

Predicting Relapse: A meta-Analysis of Sexual Offender Recidivism Studies

R. Karl Hanson and Monique T. Bussière,
Department of the Solicitor General of Canada.

In: Journal of Consulting and Clinical Psychology,
edited by the American Psychological Association,
1998, Vol. 66, No. 2, pp 348-362

R. Karl Hanson and Monique T. Bussière, Corrections Research, Department of the Solicitor General of Canada.

Content

Abstract

Acknowledgements

Article

Method

Sample

Quality of Studies

Coding procedure

Index of Predictive Accuracy

Aggregation of Findings

Generalizability of Findings

Results

How to read the Tables

Predictors of Sexual Offense Recidivism

Table 1: Predictors of Sexual Recidivism

Predictors of Nonsexual Violent Recidivism

Table 2: Predictors of Nonsexual Violent Recidivism

Predictors of General (Any) Recidivism

Table 3: Predictors of General Recidivism

Comparison Across Domains

Table 4: Correlations with Recidivism for Each Category of Predictor

Combined Risk Scales

Table 5: combined risk scales

Influence of Recidivism Methods

Discussion

References

Abstract

Evidence from 61 follow-up studies was examined to identify the factors most strongly related to recidivism among sexual offenders. On average, the sexual offense recidivism rate was low (13,4%; $n=23,393$). There were, however, subgroups of offenders who recidivated at high rates. Sexual offense recidivism was best predicted by measures of sexual deviancy (e.g., deviant sexual preferences, prior sexual offenses) and, to a lesser extent, by general criminological factors (e.g., age, total prior offenses). Those offenders who failed to complete treatment were at higher risk for re-offending than those who completed treatment. The predictors of nonsexual violent recidivism and general (any) recidivism were similar to those predictors found among nonsexual criminals (e.g., prior violent offenses, age, juvenile delinquency). Our results suggest that applied risk assessments of sexual offenders should consider separately the offender's risk for sexual and nonsexual recidivism.

Acknowledgements

We thank Margaret Alexander, Lita Furby, Gordon Hall, Roxanne Lieb, Rober Freeman-Longo, Robert Prentky, Mark Weinrott and Sharon Williams for help in locating articles for this review. The comments of Jim Bonta, Bill Marshall, Andrew Harris, and Robert McGrath on an earlier version are also appreciated. As well, we thank Jean Proulx, John Reddon, and David Thornton for access to their original data sets.

The views expressed are those of the authors and do not necessarily represent the views of the Ministry of the Solicitor General of Canada.

Correspondence concerning this article should be addressed to R. Karl Hanson, Corrections Research, Department of the Solicitor General of Canada, 340 Laurier Avenue West, Ottawa, Ontario, Canada K1A 0P8. Electronic Mail be sent to hansonk@sgc.gc.ca.

Article

Assessing chronicity is crucial for clients whose sexual behaviors have brought them into conflict with the law. Many exceptional criminal justice policies, such as post sentence detention (e.g., Anderson & Masters, 1992), lifetime community supervision, and community notification, target those sexual offenders likely to re-offend. Clinicians need to judge whether the client's behaviors are truly atypical of the individual (as the client would like us to believe) or whether the client merits a virtually permanent label as a sexual offender.

Sexual assault is a serious social problem, with high victimization rates among children (10% of boys and 20% of girls; Peters, Wyatt & Finkelhor, 1986) and adult women (10-20%; Johnson & Sacco, 1995; Koss, 1993a). Given the large number of victims, it is not surprising that a significant portion (10-25%) of male community samples (e.g., university students, hospital staff) admit to sexual offending (Hanson & Scott, 1995; Lisak & Roth, 1988; Templeman & Stinnett, 1991).

One of the simplest and most defensible approaches to recidivism prediction is to identify a stable pattern of offending. Behavior is influenced by a variety of internal and external factors that can do change over the life course. Nevertheless, it does not require any special expertise to predict with confidence the continuation of any behavior that has occurred frequently, in many different contexts, and despite the best efforts to stop it.

Some sexual offenders report a well-established, chronic pattern of offending (e.g., Abel, Becker, Cunningham-Rathner, Rouleau, & Murphy, 1987). More typically, however, sexual offenders recurrent deviant sexual interests or behavior (Kennedy & Grubin, 1992; Langevin, 1988). In the absence of an established pattern, risk assessments need to rely on other, relevant information. Determining what is "relevant" requires theoretical assumptions about the nature of sexual offending.

