

The Validity and Appropriateness of Methods, Analyses, and Conclusions in Rind et al. (1998): A Rebuttal of Victimological Critique From Ondersma et al. (2001) and Dallam et al. (2001)

Bruce Rind
Temple University

Philip Tromovitch
Tokyo Medical and Dental University

Robert Bauserman
State of Maryland

The authors respond to 2 victimological critiques of their 1998 meta-analysis on child sexual abuse (CSA). S. J. Dallam et al. (2001) claimed that B. Rind, P. Tromovitch, and R. Bauserman (1998) committed numerous methodological and statistical errors, and often miscoded and misinterpreted data. The authors show all these claims to be invalid. To the contrary, they demonstrate frequent bias in Dallam et al.'s criticisms. S. J. Ondersma et al. (2001) claimed that Rind et al.'s study is part of a backlash against psychotherapists, that its suggestions regarding CSA definitions were extrascientific, and that the moral standard is needed to understand CSA scientifically. The authors show their suggestions to have been scientific and argue that it is Ondersma et al.'s issue-framing and moral standard that are extrascientific. This reply supports the original methods, analyses, recommendations, and conclusions of Rind et al.

In 1998, we published a meta-analytic review (Rind, Tromovitch, & Bauserman, 1998) of 59 college studies that questioned victimological assumptions regarding the mental health consequences of child sexual abuse (CSA). Dallam, Gleaves, Spiegel, and Kraemer (1999) and Ondersma, Chaffin, and Berliner (1999) responded by questioning the integrity of our study. In the current article, we respond to updated versions of these critiques (Dallam et al., 2001; Ondersma et al., 2001), which stem from the victimological viewpoint.

Sexual victimology, one of several types of victimology to emerge in the wake of the various protest movements in the 1960s and early 1970s, all of which had victimization as a central theme, is a blend of social science, criminology, and victimization-based feminism that advocates social and legal reform (Best, 1997; Schultz, 1980; Sommers, 1995). As with other forms of victimology, sexual victimology holds as a basic tenet that victimization, which is defined in increasingly broad terms, typically produces lasting psychological damage; this view invited the medicalization of victimization, which prompted expansion of therapeutic services that embraced victimological assumptions as a basis for treatment (Best, 1997; Dineen, 1998; Sarnoff, 2001). For victimological therapists, CSA has become an essential component of their view of the cause of psychological maladjustment.

To avoid misunderstanding, we note that there is no doubt that some persons are harmed, and severely so, by CSA. In fact, we stated as much in our earlier article (Rind et al., 1998). Thus, the issue is not whether CSA can be harmful, but how often, to what degree, and under what circumstances, which contrasts with the more sweeping assumptions of sexual victimology just discussed. Additionally, we repeat from our earlier article that questioning psychological harmfulness should not be confused with questioning wrongfulness. The two issues are separate; determining that various types of CSA are not related to harm, for example, does not imply that they are not wrongful.

The critique by Dallam et al. (2001) focuses mainly on methodological and statistical issues. We consider all their major criticisms, including issues in (a) external validity, (b) definitions, attenuation, and moderators, (c) internal validity, and (d) qualitative analysis, and argue for the validity and accuracy of our methods and analyses. The critique by Ondersma et al. (2001) is concerned with the "moral standard" (p. 711), dangers that our article allegedly poses, and a perceived backlash against psychotherapy. They claimed that we misused science by suggesting that researchers use morally neutral terminology for some CSA experiences, which they regard as being "extrascientific" and "an attempt to erode current societal views regarding CSA" (Ondersma et al., 2001, p. 710). We show that our recommendations regarding terminology were scientific, not extrascientific, as they were based on the attempt to improve construct validity. They warned that our research will be co-opted by groups that intend "to support predetermined advocacy positions" (Ondersma et al., 2001, p. 713). We argue that warnings about possible negative consequences of research are scientifically inappropriate, as they represent an instance of the argument from adverse consequences fallacy (Lilien-

Bruce Rind, Department of Psychology, Temple University; Philip Tromovitch, College of Liberal Arts and Sciences, Tokyo Medical and Dental University, Tokyo, Japan; Robert Bauserman, Department of Health and Mental Hygiene, State of Maryland.

Correspondence concerning this article should be addressed to Bruce Rind, Department of Psychology, Temple University, Philadelphia, Pennsylvania 19122. Electronic mail may be sent to rind@vm.temple.edu.

feld, in press; Sagan, 1995). They classified our study as part of the backlash that continues to have an impact on their profession, and expressed concern that our study has or may hurt perceptions of the American Psychological Association (APA), which will impact on public trust and in turn negatively affect their ability to "service victims of child maltreatment" (Ondersma et al., 2001, p. 708). We dispute the backlash classification as inappropriate advocacy rhetoric, based as it is on dubious historical perspective. We add that, contrary to their concern regarding public trust and its effects on therapy, the only legitimate avenue to public trust and sound psychotherapeutic practice is integrity in science (Dineen, 1998; Sarnoff, 2001).

In the spring of 1999, our study came under intense attack by social conservatives, culminating in congressional condemnation. Ondersma et al. (2001) incorrectly described the origins of the attacks. In fact, they began in December 1998, when the National Association for the Research and Therapy of Homosexuality, a psychoanalytically oriented organization of therapists dedicated to the cure and prevention of homosexuality, critiqued our study, arguing that it would normalize pedophilia as had previously happened with homosexuality. Their comments were then repeated in a religious newsletter in March 1999, which a Philadelphia radio host used as a basis for his own attacks. A listener contacted "Dr. Laura" Schlessinger, who initiated a public furor over it (see Rind, Tromovitch, & Bauserman, 2000a, for a complete description).

Dr. Laura characterized our study as "garbage research with a dangerous statement at the end" and criticized our use of meta-analysis, stating, "I frankly have never seen this in general science. . . . This [pooling of studies] is so outrageous" (Schlessinger, 1999). She soon added to her methodological criticisms those provided to her by the Leadership Council for Mental Health, Justice and the Media, an organization of professionals advocating the validity of repressed memories and related therapies. The Leadership Council's critique was authored by Dallam et al. (1999) and was eventually provided to certain members of Congress, who used it as a basis for condemning our study. In their current critique, Dallam et al. (2001) omitted this information, though it is central to the political attacks. Also important is that elsewhere we thoroughly rebutted the 1999 Dallam et al. critique (Rind, Tromovitch, & Bauserman, 2000a). Thus, the current debate between us is important, as it reflects further on the validity of their earlier critique and the political attacks it supported.

Ondersma et al. (2001) mentioned the American Association for the Advancement of Science independent review of our study but did not sufficiently summarize their reply. The most relevant part in terms of evaluating our study versus the criticisms of it is the following:

After examining all the materials available to the Committee, we saw no clear evidence of improper application of methodology or other questionable practices on the part of the article's authors. . . . The Committee also wishes to express its grave concerns with the politicization of the debate over the article's methods and findings. In reviewing the set of background materials available to us, we found it deeply disconcerting that so many of the comments made by those in the political arena and in the media indicate a lack of understanding of the analysis presented by the authors or misrepresented the article's findings. All citizens, especially those in a position of public trust, have a responsibility to be accurate about the evidence that informs their public statements. We see little indication of that from the most

vocal on this matter, behavior that the Committee finds very distressing. (quoted in McCarty, 1999, pp. 2-3)

The validity of a research study is of course not decided by who praises or condemns it but rather by the methodological quality of the study and the empirical basis for its conclusions. As a consequence, we devote most of our response to addressing methodological criticisms. The reader may feel that these next several sections seem tedious at times, focusing on minute details. However, we feel it is both necessary and important to demonstrate exactly how and why so many of our critics' claims are either factually in error, debatable, or trivial and do not affect our basic conclusions even if accurate.

External Validity

Dallam et al. (2001) questioned the external validity of our review based on college samples in several ways (the methodological complaints of Ondersma et al., 2001, essentially overlap Dallam et al.'s and will also be addressed). They claimed that (a) its generalizability is limited by sample bias, arguing that college samples exclude the most severely affected; (b) college samples underestimate adverse effects of CSA; (c) we dismissed conclusions from previous reviews by claiming that the clinical and forensic samples on which they were typically based are not representative of the general population, yet argued, they went on to claim, that results from college samples "should be considered generalizable to the population as a whole" (Dallam et al., 2001, p. 717); and (d) our comparisons between college and national sample studies were flawed. These claims range from being highly debatable at best to simply wrong. In rebutting them, we present additional analyses to support the wider relevance of the college data, reinforcing the importance of our original findings.

Clinical Samples and External Validity Bias

External validity bias is a major weakness of the clinical perspective. The research of Kinsey, Pomeroy, and Martin (1948) was groundbreaking, not because its results are definitive (they are not), but because it set a new standard for approaching sexual behavior scientifically. Kinsey et al. argued that previous sex research, especially that conducted by psychiatrists and psychoanalysts, was severely limited in generalizability because it focused on clinical case studies—a problem exacerbated by clinicians' seeming lack of awareness of this limitation. They sampled a large and diverse segment of the general population to reduce this bias. Ford and Beach's (1951) work, a second milestone in sex research, went even further, arguing that even the most comprehensive survey of Americans would not adequately describe human sexuality because of the profound influence of culture. They reviewed data on numerous other cultures to seek patterns, which they elucidated through comparisons with cross-species data. The works of these researchers contradicted much of the prevailing conventional wisdom regarding human sexuality and showed that a comprehensive, scientific understanding of human sexuality is weakest when based primarily on clinical case studies, stronger with broad and diverse sampling within a society, and strongest with broad and diverse sampling of other cultures and species.

External validity bias is most prominent when the sexual behavior of interest is taboo. Then, authoritative opinion is typically deferred to the clinician, whose main goal is to treat, prevent, and cure rather than understand (Greenberg, 1988). Data from non-clinical samples or other cultures or species are then either not sought, labeled irrelevant, or ignored. This bias predominated when female sexual desire, masturbation, and homosexuality were viewed as pathological (Szasz, 1990). External validity bias has been prominent in research on CSA, a highly taboo form of sex. Victimologists concerned with CSA have frequently exhibited this bias in viewing clinical and legal research as informative well beyond these populations, while paying relatively little attention to nonclinical research and rejecting cross-cultural and cross-species perspectives on CSA (e.g., Olafson, Corwin, & Summit, 1993; Ondersma et al., 2001). But the need to move beyond clinical samples is essential, as Finkelhor (1984) noted

In spite of the number of victims judged by clinicians to have been badly affected by childhood victimization, it is possible that these may be skewed samples and not representative of the vast majority of children who have sexual contact with adults. . . . Critics are correct to point out that studies of nonclinical cases are needed, hopefully studies that allow a comparison between victimized and nonvictimized children. (p. 189)

The Logic of Using College Samples

In our literature reviews of CSA, we explicitly dealt with the problem of external validity bias by focusing on nonclinical samples (Bauserman & Rind, 1997; Rind & Tromovitch, 1997; Rind et al., 1998). Previous reviews typically drew general conclusions about CSA while restricting themselves mostly to clinical and legal samples. Before conducting our contested review of college samples, we examined studies based on national probability samples, far more representative of the general population than clinical samples (Rind & Tromovitch, 1997). Because the national samples generally lacked data relevant to issues such as confounding and causality, we extended that review by analyzing college samples, which are much richer in such data (Rind et al., 1998). However, we did not present these samples as representative of the general population. Rather, we argued that the college data were relevant to the general population—an empirically derived conclusion based on various important similarities in prevalence, severity, and correlates—rather than representative of it. When we generalized our results, we did so to the college population.

At the outset of our *Psychological Bulletin* (1998) review, we stated that our goal was to examine whether CSA typically causes pervasive harm of an intense nature that is equivalent for both sexes in the population of persons with CSA experiences. We argued that this view has been promoted by not only the media, as Dallam et al. (2001) noted we claimed in their current critique, but also by mental health professionals. To illustrate the latter, consider the Sidran Foundation's (1994) characterization of CSA—this foundation focuses on multiple personality disorder and recovered memory and provides related literature to therapists and patients. In its brochure, the foundation equates a child's "sexual activity with an adult" (p. 2) with serious threats to one's life, rape, military combat, natural or accidental disasters, and torture in terms of traumatic impact. One of the Dallam et al. coauthors separately has compared CSA to head injuries from a car accident

(Spiegel, 2000b, in press). If these dramatic analogies provided by mental health professionals are valid, then it should follow that in any population sampled—drug addicts, psychiatric patients, or college students—persons who have experienced CSA should show strong evidence of the assumed properties of CSA (even if some populations show stronger evidence than others). If we do not find such evidence in even one of these populations, then the broad and unqualified claims about the properties of CSA are contradicted. Thus, the representativeness of college samples is in fact irrelevant to the stated goals and conclusions of our study. As we stated in summing up our findings in our original article,

These results imply that, in the college population, CSA does not produce pervasive and intensely negative effects regardless of gender. Therefore, the commonly assumed view that CSA possesses basic properties regardless of population of interest is not supported. (Rind et al., 1998, p. 42)

The Relevance of the College Findings

Even though Dallam et al.'s (2001) argument that college samples are biased is not relevant to our basic goals and conclusions, it is nevertheless important to examine the claims that we missed those most affected, underestimated CSA effects by excluding relevant outcomes, and overstated the similarity between national and college samples. Each of these claims is questionable or incorrect.

Missing those most affected. The claim that college samples understate correlates of CSA because they miss those most affected is essentially a claim that college samples produce lower (i.e., biased) estimates of CSA-symptom relations than in the general population. We demonstrated in our original article a strong similarity between college and national samples in prevalence, severity, and correlates (Rind et al., 1998). Although Dallam et al. (2001) disputed this similarity, we later support our original conclusions. For now, we consider a second set of samples, those of junior high and high school students cited by Dallam et al. (2001). If junior and senior high school students who experience CSA tend not to make it to college because of the CSA, then we should expect that the magnitude of CSA-symptom associations would be substantially higher in these precollege samples because they would include more of those most affected. If this is not the case, then it is less likely that college samples introduce serious bias in terms of effect size estimates. We calculated effect sizes from these samples and meta-analyzed them.

Table 1 presents effect sizes for the samples from the 14 high school and junior high school studies cited by Dallam et al. (2001), computed separately for emotional and behavioral problems. Results are not provided for the Chandy, Blum, and Resnick (1996) study done on Minnesota students because it did not compare CSA students with controls; instead, we included two proxy studies, also done in Minnesota, that did contrast CSA and control students (Hernandez, Lodico, & DiClemente, 1993; Lodico, Gruber, & DiClemente, 1996). When a study reported separate statistics for the two sexes, that study was broken down into two samples. The Kendall-Tackett, Williams, and Finkelhor (1993) study, which included a meta-analysis of a few of its samples, did not provide data on sample sizes, so our meta-analyses were conducted with and without this study. Table 2 presents results of the meta-analyses.

