The Discriminability of Rapists from Non-Sex Offenders Using Phallometric Measures: A Meta-Analysis

MARTIN L. LALUMIÈRE and VERNON L. QUINSEY

Criminal Justice and Behavior 1994; 21; 150
DOI: 10.1177/0093854894021001010

THE DISCRIMINABILITY OF RAPISTS FROM NON-SEX OFFENDERS USING PHALLOMETRIC MEASURES

A Meta-Analysis

MARTIN L. LALUMIÈRE
VERNON L. QUINSEY
Queen's University, Kingston

The authors examined how well identified rapists could be discriminated from non-sex offenders using phallometric assessments, what variables might moderate this discrimination, and whether rapists respond more to descriptions of rape than to consenting sex. Eleven primary and five secondary phallometric studies involving 415 rapists and 192 non-sex offenders were examined using meta-analytic techniques. Study effect sizes averaged 0.82 (95% confidence interval 0.16 to 1.49). Only stimulus set was a statistically significant moderator of effect size: Stimulus sets that contained more graphic rape descriptions produced better discrimination between rapists and non-sex offenders. There was a trend for stimulus sets that contained more exemplars of rape descriptions to achieve better discrimination. Also, rapists responded more to rape than to consenting sex cues in 9 of the 16 data sets and in all 8 of those using the more effective stimulus sets.

The etiology of rape remains controversial (e.g., Palmer, 1988). Among the central issues in this controversy is whether there are individual differences among men in their propensity to engage in sexual coercion. If there were no such individual differences, then

AUTHORS' NOTE: *This research was supported by a scholarship from the Social Sciences and Humanities Research Council of Canada to the first author and by a contract between the second author and the Kingston Psychiatric Hospital. We wish to thank H. E. Barbaree, T. C. Chaplin, C. M. Earls, A. Eccles, K. Freund, A. Gordon,*

CRIMINAL JUSTICE AND BEHAVIOR, Vol. 21 No. 1, March 1994 150-175

explanations of rape would involve only natural and sexual selection as ultimate causes and societal or situational variables as proximal causes (Palmer, 1991; Quinsey, 1984, 1992; Thornhill & Thornhill, 1992; Thornhill, Thornhill, & Dizinno, 1986).

Even if it is accepted that there are individual differences among men in their propensity to commit sexual assaults against women, there is still controversy about the nature of these differences. They could involve risk factors for general criminality (Rice, Harris, & Quinsey, 1990); personality variables (Quinsey, Arnold, & Pruesse, 1980; Walker, Rowe, & Quinsey, 1993); attitudes toward women, sex, and rape (Briere & Malamuth, 1983; Costin & Schwarz, 1987; Stermac, Segal, & Gillis, 1990; Walker et al., 1993); social skill deficits (Overholser & Beck, 1986; Stermac & Quinsey, 1986); or sexual interest in coercion, humiliation, and violence (Quinsey, Chaplin, & Upfold, 1984).

These different sorts of individual differences that potentially underlie the propensity of men to rape are not mutually exclusive and, indeed, might well be correlated with each other. In addition, it is unlikely that their relevance to rape proneness is uniform across all rapists (Prentky & Knight, 1991). For example, situational factors are likely to be more relevant and sadistic sexual interest less relevant for date rapists (Koss, Gidycz, & Wisniewski, 1987) than for predatory serial rapists (MacCulloch, Snowden, Wood, & Mills, 1983).

Sexual interest in coercive sex as a motive for rape is often measured in phallometric assessments. These assessments typically involve the measurement of penile tumescence changes occasioned by audiotaped descriptions of consenting sex and rape. The phallometric investigation of individual differences in the amount of sexual interest in rape has been guided by the sexual-preference hypothesis—that is, that rapists prefer coercive to consenting sex. In phallometric assessments, this hypothesis leads to the prediction that rapists should exhibit

---

G. N. Hall, J. W. Hinton, R. Langevin, J. Looman, B. Malcolm, W. L. Marshall, J. Proulx, S. J. Wormith, and the Correctional Services of Canada for providing us with information about or raw data from their studies, and J. Atkinson, C. M. Earls, G. T. Harris, and M. E. Rice for their comments on an earlier version of this article. *All correspondence should be addressed to Vernon L. Quinsey, Department of Psychology, Queen's University, Kingston, Ontario, Canada, K7L 3N6.*

penile responses of larger magnitude in response to depictions of rape than they do to depictions of consensual sex. Rapists who show this pattern are said to prefer rape to consenting sex.

Although a within-subject criterion can be used to establish preference, it cannot be used to determine whether a pattern of responding is sexually deviant. To establish whether a pattern of phallometric responding is sexually deviant, it must be compared with patterns obtained from "normal" or non-sex-offending men. If both rapists and non-sex offenders equally preferred rape stimuli to consenting sex stimuli, such a preference would be lamentable but not deviant. Thus a within-subject difference can be used to examine preference, but a between-subject difference must be used to examine deviance.

How does the within-subject difference relate to the between-subject difference? That is, how does sexual preference relate to the discriminability of rapists from non-sex offenders with phallometric measures? Two forms of relationship could be posited. The strong form of the sexual-preference hypothesis requires that rapists, as a group, show a clear preference for rape over consenting sex and that non-sex offenders, as a group, show the reverse preference. According to the strong form of the sexual-preference hypothesis, rapists are said to be sexually deviant (using a between-subject criterion) and to prefer rape to consenting sex (using a within-subject criterion).

The weak form of the sexual-preference hypothesis requires only that rapists show a greater preference for rape over consenting sex than do non-sex offenders, a lesser preference to consenting sex over rape, or indifference to the two stimulus categories if non-sex offenders show a preference for consenting sex. Thus, according to the weak form of the sexual-preference hypothesis, rapists are sexually deviant (using a between-group criterion) but may or may not prefer rape over consenting sex (using a within-group criterion).

A number of methodological issues make it difficult to determine which, if either, of the sexual-preference hypotheses is correct. The first pertains to subject selection. Because rapists are heterogeneous in their characteristics (Prentky & Knight, 1991), it is possible that sexual preference is relevant to only some rapists, thus precluding a clear test with a mixed sample. Similarly, groups of "normal" subjects undoubtedly contain a proportion of undetected rapists (Koss et al., 1987; Walker et al., 1993), some of whom may hold deviant sexual

preferences (Malamuth, 1981), and, to the extent that this proportion is high, a test of the sexual-preference hypothesis is compromised.

