general recidivism rates were 21.4%, 43.6%, and 58.2%, respectively (total sample size of 724 for all analyses). Based on the complete follow-up period, the unadjusted sexual recidivism rates were 21.1% (85 out of 403) for offenders who received treatment compared to 21.8% (70 out of 321) for the comparison group (odds ratio of .96, 95% C. I. of .67-1.37; see Table 3). The violent (42.9% vs. 44.5%) and general (56.6% vs. 60.4%) recidivism rates were also similar for the treatment and comparison groups. None of these differences approached statistical significance. Given that the treatment group had a shorter follow-up time ($M = 11.7$ years, $SD = 1.9$) than the comparison group ($M = 13.2$ years, $SD = 2.7$), Table 3 also reports the recidivism rates after a fixed 7-year follow-up period. Again, none of the differences were

statistically significant (sexual recidivism: odds ratio of 1.07, 95% C. I. of .70 to 1.61; violent recidivism: odds ratio of 1.14, 95% C. I. of .83 to 1.56; general recidivism: odds ratio of .98, 95% C. I. of .73 to 1.32).

The next set of analyses (see Table 4) control for individual differences in static risk factors and for the year of release. Cox regressions were calculated for each outcome criteria (sexual, violent, any) using three predictor variables: static risk score (0-10), release date (year), and treatment condition (1 = CSOP; 0 = comparison). Risk was significantly related to all three recidivism criteria ($p < .001$). Release dates were significantly related to the probability of any recidivism, with those released early being the most likely to recidivate (assuming equal time at risk). Attendance in the CSOP program was unrelated to sexual (odds ratio = 1.00), violent (odds ratio = .99) or any recidivism (odds ratio = .90).

The second set of comparisons examined the recidivism rates of those who received “better” and “worse” treatment. The overall recidivism rates for those who received better treatment were not significantly different than the rates for those who received worse treatment: unadjusted overall sexual recidivism rates of 18.5% versus 21.8%, violent recidivism rates of 35.9% versus 45.1%, and general recidivism rates of 54.3% and 57.5% for those who received better versus worse treatment, respectively (see Table 5). When Cox regression was used to control for time-at-risk, year of release, and static risk scores, the differences between the groups were still nonsignificant. Contrary to expectation, the direction of the effect now favoured those who received the “worse” treatment: The odds ratio was 1.23 for sexual recidivism, 1.03 for violent recidivism, and 1.34 for general recidivism (see Table 6).

Discussion

This study had two aims. The first aim was to evaluate a specific sex offender treatment program and thereby contribute to the cumulative knowledge of treatment effectiveness. The second aim was to demonstrate ways to improve upon previous cohort studies. The major virtue of the current study was the continuous, unselected sample – an opportunity rarely available. The other innovations, however, can be easily applied in cohort designs. Researchers can check whether the recidivism rates vary according to year released, and they can control for a substantial amount of the individual differences in risk level by using a simple scale coded from criminal history records.

The treatment program examined in this study did not appear to be effective in reducing recidivism.

Although some analyses slightly favoured one group or the other, the differences between the treated and untreated groups was virtually zero after controlling for year of release, follow-up time, and static risk factors. The sample size was sufficiently large (300 to 400 in each group) that the 95% confidence intervals rule out any large effects of treatment in this sample. Given that the absolute differences in the 7-year recidivism rates was only + 0.8% for sexual recidivism, + 2.9% for violent recidivism, and - 0.4% for general recidivism, it is reasonable to conclude that the overall program did not have any meaningful effect on recidivism rates. The average sexual recidivism rate (21.4% after 12.5 years) was similar to that found in samples of untreated sexual offenders (e.g., 22% after 10 years in Hanson & Thornton, 2000).

Examination of the individual treatment programs did not yield any significant differences in recidivism rates. Treatment programs that were judged to approximate contemporary standards had slightly better outcomes than the “worse” programs (median odds ratio of .87), but any apparent effects disappeared after controlling for risk (median odds ratio of 1.23). Limited confidence, however, can be placed in the judgments of treatment quality given that they were based on the summaries completed for the previous evaluation (CS/RESORS, 1991) and telephone interviews conducted more than 10 years later. Nevertheless, the information was sufficient to determine that none of the programs would meet CSC’s current accreditation standards. The intensity of treatment was determined by sentence length, not the offenders’ criminogenic needs. As well, none of the programs had formal methods for monitoring treatment delivery or systematically evaluating treatment gains.