Table 1

Studies Based on Junior and Senior High School Samples Cited by Dallam et al. (2001): Emotional and Behavioral Correlates of CSA

Study	Key phrase operationally defining CSA	Gender	N	<i>r</i>	
				Emotional	Behavioral
Bensley, Spieker, Van Eenwyk, & Schoder (1999)	Touched when "you did not want" this	F,M	4,754	—	.11
Bensley, Van Eenwyk, Spieker, & Schoder (1999)	Touched when "you did not want" this	F,M	4,754	.16	—
Boney-McCoy & Finkelhor (1995)	"Threaten, force, or trick" into sex "you didn't want"	F	911	.20	.20
Edgards & Ormstad (2000)	"Threaten, force, or trick" into sex "you didn't want"	M	987	.11	.11
	"Against your will" w/≥5 years older	F	1,129	.14	.11
	"Against your will" w/≥5 years older	M	814	.05	.09
	"When you did not want to"	F,M	1,140	.11	.14
Erickson & Rapkin (1991)	"Nonconsensual sexual contact before age 13"	F	957	.21	.08
Fiscella, Kitzman, Cole, Sidora, & Olds (1998)					
Garnefski & Arends (1998)	"Ever sexually abused, for example, forced . . . assaulted or raped"	F	1,188	.24	.18
	"Ever sexually abused, for example, forced . . . assaulted or raped"	M	302	.31	.31
Harrison, Fulkerson, & Beebe (1997)	Touched "against your wishes"; "forced you" to touch; incest	F	60,638	—	.10
Hernandez et al. (1993)	Touched "against your wishes"; "forced you" to touch; incest	M	60,619	—	.11
	Touched "against your wishes"; "forced you" to touch; incest	M	2,973	—	.10
Hibbard, Brack, Rauch, & Orr (1988)	"I have been sexually abused"	F,M	706	.04	.14
Hibbard, Ingersoll, & Orr (1990)	"I have been sexually abused"	F	1,926	.10	.12
	"I have been sexually abused"	M	1,926	.06	.22
Kendall-Tackett et al. (1993)	[A literature review; studies with varying definitions]	F,M	n/a	.57	.63
Lodico et al. (1996)	Touched "against your wishes"; "forced you" to touch; incest	F	2,986	—	.17
	Touched "against your wishes"; "forced you" to touch; incest	M	3,238	—	.12
	Forced sex (intercourse)	F	1,406	.05	.06
Nagy, DiClemente, & Adcock (1995)					
Stock, Bell, Boyer, & Connell (1997)	Touched when "you did not want" it	F	3,120	—	.15

Note. Dashes indicate that correlate was not assessed. CSA = child sexual abuse; F = female; M = male; n/a = not applicable.

For both emotional and behavioral problems, the overall unbiased effect size estimates, .13 and .11, respectively, were quite similar to that in our meta-analysis on college students ($r_u = .09$). The unweighted mean correlations for emotional and behavioral

problems without the Kendall-Tackett et al. (1993) study were only slightly larger ($r_s = .14$). These unweighted values rose to .17 when including Kendall-Tackett et al. The Kendall-Tackett et al. results, however, are anomalous: The mean emotional effect size

Table 2

Meta-Analyses of Sample-Level Effect Sizes Assessing CSA-Adjustment Relations in Junior and Senior High School Students

Symptom class/level	<i>k</i>	<i>N</i>	r_u	95% CI	χ^2	r_{unw}
Emotional problems						
All	14					.17
All (w/o KWF)	13	18,146	.13	.12-.15	74.40*	.14
Male	4	4,029	.09	.06-.12	19.32*	.13
Female	6	7,517	.15	.12-.17	35.74*	.16
Behavioral problems						
All	20					.17
All (w/o KWF)	19	151,820	.11	.11-.12	81.41*	.14
Male	7	70,859	.11	.11-.12	37.81*	.15
Female	9	74,261	.11	.10-.11	40.35*	.13

Note. CSA = child sexual abuse. *k* = number of effect sizes (samples); *N* = total number of participants in the *k* samples; r_u = unbiased effect size estimate (positive values indicate better adjustment for control participants); CI = confidence interval for r_u ; χ^2 = within-group homogeneity statistic (based on $df = k - 1$); r_{unw} = unweighted mean effect size. "All" indicates all samples were used, including Kendall-Tackett et al. (1993), in which some data were missing and statistics could not be computed, as indicated by blank spaces; "All (w/o KWF)" indicates all samples were used except for Kendall-Tackett et al. (1993).

* $p < .05$, indicating nonhomogeneity of effect sizes.

($r = .57$) was 2.86 *SDs* above the mean of the other effect sizes for emotional problems; the mean behavioral effect size, .63, was even more deviant ($z = 3.77$). The Kendall-Tackett et al. study was based on sexual abuse treatment samples and is thus not comparable to the other samples, which were nonclinical and nonlegal. Treating this study as an outlier, justified by its sampling and results, we are left with effect sizes nearly the same on average as in the college population. It is important to note that all of these studies examined unwanted CSA only; thus, they may overestimate correlates of sociolegally defined CSA, which includes willing sex involving age discrepancies. Thus, Dallam et al.'s (2001) argument that college samples are biased because those with a history of CSA more often do not make it to college is inconsistent with the very research they cite for this claim.

In suggesting that CSA debilitates academic performance, Dallam et al. (2001) cited a number of studies (many included in Table 1) that ignored or only weakly controlled for confounding variables. In our review we cited a study (Eckenrode, Laird, & Doris, 1993) that strongly controlled for confounding variables by categorizing a representative community sample into groups on the bases of CSA, physical abuse, neglect, and combinations of these (Rind et al., 1998). In this study CSA was not associated with academic problems, but physical abuse and neglect were. Another Dallam et al. (2001) citation was a study based on a nationally representative sample of Swedish 17-year-olds, which examined not only students but also dropouts (Edgardh & Ormstad, 2000). Consistent with Dallam et al.'s (2001) argument, the study-level effect size for the association between CSA and problem areas was larger for female dropouts ($r = .22$) than students ($r = .13$). However, in contradiction to their argument that CSA causes students to drop out, Edgarh and Ormstad reported that dropping out was confounded with being in foster care ($r = .21$ for females). These results highlight the weakness in causal assertions based on correlational data.

Exclusion of relevant outcomes. Both Dallam et al. (2001) and Ondersma et al. (2001) argued that we either excluded relevant outcomes or defined harm too narrowly. Regarding Ondersma et al.'s claim of narrowness, we do not believe that the 18 separate categories of mental health measures included in our review were inadequate to examine harm. Indeed, CSA has been depicted as a "special destroyer of adult mental health," as Seligman (1994, p. 232) noted, and victimologists often provide long lists of symptoms labeled as *effects* of CSA that have included every category of symptoms examined in our review (e.g., depression, dissociation, eating disorders, and sexual maladjustment).

Dallam et al. (2001) claimed that we underestimated some adverse effects of CSA (i.e., posttraumatic stress disorder [PTSD] and behavioral problems) by using college samples. In fact, we examined all psychological correlates that appeared in at least two studies. Nevertheless, we believe that examining every imaginable correlate of CSA is not as informative as examining the magnitude of the relationship between CSA and psychological adjustment. Creating long lists of correlates capitalizes on statistical dependency between outcome measures and inflates apparent impact, unless specific measures are differentially correlated with CSA. In our review they were quite consistently related, which indicates that one measure (e.g., a general measure of adjustment) can act as a proxy for unmeasured correlates. The key issue is the magnitude

of the association. In comparison with national and high school samples, the college samples do not underestimate magnitude.

Regarding PTSD, general measures of adjustment are adequate proxies for PTSD because of similar effect sizes. Neumann, Houskamp, Pollock, and Briere's (1996) meta-analysis yielded general indices of adjustment ($d = .46$ or $r = .19$), based on 11 nonclinical and clinical samples, that were comparable to PTSD ($d = .52$ or $r = .22$), based on 4 clinical samples. A national sample of women reporting child rape, cited by Ondersma et al. (2001), had a PTSD $r = .12$ (Saunders, Kilpatrick, Hanson, Resnick, & Walker, 1999). One of our college samples did assess PTSD (Brubaker, 1994); the effect size ($r = .10$) was comparable to all other measures that we meta-analyzed. In short, available evidence shows that the CSA-PTSD relationship is comparable in magnitude to CSA's relationship with many other adjustment measures.

Dallam et al. (2001) also stated that we underestimated effects involving behavior problems. Table 2 shows this is incorrect in terms of magnitude of association; emotional correlates are adequate proxies. Moreover, high school student studies do not support the statement that CSA "has a particularly negative impact on the behavior of adolescent males" (p. 717), as seen in the meta-analytic results for boys and girls in Table 2. The specific research they cited, rather than contradicting our earlier results, is consistent with those results.

Comparison to national samples. Dallam et al. (2001) made the serious charge that we often either misreported or "presented the [abuse severity] data in a misleading manner" (p. 717) in comparing college and national samples. They claimed that CSA in our college samples was categorized by highest level of abuse severity (where *severity* increases from noncontact sex such as exhibitionism to sexual touching to intercourse), whereas CSA in our national samples was categorized by simple frequency counts, rendering comparisons between the college and national samples invalid. However, only some college samples were categorized by highest level of severity, as is clearly indicated in our comment:

Because a number of studies categorized SA [sexual abuse] participants exclusively into the most "severe" type of CSA experienced, the prevalence of less severe types is likely to be underestimated (Rind et al., 1998, p. 30).

When possible, we obtained and averaged simple frequency counts to assess the degree of each type, rather than the extent of the most severe type. This contradicts Dallam et al.'s (2001) assertion that our numbers are not comparable, as well as their claim that we misreported the data from López, Carpintero, Hernández, and Fuertes (1995), where exhibitionism was 33% of cases based on nonmutually exclusive categories.

Dallam et al. (2001, Table 1) presented their own table of the prevalence of CSA types, based on classification by highest level of severity. However, they left out the most reliable data available, those of Laumann, Gagnon, Michael, and Michaels (1994), which are based on a face-to-face interviewing format, an important methodological strength. Instead, they cited as more relevant to the general population Finkelhor, Hotaling, Lewis, and Smith (1990), a telephone interview study that found that 62% of men and 49% of women reported actual or attempted intercourse. On the basis of this single result, which doubled and quadrupled the rates we found in the college samples, they claimed our samples vastly

underestimated severity. However, this study is an outlier compared to all other national studies, probably because of the ambiguous screening question assessing intercourse. Even Finkelhor et al. (1990) stated that "some problems with methodology, however, particularly the imprecision of the screening questions, do caution against relying on findings from this study alone in absence of supporting evidence from other research" (p. 27). The screening question on intercourse was

When you were a child [elsewhere indicated to be age 18 or under], can you remember having any experience you would now consider sexual abuse—like someone trying or succeeding in having any kind of sexual intercourse with you, or anything like that? (Finkelhor et al., 1990, p. 20)

This question is clearly ambiguous. The main query is about any experience one now considers sexual abuse, and phrasing such as "any kind of" and "anything like that" is ambiguous and does not exclude nonintercourse CSA. This is clearly not a valid measure of intercourse.

In Table 3 we present percentages from five national samples, including the Edgardh and Ormstad (2000) study cited by Dallam et al. (2001). Finkelhor et al. (1990) is clearly an outlier, with percentages 2 to 12 times larger than the other percentages for men and 3 to 10 times larger for women. Extent of intercourse for men is higher in the college than in the national samples, with or without Finkelhor et al. For women, it is equal in the college and national samples without Finkelhor et al. but half as much including Finkelhor et al. In conclusion, Dallam et al. (2001) misinterpreted our analysis of severity, failed to note a serious validity issue regarding the Finkelhor et al. results, and focused on a dubious estimate coming from this single study. CSA was not less severe in the college compared to national samples, assuming the common belief that intercourse is the most severe type.

Dallam et al. (2001) also disputed our comparison of effect size estimates in the college and national samples, listing three national samples in their Table 2. Although they described Finkelhor et al. (1990) as "more relevant to the U.S. general population" (p. 717) when using it for prevalence of severity, they did not include this study in their effect size analysis. Additionally, they did not include the other U.S. national sample (Bigler, 1992) that we included in our meta-analysis of national samples (see Rind & Tromovitch, 1997). In that meta-analysis, we stipulated as an

inclusion criterion that a national sample had to report data "separately for male and female respondents" (p. 241), given that one goal of that study was to analyze results separately by gender. In the López et al. (1995) study, the only data on mental health problems reported separately by gender were presented in López et al.'s Table 7. Dallam et al. (2001) computed an effect size (shown in their Table 2) based not on López et al.'s Table 7 but rather on López et al.'s Table 8, which did not separate the sexes and was predominated by behavioral rather than psychological correlates—the mean effect size without base-rate correction for these measures was .13. In Table 4, we present our original effect sizes (see Rind & Tromovitch, 1997, p. 248), along with those that Dallam et al. omitted in their Table 2, as well as those of Edgardh and Ormstad (2000).

Meta-analyses of the male and female effect sizes resulted in the same effect size estimates for males ($r_u = .07$) and females ($r_u = .10$) that we reported previously (Rind & Tromovitch, 1997; Rind et al., 1998). We did not correct for base rates as Dallam et al. (2001) did, for reasons to be discussed later. For purposes of debate, however, we examined Dallam et al.'s (2001) Table 2 effect sizes with base-rate corrections. They argued that there was "little support for the claim that [our] findings should be considered generalizable to the population as a whole" (p. 718). We meta-analyzed the two effect sizes from the college data ($r_u = .11$, 95% confidence interval [CI] = .09–.12), $\chi^2(1, N = 14,578) = .24, p > .05$, as well as the five national effect sizes they provided ($r_u = .15$, 95% CI = .12–.17), $\chi^2(4, N = 6,638) = 10.17, p > .05$, and then contrasted these two effect size estimates. The contrast was statistically significant ($z = 2.74, p < .01$, two-tailed), but the effect size was very small ($r = .02$). We believe this exceedingly small association between sample type (college vs. national) and adjustment, based on *selected* national samples rather than all those available to Dallam et al. (2001), does not justify dismissing the college data as irrelevant to the population as a whole.

In sum, Dallam et al. (2001) did not demonstrate that the college data are biased in terms of underestimating severity of CSA or CSA–adjustment relations. In their effort to demonstrate this, they selectively used particular results while ignoring others—the kind of confirmation bias that quantitative (i.e., meta-analytic) reviews are designed to counter, by taking into account all the data that meet pre-specified criteria rather than just the data that bolster one's argument. It is important to add that the effect size estimates

Table 3
Percentage of CSA Involving Actual or Attempted Intercourse in National Samples

Study	Method	Males		Females		Total	
		%	n	%	n	%	N
Baker & Duncan (1985)	FTF	5	136	5	154	5	290
Edgardh & Ormstad (2000)	Q	24	25	18	126	19	151
Finkelhor et al. (1990)	Phone	62	169	49	416	53	585
Laumann et al. (1994)	FTF	32	176	14	293	21	469
López, Carpintero, Hernández, & Fuertes (1995)	FTF, Q	6	134	13	203	10	337
Total with Finkelhor et al. (1990)		28	640	25	1,192	26	1,832
Total without Finkelhor et al. (1990)		16	471	13	776	14	1,247
Rind et al., 1998 (College samples)		33	506	13	2,172	17	2,678

Note. CSA = child sexual abuse; FTF = face to face interview; Q = questionnaire; Phone = telephone interview.

Table 4
Psychological Correlates of CSA in National Samples

Study	Males		Females	
	<i>n</i>	<i>r</i>	<i>n</i>	<i>r</i>
Bigler (1992)	140	.07	174	.17
Boney-McCoy & Finkelhor (1995)	987	.12	911	.20
Edgardh & Ormstad (2000) ^a	814	.06	1,243 ^b	.14 ^b
Finkelhor et al. (1990)	1,142	.05	1,476	.07
Laumann et al. (1994)	1,311	.07	1,608	.05
López et al. (1995)	462	.04	433	.09
<i>N</i> and <i>r_u</i>	4,856	.07	5,845	.10
95% Confidence interval		.04–.10		.08–.13
χ^2 (homogeneity statistic)		3.46		17.77*

Note. CSA = child sexual abuse; *r_u* = the unbiased effect size estimate.
^a Combines psychological and behavioral correlates. ^b Based on high school students and dropouts combined.

* *p* < .05.

from the high school, college, and national samples are all quite similar and small in magnitude—a result that bolsters, rather than disconfirms, our original conclusions.