The second set of methodological issues pertain to the stimuli used in phallometric assessment. Because male sexual fantasies are to some extent idiosyncratic and because the determination of which particular aspects of rape stimuli are differentially sexually arousing to rapists is an empirical issue, it is possible that the stimuli chosen in a particular study are simply irrelevant or inadequate.

The third set of methodological issues involves the setting or demand characteristics of the phallometric assessment. Most subjects do not wish to show sexual arousal to rape stimuli. However, for rapists, who are usually assessed as part of their involvement in the criminal justice system, exhibiting such arousal is likely to have unpleasant consequences. Most subjects, but especially rapists, therefore, attempt to respond more to consenting sex and less to rape stimuli. To the extent that they are successful in dissimulating by modifying their responses (e.g., Quinsey & Chaplin, 1988; Wydra, Marshall, Earls, & Barbaree, 1983), the utility of phallometric assessment and the ability to test the sexual-preference hypothesis is compromised. Because of the nature of the testing situation, therefore, phallometrically measured preference for rape stimuli among rapists is generally easy to interpret, but preference for consenting sexual stimuli is more difficult because it may result from faking. Similar issues are likely responsible for the lower sensitivity than specificity of phallometric assessments in differentiating child molesters from non-child molesters (Freund & Blanchard, 1989).

These methodological issues all result in an underestimation of the degree of rapists' sexual interest in rape and in the difference in such interest observed between non-sex offenders and rapists. Although the true pattern of sexual preferences among rapists can be studied best by employing subjects who are not attempting to deceive the assessor, the utility of phallometric assessment in ordinary assessment situations depends on how well it discriminates rapists from non-sex offenders in general. If the sexual-preference hypothesis is correct, rapists should on average respond relatively more to rape than to consenting sexual stimuli in comparison to non-sex offenders. This does not require that responding to rape be greater than to consenting sex.

The issue of whether rapists are sexually deviant, that is, whether they can be discriminated from non-sex offenders, has generated disagreement in the literature. On the one hand, Quinsey (1984) asserted that "rapists have been shown to be differentiable from nonrapists on the basis of their penile responses to audiotaped descriptions of consenting and forced intercourse in studies employing different stimuli, instructions, rapist samples . . . and comparison groups" (pp. 106-107). On the other hand, Blader and Marshall (1989) concluded that assessment of sexual arousal in rapists is in general not worthwhile and that "arousal patterns do not reliably distinguish rapists from non-rapists" (p. 569). A definitive resolution of this controversy would have obvious theoretical and practical significance.

The present meta-analysis was conducted to determine how well identified rapists can be discriminated from non-sex offenders with phallometric measures; what variables, if any, affect the degree of this discriminability; and to what extent the sexual-preference hypothesis is consistent with the observed differences between rapists and non-sex offenders.

Three variables were investigated for their possible moderator effect on the degree of discriminability of rapists from non-sex offenders. First, recent findings (Proulx, 1992; Quinsey & Chaplin, 1984; Rice, Chaplin, Harris, & Coutts, 1990) showed that the content of rape episodes affects the degree of group discrimination. For example, audiotaped stories that include victim suffering discriminate rapists from non-sex offenders better than do stories that depict the victim as enjoying the rape. The phallometric literature contains three different stimulus sets that have been tested on enough subjects to be compared with each other. The differential effect of these stimulus sets were investigated in this study.

Second, from measurement theory it is expected that the number of exemplars or stimuli per category affects the reliability of phallometric measures and, therefore, their discriminative potential (Anastasi, 1988). Thus number of rape stimuli was also used as a moderator variable.

Finally, phallometric studies have been conducted in different experimental and clinical settings, which can be grouped into the two major categories of mental health and correctional settings. The effect

of this variable was also investigated. In addition, we also included publication status and year of publication as potential moderator variables.

METHOD

STUDIES

Literature Search

Studies were identified using *Psychological Abstracts* and *Social Science Citation Index* computerized databases and by examining reference lists of relevant studies and review articles. Senior authors who had conducted phallometric studies of rapists were contacted to locate unpublished studies.

Inclusion Criteria

Only studies meeting the following five criteria were included.

1. Subjects. At least one group of identified rapists and at least one group of non-sex offenders had to have been included. Rapists were defined as men who had committed a sexual offense against an adult female involving forceful physical contact. Victims were predominantly over 16 years of age. For the purpose of this meta-analysis, a non-sex offender subject was defined as an individual with no record of sexual offending at all and who reported committing no sexual assaults against women; this definition embraces subjects from the student or general population, and criminal offenders who have committed no sex offenses.

2. Stimulus categories. At least one stimulus category of mutually consenting sex between an adult male and an adult female and at least one stimulus category depicting a sexually aggressive interaction between a male protagonist and a nonconsenting female had to have

been included. Each subject within an individual study had to have been exposed to the same stimuli.

3. Stimulus modality. Audiotaped or videotaped stimuli had to be used. They were considered the best modalities of presentation of the two stimulus categories described above.

4. Instructions to subjects. Studies of voluntary control of penile responses or studies using nonstandard instructions, such as "maintain arousal" and "suppress arousal," were excluded, as were groups within studies given special instructions or special experimental conditions (Quinsey, Chaplin, & Varney, 1981; Wydra et al., 1983).

5. Dependent variable. Sufficient details about penile responses had to be included to allow the computation of a mean (and standard deviation) rape index (see below) for rapists and non-sex offenders separately; missing information was requested from authors when necessary.

Selected Studies

Table 1 shows the 12 studies meeting all five criteria (hereinafter referred to as *primary studies*). Ten of these are published and all used auditory stimuli.

Group means and standard deviations in Study 4 (Freund, Sher, Racansky, Campbell, & Heasman, 1986) were not available. The effect size value (see below) of that study could be derived easily (see Rosenthal, 1984) from the phi coefficient value (0.53) calculated using the proportion of rapists (58%) and normals (8%) showing a rape index greater than 1 (Freund et al., 1986, Table 1).