One approach is to assume that sexual offending is similar to other criminal behavior (e.g., theft, assault, drugs) with relatively little specialization (M.R. Gottfredson & Hirschi, 1990). Because many sexual offenders also engage in nonsexual criminal activities (Broadhurst & Maller, 1992; Hanson, Scott & Steffy, 1995), the same factors that predict general recidivism among nonsexual criminals may also predict sexual recidivism among sexual offenders.

The extensive research on the prediction of recidivism among nonsexual criminals (Champion, 1994; D.M. Gottfredson & Tonry, 1987) has identified a reliable set of both static (historical) and dynamic (changeable) risk factors (e.g., Bonta, 1996; Gendreau, Little & Goggin, 1996). Specifically, the persistent criminal tends to be young, have unstable employment, abuse alcohol and drugs, hold pro-criminal attitudes, and associate with other criminals (Gendreau et al., 1996). These characteristics can be considered to define a "criminal lifestyle", a concept similar to *Diagnostic and Statistical Manual of Mental Disorder's* (4th edition) "antisocial personality" (American Psychiatric Association, 1994), Hare's "psychopathy" (Hare et al., 1990), and M.R. Gottfredson and Hirschi's (1990) "lack of self control".

Some evidence, however, suggests that sexual offending may be different from other types of crime. Although sexual offenders frequently commit nonsexual crimes, nonsexual criminals rarely recidivate with sexual offenses (Bonta & Hanson, 1995b; Hanson et al., 1995). As well, many persistent sexual offenders are judged to be low risk by scales designed to predict general criminal recidivism (Bonta & Hanson, 1995b).

Rather than emphasizing general criminological risk factors, sexual offender risk assessments may concentrate on sexual deviance. All sexual offending is, by definition, socially deviant, but not all sexual offenders have deviant sexual interests or preferences. Some date rapists, for example, may prefer consensual sexual activities but misperceive their partner's sexual interests (e.g. "No" means "yes") (Hanson & Scott, 1995; Malamuth & Brown, 1994). In contrast, the sexual lives of some boy-object pedophiles may be completely focused on their preferred victim type (Freund & Watson, 1991; Quinsey, 1986).

Because self-reports are highly vulnerable to self-presentation biases, the assessment of

deviant sexual interest is best supplemented by other sources of information, such as a sexual offense history and phallometric assessment (i.e., direct monitoring of penile response; see Launay, 1994). In general, offenders with the most deviant sexual histories tend to show deviant or abnormal sexual interests on phallometric assessments (Barbaree & Marshall, 1989; Freund & Watson, 1991; Quinsey, 1984, 1986). Specifically, deviant sexual interests are most prevalent among those who victimize strangers, use overt force, select boy victims, or select victims much younger (or much older) than themselves (Barbaree & Marshall, 1989; Freund & Watson, 1991; Quinsey, 1984, 1986).

A further consideration in the assessment of sexual offenders concerns symptoms of general psychological maladjustment. Sexual offenders rarely meet diagnostic criteria for major mental illness, but they often show signs of low self-esteem, substance abuse problems, and assertiveness deficits (W.L. Marshall, 1996). Much of the current treatment and theory concerning sexual offending emphasizes poor coping strategies and negative emotional states as precursors to offending (Laws, 1989, 1995; Pithers, Beal, Armstrong & Petty, 1989). There have been few attempts, however, to examine empirically the relevance of these psychological symptoms to sexual recidivism. Empirically examining the assumed relationship between distress and sexual offending is important because, for nonsexual recidivism, subjective distress has either no relationship or a negative relationship with recidivism (Gendreau et al., 1996).

Once detected, sexual offenders' motivation to change may also be related to recidivism. Those offenders who accept responsibility, express remorse, and comply with treatment (good clinical presentation) should be at lower risk than those who deny any problems and actively resist change (poor clinical presentation). Motivation to change is difficult to assess, however, because there are clear benefits to "appearing" willing to change, and many sexual offenders have the social skills necessary to gain the confidence of sympathetic clinicians.

In agreement with Furby, Weinrott, and Blackshaw (1989), we believe that group comparison within follow-up studies provide "by far the best sources of data" for the identification of recidivism factors (Furby et al., 1989, Pg. 27). The absolute recidivism rates vary across studies as a result of differences in follow-up periods, and local criminal practices. These factors are controlled, however, when the recidivist-non-recidivist comparisons are made within a single follow-up study.

Previous narrative reviews have examined a small number of studies and risk factors (Furby et al., 1989; Hall, 1990; Quinsey, Lalumière, Rice & Harris, 1995). Their conclusions have been tentative and, at times, contradictory. Rapists were considered high risk by Furby et al. (1989), but not by Hall (1990) or Quinsey, Lalumière, et al. (1995). Quinsey, Lalumière, et al. (1995) did, however, report that general criminality (nonsexual offenses) and sexual deviancy (prior sex offenses) predicted sexual recidivism. None of the previous reviews have considered general psychological adjustment or clinical presentation variables as predictors.