Definitions, Attenuation, and Moderators

The next set of criticisms we examine centers on the methods of our review. Dallam et al. (2001) claimed that we (a) compounded the lack of standardization of definitions of CSA in the primary studies by including certain studies they believed questionable and excluding others they believed appropriate; (b) failed to account for possible attenuation of effect sizes by our use of the Pearson *r* effect size, as opposed to Cohen's *d*; and (c) mishandled our analyses of moderators of the CSA–symptoms relationship. We show these criticisms to be factually incorrect in most cases and highly debatable in the rest.

Operational Definition of CSA

Operational definitions of CSA have varied widely, a problem for all reviews of CSA research. Dallam et al. (2001) claimed that we “compounded” this problem by including studies “that did not even purport to examine the effects of CSA” (pp. 718–719), citing three instances: Landis (1956), Schultz and Jones (1983), and Sedney and Brooks (1984). In fact, Landis consistently wrote of the “child” and the “offender,” referring to interactions with a “sexual deviate,” someone considerably older. He reported the median age of this experience to be 11.93 for girls and 14.92 for boys, and that for girls “41.9 percent were under the age of 11 when they had their deviate experience” (p. 95). About 90% of females and 80% of males were 18 or younger at the time of their experience. Finally, the Landis study is used as an example of an early CSA study by many other researchers (e.g., Finkelhor, 1979a; Fishman, 1991; Fromuth & Burkhart, 1989; Sarbo, 1985). Clearly, Landis did purport to study what is now called CSA, actually did study CSA, and is recognized for having done so by many other researchers.

As for Schultz and Jones (1983), Dallam et al. (2001) mentioned that they “looked at all types of ‘sexual acts’ before age 12” (Dallam et al., 2001, p. 719), but failed to add that respondents

were asked “if their experience was with a person over the age of 16” (Schultz & Jones, 1983, p. 100) and that Schultz and Jones’s discussion and conclusions were based entirely on CSA experiences, not peer sex. Finally, Dallam et al. (2001) noted that Sedney and Brooks (1984) “examined all types of ‘sexual experiences’ during childhood” (p. 719), but Sedney and Brooks themselves explained that their “rather broad definition of ‘child sexual abuse’ was used because of the difficulty posed by a priori decisions about what type of sexual experiences are ‘problems’” (p. 215). Dallam et al. (2001) cited Neumann et al. (1996) as researchers who did not include this study because of its broad definition. However, Dallam et al. (2001) failed to note that a variety of other researchers whom they also cited elsewhere to buttress other arguments referred to Sedney and Brooks as an important early nonclinical study on CSA effects (e.g., Fergusson, Horwood, & Lynskey, 1996; Garnefski, & Arends, 1998; Mullen, Martin, Anderson, Romans, & Herbison, 1993). Last, whereas the Landis (1956) and Schultz and Jones (1983) studies were not included in our meta-analyses of CSA–adjustment relations, the Sedney and Brooks study was. Its effect size (*r* = .15) was higher than the mean and thus did not bias results in terms of underestimating the CSA–adjustment association. In short, Dallam et al.’s (2001) claim that these studies did not purport to examine CSA effects is incorrect; their inclusion did not bias our meta-analysis of CSA–symptom relations.

Dallam et al. (2001) next argued that some of our studies included sexual experiences that occurred after age 17, again citing three instances: Greenwald (1994), Landis (1956), and Sarbo (1985). In Sarbo’s “purified” sample (p. 45), analyses were done on CSA occurring before age 17; these are the data we used to compute effect sizes. From Greenwald’s study, we computed effect sizes on the basis of CSA under age 16 (Greenwald, 1994). In Landis’s article, as we already noted, the vast majority of experiences with “deviates” occurred during childhood up to age 18 (89% of male and female cases combined). Thus, two of three instances cited by Dallam et al. (2001) are incorrect, and we argue the third is trivial (again, the Landis data were not part of the meta-analyses, and eliminating them from the analysis of reactions actually substantially increases the proportion of positive and neutral reactions). It is notable that some of the CSA studies Dallam et al. (2001) cited elsewhere in their critique to bolster their own arguments also include cases up to age 18 or 19 (e.g., Erickson & Rapkin, 1991; Finkelhor et al., 1990; Garnefski & Arends, 1998; Kendall-Tackett et al., 1993).

Dallam et al. (2001) continued by calling our exclusion of Roland, Zelhart, and Dubes (1989) and Jackson, Calhoun, Amick, Maddever, and Habif (1990)—both of which consisted mainly of incest experiences—“quite baffling” (p. 719). In a footnote right before this criticism, however, they cited Neumann et al. (1996) to argue that we should not have included the Sedney and Brooks (1984) study. However, they failed to note that Neumann et al. also excluded Roland et al. as an outlier, even among clinical studies, a point noted in our original article (Rind et al., 1998, p. 31). Our treatment of this issue was unbiased: We reported results with and without the outliers, finding that the unbiased effect size estimate was not affected by excluding the outliers, and our exclusion of these studies was based solely on statistical grounds.

Finally, Dallam et al. (2001) argued that we “erroneously coded” (p. 719) Silliman (1993) in the wrong direction. Had we

not "miscalculated" Silliman, they argued, then the two excluded studies just discussed would not have been outliers. Here is what Silliman (1993) wrote regarding the self-esteem measure in question:

It was hypothesized that the women who recalled sexual abuse during childhood would score significantly more externally on locus of control and score significantly lower on the self-esteem measure than the control group; however, contrary to expectation, the hypotheses were not confirmed ($t_{64} = .16$ and -4.50 , $p > .05$, respectively). (p. 1294)

Because it was predicted that abused participants would have lower self-esteem, but they did not, it follows that self-esteem was higher for the abused group, because the large t was in the opposite of the hypothesized direction, indicated by the nonsignificant p value. Thus, given the results as reported by Silliman (1993), we calculated the effect size appropriately. However, Dallam et al. (2001) reported that they contacted Silliman and found that self-esteem was lower for the abused group, which means that $t = -4.50$ would have been statistically significant ($p < .05$, not $p > .05$). Dallam et al. (2001) then noted that the entire distribution of effect sizes "would have shifted slightly" (p. 719) and that the two incest studies would not have been outliers. It would have been informative had they specified how slight the shift would have been. We computed the unbiased effect size estimate in three ways: (a) using the corrected value of $r = .255$ for Silliman and including all 54 samples; (b) using the incorrect value for Silliman of $r = -.25$ and including all 54 samples; and (c) using only 51 samples, excluding Silliman (1993), Roland et al. (1989), and Jackson et al. (1990). We obtained, respectively, $r_u s = .0969$, $.0948$, and $.0921$. In short, the coding of Silliman and the exclusions have no meaningful impact on the effect size estimate.

Attenuation

Dallam et al. (2001) argued that our use of the Pearson r effect size, as opposed to Cohen's d , in studies with unequal proportions in the comparison groups "created a situation in which clinically large effects could be represented by what appear to be small r values" (p. 720), citing Cohen's (1977) description of $d = .80$ as a large effect. We addressed this criticism in a previous rebuttal (Rind, Tromovitch, & Bauserman, 2000a) to Dallam et al. (1999), in which we used R. Rosenthal's (1984) formula that takes into account population prevalences to convert our r s to d s for each sample. Assuming prevalences are 50-50 for CSA and control populations, we obtained the following mean d s for our 14 male and 33 female samples: $d_s = .22$ and $.25$, respectively. Both of these d s were small, not large, according to Cohen's (1977) guidelines, which suggested that $d = .20$ is small. These findings contradict the thrust of Dallam et al.'s (2001) argument about attenuation of effect sizes.

Base rate differences for men and women. On the basis of the argument that our effect sizes were attenuated owing to unequal sample size, Dallam et al. (2001) used Becker's (1986) formula to correct our effect size estimates for men ($r_u = .07$) and women ($r_u = .10$), finding them to be $.10$ and $.11$, respectively. They argued from this that our computations "created the appearance of gender-related differences in CSA adjustment, when effect sizes for men and women were actually equivalent" (p. 721), but we

never concluded that the r_u s of $.07$ and $.10$ for men and women, respectively, were significantly different statistically. In fact, we reported that the contrast between these effect size estimates "was nonsignificant, $z = 1.42$, $p > .10$, two-tailed" (Rind et al., 1998, p. 33). Later, Dallam et al. (2001) provided this same quote to incorrectly argue a different point on moderator analysis involving gender (see below). What we did report as significantly different was the contrast between male and female effect size estimates for the all-types-of-consent groups, where $r_u s = .04$ and $.11$, respectively. If we follow Dallam et al. (2001) and apply Becker's correction formula to these values, they become $r_u s = .06$ and $.12$ for men and women, respectively. The contrast is still statistically significant ($z = 2.68$, $p < .01$, two-tailed), contrary to Dallam et al.'s (2001) claim.

Dallam et al. (2001) further argued that, although we wrote that effect sizes would increase at most by $.03$ for a 50-50 split (Rind et al., 1998), the effect sizes in some cases increased much more. They made this point despite quoting us two sentences earlier as having written that "an $r = .07$ based on a 14-86 split [for male samples] increases at most by $.03$ (to $r = .10$) in a 50-50 split (Rind, Tromovitch, & Bauserman, 2000a, p. 29)" (Dallam et al., 2001, p. 721), which makes it clear that we were referring to a maximum $.03$ increase in the overall effect size, not individual effect sizes.

For the sake of argument in the discussion above, we corrected our r s on the basis of samples with naturally occurring unequal base rates in the population. It is important to note, however, that many statistical experts believe that correction in this type of situation is inappropriate, including Becker (1986) himself. Although Dallam et al. (2001) cited Becker and used his formula, they failed to note his qualification of valid use of this formula. He noted that it "is appropriate to correct for unequal sample size when populations represented by the samples can be assumed to be equally numerous" (Becker, 1986, p. 5), as in randomized or block experiments. But he noted further that when populations are not equal in size, "the inequality should be reflected in estimating degree of relationship" (Becker, 1986, p. 6). As examples, he gave people with schizophrenia versus controls and people with double-recessive genes versus all carriers of a given gene, among others. Regarding these cases, he commented that "a correction would be inappropriate. Indeed, it can be argued that for such populations, when sample size is constrained to be equal, the resulting r should be corrected in the other direction" (Becker, 1986, p. 6).

Hunter and Schmidt (1990), who provided the same formula that Becker did, similarly noted that its intended use was for situations in which equal sample sizes could or should have been obtained (e.g., as in experiments) but were not. In natural settings with populations inherently unequal in size, however, they argued that correlations should reflect this difference. To illustrate, they noted that the association between race (White vs. Black) and achievement test performance among U.S. students is $.45$. This value is based on transforming a racial performance difference of 1 SD to Pearson's r , and assumes equal numbers of Whites and Blacks. Noting, however, that Blacks compose only 13% of the total, they argued that "this value *must* be adjusted to reflect that fact" (p. 276, italics added); they deemed the reverse-adjusted correlation of $.32$ (i.e., the estimated point-biserial correlation) to be the appropriate value. R. Rosenthal (1984) also opined that, when comparing naturally occurring groups with different base rates, the ap-

appropriate correlation is the one reflecting this difference (i.e., the uncorrected correlation). He added that correction for unequal sample sizes is appropriate if one is trying to estimate what one might find in a future study drawn from a population of equal numbers (i.e., $p = q = .5$), but is inappropriate if the goal is to estimate results from a future study drawn from the same population with unequal numbers—that is, to obtain a real world estimate (R. Rosenthal, personal communication, April 12, 2001).

This statistical opinion supports our noncorrection for base-rate differences. Nevertheless, we acknowledge the intuitive appeal for arguing for correction—that effect sizes should be base-rate independent (which Pearson's r is not, but Cohen's d is), especially if one is comparing effect sizes across groups. But as we showed, even from this perspective, group differences between the genders that we identified in our 1998 review remain different upon reanalysis, contrary to Dallam et al.'s (2001) contention, which undermines their reason for raising this issue to begin with. In short, our handling of Pearson's r in the face of base-rate differences was methodologically proper and produced no important bias, if any at all. Dallam et al. (2001), on the other hand, exhibited bias in their criticisms, selectively ignoring key clarifying quotes by us but citing them elsewhere in their critique to argue different points, and ignoring or overlooking a key caveat by Becker (1986) regarding appropriate use of his correction formula.

Dichotomization. Most studies on CSA, including many of the ones we reviewed (Rind et al., 1998), have classified participants dichotomously as either having experienced or not having experienced CSA. Dallam et al. (2001) argued that CSA as a variable has an underlying continuum and that dichotomizing it attenuates the CSA–adjustment association. They argued that a biserial correlation or a tetrachoric correlation is then appropriate, depending on whether only CSA is dichotomized or both CSA and the dependent measure are. Biserial and tetrachoric correlations are not product-moment correlations; rather, they are intended to estimate them, but can fail rather badly (Nunnally, 1978; R. Rosenthal, personal communication, April 12, 2001; Sheskin, 1997). R. Rosenthal (personal communication, April 12, 2001) gave an example in which X values are 0, 0, 1, 1, 2, 2, 3, and 3, and Y values are 0, 1, 0, 1, 2, 3, 2, and 3. Median splits on each produce $r = 1.00$, and dichotomizing on X only produces $r = .89$. Both of these r s based on dichotomization, however, are larger, not smaller, than the nondichotomized $r = .80$ based on continuous data. R. Rosenthal added that his example, in which dichotomization increases rather than attenuates the correlation, is not far-fetched and can easily occur when there is a third variable problem, in which X and Y are uncorrelated in levels A and B of a third variable, but A and B differ in their mean X and mean Y scores. Then, combining A and B can yield a large positive or negative overall correlation. We argue that this is just the kind of situation that could obtain in CSA studies. In sum, R. Rosenthal continued, although dichotomizing continuous data can decrease the magnitude of Pearson's r under some conditions, it can increase it in others; without having the continuous data, one cannot be sure which situation obtains. This argues against use of the biserial and tetrachoric correlations.

Furthermore, both corrections assume underlying continuous and normal distributions (Glass & Hopkins, 1996; Nunnally & Bernstein, 1994; Sheskin, 1997). For CSA, evidence suggests that these assumptions do not hold in the college population. Continuous CSA is assumed to correspond to severity, where noncontact

events are the least severe, touching is a level up, oral sex is even more severe, and intercourse is the most severe. It is clear that, in the case of adult–adult sex, these levels do not form a continuum of “severity” (i.e., seriousness or negativity). Reactions and levels of “severity” cannot be assumed to correspond, unless the sex is forced. In the case of CSA, numerous college studies have found no relation between “severity” and reactions. Finkelhor (1979a), on the basis of his college sample, observed

People . . . still use the standard of intercourse to judge the seriousness of a child's sexual experience. In other words, they presume that experiences involving intercourse are the most traumatic. . . . However, our data show the opposite; that is, the seriousness of sexual activity as it is usually understood does not seem related to greater trauma in children. . . . It suggests that the actual sexual activity involved is less important than its context. (p. 103)

West and Woodhouse (1993) presented interviews of 24 male students, 7 of whom had oral sex or intercourse with adults when they were minors. Six experienced these encounters positively, whereas only 1 reacted negatively. Two had heterosexual encounters, 4 had homosexual encounters, and 1 had both. Negative or neutral reactions among the 24 students occurred almost exclusively in less severe cases involving noncontact sexual approaches or fondling. Condy, Templer, Brown, and Veaco (1987) examined sexual relations between boys under age 16 and females at least age 16 and 5 years older than the boys. These relations were predominantly of a “severe” nature: Among the college participants, 68% involved intercourse. Relatively few men reacted negatively or felt they had been harmed. Most reported that they consented or even initiated the encounters. Negative feelings and self-reported effects were associated with perceived lack of consent and incest. In Fromuth and Burkhart's (1987) male college sample, about 30% of the CSA involved oral sex and another 25%, intercourse. Despite this large degree of “severity,” only 15% reported negative effects, whereas 39% and 46% reported positive or neutral effects, respectively. Among teenagers aged 13 to 16, only 3% reported negative effects, with 60% reporting positive effects. In a recent study based on a sample of gay and bisexual males who were mostly college students (Rind, 2001), no relation was found between level of “severity” and reaction for boys aged 12 to 17 involved in contact sex with adult males; what mattered was context, particularly level of willingness. To repeat Finkelhor (1979a), it is the context that is important.