Also included in Table 1 are studies meeting all criteria except the inclusion of a control group of non-sex offenders (hereinafter referred to as *secondary studies*). However, rapists from these studies could be compared with control subjects from primary studies that had the same setting, stimulus set, and instructions. The comparison figures are italicized. Rapists from Studies 13, 15, 16, and 17a were compared with a combined group of non-sex offenders from Baxter, Barbaree, and Marshall (1986), and Barbaree, Marshall, and Lanthier (1979).

TABLE 1: Statistics and Coding Values for Selected Studies[a]

Authors/Date	Rape Index[b] Rapists	Rape Index[b] Controls	*d*	r_b	CL	Stimulus Set[c]	Setting[d]	Number of Exemplars
Primary studies								
1. Rice et al. (1990)	1.55 (0.47) 10	0.42 (0.26) 13	3.14	0.85***	.98	OR	MH	16
2. Quinsey & Chaplin (1984)	1.44 (0.62) 15	0.42 (0.40) 15	1.97	0.71***	.92	OR	MH	16
3. Earls & Proulx (1986)	1.64 (0.98) 10	0.54 (0.29) 10	1.51	0.62**	.86	A	MH	2
4. Freund et al. (1986)	NA NA 12	NA NA 12	1.25	0.55**	NA	O	C + MH	4
5. Quinsey et al. (1981)	1.28 (0.74) 20	0.56 (0.42) 20	1.20	0.52***	.80	OR	MH	5
6. Quinsey et al. (1984)	2.09 (2.04) 20	0.37 (0.24) 20	1.19	0.52***	.80	OR	MH	4

continued

TABLE 1: continued

Authors/Date	Rape Index[b] Rapists	Rape Index[b] Controls	*d*	r_b	*CL*	Stimulus Set[c]	Setting[d]	Number of Exemplars
7. Proulx et al. (1992)	1.47 (1.06) 10	0.60 (0.25) 10	1.13	0.51*	.79	A	MH	4
8. Marshall (1988)	0.71 (0.30) 23	0.43 (0.20) 24	1.10	0.49***	.78	B	C	3
9. Abel et al. (1978)	1.41 (1.29) 19	0.59 (0.65) 15	0.78	0.37*	.72	A	O	2
10.Barbaree et al. (1979)	0.70 (0.47) 10	0.41 (0.28) 10	0.75	0.37	.70	B	C	6
11. Baxter et al. (1986)	0.57 (0.27) 50	0.47 (0.33) 31	0.32	0.16	.59	B	C	6
12. Langevin et al. (1985)	1.66 (2.88) 16	1.41 (2.26) 15	0.10	0.05	.53	O	MH	5

Secondary studies[e]								
13. Wydra et al. (1983)	0.89	0.46	1.36	0.58**	.83	B	C	4
	(*0.31*)	(*0.32*)						
	10	10						
14. Quinsey & Chaplin (1982)	1.64	0.56	1.33	0.56***	.83	OR	MH	5
	(1.03)	(*0.42*)						
	24	*20*						
15. Eccles (1990)	0.62	*0.41*	0.62	0.28*	.66	B	C	6
	(0.40)	(*0.32*)						
	19	*41*						
16. Malcolm (1992)	0.61	*0.41*	0.56	0.26**	.66	B	C	3
	(0.39)	(*0.32*)						
	83	*41*						
17a. Looman (1989)	0.57	*0.41*	0.35	0.17	.61	B	C	3
	(0.50)	(*0.32*)						
	68	*41*						
17b. Looman (1989)	1.10	*0.56*	0.30	0.12	.60	OR	C	5
	(2.01)	(*0.42*)						
	79	*20*						

NOTE: d = effect sizes; r_b = point-biserial correlation; *CL* = common language effect size.
a. Studies 1, 7, 15, 16, 17a, and 17b are unpublished studies.
b. Numbers represent means, standard deviations (in parentheses), and sample size, respectively.
c. B = Barbaree series, OR = Oak Ridge series, A = Abel series, and O = other.
d. MH = mental health setting, C = correctional setting, and O = other.
e. Italicized numbers represent values obtained from control subjects from primary studies.
*$p < .05$; **$p < .01$; ***$p < .001$.

Rapists from Studies 14 and 17b were compared with non-sex offenders from Quinsey et al. (1981). Studies 13 and 14 are published studies.

Data from Study 13 (Wydra et al., 1983) are from Experiment 2, normal instructions. Standard deviations could not be found and were inferred from the data of a large sample of rapists and non-sex offenders who were assessed at the same time period using the same stimulus set and instructions (Barbaree et al., 1979, combined with Baxter et al., 1986). This study therefore was considered a secondary study. Data from Study 15 (Eccles, 1990) are from Experiment 5 using the stimulus set of Barbaree et al. (1979). Data from Study 16 (Malcolm, 1992) are from 83 rapists that were assessed pretreatment between 1983 and 1985 using Barbaree's stimulus set. Data from Studies 17a and 17b (Looman, 1989) are from rapists who were assessed pretreatment in the 1980s; 68 rapists (Study 17a) were assessed using the stimuli of Barbaree et al. (1979), and 79 rapists (Study 17b) were assessed using the stimuli of Quinsey et al. (1981).

Studies Excluded

Five studies included a group of identified rapists and a group of non-sex offenders but did not meet one of the other inclusion criteria: Hinton, O'Neil, and Webster (1980, Criteria 2 and 5); Kercher and Walker (1973, Criteria 2 and 3); Murphy, Krisak, Stalgaitis, and Anderson (1984, Criterion 4[1]; Wormith (1986, Criteria 2 and 3); and Wormith, Bradford, Pawlak, Borzecki, and Zohar (1988, Criterion 5, no standard deviations available). A sixth study of rapists (Hall, 1989) met all criteria except the inclusion of a control group of non-sex offenders; rapists in that study could not be compared with controls from other studies that had the same setting, stimulus set, and instructions. As explained next, four other studies were excluded because of sample replication.