The present study provides a quantitative review of the sexual offender recidivism literature. All the participants were sexual offenders, but we examined three types of recidivism: sexual, non-sexual violent, and general (any). Sexual and non-sexual violent recidivism were considered separately, because preliminary evidence suggested that

they may be predicted by different sets of characteristics (Hanson et al., 1995; Marques, Nelson, West & Day, 1994). Quantitative summaries, or meta-analyses, have become a standard feature of research reviews in psychology and medicine (Rosenthal, 1995, Spitzer, 1995). Meta-analyses have several advantages over the traditional narrative forms of review:

- a. Many studies can be considered simultaneously;
- b. large, pooled samples yield high statistical power;
- c. numerical estimates indicate the relative magnitude of effects; and
- d. the generalizability of findings across studies can be tested.

It was expected that the best predictor of sexual offense recidivism would be a history of sexual deviancy. On the basis of previous reviews (Quinsey, Lalumière, et al., 1995), it was also expected that criminal lifestyle variables would be related to both sexual and non-sexual recidivism. The relevance of the other factors, namely, psychological symptoms and clinical presentation, was less clear. Psychological symptoms have been unrelated to recidivism among general criminal populations (Gendreau et al., 1996), but sexual offending may be a special case. The clinical presentation variables may also have little predictive value given the difficulty identifying sincere remorse and genuine motivation to change as well the active debate concerning the efficacy of treatment for sexual offenders (W.L. Marshall & Pithers, 1994); Quinsey, Harris, Rice & Lalumière, 1993).

Method

Sample

Computer searches of both PsycLIT and the National Criminal Justice Reference System were conducted using the following key terms: Sex(ual) offender, rape, rapist, child molester, pedophile, pedophilia, exhibitionism, sexual assault, incest, voyeur, frotteur, indecent exposure, sexual deviant, paraphilia (c), predict, recidivism, recidivist, recidivate, re-offend, reoffense, relapse, and failure. Reference lists were searched for additional articles. Finally, letters were sent to 32 established sexual offender researchers requesting overlooked or as yet unpublished articles or data.

To be included, studies needed to follow up a sample of sex offenders; report recidivism information for sexual offenses, nonsexual violent offenses, or any offenses; and include sufficient statistical information to calculate the relationship between a relevant offender characteristic and recidivism. Further description of the selection criteria can be found in Hanson and Bussière (1996) and in the complete coding manual, which is available from the authors.

As of December 31, 1995, our search yielded 87 usable documents (e.g., published articles, books, government reports, unpublished program evaluations, conference presentations). When the same data set was reported in several articles, all the results were considered to come from the same study. Consequently, the 87 documents represent 61 different studies (country of origin: 30 United States, 16 Canada, 10 United Kingdom, 2 Australia, 2 Denmark, 1 Norway; 45% unpublished; produced between 1943 to 1995, with median of 1989; mean sample size of 475, median of 198, range of 12-15,000.

Most of the studies examined mixed groups of adult sexual offenders (55 mixed offense types, 6 child molesters only; 52 samples of adults, 6 adolescents only; 3 both adolescents and adults). The offenders became from institutions (48%), the community (25%), or both (27%). Nineteen studies focused exclusively on correctional samples, 11 on samples from secure mental health facilities, and the remaining from other sources (private clinics, courts, mixture of sources). Approximately one half of the samples (48%) were from sexual offender treatment programs. When demographic information was presented, the offenders were predominantly Caucasian (27 of 28 studies) and of lower socioeconomic status (27 of 29 studies).

The most common measures of recidivism were reconviction (84%), arrests (54%), self-reports (25%), and parole violations (16%). Multiple indexes of recidivism were used in 27 of 61 studies (44%). The most common sources of recidivism information were national criminal justice records (41%), state or provincial records (41%), records from treatment programs (29%), and self-reports (25%). Other sources (e.g., child protection records) were used in 25% of the studies. In 43% of the studies, the source of the recidivism information was not reported. The reported follow-up periods ranged from 6 months to 23 years (median = 48 months; mean = 66 months).