Sheskin (1997) noted that the accuracy of the biserial correlation is “highly dependent on the assumption of normality, and it should not be employed unless there is empirical evidence to indicate that the distribution underlying the dichotomous variable is normal” (p. 583)—he provided the same caveat for the tetrachoric correlation. Similarly, Nunnally (1978) noted that both “these correlations very much depend on a *strict* assumption of the normality of the continuous variables” (p. 137, *italics added*) and noted further that when the assumption of normality is not met, estimates can be off by more than 20 points of correlation. Evidence indicates that CSA is not normally distributed, even if one assumes continuity. For example, consider the case of a male sample in which 14% had and 86% did not have CSA. If we assume a continuum of severity, where not having CSA puts one at the low end, then the low end is the mode—a very dominating one—and the positively skewed distribution deviates markedly from normality. Sheskin added that

if there is reason to believe the normality assumption for the dichotomous variable has been violated, most sources recommend computing the point-biserial, rather than biserial, correlation, because the latter may be a spuriously inflated estimate of the underlying population correlation.

In summary, the assumption that CSA in the college population is a continuum of "severity" according to increasing physical intimacy, irrespective of context, is not empirically supported; the assumption of normality even if continuous is also not supported. These facts add to the issue of unreliability of the biserial and tetrachoric correlations to argue against their use in this population.

Interpretation of effect sizes. Both sets of critics raised concerns about our discussion of effect sizes. Dallam et al. (2001) claimed that we interpreted r^2 as the measure of effect size, rather than r , by quoting us as writing "according to Cohen's (1988) guidelines; in terms of variance accounted for, CSA accounted for less than 1% of the adjustment variance." What we actually wrote was

The resulting unbiased effect size estimate . . . was $r_u = .09$ This difference in adjustment between SA and control students was small, however, according to Cohen's (1988) guidelines; in terms of variance accounted for, CSA accounted for less than 1% of the adjustment variance. (Rind et al., 1998, p. 31)

The partial quote taken out of context by Dallam et al. (2001) gives the impression that we cited Cohen to conclude that a 1% variance is a small effect size. Clearly, our citation of Cohen refers to our reporting of the unbiased effect size, $r_u = .09$.

More important, both Ondersma et al. (2001) and Dallam et al. (2001) argued that small effect sizes can have huge personal and social costs. Dallam et al. (2001) cited R. Rosenthal and Rubin's (1982) binomial effect size display (BESD) as one means of indicating this. In regard to our study, the BESD would categorize 100 students as having had CSA and another 100 as not having had CSA; additionally it would categorize 100 students as being worse adjusted and another 100 as being better adjusted, producing four cells in a 2×2 matrix. Categorization into CSA \times Adjustment combinations would be based on Pearson's r , the association between CSA and adjustment (i.e., the effect size). Specifically, for CSA-worse adjusted (Cell A) and for no CSA-better adjusted (Cell D), the cell frequencies would be $100(.500 + r/2)$; for CSA-better adjusted (Cell B) and for no CSA-worse adjusted (Cell C), the frequencies would be $100(.500 - r/2)$. With the overall effect size in our meta-analysis of .09, the four values above would be as follows: (A) 54.5, (B) 45.5, (C) 45.5, and (D) 54.5. From this, one might conclude that exposure to CSA for every 100 persons produces a decrease in good adjustment from 54.5 to 45.5; in other words, 9 persons per 100 exposed to CSA are now more poorly adjusted. This interpretation strongly suggests that the small effect size of .09, according to Cohen's (1988) guidelines, is nevertheless an important effect. As we discuss next, however, this interpretation would be misleading.

This interpretation is based on the assumption that the increase in more poorly adjusted persons is due to CSA. Both Ondersma et al. (2001) and Dallam et al. (2001) implied that it is. Ondersma et al. (2001), by analogy, cite the small effect size ($r = .03$) in the well-known aspirin versus heart attack exper-

iment to argue that "even miniscule *effects* can have huge personal and societal costs when one extrapolates to a societal level" (p. 709, italics added). Dallam et al. (2001) cited Ondersma et al.'s (1999) meta-analysis of 14 studies on smoking and lung cancer, which showed an effect size of .17 to argue that the relationship we found between CSA and symptoms was roughly comparable to "the *effect* of cigarette smoking on lung cancer in the general population" (Dallam et al., 2001, p. 729, italics added). The *effects* for aspirin are true effects—we know this because this study used the experimental design. Similarly, the *effects* of cigarette smoking are now confirmed as true effects through an enormous amount of research that has identified at least 63 distinct cancer-causing agents in cigarettes. Although less than 15% of regular smokers develop lung cancer (thus the low effect size), smoking is directly responsible for 87% of lung cancer cases (American Lung Association, 2000). Such a dramatic relation has no empirical parallel in CSA research, vitiating the CSA-maladjustment/smoking-lung cancer analogy. The comparability with CSA is also much less clear than Dallam et al. (2001) asserted, because some, much, or all of the difference between CSA-adjustment combinations reflects confounding rather than actual effects. To highlight this point, consider the BESD for family environment (FE; better vs. worse) and adjustment (better vs. worse). In our meta-analysis, we found an unbiased effect size estimate of .29. Thus, in the case of poor family environment, for every 35.5 persons better adjusted, there would be 64.5 worse adjusted, which represents an increase of 29 persons per 100, a substantially greater number compared with CSA. Given that our meta-analysis showed that CSA was confounded with FE, and statistical control often eliminated significant CSA-adjustment relations, the BESD results regarding CSA should not be interpreted causally or compared with similar effect sizes that are known to be causal. Additionally, being put into the worse versus better adjusted categories in the CSA example reflects very minor differences between subjects on either side of the median for adjustment and near to it. By contrast, in the heart attack and lung cancer cases, differences in categorization are always hugely important. This further weakens our critics' analogies.

Moderators

We examined several variables as moderators of the CSA-adjustment relationship. Dallam et al. (2001) questioned our examination of these variables, including contact versus noncontact experiences, gender, and willingness or consent.

Contact. Dallam et al. (2001) cited a number of studies that separated contact and noncontact sex to argue that it is "remarkably consistent" (p. 722) that adjustment is poorer for contact groups and to argue that it is very questionable whether our findings apply to more serious forms of CSA. Ondersma et al. (2001) also questioned the use of studies that combine contact and noncontact CSA. As discussed above, however, much research indicates that it is the context that really seems to matter, not contact versus noncontact or level of contact. Context has much to do with the degree of force versus willingness, or at least absence of coercion. According to Finkelhor (1979a),

Unlike force, sexual activity and duration both are ambiguous in their implications. A longer relationship and one involving intercourse

indicate greater intensity. Intensity may be more harmful, but it could also be an indicator in some cases of a positive, or at least, an ambivalent, bond. In contrast, presence of force would almost always signal something negative about the relationship. It is a concise symptom of a whole negative context—the reluctance of the child, the pressure exerted by the partner, the difference in power and control. The primary recollection of the child is of the coercion. That there was sex involved is perhaps less important than the fact that there was aggression. (pp. 104–105)

In our college studies, meta-analysis of the 11 studies that confined CSA to contact experiences yields a small unbiased effect size estimate ($r_u = .10$, 95% CI = .06–.15), $\chi^2(10, N = 1,776) = 7.64, p > .05$. This estimate is itself consistent (i.e., the effect sizes were homogeneous) and is not different from the overall effect size estimate ($r_u = .09$) based on a majority of studies including noncontact cases. In the Laumann et al. (1994) national study, which included only contact cases of CSA, effect sizes were small for both male ($r = .07$) and female ($r = .05$) participants. Laumann et al. created a severity index based on level of contact (kissing or genital touching vs. oral sex or intercourse); the index failed to moderate current adjustment. These results show that the evidence is not as consistent as our critics claim.

Moreover, as critics have argued (e.g., Best, 1997; Dineen, 1998; Jenkins, 1998; Sarnoff, 2001), CSA researchers have extensively used broad definitions of CSA that include noncontact experiences. For example, Boney-McCoy and Finkelhor (1995), commenting on the finding that girls in their study with noncontact CSA were symptomatic, stated,

This finding highlights the noxious correlates of even interrupted forms of predatory sexual behavior. At present, these results warrant advising clinicians and others who work with young people to be aware of the potentially harmful consequences of what may appear to be “minor” sexual victimization experiences. (p. 733)

It seems inconsistent for CSA researchers following a victimological model to uncritically combine contact and noncontact experiences in their own research, then criticize us for including studies that combine both contact and noncontact CSA in a meta-analytic review.

Gender. Dallam et al. (2001) claimed that our finding that gender moderated adjustment was not reflected in our moderator analysis. To back this claim, they noted that we reported that the contrast between the male and female effect size estimates (.07 and .10, respectively) was nonsignificant. First, this citation is selective, as they ignored it earlier when they argued that we used these same effect sizes to create the appearance of gender differences, when we clearly did not. Second, their current claim is simply incorrect. Our regression analysis, the first step in our moderator analysis, showed that gender and the Consent \times Gender interaction both moderated the effect sizes across samples. On the basis of this significant interaction, we computed main effects and interaction contrasts, which yielded no main effects (i.e., no difference between .07 and .10) but did yield a significant interaction involving gender and level of consent.

Consent. Dallam et al. (2001) repeatedly encased *consent* and *willing* in quotation marks and questioned our use of this construct more vigorously in their other critiques (e.g., Dallam et al., 1999; Spiegel, 2000a, 2000b, in press). Their position is that consent is not possible because CSA is immoral, and therefore scientific use

of *consent* is invalid; Spiegel (2000a) called such use a “moral outrage” (p. 66). They argued that we were wrong to assume that studies that asked students about any sexual experiences with older persons, as opposed to unwanted experiences only, contained much in the way of willing sex. We focus now on males, because it was with males that willingness moderated effect sizes in our study.

In studies involving male participants who were asked about all types of sexual experiences instead of just unwanted experiences, many have reported encounters that they themselves define as *willing*. In Condy et al.’s (1987) study of boys involved with women, where two thirds of the cases involved intercourse, only 14% of the sex acts fell into the “female forced male” category, whereas 49% fell into the “male wanted, female agreed” category and 67% fell into the “female wanted, male agreed” category. A. Nelson and Oliver (1998), in their college sample, found that 75% of sexual experiences between boys younger than 16 and adults 18 or older and at least 4 years older than the boys were “consensual” (their term). College studies that have included narratives along with quantitative data provide additional clear evidence for willingness (e.g., Fishman, 1991; Rind, 2001; West & Woodhouse, 1993). In West and Woodhouse’s interviews, as discussed previously, the more “severe” the sex was, the more willing the boy tended to be and the more positively the sex tended to be experienced. In a gay and bisexual male sample of mostly college students, 26 participants out of 129 had age-discrepant sexual relations between ages 12 and 17 with men at least age 18 and at least 5 years older than themselves (Rind, 2001). Twenty-three percent initiated their sexual contacts, while another 69% mutually consented. Positive reactions were strongly related with higher levels of consent ($r = .43$). Finally, in a recent large-scale, non-clinical study, Coxell, King, Mezey, and Gordon (1999) examined a sample of 2,474 men aged 18 to 94 in Great Britain recruited from general medical practices. Participants were asked about CSA occurring before 16 with someone at least 5 years older. In the entire sample, 7.7% of participants had “consensual” CSA—the authors’ term—whereas 5.3% had nonconsenting CSA. Thus, 59% of CSA was consenting. When asked whether they ever had a psychological problem of at least 2 weeks duration, nonconsenters had statistically significantly more problems than controls ($r = .10$) but consenters did not ($r = .02$). Sandfort (1992) examined a Dutch sample of 283 young adults aged 18 to 23, consisting of both students and working people. CSA was restricted to contact sex before age 16 with someone at least 5 years older. Most men who had experienced CSA were “consenting” (71%)—Sandfort’s term. Consenting participants were as well adjusted as controls. In sum, empirical evidence shows the existence of a substantial degree of consenting CSA among males, supporting our inference that studies asking male respondents about all experiences are likely to contain a nontrivial proportion of willing cases. It is important to recognize that willingness is not the same as informed consent, a point to which we return later.

Dallam et al. (2001) went on to calculate effect sizes for the “objective” measures for the male studies. For example, in the Fishman (1991) study (as indicated in their footnote 9), they only used one scale (sexual adjustment) to come up with their value ($r = .07$) instead of all scales for which results were reported, as we did, to come up with our value ($r = -.04$). This selective

approach lends itself to data picking and researcher bias, risks well known to meta-analysts.

Dallam et al. (2001) next contended that we were incorrect to conclude that there was a statistically significant difference in CSA-symptom relations for males in the two different levels of consent categories. They based this argument on correcting the effect sizes and then re-meta-analyzing. Correction for base rates is unnecessary and even inappropriate, as we argued previously in some detail. For argument's sake, however, we consider their analysis and present our own. Table 5 (a) repeats their meta-analysis shown in their Table 7, which was based on their corrected values that they presented in their Table 6; (b) presents our own meta-analysis of these same corrected values; and (c) presents our meta-analysis of corrected values from our original effect sizes. As can be seen, our meta-analysis of their corrected values differs from their presentation. In particular, ours shows that the 95% confidence intervals for all types of CSA and for unwanted CSA for males do not overlap, implying that these groups differ. Additionally, our meta-analysis, based on our corrections of our original effect sizes, shows that the 95% confidence intervals do not overlap. This further reinforces the validity of our original results.

Although nonoverlapping 95% confidence intervals imply that groups differ, overlapping confidence intervals do not imply that they are statistically the same, contrary to Dallam et al.'s (2001) remark that we "disregarded" (p. 724) overlapping intervals in forming our conclusions. Consider Glass and Hopkins' (1996) illustration: for Group 1, $r = .83$, $n = 30$ pairs, 95% CI = .67-.92; for Group 2, $r = .93$, $n = 83$ pairs, 95% CI = .89-.95. Although the confidence intervals overlap, the correlations are clearly different, as revealed by testing the significance of the difference between independent correlations ($z = 2.11$, $p < .05$, two-tailed). What matters is the 95% confidence interval of the difference between independent correlations, not the confidence intervals of

each independent correlation. Far from disregarding the overlapping intervals, we appropriately performed contrasts (analogous to the Glass and Hopkins example just given) demonstrating significant differences. In fact, it was Dallam et al. (2001) who disregarded the relevant statistics (our contrasts) for evaluating differences. Finally, Dallam et al. (2001) disagreed with our conclusion that "adjustment was associated with level of consent for men, but not for women" (Rind et al., 1998, p. 34), attempting to support their position with the irrelevant point that their corrected values for men and women in the all-levels-of-consent groups were nearly the same. Clearly, adjustment was associated with level of consent for men, indicated by the contrasts shown in Table 5 ($z_s = 3.63$ and 3.45 , $ps < .001$, two-tailed, for our meta-analyses of their and our corrected values, respectively).

To sum up, there is ample justification for our handling of contact sex, gender, and consent or willingness as moderator variables. The arguments raised against them by Dallam et al. (2001) are questionable or simply incorrect both in terms of empirical evidence and statistical practice.