Sample Replication

Each study except Studies 17a and 17b is based on independent samples. In the original papers, the 10 rapists and 10 controls from Barbaree et al. (1979) were also used in Baxter et al. (1986), and the 20 rapists in Quinsey et al. (1981) were also used in Quinsey and

Chaplin (1982); therefore, in this meta-analysis, Baxter et al. data are from the 81 new subjects, and Quinsey and Chaplin data are from the 24 new rapists. Four other studies were excluded because subjects were subsequently used in other studies: Rapists in Baxter, Marshall, Barbaree, Davidson, and Malcolm (1984) were reported on in all or in part in Barbaree, Baxter, and Marshall (1989); Davidson and Malcolm (1985); and Study 11 (Baxter et al., 1986). The rapists in Abel, Barlow, Blanchard, and Guild (1977) were reported on in Study 9 (Abel, Blanchard, Becker, & Djenderedjian, 1978). Note that for the latter study, all subjects were assessed using standardized stimuli (Abel et al., 1978, p. 327). Studies 17a and 17b (Looman, 1989) are presented separately for the purpose of meta-analyses of moderator variables (see below) but, in fact, the two data sets overlap. The bare-bones meta-analysis used a weighted average effect size (0.32) and sample size ($n = 74$) from these two data sets. There are therefore 12 independent primary studies in Table 1 and an additional 5 independent secondary studies.

SUBJECT EXCLUSION

Researchers have traditionally excluded subjects based on low responding. In this meta-analysis, only subjects used in the statistical analyses of the original papers were included.

RAPE INDEX

The dependent measure was the rape index, originally developed by Abel and his associates (Abel et al., 1977). This index provides a measure of arousal to rape stimuli relative to consenting stimuli. The rape index was calculated by dividing the average response to all heterosexual rape stimuli by the average response to all mutually consenting heterosexual stimuli. A rape index greater than 1.00 indicates greater responding to rape stimuli.

Although other indexes have greater discriminative power (e.g., difference scores calculated from standard scores; see Harris, Rice, Quinsey, Chaplin, & Earls, 1992), this particular rape index was used because it was available from all studies. Indeed, standard scores could be unstable when non neutral stimuli were used, and raw difference

scores could not be calculated when individual data were not available (Studies 4, 8, 12, and 13). It is likely, therefore, that effect sizes reported in this meta-analysis underestimate between-group differences.

Another potentially useful approach would have been to use study F values for Stimulus Category × Subject Status interaction (e.g., Benassi & Belli, 1989; Rosenthal, 1984). This could not be done here because of lack of information in some studies and design differences between studies.

EFFECT SIZE

A standardized statistic of effect size, d, was calculated for each study (Cohen, 1988). This statistic expresses between-group differences in standard deviation units and is calculated by subtracting the mean rape index of the non-sex offender group (m_2) from the mean rape index of the rapist group (m_1) and then dividing this difference by the pooled standard deviation (s_w):

$$d = \frac{m_1 - m_2}{s_w}$$

s_w was used because it is a better estimate of the true population variability when there is no experimental or treatment manipulation of subjects (Hunter & Schmidt, 1990). A positive d means that rapists have a higher rape index than do non-sex offenders.

CODING OF STUDY CHARACTERISTICS

Five study characteristics were coded for their potential effect as moderator variables: stimulus set, setting of the experiment, number of rape exemplars used for calculating individual rape indexes, year of publication, and publication status. The coding values are presented in Table 1.

META-ANALYTIC STRATEGY

Following Hunter and Schmidt (1990; see also Durlak & Lipsey, 1991, for a simpler description), the procedure involved three steps.

First, a bare-bones meta-analysis was conducted in which an average of all study effect sizes, δ, was calculated, and 95% confidence limits were calculated to assess statistical significance; if the limits included zero, δ was considered to be nonsignificant.

Second, the expected sampling error variance (V_e; Hunter & Schmidt, 1990) of the mean of the *d*s was calculated and compared with the observed variance (V_o) across the *d*s. Following the Hunter-Schmidt 75% rule, if V_e accounted for less than 75% of V_o, it was concluded that one or more moderator variables produced variance among studies. If, however, V_e accounted for 75% or more of the V_o, then it was concluded that all residual variance was a product of artifacts such as error of measurement. This 75% rule has been found to be a powerful method for detecting moderator variables (Sackett, Harris, & Orr, 1986).

Third, a search for moderator variables was undertaken when the 75% criterion was not met. Continuous candidate variables (year of publication, number of rape stimulus exemplars) were assessed using the Pearson *r* product-moment correlation coefficient, and discrete candidate variables (stimulus set, setting, publication status) were assessed by calculating confidence limits around the mean of *d*s for each level of each variable. If the mean of *d*s of a level of a particular candidate variable was not included in the confidence limits of another level of the same variable, its status as a moderator variable was inferred. Tests of significance for each level of the moderator variable, if present, were conducted as well by noting if the confidence limits included zero.

Ninety-five percent confidence limits were calculated by adding and subtracting to δ the product of 1.96 and the population standard deviation, S_δ which is the observed standard deviation minus the expected sampling error standard deviation. A negative S_δ was considered to be zero.

PRELIMINARY ANALYSIS

A Pearson correlation coefficient was calculated between primary study effect size and the total number of subjects in each study. There

was a nonsignificant trend for studies with larger effect sizes to have a smaller number of subjects, $r(10) = -0.41$. Because of this trend, the results presented later are weighted by sample size. However, all analyses were also conducted without weighting by sample size, and the results were identical, with one exception noted later.

Outliers

A search for outliers was conducted using the Fourth-spread method (Hoaglin, Mosteller, & Tukey, 1983). Effect sizes that deviated by a proportion of 1.5 of the interquartile-range from the first and third quartile were considered outliers. Two searches were conducted, one on the 12 primary studies only, and one on all 17 studies (one data point for Studies 17a and 17b). Both searches found that one effect size (Study 1) deviated from the cutoffs (–0.37 and +2.64, for primary studies). Study 1 was therefore excluded from further analyses. This resulted in smaller δ and S_δ. The correlation between primary study effect size and sample size based on the 11 remaining studies was unaltered.

RESULTS

Mean rape indexes, standard deviations, and sample sizes are presented in Table 1. Studies are presented in decreasing order of their effect size (*d*). For illustrative purposes, the biserial correlation (r_b) between group membership and individual rape index, and the corresponding significance level, are also reported; r_b can be directly obtained from *d* (Hunter & Schmidt, 1990). Also included is the newly developed common language effect size statistic (*CL*; McGraw & Wong, 1992); *CL* is the probability that a randomly chosen member of the rapist group has a higher rape index than a randomly chosen member of the non-sex offender group. In the absence of between-group differences, *CL* approaches .50. The 16 independent studies (excluding Study 1) contained 415 rapists and 192 non-sex offenders.