The study does not allow conclusions about what was effective or ineffective in the CSOP interventions. The findings do suggest, however, that some highly plausible interventions may have little overall effect. The CSOP program was well regarded at the time, and the initial evaluation was encouraging (Stephenson, 1991). Given that very similar programs are still being offered in other jurisdictions, we still have much to learn about how best to intervene with sexual offenders.

The findings of the current study contrast with the positive effects of cognitive-behavioural treatment found in previous reviews (Gallagher et al., 1999; Hanson et al., 2002). No single study is sufficient to determine whether treatment works or not. Those inclined to believe that treatment is effective will emphasize that the interventions examined in the current study did not meet contemporary standards;

those inclined to doubt the effectiveness of treatment will emphasize the methodological weaknesses in the other studies that have shown positive treatment effects. The current study is unique, to our knowledge, for the lack of selection bias in the treatment and comparison groups. Understanding of sex offender treatment will advance as individual studies improve, and the cumulative results of these studies are meaningfully integrated through meta-analyses.

Another objective of the current study was to demonstrate practical ways of improving commonly used cohort designs. These studies compare treated sex offenders to an untreated group released prior to the implementation of the treatment program (e.g., Bakker, Hudson, Wales, & Riley, 1999; Martin, 1998; Proctor, 1996). Assuming equal follow-up times, there is no a priori reason for expecting group differences; however, such difference may still exist due to year-by-year variation in the sample or in the response of the criminal justice system. Friendship and Thornton (2001), for example, found substantial differences in sexual offender reconviction rates based on the year of release.

In the current study, two opposing cohort effects were detected: a) offenders released in earlier years were more likely to be recidivists than those released later (controlling for follow-up time and static risk); and b) those released later were higher risk on static risk factors than those released earlier. The first cohort effect could be attributable to the selective attrition of inactive records (Hanson & Nicholaichuk, 2000). The second cohort effect awaits explanation, but does suggest that sexual offenders in CSC changed during the 1980s. Both effects, however, reinforce the need to consider the potential changes (other than the introduction of treatment) that could influence the typical cohort design. The information needed to conduct these analyses is readily available, if rarely used.

The current study addressed these cohort effects by statistically controlling for follow-up time, year of release, and eight static risk factors. Although any number of unmeasured factors could still be influencing the results, these control variables accounted for a significant amount of variability of the recidivism criteria. Given the ease with which they can be collected, routine consideration of these (or similar) control variables should be considered the minimum requirements for future studies of sexual offender treatment outcome.

The views expressed are those of the authors and do not necessarily represent the views of the Department of Public Safety and Emergency Preparedness Canada.

Correspondence should be addressed to R. K. Hanson, Corrections Research, Department of Public Safety and Emergency Preparedness Canada, 340 Laurier Avenue, West, Ottawa, Ontario, Canada K1A 0P8 (E-mail: hansonk@sgc.gc.ca).