Internal Validity

That correlation is not equivalent to causation is a fundamental tenet of research methodology. Using terms such as *effect* or *impact* is inappropriate in correlational research (especially when important confounds are present), as these terms imply causation (R. Rosenthal, 1994). Despite this, Dallam et al. (2001) used such causal language throughout their critique. They criticized our suggestion that the relation between CSA and adjustment might be spurious because of confounding, arguing that we made this suggestion "despite finding that students who reported a history of CSA were less well adjusted in 17 of 18" measures (Dallam et al., 2001, p. 725). However, this argument is unconvincing because these correlations were all small in magnitude and very similar to

Table 5
Level of Consent and Psychological Correlates for CSA in Male College Students Using Corrected r s by Analysis

Consent	k	N	r_c	95% CI	χ^2	z_{contrast}	N_{contrast}	p
Dallam et al. (2001) ^a								
All types	10	1,525	.11	.05-.19	7.92			
Unwanted	4	826	.22	.13-.30	1.93			
Dallam et al. (2001) redone ^b								
All types	10	1,855	.10	.06-.15	21.37*	3.63	2,845	.0003
Unwanted	4	990	.24	.18-.30	3.30			
Rind, Tromovitch, & Bauserman (current) corrected r s ^c								
All types	10	1,855	.08	.03-.12	17.58*	3.45	2,845	.0006
Unwanted	4	990	.21	.15-.27	1.06			

Note. CSA = child sexual abuse; k = number of samples in the meta-analysis; N = total number of participants in the k samples; r_c = unbiased effect size estimate based on base-rate corrected r s; CI = confidence interval for r_c ; χ^2 = within-group homogeneity statistic (based on $df = k - 1$); z_{contrast} = comparison of the all-types and no-consent effect size estimates; N_{contrast} = total number of participants in the all-types and unwanted levels; p = p value associated with z_{contrast} .

^a Reproduces Dallam et al.'s own analysis in their Table 7 based on their corrected r s from their Table 6. ^b Represents our computation of their Table 6 corrected r values. ^c Represents analysis of corrected r s based on our original effect sizes from Rind et al. (1998) in the Appendix.

* $p < .05$, in homogeneity analysis.

each other, as one might expect of correlated measures reflecting a single underlying construct; their consistency therefore does not add weight to the argument for causation. Ondersma et al. (2001) questioned whether efforts to control for FE are appropriate at all, claiming this is problematic when "environment and CSA events are so thoroughly intertwined" (p. 709). However, the overall relation between FE and CSA was .13, which, although representing confounding, does not constitute being thoroughly intertwined. We now respond to arguments that our approach to statistical control was invalid, and present new data and analyses that further question the proposition that the CSA-adjustment association can safely be assumed to be causal.

Family Environment

Dallam et al. (2001) argued that the "deck was stacked against" (p. 725) CSA when compared with FE in accounting for adjustment because of the way CSA was treated statistically. CSA was often a dichotomous measure, whereas FE was often continuous, and they argued that our discussion of continuous CSA measures was unconvincing because these measures had no reliability or validity data and were retrospective. These latter criticisms apply to all CSA research, including that cited by our critics to bolster their arguments, and can hardly be selectively applied to our studies. We cited Wisniewski (1990) as a study that used a continuous measure for CSA yet found FE substantially more important in predicting current adjustment. Wisniewski constructed the continuous measure from degree of felt victimization and negative reactions, a face-valid measure of actual severity that would also have face validity for adult-adult sex, as opposed to severity based on a hierarchy of types of sex from exhibitionism to intercourse. Moreover, she used a large sample ($N = 3,187$) of female college students drawn from 32 colleges and universities fairly representative of those across the United States, increasing the generalizability of her inferences to the college population. In a path analysis, she found that family violence, not CSA, was behind current adjustment problems. Further, her zero-order effect size for CSA ($r = .11$) was consistent with most other college studies. In their national study, Laumann et al. (1994) created a more conventional severity index based on a hierarchy. They found that this continuous measure did not improve prediction over their dichotomous measure. Like the Wisniewski study, the Laumann et al. study is important because of its representativeness (the U.S. adult population) and large sample size. These findings, cited in our original review, are an important counterweight to the "stacked deck" argument.

Dallam et al. (2001) went on to argue that nonclinical samples do not have many severe cases. If *severity* means intercourse, we have already shown that this claim is incorrect. They next argued that our statistical control came predominantly from analyses of covariance (ANCOVAs) and hierarchical regression, where the covariate (FE) should be unaffected by the independent variable (CSA) for valid control. From this, they claimed that CSA and FE would have to be uncorrelated. This last claim is incorrect. Partialing a third variable from two primary variables, both of which are correlated with it, is valid if one assumes the third variable is either a mediator of the two primary variables (i.e., caused by one and causing the other in turn) or a common cause of both of them (Pedhazur, 1997). Partialing is problematic when, for example, the

third variable is affected (caused) by the independent variable, which affects the dependent variable, and the dependent variable is also affected by the third variable. Then, partialing may remove too much of the relation between the primary variables—for example, if CSA causes FE, which causes symptoms, and CSA directly causes symptoms, then controlling for FE removes part of the CSA effect. The latter is the critical case and is what Briere (1988) and others have argued happens in the real world: CSA causes harm and disrupts families, which results in more harm. We addressed this issue in our original article (Rind et al., 1998), presenting evidence to argue that, although Briere's concern may be valid in incest cases, it was unlikely to apply in the bulk of the cases we examined in the college samples, where close family CSA was rare and CSA involving a parent even rarer. Finkelhor et al.'s (1990) national sample findings are also relevant to this issue. They conducted a series of analyses that indicated that "unhappy family life is a true risk factor [for CSA] and not simply a distorted perception that a victim develops as a result of having been abused" (p. 24). In other words, contrary to Briere's concern, their findings favored causality going from FE to CSA and not the reverse. This is consistent with Pedhazur's criteria for valid partialing.

Dallam et al. (2001) next claimed that our findings are at variance with other large-scale nonclinical studies "whose methods are *more appropriate* for disentangling the effects of CSA from those of family dysfunction" (p. 725, italics added) and presented these studies in their Table 8. We find it curious that the methods of ANCOVA and hierarchical regression were labeled *inappropriate* in the studies we used but were considered more appropriate when used in 5 of 13 samples cited by Dallam et al. (see our Table 6). In any case, the studies cited are quite mixed in terms of degree of including relevant covariates: In half the samples, these control variables were weak (see Table 6), not including factors such as physical abuse and emotional neglect. Effects of statistical control were mixed and were compromised in the studies with weak covariates. On a further note, we computed sample-level effect sizes for each sample, combining emotional and behavioral measures. Meta-analytic results for all samples combined yielded $r_u = .12$, $k = 11$, 95% CI = .10–.14, $\chi^2(10, N = 14,720) = 24.26$, $p < .05$. Results for the male samples were $r_u = .08$, $k = 3$, 95% CI = .05–.11, $\chi^2(10, N = 4,390) = 2.10$, $p > .05$. Results for the female samples were $r_u = .13$, $k = 6$, 95% CI = .11–.15, $\chi^2(5, N = 8,732) = 12.56$, $p < .05$. The female unbiased effect size estimate was greater than the male value ($z = 2.74$, $p < .02$, two-tailed). As with the results of the high school and junior high samples discussed previously, these results are consistent with those from the college and national samples.

Precocious Sex

Clearly, meta-analyses of nonclinical samples show that both men and women with a history of CSA are slightly less well adjusted than controls. However, in our society, minors in general who have *precocious sex*—for example, willing peer intercourse at a young age—are also less well adjusted (e.g., Jessor, Costa, Jessor, & Donovan, 1983; Ketterlinus, Lamb, Nitz, & Elster, 1992; Resnick & Blum, 1994; D. Rosenthal, Smith, & de Visser, 1999). Early sex is nonconventional in our society, but not cross-culturally (Ford & Beach, 1951), and reflects a package of emo-

Table 6

Study-Level Effect Sizes for Dallam et al.'s (2001) "Well-Designed Nonclinical Studies That Controlled for Family Dysfunction"

Study	Key phrase for CSA operational definition	Control method	Quality of FE control	Gender	N	r
Boney-McCoy & Finkelhor (1995)	"Threaten, force, or trick" into sex "you didn't want"	ANCOVA	Poor	F	911	.20
Boney-McCoy & Finkelhor (1996)	"Threaten, force, or trick" into sex "you didn't want"	ANCOVA	Poor	M	987	.12
	"Threaten, force, or trick" into sex "you didn't want"	ANCOVA	Poor	F	686	.22
	"Threaten, force, or trick" into sex "you didn't want"	ANCOVA	Poor	M	771	.14
Dinwiddie et al. (2000)	"Forced into sexual activity"	Logistic regression, twin control	Good	F	3,868	.12
	"Forced into sexual activity"	Logistic regression, twin control	Good	M	2,078	.08
Fergusson et al. (1996)	"Did not want to happen"	Logistic regression	Good	F,M	1,019	.16
Fleming, Mullen, Sibthorpe, & Bammer (1999)	<12/5+ or 12-16/5+ "unless wanted or not distressing"	Logistic regression	Good	F	710	.05
Johnson, Cohen, Brown, Smailes, & Bernstein (1999)	"Older person touched them or forced them to touch"	ANCOVA	Good	F,M	579	.14
Kendler et al. (2000)	"Unwanted"	Logistic regression, twin control	Good	F	1,411	.17
Mullen et al. (1993)	"Unwanted sexual advances"	Hierarchical regression	Good	F	474	.16
Stein, Golding, Siegel, Burnam, & Sorenson (1988)	"Contact from pressure or force"	Logistic regression	Very poor	F	1,358	.12
	"Contact from pressure or force"	Logistic regression	Very poor	M	1,325	.06

Note. Quality of family environment (FE) control is poor if nonsexual abuse factors are omitted, and very poor if only demographics are controlled for. CSA = child sexual abuse; ANCOVA = analysis of covariance; F = female; M = male.

tional, behavioral, familial, and social correlates that are maladaptive according to our society's standards. Jessor et al. (1983) described this package in terms of *behavior problem theory*, which is composed of three systems that promote nonnormative behavior. In the personality system, proneness is represented in the motivational structure; for example, the young person places a lower value on academic achievement, is more tolerant toward deviance, and is less religious. In the environment system, proneness comes from lesser parental control, greater peer influence, and social models for problem behaviors. In the behavior system, proneness is reflected in greater involvement in other problem behaviors and simultaneous less involvement in conventional behaviors, such as doing well in school. Early sex is seen as originating in these systems rather than causing them. This point is relevant to cause and effect regarding CSA.

Using the four studies just cited, we meta-analyzed various correlates of precocious consensual sex that did not involve CSA (*consensual* is the authors' term). Table 7 provides the results for each study. Precocious sex was associated with more school problems ($r_u = .21$), greater use of drugs ($r_u = .18$), lesser religious involvement ($r_u = .14$), greater peer influence and peer models for nonconventional behavior ($r_u = .34$), and poorer family environment ($r_u = .16$). Jessor et al. (1983) found that precocious sex was strongly associated with general problem behavior ($r = .45$). Resnick and Blum (1994) found that precocious sex, defined as *consensual* intercourse before age 10 (the authors' term), was associated with mental health problems ($r = .10$). In short, this research shows that early sex is associated with a wide range of problems, emotional and behavioral, and/or nonconventional behavior. Of importance is that there is no indication or argument by these researchers that precocious sex causes other problems. In-

Table 7
Psychological Correlates of Precocious Sex

Study	n	School	Drugs	Religion	Peers	Home
Jessor et al. (1983)	254	.20	—	.22	.45	.19
Ketterlinus et al. (1992)	3,760	.25	.18	.13	—	—
Resnick & Blum (1994)	1,866	.13	—	—	.32	.16
D. Rosenthal et al. (1999)	241	—	.15	—	—	—
r_u		.21	.18	.14	.34	.16
N		5,880	4,001	4,014	2,120	2,120
95% Confidence interval		.19-.23	.15-.21	.11-.17	.30-.37	.12-.20
χ^2 (homogeneity statistic)		19.40*	0.21	2.03	5.18*	0.21

Note. Table contains effect sizes (r); positive values indicate more problems or greater nonconventional behavior for the precocious sex group than for controls. Dashes indicate that the study in question did not assess the behavior in question. r_u is the unbiased effect size estimate.

* $p < .05$.

stead, it is a correlate of other problems with the common cause of personality, home, and social environmental factors.

Behavioral and emotional correlates of CSA look very similar to those of peer precocious sex. Therefore, causal inferences about CSA effects need to take into account the findings from research on precocious sex with peers. Controlling for FE alone in CSA research may in fact insufficiently control for other key influences on the CSA–adjustment relationship. Problem behavior theory questions even further the degree to which CSA, particularly a single event occurring many years previously and not perceived as especially important at the time, can be labeled the sole or even primary cause of current adjustment difficulties.

Other Issues

Finally, it is important to add that correlational research has serious limitations in drawing causal inferences. This is much less of a problem, however, when zero-order correlations are nonsignificant or valid statistical control eliminates significant relations. In these situations, inferences can be made, especially when results are replicable, because the absence of correlation implies a lack of causation. When correlations remain significant after statistical control, however, causation cannot be firmly assumed, given the problem of unexamined third variables. Future research, to more adequately address causality in the case of CSA and adjustment, needs to focus more on alternative research strategies, such as prospective studies, studies in which family and social environment variables are carefully assessed rather than self-reported, national stratified samples, cross-cultural studies (especially in societies that do not hold modern Western attitudes about CSA in particular and teenage or child sexuality in general), and genetically informed research. Regarding the last of these, the Dinwiddie et al. (2000) and Kendler et al. (2000) studies using twin controls are a start, one finding an elimination of significant CSA–symptom relations and the other, not. Such research would be improved if it included cases in which the twins perceived themselves as willing or assenting in the CSA, as the two studies just cited included unwanted CSA only yet assumed to be addressing the more general issue of sociolegal CSA. Also, such studies could examine precocious sex as discussed above to further clarify the issue of CSA and causality.

Comparison of Kendall-Tackett et al. (1993) and Rind et al. (1998)

The last review article on CSA to appear in *Psychological Bulletin* before ours was that by Kendall-Tackett et al. (1993). This review has been widely cited in the psychology and psychiatry fields as key evidence for the pervasive and intensely harmful effects of CSA. In the same vein, both Dallam et al. (2001) and Ondersma et al. (2001) cited this review as authoritative. Although they have scrutinized the external and internal validity of our review, they did not offer, nor have they offered in the past to our knowledge, any criticism of the Kendall-Tackett et al. review; rather, they accepted its methods and findings uncritically. We argue that this lack of scrutiny of Kendall-Tackett et al. and simultaneous intense scrutiny of our review represents selective criticism, for much is to be criticized in the Kendall-Tackett et al. review. Now that we have responded to criticisms of our external

and internal validity, we examine Kendall-Tackett et al. along these same dimensions. This comparison will achieve two things: First, it will support our contention that Dallam et al.'s (2001) and Ondersma et al.'s (2001) criticisms of our review were selective; second, it will show the advancements our review offers to the field, with its more careful attention and sounder approach to external and internal validity, two substantial indicators of scientific quality.

External Validity

As we noted previously, the Kendall-Tackett et al. (1993) mean effect sizes were .57 for emotional and .63 for behavioral problems. These results were based on sexual abuse treatment samples, not nonclinical samples. Compared with nonclinical junior and senior high school students, however, these effect sizes were highly anomalous, being 2.86 and 3.77 SDs above the mean effect sizes for all studies combined. Clearly, the Kendall-Tackett et al. samples were outliers, highly unrepresentative of the general population of minors. The results from our meta-analyses of national and college samples (Rind & Tromovitch, 1997; Rind et al., 1998), on the other hand, were almost identical to the unbiased effect size estimates of the nonclinical junior and senior high students, as they were to the nonclinical samples that Dallam et al. (2001) mentioned in their Table 8, which we summarized in our Table 6. In our review, we paid explicit attention to the issue of external validity, making appropriate comparisons between college and national samples. By contrast, Kendall-Tackett et al. completely ignored the issue of external validity, except in a single footnote near the beginning of their review. Consigning such reference to a single footnote and ignoring this important issue entirely in their Discussion section created the impression that their findings were more broadly relevant than they actually were. In sum, our review was relatively strong in its treatment of external validity, whereas theirs was weak. Critics of our review on this issue would have more balanced arguments if they applied their criticisms equally to reviews that favor their point of view, rather than accepting them uncritically.