Eight of the 11 primary studies and 4 of the 6 secondary studies showed statistical significance in the direction of positive discrimination. All studies had positive effect sizes. Also, 9 of the 16 studies

(excluding Freund et al., 1986, for lack of information) reported an average rape index greater than 1.00 for rapists, and all but one study reported an average rape index smaller than 1.00 for non-sex offenders.

Study 12 (Langevin et al., 1985) stands alone in Table 1 by the fact that non-sex offenders show greater responding to rape stimuli compared to mutually consenting stimuli. It is also the only study in which standard deviations for both groups were greater than the corresponding averages. Excluding Study 12, the range of the rape index for non-sex offenders was quite narrow: 0.37 to 0.60. The corresponding value for rapists was much wider: 0.57 to 2.09. Interestingly, larger effect sizes (between-group differences) seemed to be due to larger rape indexes among the rapist groups rather than to variation among the non-sex offender groups.

BARE-BONES META-ANALYSES

Two meta-analyses were conducted, one including only the 11 primary studies, and one adding the 5 independent secondary studies (one data point for Studies 17a and 17b). The sample size values used in the meta-analysis of the primary and secondary studies differed. In the primary studies, sample size was the total number of subjects, and in the secondary studies (except for Study 13, which contains a control group), sample size was the total number of rapists. Sample duplication was avoided.

Weighted average effect size values, confidence limits, and V_es are presented in Table 2. Both analyses showed that the confidence limits do not include zero and that there were likely one or more moderator variables producing variance across studies. An effect size of 0.82 and larger is considered to be large (Cohen, 1988). Although one or more moderator variables can explain part of the variance between studies, these results indicate that rapists can be discriminated from non-sex offenders.[2]

ANALYSES OF MODERATOR VARIABLES

Only two of the postulated moderator variables, stimulus set and number of rape exemplars, were found to have explanatory value.

TABLE 2: Weighted Average *d* Values, 95% Confidence Limits, and Percentage of Variance Due to Sampling Error for the Bare-Bones Meta-Analyses and as a Function of Stimulus Set and Setting

Primary Studies	*All Studies*
Bare-bones meta-analysis	
11 studies	16 studies
205 rapists, 182 non-sex offenders	415 rapists, 192 non-sex offenders
0.22 < 0.93 < 1.64	0.16 < 0.82 < 1.49
50.6%	51.1%
Oak Ridge stimuli	
3 studies	5 studies
55 rapists, 55 non-sex offenders	158 rapists, 55 non-sex offenders
1.40	0.05 < 0.99 < 1.92
100%	32.6%
Barbaree stimuli	
3 studies	7 studies
83 rapists, 65 non-sex offenders	263 rapists, 75 non-sex offenders
0.25 < 0.63 < 1.00	0.38 < 0.60 < 0.82
70.5%	87.7%
Abel stimuli	
3 studies	
39 rapists, 35 non-sex offenders	
1.07	
100%	
Mental health setting	
6 studies	7 studies
91 rapists, 90 non-sex offenders	115 rapists, 90 non-sex offenders
0.40 < 1.16 < 1.92	0.53 < 1.18 < 1.83
52.8%	61.3%
Correctional setting	
3 studies	7 studies
83 rapists, 65 non-sex offenders	269 rapists, 75 non-sex offenders
0.25 < 0.63 < 1.00	0.33 < 0.59 < 0.84
70.5%	84.3%

Stimulus Set

There were three separate stimulus sets. One from Quinsey, Rice, Harris, and their colleagues (hereinafter referred to as Oak Ridge); one from Barbaree and his colleagues; and one from Abel and his colleagues. The last stimulus set has also been translated into French (Earls & Proulx, 1986; Proulx, Aubut, McKibben, & Côté, 1992). Note that, as part of their study, Proulx et al. (1992) used a translation of

Abel's stimulus set as well as their own stimuli. The rape index calculated from that study is based on the average of all rape stimuli; we still included that study in the Abel group. Stimulus sets from Studies 4 and 12 were not used in more than one study, and these studies therefore are not included in this analysis.

Weighted average *d* values for each stimulus set, confidence limits, and V_es also are presented in Table 2. None of the confidence limits included zero, indicating that rapists can be discriminated from non-sex offenders with all stimulus sets.

In the case of primary studies only, all observed variance in the Oak Ridge and Abel stimulus sets was due to sampling error alone, indicating that there is no other moderator variable present. Regarding the Barbaree stimulus set, the 75% criterion was almost reached; note that the unweighted analysis (not presented in Table 2) showed that sampling error accounted for 87.7% of the variance. Confidence limits show that the Oak Ridge and Abel stimulus sets produced significantly higher δs than the Barbaree stimulus set.

When secondary studies were included, a similar picture emerged, except that there was likely one or more moderator variables in the Oak Ridge stimulus set. The small number of studies in that group prevented further analyses, although it is apparent that Study 17b (Looman, 1989; $d = 0.30$) was responsible for the observed variance. This study comes from a correctional setting. When Looman (1989) is excluded, all observed variance resulted from sampling error and δ equalled 1.42.

CL values for the three stimulus sets varied greatly. The Oak Ridge stimulus set yielded weighted values of .82 (primary studies only) and .74 (all studies); Barbaree yielded values of .59 and .58; and Abel produced a value of .78 (primary studies).

Number of Exemplars

A Pearson *r* was calculated between effect sizes and number of rape stimuli exemplars. For primary studies, there was a nonsignificant trend for larger effect sizes to be associated with greater number of exemplars, $r(9) = 0.40$. The same pattern was observed when all studies were included, $r(15) = 0.37$. Statistical power obviously is lacking here.

Setting

Table 2 also presents results for two types of setting, mental health and correctional. Studies that could not be classified unambiguously were excluded (Studies 4 and 9). None of the confidence limits were found to include zero. There was overlap between the confidence limits of the two settings, indicating that they did not produce different results.[3] There was likely another moderator variable for studies in mental health settings. However, when Study 12 (Langevin et al., 1985; other stimulus set) was deleted, all observed variance in the analyses of both the primary studies and all studies together could be accounted for by sampling error alone and δ equalled 1.39.