Quality of Studies

An important concern in meta-analytic reviews is the quality of the studies reviewed. This issue was less of a concern in the current review, however, because all studies used the best available design (i.e., the matched, longitudinal follow-up design; Furby et al., 1989). Nevertheless, variation in the assessment of the predictor variables and of recidivism could affect the results. In most cases, the predictor variables were sufficiently explicit that there was little concern about reliability or validity (e.g., age, criminal history, victim type). There was, however, enough variability in the recidivism measures to justify further analysis. Consequently, the thoroughness of the recidivism search was coded for each study using a 7-point scale ranging from (1) *questionable methods* (e.g., mail-in questionnaires only) / *inadequate follow-up periods* (<6 months) to (7) *multiple, credible data sources* (e.g., local and national records, collateral sources) / *long follow-up periods* (>10 years). Each study was rated by the two authors, and differences were resolved by discussion. In a sample of 20 independent ratings, the rater reliability was .72, using equation ICC (2,1) from Shrout and Fleiss (1979).

The ratings of the adequacy of the recidivism information ranged from 3 to 7 ($M = 4.6$, $SD = 1.0$), indicating overall acceptable levels of diligence in identifying recidivists. No studies used only self-reported or wholly inadequate recidivism detection methods.

Coding procedure

Each document was coded separately by R. Karl Hanson and Monique T. Bussière using a coding manual. When disagreement occurred, most involved calculation errors were immediately corrected. In rare cases of differences of interpretation, advice was sought from colleagues familiar with forensic meta-analytic reviews.

Only one finding of each type of predictor variable was coded from any one study (data

set). Given several related variables, the variable that best represented the category was selected first (e.g., for the category of "any prior sex offenses", "all prior sex offenses" was selected before "prior child molesting offenses"). Next, given conceptually equivalent findings, the selection was based on sample size and completeness of information. Finally, in those rare cases in which several options remained, we simply selected the median value. Further details on the coding and selection procedure are available in het coding manual.

To illustrate the coding procedure, Rice, Harris and Quinsey (1990; Rice, Quinsey & Harris, 1991; Quinsey, Rice & Harris, 1995) reported on the relationship between age and recidivism in at least three studies from the same setting (Mental Health Centre, Penetanguishene, Ontario). One study examined rapists ($n = 54$; Rice et al., 1990), another examined child molesters ($n = 136$; Rice et al., 1991), and a third examined the combined sample ($n = 178$; Quinsey, Rice & Harris, 1995). We only used the findings from Quinsey, Rice and Harris (1995, Table 2) because it was based on the largest sample size and the longest follow-up period.

Even though each study could contribute only one finding per predictor, studies frequently reported on more than one predictor variable. Consequently, it is possible that the correlations within each study are themselves correlated. Although we ignored these potential intercorrelations, the major consequence of this approach was to make the test of differences between predictors conservative. Given that the sample sizes were generally large, the potential loss of a small amount of statistical power was of little concern.

Index of Predictive Accuracy

Predictive accuracy was calculated using r because it is readily understood and the statistical procedures for aggregating r s are well documented (Hedges & Olkin, 1985; Rosenthal, 1991). The magnitude of a correlation can be interpreted as an approximation of the percentage difference in recidivism rates between offenders with or without a particular characteristic (Farrington & Loeber, 1989; Rosenthal, 1991). If, for example, the overall recidivism rate was 25% and "blue eyes" correlated .20 with recidivism, there would be a 20 percentage point difference in the rates between the groups (35% blue eyed vs. 15% non-blue eyed). Except with extreme distributions, this span of 20 percentage points should be centered around the base rate (i.e., 25 ± 10 percentage points).

Formulas for converting study statistics (F , t , significance levels) into r were drawn from Rosenthal (1991). The correlations were calculated from the most direct data available. If a study reported both the raw frequencies and a chi-square, for example, the correlation was calculated from the raw frequencies. "Nonsignificant" findings were assigned an r value of 0 (7.3% of findings). For five studies (Bonta & Hanson, 1995a; Hanson, Steffy & Gauthier, 1993b; Proulx, Pellerin, McKibben, Aubut & Quimet, 1995; Reddon, 1995; Thornton, 1995), the correlations were calculated directly from the original raw data sets. Some of the information from these unpublished data sets have been reported previously (Bonta & Hanson, 1995b; Hanson et al., 1995; Hanson, Steffy & Gauthier, 1992, 1993a; Proulx et al., 1997; Studer, Redder, Roper & Estrada, 1996).

Aggregation of Findings

Two methods were used to aggregate findings. The first was the median r value across studies. Median values have been recommended for meta-analysis (Slavin, 1995) because they are relatively insensitive to outliers and are easy to calculate and interpret. On the other hand, statistics for estimating the variability of median values are not readily available. Median values do not take into account factors that may influence the results, such as recidivism base rates and sample size. Consequently, a second method (the weighted averaged r) was also used.