Internal Validity

Our review of the college studies was a critical review of causality. It added to previous meta-analytic reviews, where causality could not be analyzed because the primary studies had provided insufficient data on third variables and statistical control (Jumper, 1995; Neumann et al., 1996). Both Jumper and Neumann et al. pointed to the need for future research to address this weakness. Our review was one such response to this problem, made possible by the fact that the college studies had sufficient relevant data. Whether our findings regarding causality hold up in future investigations is less important than the fact that they formed a central focus of our presentation, as they should have. The Kendall-Tackett et al. (1993) review, on the other hand, has been par for the course of victimological research on CSA, accepting more or less uncritically CSA's causal role. This review included various studies based on daycare satanic ritual abuse (SRA), such as one on the McMartin preschool children (Kelly, 1993), in which nearly half the children fell in the clinical range of PTSD symptomatology. However, the McMartin case has been so

thoroughly discredited as a case of implanted memories of abuse rather than real abuse (e.g., Nathan, 1990; Nathan & Snedeker, 1995) that it seems negligent for Kendall-Tackett et al. not to have informed their readers specifically that these were McMartin data and must be viewed with skepticism. The dramatic effects in this case were attributed to CSA, when in fact they were in greatest likelihood iatrogenic. They also included a review of SRA cases by Kelley (1989), in which again nearly half the children were in the clinical range of PTSD symptomatology, and once again we are given the impression CSA has dramatic effects, when clear alternative explanations are apparent. For instance, Kelley (1990) showed that parents of these children were highly disturbed, which suggests that they may have been passing the anxiety on to their children or seeing it in them even if not there. Kendall-Tackett et al. themselves acknowledged that the mothers' judgments about their children's symptoms were highly related to their own level of distress and willingness to believe the children—most reports on child symptoms came from parent-completed checklists.

Kendall-Tackett et al. (1993) dismissed parental reporting bias, noting that therapist judgments were similar, although children's self-reports were much less negative and mothers' reports were poorly related to their children's reports. But because of researcher bias, including demand characteristics and expectancy effects (R. Rosenthal & Rosnow, 1969; Rosnow & Rosenthal, 1997), one cannot assume the validity of therapist judgment, especially given Kendall-Tackett et al.'s note of biased reporting: "Unfortunately, few investigators have reported on . . . asymptomatic children, perhaps out of concern that such figures might be misinterpreted or misused" (p. 168). Aside from discussing this issue of measurement validity, which is relevant to internal validity, Kendall-Tackett et al. never directly addressed the issue of causality; rather, they just assumed it. They assumed it so strongly that they attributed the large percentage of asymptomatic children in their studies to insensitive measures rather than lack of caused harm.

Additionally, Kendall-Tackett et al. (1993) inflated the impression that CSA causes harm by calling sexualized behavior a *symptom*—this was the most common symptom they found. But sexualized behavior is not a symptom of disease or distress—arguing that it is constitutes a value judgment. As Ford and Beach (1951) observed in their seminal review of cross-cultural and cross-species data, "as long as the adult members of a society permit them to do so, immature males and females engage in practically every type of sexual behavior found in grown men and women" (p. 197). From their review they concluded that "tendencies toward sexual behavior before maturity and even before puberty are genetically determined in many primates, including humans" (p. 198). Regarding the relative lack of sexuality in juveniles in our and similar societies compared with others, they noted that this is a product of the restrictive nature of these societies rather than the nature of juveniles:

The extreme pains to which adults in these societies are forced to go in order to control the sexual behavior of young people is an eloquent expression of the strength of the tendency on the part of older children and adolescents to engage in such activity. (p. 182)

Ford and Beach's findings argue against labeling sexualized behavior a symptom on par with depression, anxiety, and suicidal ideation. Such behavior may be inappropriate or undesirable according to social norms or practical concerns in our culture, but

these facts are not a valid basis for medicalizing it (cf. Szasz, 1990).

In sum, although our review should not be viewed by any means as definitive regarding internal validity, it was nevertheless a much needed critical review that directly dealt with the issue. Although inferring causality from correlational data is fraught with problems because of unexamined third variables, one is on much stronger ground in inferring lack of support for causality when factors are no longer correlated after statistical control, because a requirement of causation is correlation. Our review, which followed this logic in assessing causality, represents an advance over the Kendall-Tackett et al. (1993) review, which by contrast paid scant attention to causality—just assuming it instead—and inflated the impression of causal effects with some questionable data, measures, and definitions of harm. Critics of our review, to be balanced, should also discuss the weaknesses of the Kendall-Tackett et al. review, as well as the many other similar reviews.

Qualitative Analysis

Dallam et al. (2001) questioned our qualitative analyses. First, they criticized our conclusion that when negative effects occur they are often temporary. Next, they argued that our conclusion that "lasting negative effects are not prevalent" applies only to studies looking at "milder forms of abuse" (p. 727). Finally, they claimed that there are "numerous" (p. 727) problems with how we coded and reported data on reactions. Similarly, Ondersma et al. (2001) challenged self-reports and argued that short-term harm is well documented. We now rebut these claims.

Temporary Versus Lasting Harm

Dallam et al. (2001) claimed that our conclusion that negative effects are often temporary is severely limited because only three studies examined this issue (Landis, 1956; Nash & West, 1985; West & Woodhouse, 1993), including very little in the way of serious abuse. It is true that only these three studies directly asked how long any effects might have lasted. Nevertheless, all 11 samples (from eight studies) with data on self-reported effects (see Rind et al., 1998, Table 8) are relevant to lasting versus temporary harm. If a certain percentage in any study reported no negative effects, then by definition negative effects were not lasting for those cases. Taken together, self-perceived lasting harm was rare in these samples. Additionally, Dallam et al.'s (2001) claim that these samples included little in the way of serious (i.e., oral sex and intercourse) CSA is incorrect. For example, as previously discussed, Condy et al. (1987) reported that 68% of the male students had intercourse CSA, yet only 16% of the entire sample reported negative effects. West and Woodhouse (1993) had a notable proportion of trivial homosexual advances, as Dallam et al. (2001) noted, but also a fair amount of serious CSA, as we discussed previously, with virtually no reports of any lasting negative effects. A college study we failed to include in the table was that of Fromuth and Burkhart (1987); we relied on their 1989 publication for our meta-analysis. In their sample, 30% of males reported oral sex and another 25% reported intercourse, yet only 15% reported negative effects (Fromuth & Burkhart, 1987).

In a previous rebuttal (Rind, Tromovitch, & Bauserman, 2000a) to Dallam et al.'s (1999) argument that Landis (1956) included too

many noncontact cases, we adjusted his results on the basis of the assumption that all negative effects came from contact sex. We found that lasting self-reported negative sexual effects at a maximum would have occurred in 4% of male and 6% of female contact cases, and lasting general negative effects would have obtained in 0% of male and 7% of female contact cases. These results suggest that the Landis study is indeed informative about lasting effects of CSA.

Finally, Dallam et al. (2001) endorsed Nash and West (1985) as the "most relevant of the three studies" (p. 727; the other two were Landis, 1956, and West & Woodhouse, 1993). They noted that Nash and West used a broad definition of CSA and found that 22% of their female participants reported they were still being adversely affected. It is unclear why the Nash and West study is more relevant than the West and Woodhouse study. Similar definitions of CSA were used in both, English college students formed the samples in each, and West oversaw both studies. The only important difference was that the first study involved women, 22% of whom felt lasting negative effects, whereas the second involved men, with only 2% feeling lasting negative effects. In comparing the female and male reactions in these two samples, West and Woodhouse noted the "particularly striking contrast" (p. 122) between them. Both female and male narratives are relevant, and it is unclear why one source should be taken as more relevant than the other, unless *relevance* is confused with the degree to which a study indicates harm.

Effects on Current Life

Dallam et al. (2001) claimed that we misreported Fishman's (1991) results on negative general and sexual effects. They noted that we reported these as 27% and 13%, respectively, but claimed, citing Fishman (1991, p. 162), that the correct figures are 47% and 23%, respectively. In fact, the latter figures refer to homosexual encounters only. The corresponding figures for heterosexual encounters were both 0%. When both homosexual and heterosexual encounters are combined, these percentages turn out exactly as we reported them.

Initial Emotional Reactions

In Dallam et al.'s (1999) initial critique, they stated that we were "ill responsible [sic]" with seeming "intent on misleading the reader" (per personal communication from R. Fowler, June 4, 1999) in how we coded the reaction data. In rebuttal, after arguing that our coding was valid, we showed that removing the disputed studies had virtually zero effect on the percentages of positive, neutral, and negative reactions (Rind, Tromovitch, & Bauserman, 2000a). In their most recent critique, Dallam et al. (2001) renewed this claim by citing three different examples (Brubaker, 1991, 1994; Fishman, 1991). They claimed that Fishman reported 53% negative reactions at the time, not the 30% value we reported. Once again, the figure they cited refers to homosexual encounters only, not all encounters that the men had; Fishman reported 0% negative reactions at the time to heterosexual encounters. The figures we reported were based on all experiences. For the Brubaker samples, they cited her findings from a unipolar scale of degree of upset feelings. This is not the measure of reactions relevant to our analysis because it did not classify reactions into

positive, neutral, and negative categories, as we stipulated a measure must do for inclusion. We used Brubaker's (1991, 1994) scale of pleasure, interest, surprise, shock, and fear, where the first two are labeled as positive reactions, the third is neutral, and the last two are negative. This classification began with Landis (1956) and was subsequently used by other researchers with college samples (e.g., Finkelhor, 1979a; Goldman & Goldman, 1988). In short, our coding was neither erroneous nor misleading, but was consistent with practices followed by other researchers.

Undersma et al.'s (2001) Critique of Self-Reports

Related to the issue of harm, Undersma et al. (2001) argued that

harm does not require that the victim perceive the experience negatively. For example, the possibility that a child might learn from an abuser that such experiences are normal and positive is one of the most concerning possible outcomes of CSA. (p. 709)

We believe this assertion is unscientific because it sets up the problem of unfalsifiability (Popper, 1961). All outcomes become evidence for harm, negative and positive alike. By the same logic, organizations that regard homosexuality as pathological (e.g., National Association for the Research and Therapy of Homosexuality) can define it as such even, and perhaps especially, if the gay person has a positive self-identity. To repeat from our original article, the wrongfulness and harmfulness of a behavior are separate issues—and the kind of harm that mental health professionals should be concerned with is mental health (e.g., depression), not attitudinal deviance.

Undersma et al. (2001) went on to cite Kendall-Tackett et al. (1993) for well-documented evidence of short-term harm and equivalence of outcomes for boys and girls. But even Kendall-Tackett et al. did not maintain that their finding of equivalence should be seen as robust, given the "bias in identification of male victims, [where] only the most symptomatic boys end up in clinical samples" (p. 170). In addition, as we discussed previously, the Kendall-Tackett et al. review has internal validity problems concerning dubious data, measures of harm, and definitions of harm, all of which weaken the authority of this review as a basis for claiming that short-term harm has been well documented.

Undersma et al. (2001) also criticized our treatment of boys' reports of positive experiences by stating that we failed to consider alternative explanations for such reports, such as the possibility that some men and boys may refuse to recognize themselves as victimized because of male socialization or successful indoctrination by the "abuser." This criticism represents a curious double standard. Our critics, and many CSA researchers, have had no difficulty accepting reports of negative experiences at face value, but have selectively denied any validity to reports of positive experiences and seek to discount them as the result of processes such as denial (Rind, Bauserman, & Tromovitch, 2001a). A more objective stance would hold that neither type of experience can be selectively dismissed or explained away. In our review, we simply summarized and reported how both men and women said they perceived their experiences, without assuming that positive or negative perceptions were inherently suspect.

To our knowledge, no one has systematically studied the sorts of pressures that might exist on individuals to redefine experiences initially seen as positive, as well as those initially seen as negative.

Although it may be true that denial plays a role in reporting some experiences as positive, it is at least equally plausible to suggest that those who initially perceive their experiences as positive are subjected to massive social pressures to redefine them as negative. Certainly, everyday messages in society about CSA experiences are uniformly negative, and anyone who perceives his or her own experience otherwise is exposed to constant messages about how damaging and traumatizing such experiences are. In the absence of well-conducted research on the types and relative strength of pressures to redefine contacts experienced as positive versus negative, there is no empirical justification for singling out self-reported positive experiences as uniquely unreliable. It is clear that boys were significantly more likely than girls to report their experiences as positive, but it is quite possible that this simply reflects the well-documented finding that males are primed for casual sex far more than females, a difference present from adolescence on. A vast and growing body of research documents such differences in male and female sexual attitudes and behaviors (Sommers, 2001).

Construct Validity

Although Dallam et al. (2001) in particular, and Ondersma et al. (2001) as well, vigorously challenged our analyses, coding, and other methods, we believe that much of the reaction to our meta-analysis stemmed from other issues. In particular, we believe that our recommendation regarding more cautious use of the term *sexual abuse* and our referral to some CSA experiences as *consenting* triggered the strongest reactions. We now turn to these issues.

A primary thesis of Ondersma et al. (2001) is that we misused science by means of "extrascientific" (p. 712) considerations, especially when suggesting that professionals use more value-neutral language when discussing some CSA experiences. We feel it is important to discuss how this section of our review came about and its scientific justification in terms of construct validity.

In our original drafts submitted to *Psychological Bulletin*, we did not make any recommendations concerning use of value-neutral terms. We did so in the final draft in response to comments by reviewers and the action editor, who noted that we had found that broad definitions of CSA resulted in poor predictive validity and, by omitting the context of a sexual contact, might fail to adequately capture the "essence" of "abuse" (K. J. Sher, personal communication, May 14, 1997). In short, the goal was to improve construct validity, such that *abuse* would be predictive of actual or probable negative outcomes. The section of our Discussion labeled "Child Sexual Abuse as a Construct Reconsidered" and the recommendations offered in our Conclusions section represented our ideas as to how best identify and study those adult-adolescent and adult-child sexual contacts most likely to be associated with harm. We never argued that the term *child sexual abuse* should not be used, rather its use should be restricted to experiences (e.g., those involving coercion or force or involving negative reactions) that are more predictive of negative outcomes. Furthermore, we specified that we were concerned with research definitions of CSA, not social and legal definitions, and specifically noted that whether an experience caused harm was not the same as whether it was wrongful. These recommendations were concerned with improving the predictive validity—and in turn, the construct validity—of

CSA. Clearly, alternative approaches are possible. For example, Brant and Tisza (1977) proposed labeling some CSA experiences *sexual misuse* rather than *abuse*, to distinguish experiences not associated with harm from those that are. However, we felt it more useful to avoid terms that carried any implications or assumptions of psychological or other harm unless specifically justified by the context.