Year of Publication

There were 16 dated reports. There was no significant association between effect sizes and the dates of publication, $r(9) = +0.12$, primary studies only; and $r(14) = -0.22$, all studies.

Publication Status

Twelve studies were published in scholarly journals or books. The average effect size value for published studies, including primary and secondary studies, was 0.96; the corresponding value for the four unpublished reports was 0.53. The difference was not significant. None of the confidence limits included zero.

DISCUSSION

Several conclusions can be drawn from the above analyses. First, rapists respond differently than non-sex offenders on phallometric tests. Second, the weak form of the sexual-preference hypothesis is supported: Rapists, as a group, respond more to rape cues than to consenting sex cues in comparison to non-sex offenders, and non-sex offenders prefer consenting sex to rape. The strong form of the sexual-preference hypothesis is also supported but only with the use of graphic and brutal rape stimuli (the Oak Ridge and Abel stimulus

set). Third, stimulus sets that employ graphic and brutal rape stimuli and multiple rape exemplars discriminate rapists from non-sex offenders more effectively.

In connection with the conclusions from these analyses, the identification of Study 1 (Rice, Chaplin, Harris, & Coutts, 1990) as an outlier is worth commenting on. This study produced an unusually high effect size (3.14), occasioned by low variance in each group. The low variance is likely caused by the fact that this study used 16 rape exemplars. The only other study that used a high number of exemplars (Quinsey & Chaplin, 1984) also obtained a very high effect size value (1.97). Note also that the best discrimination between rapists and non-sex offenders was obtained in Rice, Chaplin, Harris, and Coutts (1990) when the rape story was told from the female point of view and when the female experienced pain. It is therefore likely that the large effect size obtained is not entirely artifactual but, rather, reflects the power of phallometric assessment when the quality and the quantity of stimuli are optimal.

LIMITATIONS

Several methodological issues potentially weaken some of the conclusions reached in this study. The first issue involves the partial confounding of stimulus set and study setting. Most of the studies employing the Oak Ridge stimulus set were conducted in a maximum-security psychiatric institution, and all of those using the Barbaree stimulus set were conducted in an outpatient or a correctional setting. Although the rapists studied in these different settings have very similar characteristics (Harris et al., 1992), and most of the rapists assessed in the maximum-security setting remain there for only a month before going to a correctional setting (Quinsey & Maguire, 1983), the issue is still of some concern. Only Looman's (1989) secondary study used both the Oak Ridge and the Barbaree stimulus sets in the same (correctional) setting. As would be expected from this meta-analysis, the rape index that Looman obtained with the Oak Ridge stimulus set was almost twice as large as the rape index obtained with the Barbaree set, $t(145) = 2.12$, $p < .05$. Nevertheless, a direct comparison of the two stimulus sets in the same study using both rapists and non-sex offenders is required to lay the issue to rest.

The second issue pertains to our interpretation of the stimulus set effect. The Oak Ridge stimulus set actually is composed of four different audiotaped stimulus series. These, however, were all developed using the same logic of maximizing the difference between consenting and rape stimuli by making the rape stimuli very brutal and graphic and of using multiple exemplars per category. The analyses reported here indicate that all of the differences between these audiotapes can be accounted for by sampling error, at least with the number of subjects tested. Our interpretation of the difference in effectiveness between the Barbaree stimulus set and the Oak Ridge and Abel stimulus sets is that the latter two contain more graphic and brutal stimuli. Although it is clear from individual studies (Quinsey & Chaplin, 1984; Quinsey et al., 1984; Rice, Chaplin, Harris, & Coutts, 1990), that more brutal rape scenarios are associated with better discrimination of non-sex offenders from rapists, no quantification of stimulus set brutality was related to discriminative ability in the present study.

Third, the question of whether the results of this study generalize to the small percentage of subjects who display low levels of responding needs to be addressed in future studies. Many studies included in this meta-analysis excluded subjects who did not achieve a minimum criterion varying between 10% and 20% of full erection. There are results that suggest, however, that sex offenders, including rapists, can be discriminated from non-sex offenders even when level of responding is low (Harris et al., 1992).

Fourth, the number of studies included in this meta-analysis was relatively small. However, the total number of subjects used in the analyses was quite large (415 rapists and 192 non-sex offenders). In addition, the residual unexplained variance was quite low after accounting for only one moderator variable (stimulus set), even without controlling for imperfect measurement reliability. Also, we have conducted sensitivity analyses based on individual data from a subset of the studies used in this meta-analysis (Lalumière & Quinsey, in press). This type of analysis is not based on the same statistical rationale as meta-analysis, but it produced very similar results: Individual studies showed consistent classification patterns (i.e., detection of deviant sexual preferences in rapists samples), and interstudy differences in classification patterns could be explained with the same

moderator variable (stimulus set). Thus, when the specificity level was set at 90% (i.e., 90% of non-sex offenders were considered to be sexually nondeviant, based on the rape index), 69%, 77%, and 20% of rapists were considered sexually deviant using the Oak Ridge, Abel, and Barbaree stimulus sets, respectively.

The last issue concerns the use of the rape index as the dependent variable. It remains to be seen how much the discriminative ability of phallometric assessments has been underestimated in this study by using the rape index ratio instead of the more optimal difference *z*-score method (Harris et al., 1992).

CONCLUSION

The analyses presented here support the idea that there are substantial individual differences between men in the amount of sexual interest they have in coercive sex and that the degree of this interest is related to the likelihood of their becoming rapists. Although the data analyzed in the present study are all postdictive and thus correlational in nature, rape indexes also predict sexual offending in follow-up studies of rapists (Rice, Harris, & Quinsey, 1990). The results of the present study support the use of phallometric assessment in identifying treatment needs and in assessing the risk of recidivism among identified rapists, provided that appropriate stimulus sets and scoring procedures are used.

Although the results of these analyses speak to the sexual motivation of rapists, they do not preclude differences in motives between rapist subtypes. Moreover, there are probably other explanatory variables at different levels (i.e., evolutionary, societal, and situational) or of different kinds (e.g., attitudinal) that must be woven together to provide a complete account of rape.

NOTES

1. Murphy, Krisak, Stalgaitis, and Anderson (1984) used alternate instructions to "maintain" and "suppress" arousal to mutually consenting and rape videotaped scenarios. Under maintain arousal instructions, the 14 rapists and 15 non-sex offenders had a mean rape index of .93 and .77, respectively. The effect size was +0.28.