Before averaging, each correlation was corrected for differences in recidivism base rates using formula 12:8 from Ley (1972). Because correlations decrease predictably with reductions in variance (i.e., base rates), the correction increased the size of correlations from studies with relatively high recidivism rates. The resulting r values were aggregated using the standard procedures recommended by Hedges and Olkin (1985). Details of the formulas used are available in Hanson and Bussière (1996).

Generalizability of Findings

Hedges and Olkin's (1995) formulas were used to create confidence intervals as a measure of the error in estimation. Specifically, there is a 1-in-20 chance that the true value is outside the bounds of the 95% confidence interval. Confidence intervals contain all the information of traditional null hypothesis testing and allow for multiple comparisons while limiting the overall Type I error rate to 5% (Schmidt, 1996).

Variability across studies was indexed by Hedges and Olkin's (1985) Q statistic. The Q statistic is distributed as a chi-square with $k - 1$ degrees of freedom, where k is the number of studies. A finding was considered an outlier if (a) it was an extreme value (highest or lowest), (b) the Q statistic was significant, and (c) it accounted for more than 50% of the value of the Q statistic. When an outlier was detected, the results were reported with and without the exceptional case.

Maletzky's study (1991, 1993) requires special mention as an outlier. Given his large sample size (4,381-5,000), even small deviations from the other studies could be statistically significant. As well, he used an unusually broad definition of recidivism (including "treatment failure" along with new sexual offenses). Rather than eliminate Maletzky's study a priori, it was considered to be an outlier only when justified according to the empirical rules mentioned previously.

Results

The 61 studies provided information on 28,972 sexual offenders, although sample sizes were smaller for any particular analysis. On average, the sex offense recidivism rate was 13.4% ($n = 23,393$; 18.9% for 1,839 rapists and 12.7% for 9,603 child molesters). The average follow-up period was 4 to 5 years. The recidivism rate for nonsexual violence was 12.2% ($n = 7,155$), but there was a substantial difference in the nonsexual violent recidivism rates for the child molesters (9.9%; $n = 1,774$) and the rapists (22.1%; $n = 782$). When recidivism was defined as any reoffense, the rates were predictably higher: 36.3%

overall ($n = 19,347$), 36,9% for the child molesters ($n = 3,363$), and 46,2% for rapists ($n = 4,017$). These averages should be considered cautiously because they are based on diverse methods and follow-up periods, and many sexual offenses remain undetected (Bonta & Hanson, 1994).

How to read the Tables

The recidivism predictors are presented separately for sexual recidivism (Table 1), nonsexual violent recidivism (Table 2), and general (any) recidivism (Table 3). Only predictor variables examined in at least three studies are presented. The primary consideration when estimating the importance of a risk predictor is the size of its correlation with recidivism, as indicated by the median r values and the weighted average (r_{+}). For the prediction of sexual offense recidivism, correlations greater than .30 would be considered large (recidivism rate differences of 30%), correlations greater than .20 moderate, and correlations in the .10 to .20 range small. Correlations less than .10 would have little practical utility in most settings.

The most reliable findings are those that have low variability across studies. If Q is significant, the variability is greater than would be expected by chance. It is important to remember, however, that with large size samples (greater than 1,000) small differences between studies can result in statistically significant Q values. Another check on the variability across studies is the similarity between the weighted average (r_{+}) and the median. When the median and mean suggest substantially different interpretations, then neither result should be considered reliable.

In general, the larger the sample size, the more closely the observed estimates should approximate population values. The influence of sample size is shown in the size of the 95% confidence intervals (the smaller, the better). When the confidence interval does not contain zero, it is equivalent to being statistically significant at $p < .05$, two-tailed. When the confidence intervals for two predictor variables do not overlap, they can be considered statistically different from each other.

Predictors of Sexual Offense Recidivism

Of the demographic variables, only age (young) and marital status (single) were related to the sexual offense recidivism. The effects were small but replicated across many studies. Employment instability and low social class predicted treatment failure only in Maletzky's (1993) study, but his study used an unusually broad measure of failure.

Criminal lifestyle variables appeared to be reliable, although modest, predictors of sexual offense recidivism. The largest of these predictors were antisocial personality disorder ($r_{+} = .14$) and the total number of prior offenses ($r_{-} = .13$).

Many of the sexual criminal history variables showed small to moderate correlations with recidivism. The risk for sexual offense recidivism was increased for those who had prior sexual offenses (.19), had victimized strangers, had an extrafamilial victim, began offending sexually at an early age, had selected male victims, or had engaged in diverse sexual crimes. Neither the degree of sexual contact, or force used, nor injury to victims were significant predictors of sexual offense recidivism.