Ondersma et al. (2001), and other critics, implied that we were extreme in suggesting that *abuse* might not be an appropriate label in some cases. However, many sexologists have argued that uncritical use of victimological language in CSA research creates problems in scientific validity (e.g., Green, 1992; Kilpatrick, 1987; Money & Weinrich, 1983; J. A. Nelson, 1989; Okami, 1990; Sandfort, 1992; West, 1998). Some authors of the college studies argued this as well (e.g., Fishman, 1991; Fromuth & Burkhart, 1987; Long & Jackson, 1993; West, 1998). For example, West (1998), a criminologist and sexologist from Cambridge University and coauthor of two of the college studies in our review, argued that professional use of terms such as *abuse*, *perpetrator*, *victim*, and *survivor* has incorrectly reinforced the idea that any kind of sexual incident with a child is likely to cause great and lasting harm, stating that this usage has introduced a moral tone "alien to scientific inquiry" (p. 539). After qualitatively reviewing the CSA literature, Green (1992), a psychiatrist, lawyer, and editor of the *Archives of Sexual Behavior*, reached similar conclusions to ours. He stated the following:

Ultimately, scientists, if no one else, must be objective in their approach to this emotional issue. Judgmental terminology regarding intergenerational sexuality is more dramatic than that in the earlier psychiatric literature on homosexuality. There, patients were labeled *perverts* and *psychopaths*. Here, the experience is always *abuse*, the children are invariably *victims*, the adults are *perpetrators*, and those who later report childhood sexual experiences are, without apology to victims of the Nazi Holocaust, *survivors*. (p. 175)

Thus, our recommendations were in line with both scientific principles of validity and sexological precedent. Emotionally loaded victimological terminology carries clear and strong implications about the nature and effects of the experiences it describes. It is reasonable to question whether such implications are always appropriate.

The further one moves away from clinical and psychiatric reports, the clearer it becomes that neutral rather than victimological language is the norm. In Laumann et al.'s (1994) national study, age-discrepant contact sex was simply labeled *sexual touching*. Many nonclinical studies that are based on convenience samples of boys' sexual experiences with older persons have avoided *abuse* terminology unless abuse clearly applied to individual cases (see Bauseman & Rind, 1997). "Abuse" terminology is almost completely absent when moving beyond psychology into anthropology, zoology, and history. Ford and Beach (1951) explicitly stated that "moral evaluations form no part of this book" (p. 14), at a time when many of the forms of sex they analyzed were taboo and/or illegal in the U.S. Hundreds of anthropological and zoological publications since Ford and Beach have avoided such terminology (e.g., Greenberg, 1988; Gregersen, 1983; Vasey, 1995). Because these researchers have not been concerned with moralizing, treating, preventing, and curing, they have been able to approach these phenomena as scientists attempting to describe and

understand, and they frequently have discussed functions such behaviors appear to serve—something that is not possible if one is constrained by a disease model. The same applies to historical research, in which scholars discussing pederasty in ancient Greece or Rome (e.g., Cantarella, 1992; Percy, 1996) or early modern Japan (e.g., Schalow, 1989) almost invariably use neutral language.

Ondersma et al.'s (2001) claim that value-neutral language is a threat to the moral fabric because "small but vigorous" (p. 712) minorities will exploit it is an example of the "argument from adverse consequences fallacy" (Lilienfeld, in press; Sagan, 1995). This is the error of evaluating the validity of an argument by considering its potential negative consequences. Use of neutral language is part of scholarship, even that which deals with CSA. Combined with selective use of terms such as *abuse*, it enhances predictive validity. Value-laden language, we argue, has the effect of framing all aspects of the issue in a particular moral and political viewpoint. In turn, this obscures the empirical questions of the presence and degree of harm.

Consent

Both Dallam et al. (2001) and Ondersma et al. (2001) rejected the notion of consent; the latter complained that we emphasized some adults' recollections of their experiences as *wanted*, which they argued implies that children and adolescents can make informed decisions about sex with adults. Because consent, like terminology, was a primary stimulus for the many attacks on our original article, it is important to address this issue.

Simple Consent

First, *consent* does not always mean informed consent. Webster's Third New International Dictionary (1981), for example, defines consent first as "compliance or approval, especially of what is done or proposed by another" and second as "capable, deliberate, and voluntary agreement to or concurrence in some act or purpose implying physical and mental power and free action" (p. 482). The first definition may be termed *simple consent*, the one used in the primary studies we examined, as well as in other nonclinical research (e.g., Condry et al., 1987; Coxell et al., 1999; A. Nelson & Oliver, 1998; Rind, 2001; Sandfort, 1992; West & Woodhouse, 1993). Simple consent, which might alternatively be labeled *willingness* or *assent*, can be observed in adolescents or children in a whole range of behaviors. As discussed previously in this rebuttal and elsewhere (Rind, Bauserman, & Tromovitch, 2000; Rind, Tromovitch, & Bauserman, 2000a), the simple consent construct has predictive utility, making it a scientifically valid construct. The second definition involves *informed consent*, which was not implied in our study or any others just cited. At no point did we claim in any way that adolescents or children, even if they perceive their sexual contact with an adult as willing, are providing informed consent in an adult sense.

Because our critics appear to view even simple consent, willingness, or assent as impossible by definition, on the basis of legal and moral arguments (e.g., Finkelhor, 1979b, 1984), they made errors in inference. For example, as shown in Tables 1 and 6, all definitions in the junior and senior high school surveys and the community surveys that Dallam et al. (2001) used to make broad

statements about CSA all specified unwanted CSA, not CSA in general (i.e., sociolegal CSA). Clearly, many participants in national, community, college, and secondary school samples seem willing to make distinctions about whether their sexual contacts were wanted or unwanted, willing or unwilling.

Informed Consent

Firm statements are made about informed sexual consent, as if this construct has been empirically studied. To our knowledge, it has not. Instead, opinion is drawn from moral philosophy and the law. Ondersma et al. (2001) cited Finkelhor (1979b) as an example of cogent thinking on this issue. Finkelhor argued that harm is not required to condemn CSA. Rather, it is wrong because children cannot consent because they do not know what they are getting into and cannot say no. These shortcomings are no problem for nonsexual behaviors, Finkelhor (1984) later argued, because CSA is more likely to be harmful. This circular reasoning is not cogent. With respect to the law, statutes vary considerably across nations. Whereas the median age of consent is 16 in the United States, it is 14 in Europe, ranging from 12 to 17 (Graupner, 2000). At times it has been set as high as 21 but historically has been considerably lower, with an age of 10 in most U.S. states before the 1880s (Jenkins, 1998). Thus, many cases considered CSA in current U.S. research are not legally such in other Western countries or even in the United States in the past. The law can also be contradictory, as in the case of teenage girls who can consent to sex with much older men in many states if married but cannot otherwise. In short, legal statutes are not a reliable guide for scientific evaluation of ability to consent.

In a related area (consent to an abortion in adolescence), the APA prepared an amicus curiae brief for the U.S. Supreme Court in October 1989 in which, on the basis of a review of cognitive, social, and moral development, they concluded

by age 14 most adolescents have developed adult-like intellectual and social capacities including specific abilities outlined in the law as necessary for understanding treatment alternatives, considering risks and benefits, and giving legally competent consent. . . . [Additionally,] there are some 11- to 13-year-olds who possess adult-like capabilities in these areas. (p. 20)

These conclusions, which were based on developmental research in many areas, cast doubt on the validity of automatic inclusion of adolescents into the category of CSA on the basis of an informed consent criterion. This validity is further weakened by the opinions of various European governmental commissions assigned to study the legal age of sexual consent, most of which recommended 14 (e.g., Austria, Denmark, Germany, Sweden, Switzerland; Graupner, 1997).

Sociohistorical Context

We believe that Ondersma et al.'s (2001) discussion of the sociohistorical context casts the issue of CSA in a victimological framework that impedes rather than promotes scientific discussion of CSA (cf. Jenkins, 1998). This section of theirs lays out the philosophical foundation of modern victimology as a struggle for truth against prejudice and obscurantism (Jenkins, 1998), leading Ondersma et al. (2001) to suggest that "no amount of explaining"

(p. 711) will suffice to make us see why sexual acts involving children are seen in a particular way by society at large. As a basis for this context, they cited Olafson et al. (1993).

Olafson et al. (1993) accepted the "discoveries" of Janet and Freud regarding CSA and memory repression as verified fact, although substantial reviews of the empirical evidence by academic researchers do not support repression (Brandon, Boakes, Glaser, & Green, 1998; Pope & Hudson, 1995). Early founders of their perspective were described in glowing terms (e.g., the "venerable intellectual lineage" traceable back to Janet; Olafson et al., 1993, p. 11), whereas opinions of skeptics were characterized as a "powerful backlash" (p. 16). They dismissed historical, cross-cultural, and cross-species perspectives on adult–juvenile sex as rationalizations for CSA, although sexology recognizes the scientific relevance of these approaches (e.g., Bullough & Bullough, 1977; Ford & Beach, 1951; Greenberg, 1988). They commented that it

remains to be seen whether the current backlash will succeed in resuppressing awareness of sexual abuse, again concealing 'vast aggregates of pain and rage' . . . and returning us to the 'shared negative hallucination' that has obscured our vision in the past. (Olafson et al., 1993, p. 19)

In short, their "historical" review blends unverified clinical opinion and victimological ideology to justify a particular social and therapeutic agenda.

Ondersma et al. (2001) did not cite Jenkins (1998), who documented the development of current stereotypes about CSA as a social construction fueled by various constituencies largely concerned with advocacy and ideology. At the beginning, middle, and end of the twentieth century, essentially the same groups—psychiatrists and therapists, social workers, feminists, religious and moral conservatives, law enforcement, and politicians—came together to raise awareness about CSA and campaign for reform. In each period, however, advocates eventually exaggerated claims, which the media sensationalized and uncritically disseminated, until the movement devolved into a "moral panic," with beliefs and responses out of all proportion to what realistic appraisal could sustain. The calmer periods between these panics were less "collective denial," as Ondersma et al. (2001, p. 708) phrased it, than a reaction to the extreme excesses of the panics. Regarding the current panic, from 1976 to present, Jenkins argues that feminist concerns some 30 years ago about injustice regarding rape shifted to incest by the mid-1970s, using the same rape theory and rhetoric of dominance and oppression to frame the issue. The Child Abuse Treatment and Prevention Act, passed in 1974, was intended to set up programs combating physical abuse and emotional neglect. Within 2 to 3 years, however, the burgeoning child abuse establishment (consisting of a loose network of social workers, therapists, and law-enforcement personnel concerned with this issue, who were in part funded by this act) refocused most of its attention on CSA, where it has stayed ever since. The incest model, based on the rape model, came to dominate thinking on all forms of sex involving minors where an age discrepancy was involved. This model, with the exemplar of the young, prepubescent daughter dependent on her guardian yet helpless to escape his unwanted sexual intrusion, with consequent psychological damage, has powerfully influenced psychiatric and psychological thought on all sociolegal CSA. Summit (1983), in his influential article on the

child sexual abuse accommodation syndrome, which was based largely on clinical incest cases, warned that his syndrome "should not be viewed as a procrustean bed which defines and dictates a narrow perception of something as complex as child sexual abuse" (p. 180). Despite this warning, the incest model–syndrome has come to dominate theory, which in turn has dictated the interpretation of sexual contacts that have little to do with the incest paradigm other than age discrepancy (Jenkins, 1998; Rind, 2001).

Ondersma et al. (2001) cited Masson (1984), who speculated on the reasons why Freud reversed himself on his views about CSA's contribution to psychopathology, but did not acknowledge well-reasoned critiques of Masson's arguments (e.g., Crews, 1998). They also attributed invested efforts to critics of their field, but acknowledged no possible bias of their own (Dineen, 1998; Jenkins, 1998; Sarnoff, 2001). For example, Dineen (1998) has argued that too often practitioners have favored marketing value over science by using techniques such as inappropriate pathologizing and generalizing to expand the numbers of "victims." Ondersma et al. (2001) suggested that a backlash continues to have an impact on their profession but did not acknowledge the possibility that victimological views may have caused harm in such instances as the SRA day-care cases (Nathan & Snedeker, 1995; Rind, Bauserman, & Tromovitch, 2001b). They cited a review of 24 introductory psychology textbooks by Letourneau and Lewis (1999) as evidence for the backlash but did not acknowledge that the premise for what these authors labeled *bias*—failure to give full credence to delayed memory recall—is clinical opinion and a subject of serious scientific debate rather than established fact (Brandon et al., 1998). They argued their profession is moderate in its views, citing Finkelhor (1979a), who emphasized that children may not be clearly harmed by CSA and may even see it positively, but did not acknowledge less nuanced claims in his more recent publications. For example, in their 1993 *Psychological Bulletin* review article, Kendall-Tackett, Williams, and Finkelhor observed that a significant number of children who experienced CSA have no measurable long-term outcomes (as Ondersma et al., 2001, noted). Rather than interpreting this as indicating that long-term harm may not be present, they argued instead that the impact is likely to be "muted or masked" (Kendall-Tackett et al., 1993, p. 168). Finally, Ondersma et al. (2001) stated that Fromuth (1986) "left untouched the basic societal value that sex with children is abuse" (p. 708) but failed to cite a later publication in which she opined that "there is much less clarity in defining the cases as abusive" (Fromuth & Burkhart, 1987, p. 250) when they involved boys' sexual contacts with women that the boys viewed predominantly positively.

In short, we believe that Ondersma et al.'s (2001) sociohistorical discussion provides only one perspective and fails to acknowledge credible alternatives. We believe that recognition and assessment of such alternatives is essential to skepticism and critical inquiry, which in turn is essential to science (Rind, Bauserman, & Tromovitch, 2000; Sagan, 1995; Sarnoff, 2001). Their characterization of our work as part of the backlash invites dismissal of our research not on the basis of its quality, but on its opposition to their perspective.

Conclusion and Commentary

Previously, we argued that the Dallam et al. (1999) critique was a "kitchen sink" attack, throwing every possible criticism our way

irrespective of relevance or importance but in the end failing to demonstrate any bias that would necessitate altering our basic conclusions (Rind, Bauserman, & Tromovitch, 2001b). We believe that we have shown that their current critique (Dallam et al., 2001), a more refined version, similarly fails to show any such bias, let alone justify or support the media and political attacks on our study as being severely flawed. We believe we successfully showed that our logic regarding use of the college samples was sound; that our interpretations of the data were appropriate; that the college samples in fact were comparable in key respects to national samples; that our treatment of definitions and our inclusion and exclusion decisions were justified, as was our noncorrection of effect sizes; that our moderator and statistical control analyses were valid; and that our coding was accurate. At the same time, we believe we successfully demonstrated that Dallam et al. (2001) repeatedly erred in charging us with such things as misreporting data, presenting them in a misleading way, erroneously coding studies, disregarding confidence intervals, giving the appearance of differences, including studies that did not purport to study what we stated they did, and interpreting the wrong statistic as an effect size. Further, we believe we have shown that they, in making their points, were frequently selective in terms of citing studies for one purpose when they supported their argument but then ignoring the same studies on other points when they did not, citing our own comments on one issue to argue a point but then ignoring the same comments on other issues for which the comments were highly relevant, and citing research that supported their arguments while ignoring equally relevant research that did not.

The conceptual criticisms are no improvement. Our decisions regarding terminology, using consent as a variable, and making cross-cultural comparisons were characterized by Ondersma et al. (2001) as extrascientific. We believe that the empirical justification for our handling of terminology and consent issues is clear, however, and that cross-cultural comparisons are scientific, not extrascientific. It is incorrect to assume that human sexual nature can be fully understood by relying on a single culture with a traditionally negative stance toward sexuality, especially juvenile sexuality (Ford & Beach, 1951). As historian Randolph Trumbach (1977) put it,

the psychological study of sexual behavior is (or should be seen to be) notoriously ethnocentric and moralistic. Psychological studies presuppose a universal human nature, but then limit their sample to Western societies. (p. 1)

We believe that calls for incorporating moral and ideological stances into research methods and interpretations, such as Ondersma et al.'s (2001) contention that the moral standard is integral to the scientific understanding of CSA, are themselves extrascientific. Expansive and all-encompassing victimological definitions of harm violate Popper's (1961) falsifiability principle, and arguing for an "emphasis on understanding CSA-related harm rather than on proving that harm" (p. 713) stems from an unexamined assumption of harm in all cases, regardless of context.