2. An alternative method of performing meta-analytic calculations, using the software package (DSTAT) prepared by Johnson (1989), confirmed the robustness of our principal analyses. This package uses an estimation of the variance of the normal distribution of d to calculate confidence limits and the chi-square test of homogeneity (Q_w) to detect moderator variables (Hedges & Olkin, 1985). The weighted mean value of d (corrected for small sample size bias) was 0.93, with a 95% confidence interval range of 0.72 to 1.14, for the 12 primary studies. The test for homogeneity was significant, $Q_w(11) = 32.36$, $p < .001$, indicating the presence of one or more moderator variables, and the largest outlier was Study 1. The comparable values without Study 1 were $d = 0.86$ (0.65 to 1.08) and $Q_w(10) = 20.33$, $p < .05$.

3. All mean comparisons presented in this article were also performed using t tests, which use pooled sampling error variance. In the cases where residual variance remained after accounting for expected sampling variance, t tests ($\alpha = 0.05$) produced the same results as confidence limits, except in the case of the comparison between mental health and correctional settings using all studies, $t(13) = 4.41$, $p < .001$.

REFERENCES

Abel, G. G., Barlow, D. H., Blanchard, E. B., & Guild, D. (1977). The components of rapists' sexual arousal. *Archives of General Psychiatry, 34*, 895-903.

Abel, G. G., Blanchard, E. B., Becker, J. V., & Djenderedjian, A. (1978). Differentiating sexual aggressives with penile measures. *Criminal Justice and Behavior, 5*, 315-332.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Barbaree, H. E., Baxter, D. J., & Marshall, W. L. (1989). The reliability of the rape index in a sample of rapists and nonrapists. *Violence and Victims, 4*, 299-307.

Barbaree, H. E., Marshall, W. L., & Lanthier, R. D. (1979). Deviant sexual arousal in rapists. *Behaviour Research and Therapy, 17*, 215-222.

Baxter, D. J., Barbaree, H. E., & Marshall, W. L. (1986). Sexual responses to consenting and forced sex in a large sample of rapists and nonrapists. *Behaviour Research and Therapy, 24*, 513-520.

Baxter, D. J., Marshall, W. L., Barbaree, H. E., Davidson, P. R., & Malcolm, P. B. (1984). Deviant sexual behavior: Differentiating sex offenders by criminal and personal history, psychometric measures, and sexual response. *Criminal Justice and Behavior, 11*, 477-501.

Benassi, V. A., & Belli, R. F. (1989). Statistics and replication experiments: Comparing interaction effects. *Cognitive Therapy and Research, 13*, 283-294.

Blader, J. C., & Marshall, W. L. (1989). Is assessment of sexual arousal in rapists worthwhile? A critique of current methods and the development of a response compatibility approach. *Clinical Psychology Review, 9*, 569-587.

Briere, J., & Malamuth, N. M. (1983). Self-reported likelihood of sexually aggressive behavior: Attitudinal versus sexual explanations. *Journal of Research in Personality, 17*, 315-323.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

Costin, F., & Schwarz, N. (1987). Beliefs about rape and women's social roles: A four-nation study. *Journal of Interpersonal Violence, 2*, 46-56.

Davidson, P. R., & Malcolm, P. B. (1985). The reliability of the rape index: A rapist sample. *Behavioral Assessment, 7*, 283-292.

Durlak, J. A., & Lipsey, M. W. (1991). A practitioner's guide to meta-analysis. *American Journal of Community Psychology, 19*, 291-332.

Earls, C. M., & Proulx, J. (1986). The differentiation of francophone rapists and nonrapists using penile circumferential measures. *Criminal Justice and Behavior, 13*, 419-429.

Eccles, T. (1990). *The clinical utility of the Rape Index.* Unpublished doctoral dissertation, Queen's University, Kingston.

Freund, K., & Blanchard, R. (1989). Phallometric diagnosis of pedophilia. *Journal of Consulting and Clinical Psychology, 57*, 100-105.

Freund, K., Sher, H., Racansky, I. G., Campbell, K., & Heasman, G. (1986). Males disposed to commit rape. *Archives of Sexual Behavior, 15*, 23-25.

Hall, G.C.N. (1989). Sexual arousal and arousability in a sexual offender population. *Journal of Abnormal Psychology, 98*, 145-149.

Harris, G. T., Rice, M. E., Quinsey, V L., Chaplin, T. C., & Earls, C. M. (1992). Maximizing the discriminant validity of phallometric data. *Psychological Assessment, 4*, 502-511.

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

Hinton, J. W., O'Neil, M. T., & Webster, S. (1980). Psychophysiological assessment of sex offenders in a security hospital. *Archives of Sexual Behavior, 3*, 205-216.

Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (1983). *Understanding robust and exploratory data analysis.* New York: Wiley.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings across studies.* Newbury Park, CA: Sage.

Johnson, B. T. (1989). *DSTAT: Software for the meta-analytic review of research literatures.* Hillsdale, NJ: Lawrence Erlbaum.

Kercher, G. A., & Walker, C. E. (1973). Reactions of convicted rapists to sexually explicit stimuli. *Journal of abnormal Psychology, 81*, 46-50.

Koss, M. P., Gidycz, C. A., & Wisniewski, N. (1987). The scope of rape: Incidence and prevalence of sexual aggression and victimization in a national sample of higher education students. *Journal of Consulting and Clinical Psychology, 55*, 162-170.

Lalumière, M. L., & Quinsey, V. L. (in press). The sensitivity of phallometric measures with rapists. *Annals of Sex Research.*

Langevin, R., Bain, J., Ben-Aron, M. H., Coulthard, R., Day, D., Roper, V., Russon, A. E., Webster, C. D., & Wortzman, G. (1985). Sexual aggression: Constructing a predictive equation: A controlled pilot study. In R. Langevin (Ed.), *Erotic preference, gender identity, and aggression in men: New research studies* (pp. 39-76). Hillsdale, NJ: Lawrence Erlbaum.

Looman, J. (1989). *Rapists/pedophiles: A distinct group?* Unpublished honours thesis, University of Saskatchewan, Saskatoon.