References

- Allison, P. D. (1984). *Event history analysis: Regression for longitudinal event data*. Beverly Hills, CA: Sage.
- Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P., & Cullen, F. T. (1990). Does correctional treatment work? A psychologically informed meta-analysis. *Criminology*, 28, 369-404.
- Bakker, L., Hudson, S., Wales, D., & Riley, D. (1999). "And there was light": An evaluation of the Kia Marama treatment programme for New Zealand sex offenders against children. Unpublished report. New Zealand.
- Barbaree, H. (1997). Evaluating treatment efficacy with sexual offenders: The insensitivity of recidivism studies to treatment effects. *Sexual Abuse: A Journal of Research and Treatment*, 9, 111-128.
- Correctional Service of Canada (August, 1998). *Criteria for program accreditation*. Ottawa, ON: Author.
- CS/RESORS Consulting (1991). *An evaluation of community sex offender programs in the pacific region*. Report presented to the Correctional Service Canada Regional Headquarters (Contract no. 21803-0-A602/01-XSB). Vancouver, BC: Author.
- Dempster, R. J., & Hart, S. D. (2002). The relative utility of fixed and variable risk factors in discriminating sexual recidivists and nonrecidivists. *Sexual Abuse: A Journal of Research and Treatment*, 14, 121-138.
- Fleiss, J. L. (1994). Measures of effect size for categorical data. In H. Copper & L. Hedges (Eds.), *The handbook of research synthesis* (pp. 245-260). New York: Sage.
- Friendship, C., & Thornton, D. (2001). Sexual reconviction for sexual offenders discharged from prison in England and Wales. *British Journal of Criminology*, 41, 284-292.
- Furby, L., Weinrott, M. R., & Blackshaw, L. (1989). Sex offender recidivism: A review. *Psychological Bulletin*, 105, 3-30.
- Gallagher, C. A., Wilson, D. B., Hirschfield, P., Coggeshall, M. B., & MacKenzie, D. L. (1999). A quantitative review of the effects of sex offender treatment of sexual reoffending. *Corrections Management Quarterly*, 3, 19-29.
- General Accounting Office (1996). *Sex offender treatment: Research results inconclusive about what works to reduce recidivism*. Report to the Chairman, Subcommittee on Crime, Committee on the Judiciary, House of Representatives. Washington, DC: I.S. General Accounting Office.
- 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., Gordon, A., Harris, A. J. R., Marques, J. K., Murphy, W., Quinsey, V. L., & Seto, M. C. (2002). 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, 169-194.
- Hanson, R. K., & Nicholaichuk, T. (2000). A cautionary note regarding Nicholaichuk et al. *Sexual Abuse: Journal of Research and Treatment*, 12, 289-293.
- Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behaviour*, 24, 119-136.
- Hanson, R. K., & Thornton, D. (2003). *Notes on the development of Static-2002*. User Report 2003-01. Ottawa, ON: Department of the Solicitor General of Canada.
- Harris, G. T., Rice, M. E., & Quinsey, V. L. (1998). Appraisal and management of risk in sexual aggressors. *Psychology, Public Policy, and Law*, 4, 73-115.
- Laws, D. R. (1989). *Relapse prevention with sex offenders*. New York: Guilford Press.
- Lipton, D. S., Thornton, D., McGuire, J., Porporino, F. J., & Hollin, C. R. (2000). Program accreditation and correctional treatment. *Substance Use & Abuse*, 35, 1705-1734.
- 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 research and practice* (pp. 79-111). Chichester, UK: Wiley.
- MacKenzie, D. L., & Hickman, L. J. (1998). *What works in corrections? An examination of the effectiveness of the type of rehabilitation programs offered by Washington Department of Corrections: Report to the State of Washington Legislature Joint Audit and Review Committee*. College Park, MD: University of Maryland.
- Marshall, W. L., & Barbaree, H. E. (1988). The long-term evaluation of a behavioral treatment program for child molesters. *Behavioural Research and Therapy*, 26, 499-511.
- Martin, I. (1998). *Efficacité d'un programme cognitif-behavioral institutionnel pour délinquants sexuels*. Unpublished doctoral dissertation, University of Montreal, Quebec.
- 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. *International Journal of Offender Therapy & Comparative Criminology*, 43, 357-374.
- Proctor, 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.
- Rosenthal, R., & DiMatteo, M. R. (2000). Meta analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, 52, 59-82.
- SPSS Inc. (1999). *SPSS advanced models 10.0*. Chicago, IL: SPSS Inc.
- Stephenson, M. (1991). A summary of an evaluation of the community sex offender program in the Pacific region. *Forum on Corrections Research*, 3(4), 25-30.
- Wierzbicki, M., & Pekarik, G. (1993). A meta-analysis of psychotherapy dropout. *Professional Psychology: Research and Practice*, 24, 190-195.

Received January 17, 2003

Revised August 7, 2003

Accepted October 2, 2003