MacCulloch, M. G., Snowden, P. R., Wood, P.J.W., & Mills, H. E. (1983). Sadistic fantasy, sadistic behaviour, and offending. *British Journal of Psychiatry, 143*, 20-29.

Malamuth, N. M. (1981). Rape proclivity among males. *Journal of Social Issues, 37*, 138-157.

Malcolm, B. (1992). *Rapists assessment data from the Sex Offender Treatment Program of the Regional Treatment Centre, Kingston.* Unpublished manuscript.

Marshall, W. L. (1988). The use of sexually explicit stimuli by rapists, child molesters, and nonoffenders. *Journal of Sex Research, 25*, 267-288.

McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin, 111*, 361-365.

Murphy, W. D., Krisak, J., Stalgaitis, S., & Anderson, K. (1984). The use of penile tumescence measures with incarcerated rapists: Further validity issues. *Archives of Sexual Behavior, 13*, 545-554.

Overholser, J. C., & Beck, S. (1986). Multimethod assessment of rapists, child molesters, and three control groups on behavioral and psychological measures. *Journal of Consulting and Clinical Psychology, 54*, 682-687.

Palmer, C. T. (1988). Twelve reasons why rape is not sexually motivated: A sceptical examination. *Journal of Sex Research, 25*, 512-530.

Palmer, C. T. (1991). Human rape: Adaption or by-product? *Journal of Sex Research, 28*, 365-386.

Prentky, R. A., & Knight, R. A. (1991). Identifying critical dimensions for discriminating among rapists. *Journal of Consulting and Clinical Psychology, 59*, 643-661.

Proulx, J. (1992), June). *Les réponses péniennes de violeurs à deux types de stimuli de viol.* Paper presented at the annual convention of the Canadian Psychological Association, Québec City.

Proulx, J., Aubut, J., McKibben, A., & Côté, M. (1992). *Penile responses of rapists and nonrapists to rape stimuli involving physical violence or humiliation.* Manuscript submitted for publication.

Quinsey, V. L. (1984). Sexual aggression: Studies of offenders against women. In D. N. Weisstub (Ed.), *Law and mental health: International perspectives* (pp. 84-121). New York: Pergamon.

Quinsey, V. L. (1992). Individual differences in the propensity to rape. *Behavioral and Brain Sciences, 15*, 400.

Quinsey, V. L., Arnold, L. S., & Pruesse, M. G. (1980). MMPI profiles of men referred for pre-trial psychiatric assessment as a function of offense type. *Journal of Clinical Psychology, 36*, 410-417.

Quinsey, V. L., & Chaplin, T. C. (1982). Penile responses to nonsexual violence among rapists. *Criminal Justice and Behavior, 9*, 372-384.

Quinsey, V. L., & Chaplin, T. C. (1984). Stimulus control of rapists' and non-sex offenders' sexual arousal. *Behavioral Assessment, 6*, 169-176.

Quinsey, V. L., & Chaplin, T. C. (1988). Preventing faking in phallometric assessments of sexual preference. In R. A. Prentky & V. L. Quinsey (Eds.), *Human sexual aggression: Current perspectives* (pp. 49-58). New York: Annals of the New York Academy of Sciences.

Quinsey, V. L., Chaplin, T. C., & Upfold, D. (1984). Sexual arousal to nonsexual violence and sadomasochistic themes among rapists and non-sex-offenders. *Journal of Consulting and Clinical Psychology, 52*, 651-657.

Quinsey, V. L., Chaplin, T. C., & Varney, G. (1981). A comparison of rapists' and non-sex offenders' sexual preferences for mutually consenting sex, rape, and physical abuse of women. *Behavioral Assessment, 3*, 127-135.

Quinsey, V. L., & Maguire, A. (1983). Offenders remanded for a psychiatric examination: Perceived treatability and disposition. *International Journal of Law and Psychiatry, 6*, 193-205.

Rice, M. E., Chaplin, T. C., Harris, G. T., & Coutts, J. (1990). *Empathy for the victim and sexual arousal among rapists and nonrapists* (Report No. 8). Penetanguishene, Ontario: Mental Health Research Centre.

Rice, M. E., Harris, G. T., & Quinsey, V. L. (1990). A follow-up of rapists assessed in a maximum security psychiatric facility. *Journal of Interpersonal Violence, 5*, 435-448.

Rosenthal, R. (1984). *Meta-analytic procedures for social research.* Beverly Hills, CA: Sage.

Sackett, P. R., Harris, M. M., & Orr, J. M. (1986). On seeking moderator variables in the meta-analysis of correlational data: A Monte Carlo investigation of statistical power and resistance to Type 1 error. *Journal of Applied Psychology, 71*, 302-310.

Stermac, L. E., & Quinsey, V. L. (1986). Social competence among rapists. *Behavioral Assessment, 8*, 171-185.

Stermac, L. E., Segal, Z. V., & Gillis, R. (1990). Social and cultural factors in sexual assault. In W. L. Marshall, D. R. Laws, & H. E. Barbaree (Eds.), *Handbook of sexual assault: Issues, theories, and treatment of the offender* (pp. 143-159). New York: Plenum.

Thornhill, R., & Thornhill, N. W. (1992). The evolutionary psychology of men's coercive sexuality. *Behavioral and Brain Sciences, 15*, 363, 375.

Thornhill, R., Thornhill, N. W., & Dizinno, G. (1986). The biology of rape. In S. Tomaselli & R. Porter (Eds.), *Rape* (pp. 102-115). New York: Basil Blackwell.

Walker, W. D., Rowe, R. C., & Quinsey, V. L. (1993). Authoritarianism and sexual aggression. *Journal of Personality and Social Psychology, 65*, 1036-1045.

Wormith, J. S. (1986). Assessing deviant sexual arousal: Physiological and cognitive aspects. *Advances in Behavior Research and Therapy, 8*, 101-137.

Wormith, J. S., Bradford, J. M., Pawlak, A., Borzecki, M., & Zohar, A. (1988). The assessment of deviant sexual arousal as a function of intelligence, instructional set and alcohol ingestion. *Canadian Journal of Psychiatry, 33*, 800-808.

Wydra, A., Marshall, W. L., Earls, C. M., & Barbaree, H. E. (1983). Identification of cues and control of sexual arousal by rapists. *Behaviour Research and Therapy, 21*, 469-476.