Both sets of critics pointed to supposed dangers coming from our study. Dallam et al. (2001) reported that our study has been used in court in a few instances, but testimony from a victimological perspective has occurred innumerable times, putting forth ideological beliefs as verified facts (Sarnoff, 2001). Continuing this theme, Dallam et al. (2001) ended by stating that attempts to

use our study to argue that an individual has not been harmed by CSA are a "serious misapplication of its findings" (p. 729). Certainly, a review such as ours examined only general patterns or trends and cannot be used to show or argue that any particular individual has or has not been harmed. Some individuals are undoubtedly harmed by CSA, a point we clearly acknowledged in our review. However, the claim that the typical person with CSA, as typically and broadly defined, suffers extreme, pervasive, and lasting harm has not been verified in a scientific sense. Ondersma et al. (2001) feared a backlash against therapists. Psychotherapy, like medical intervention, must proceed from sound science. Psychotherapy, unlike medical intervention, has traditionally not been held up to the same standard of scientific rigor in verification of treatment efficacy and concern with potential side effects. Any research, such as ours, that can improve therapists' knowledge of who needs to be treated and to what degree and for whom treatment is unnecessary or even contraindicated, should be welcome.

Finally, our review and the issues raised in our current rebuttal offer a number of guideposts for future research. First, in terms of causality, researchers should take into account as possible confounds not only family background but also peer variables and personality dispositions, as problem behavior theory has identified these latter factors as also highly relevant to precocious sex and its correlates. Second, contextual factors should be taken into account. Aside from the obvious factors of frequency, duration, relatedness, and presence or absence of force, level of willingness should be examined, with the understanding that the issue in this case is perceived willingness, assent, or simple consent, not informed consent. It is important for psychologists to uncover and examine participants' own perceptions of their experiences, not project their own point of view onto participants' responses. Third, if researchers choose to examine only unwanted CSA, as many have, then they should be clear in their interpretation of the results that their findings do not apply to all CSA but just unwanted CSA for a particular population, a qualification that researchers currently generally do not provide. The Coxell et al. (1999) study is a clear illustration of the benefits of categorizing participants by level of consent and the inferential biases that can occur if only unwanted CSA is assessed but then generalized to all sociolegal CSA. Fourth, informed sexual consent should be formally studied by psychologists, which it never has been to our knowledge. It is a construct with important legal and social implications. Across societies, those in politics and government vary greatly in proclaiming at what age it is possible, not infrequently alluding to psychological opinion. Yet, such opinion is almost always based recursively on legal and social criteria. To break this circularity, researchers should begin by defining this ability in scientific terms and examining it scientifically. Fifth, related to uncovering participants' actual perceptions, researchers should assess and report reactions from negative to positive, along with perceived costs or benefits. Many college studies did this, in sharp contrast to studies based on other populations (e.g., clinical, forensic), whose assessments have often structurally disallowed reports of positive reactions or benefits (Okami, 1990). Furthermore, negative reactions are typically labeled by CSA researchers as *traumatic* (e.g., Finkelhor, 1979a), when in fact they may range from slightly unpleasant to truly horrific. To avoid overstatement, researchers should draw distinctions among types of negative response. Sixth, to understand sociolegal CSA more comprehensively, the further researchers

move away from clinical samples, the more scientifically valid their findings will be both in terms of external and internal validity. Examination of participants in non-Western cultures in which victimological values have not taken hold, for example, has shown striking differences in perceived reactions and effects (e.g., Williams, 1996). The interaction of culture and sexual response has for too long been ignored in CSA research (cf. Ford & Beach, 1951). Finally, the victimological perspective has dominated almost all research in this area for the past quarter-century. Victimology has its place but contains a heavy degree of ideology. Researchers should not feel obligated to restrict design, analysis, and interpretation to a victimological perspective, but rather they should consider other models. All of these approaches can help move research on CSA and its correlates beyond the current paradigm in this field.

References

- American Lung Association (2000). *American Lung Association fact sheet on smoking*. Available at: <http://www.lungusa.org/tobacco/index.html>.
- American Psychological Association. (1989, October) Brief for amici curiae in *Ohio v. Akron Center for Reproductive Health et al.* [United States Court of Appeals for the Sixth Circuit, in Support of Petitioners/Cross-Respondents in Nos. 88-1125 and 88-1309 and in Support of Appellees in No. 88-805]. Washington, DC: Author.
- Baker, A. W., & Duncan, S. P. (1985). Child sexual abuse: A study of prevalence in Great Britain. *Child Abuse & Neglect*, 9, 457-467.
- Bauserman, R., & Rind, B. (1997). Psychological correlates of male child and adolescent sexual experiences with adults: A review of the nonclinical literature. *Archives of Sexual Behavior*, 26, 105-141.
- Becker, G. (1986). Correcting the point-biserial correlation for attenuation owing to unequal sample size. *Journal of Experimental Education*, 55, 5-8.
- Bensley, L. S., Spieker, S. J., Van Eenwyk, J., & Schoder, J. (1999). Self-reported abuse history and adolescent problem behaviors: II. Alcohol and drug use. *Journal of Adolescent Health*, 24, 173-180.
- Bensley, L. S., Van Eenwyk, J., Spieker, S. J., & Schoder, J. (1999). Self-reported abuse history and adolescent problem behaviors: I. Anti-social and suicidal behaviors. *Journal of Adolescent Health*, 24, 163-172.
- Best, J. (1997, May/June). Victimization and the victim industry. *Society*, 9-17.
- Bigler, M. O. (1992). *Childhood sexual trauma as a predictor of adult sexual self-esteem and sexual dysfunction*. Unpublished doctoral dissertation, New York University.
- Boney-McCoy, S., & Finkelhor, D. (1995). Psychosocial sequelae of violent victimization in a national youth sample. *Journal of Consulting and Clinical Psychology*, 63, 726-736.
- Boney-McCoy, S., & Finkelhor, D. (1996). Is youth victimization related to trauma symptoms and depression after controlling for prior symptoms and family relationships? A longitudinal, prospective study. *Journal of Consulting and Clinical Psychology*, 64, 1406-1416.
- Brandon, S., Boakes, J., Glaser, D., & Green, R. (1998). Recovered memories of childhood sexual abuse. *British Journal of Psychiatry*, 172, 296-307.
- Brant, R. S. T., & Tisza, V. B. (1977). The sexually misused child. *American Journal of Orthopsychiatry*, 47, 80-90.
- Briere, J. N. (1988). Controlling for family variables in abuse effects research: A critique of the "partialing" approach. *Journal of Interpersonal Violence*, 3, 80-89.
- Brubaker, D. (1991). *Relationship of dissociation and stress response among survivors of childhood sexual abuse*. Unpublished master's thesis, Texas A&M University.
- Brubaker, D. (1994). Stress response and dissociation among survivors of childhood sexual abuse (Doctoral dissertation, Texas A&M University, 1993). *Dissertation Abstracts International*, 54, 4382.
- Bullough, V., & Bullough, B. (1977). *Sin, sickness, and sanity: A history of sexual attitudes*. New York: Meridian.
- Cantarella, E. (1992). *Bisexuality in the ancient world*. New Haven, CT: Yale University Press.
- Chandy, J. M., Blum, R. W., & Resnick, M. D. (1996). Gender-specific outcomes for sexually abused adolescents. *Child Abuse & Neglect*, 20, 1219-1231.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Cohen, J. (1988). *Statistical power analyses for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Condy, S., Templer, D., Brown, R., & Veaco, L. (1987). Parameters of sexual contact of boys with women. *Archives of Sexual Behavior*, 16, 379-394.
- Coxell, A., King, M., Mezey, G., & Gordon, D. (1999). Lifetime prevalence, characteristics, and associated problems of non-consensual sex in men: Cross sectional survey. *British Medical Journal*, 318, 846-850.
- Crews, F. (1998). *Unauthorized Freud: Doubters confront a legend*. New York: Penguin Books.
- Dallam, S. J., Gleaves, D. H., Cepeda-Benito, A., Silberg, J. L., Kraemer, H. C., & Spiegel, D. (2001). The effects of child sexual abuse: Comment on Rind, Tromovitch, and Bauserman (1998). *Psychological Bulletin*, 127, 715-733.
- Dallam, S. J., Gleaves, D. H., Spiegel, D., & Kraemer, H. (1999). *An analysis of Rind et al.'s meta-analysis of the long-term effects of child sexual abuse*. Unpublished manuscript.
- Dineen, T. (1998). *Manufacturing victims* (2nd ed.). Montreal, Canada: Robert Davies Publishing.
- Dinwiddie, S., Heath, A. C., Dunne, M. P., Bucholz, K. K., Madden, P. A. F., Slutske, W. S., Bierut, L. J., Statham, D. B., & Martin, N. G. (2000). Early sexual abuse and lifetime psychopathology: A co-twin-control study. *Psychological Medicine*, 30, 41-52.
- Eckenrode, J., Laird, M., & Doris, J. (1993). School performance and disciplinary problems among abused and neglected children. *Developmental Psychology*, 29, 53-62.
- Edgards, K., & Ormstad, K. (2000). Prevalence and characteristics of sexual abuse in a national sample of Swedish seventeen-year-old boys and girls. *Acta Paediatrica*, 89, 310-319.
- Erickson, P. I., & Rapkin, A. J. (1991). Unwanted sexual experiences among middle and high school youth. *Journal of Adolescent Health*, 12, 319-325.
- Fergusson, D. M., Horwood, L. J., & Lynskey, M. T. (1996). Childhood sexual abuse and psychiatric disorder in young adulthood: II. Psychiatric outcomes of childhood sexual abuse. *Journal of the American Academy of Child and Adolescent Psychiatry*, 34, 1365-1374.
- Finkelhor, D. (1979a). *Sexually victimized children*. New York: Free Press.
- Finkelhor, D. (1979b). What's wrong with sex between adults and children? Ethics and the problem of sexual abuse. *American Journal of Orthopsychiatry*, 49, 692-697.
- Finkelhor, D. (1984). *Child sexual abuse: New theory and research*. New York: Free Press.
- Finkelhor, D., Hotaling, G., Lewis, I. A., & Smith, C. (1990). Sexual abuse in a national survey of adult men and women: Prevalence, characteristics, and risk factors. *Child Abuse & Neglect*, 14, 19-28.
- Fiscella, K., Kitzman, H. J., Cole, R. E., Sidora, K. J., & Olds, D. (1998). Does child abuse predict adolescent pregnancy? *Pediatrics*, 101, 620-624.
- Fishman, J. (1991). Prevalence, impact, and meaning attribution of childhood sexual experiences of undergraduate males (Doctoral dissertation, University of Massachusetts at Amherst, 1990). *Dissertation Abstracts International*, 52, 114.

- Fleming, J., Mullen, P. E., Sibthorpe, B., & Bammer, G. (1999). The long-term impact of childhood sexual abuse in Australian women. *Child Abuse & Neglect*, 23, 145-159.
- Ford, C. S., & Beach, F. A. (1951). *Patterns of sexual behavior*. New York: Harper & Row.
- Fromuth, M. (1986). The relationship of childhood sexual abuse with later psychological and sexual adjustment in a sample of college women. *Child Abuse & Neglect*, 10, 5-15.
- Fromuth, M., & Burkhart, B. (1987). Sexual victimization among college men: Definitional and methodological issues. *Violence and Victims*, 2, 241-253.
- Fromuth, M., & Burkhart, B. (1989). Long-term psychological correlates of childhood sexual abuse in two samples of college men. *Child Abuse & Neglect*, 13, 533-542.
- Garnefski, N., & Arends, E. (1998). Sexual abuse and adolescent maladjustment: Differences between male and female victims. *Journal of Adolescence*, 21, 99-107.
- Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in education and psychology* (3rd ed.). Boston: Allyn & Bacon.
- Goldman, R., & Goldman, J. (1988). The prevalence and nature of child sexual abuse in Australia. *Australian Journal of Sex, Marriage & Family*, 9, 94-106.
- Graupner, H. (1997). *Sexualität, Jugendschutz & Menschenrechte: Über das Recht von Kindern und Jugendlichen auf sexuelle Selbstbestimmung* [Sexuality, protection of youth, & human rights: On the rights of children and youths in sexual self-determination]. Frankfurt, Germany: Peter Lang, Europäischer Verlag der Wissenschaften.
- Graupner, H. (2000). Sexual consent: The criminal law in Europe and overseas. *Archives of Sexual Behavior*, 29, 415-461.
- Green, R. (1992). *Sexual science and the law*. Cambridge, MA: Harvard University Press.
- Greenberg, D. (1988). *The construction of homosexuality*. Chicago: University of Chicago Press.
- Greenwald, J. (1994). Romantic and sexual self-esteem, a new construct: Its relationship to current and early childhood sexual experiences (Doctoral dissertation, University of Nebraska—Lincoln, 1993). *Dissertation Abstracts International*, 54, 3852.
- Gregeren, E. (1983). *Sexual practices: The story of human sexuality*. New York: Franklin Watts.
- Harrison, P. A., Fulkerson, J. A., & Beebe, T. J. (1997). Multiple substance use among adolescent physical and sexual abuse victims. *Child Abuse & Neglect*, 21, 529-539.
- Hernandez, J., Lodico, M., & DiClemente, R. (1993). The effects of child abuse and race on risk-taking in male adolescents. *Journal of the National Medical Association*, 85, 593-597.
- Hibbard, R. A., Brack, C. J., Rauch, S., & Orr, D. P. (1988). Abuse, feelings, and health behaviors in a student population. *American Journal of Diseases of Childhood*, 142, 326-330.
- Hibbard, R. A., Ingersoll, G. M., & Orr, D. P. (1990). Behavioral risk, emotional risk, and child abuse among adolescents in a nonclinical setting. *Pediatrics*, 86, 896-901.
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis*. Newbury Park, CA: Sage.
- Jackson, J. L., Calhoun, K. S., Amick, A. E., Maddever, H. M., & Habif, V. L. (1990). Young adult women who report childhood intrafamilial sexual abuse: Subsequent adjustment. *Archives of Sexual Behavior*, 19, 211-221.
- Jenkins, P. (1998). *Moral panic: Changing concepts of the child-molester in modern America*. New Haven, CT: Yale University Press.
- Jessor, R., Costa, F., Jessor, L., & Donovan, J. (1983). Time of first intercourse: A prospective study. *Journal of Personality and Social Psychology*, 44, 608-626.
- Johnson, J. G., Cohen, P., Brown, J., Smailes, E. M., & Bernstein, D. P. (1999). Childhood maltreatment increases risk for personality disorders during early adulthood. *Archives of General Psychiatry*, 56, 600-606.
- Jumper, S. (1995). A meta-analysis of the relationship of child sexual abuse to adult psychological adjustment. *Child Abuse & Neglect*, 19, 715-728.
- Kelley, S. (1989). Stress responses of children to sexual abuse and ritualistic abuse in day care centers. *Journal of Interpersonal Violence*, 4, 502-513.
- Kelley, S. (1990). Parental stress response to sexual abuse and ritualistic abuse of children in day care centers. *Nursing Research*, 39, 25-29.
- Kelly, R. J. (1993). Overall level of distress. In J. Waterman, R. J. Kelly, M. K. Oliveri, & J. McCord (Eds.), *Behind the playground walls: Sexual abuse in preschools* (pp. 79-92). New York: Guilford Press.
- Kendall-Tackett, K. A., Williams, L. M., & Finkelhor, D. (1993). Impact of sexual abuse on children: A review and synthesis of recent empirical studies. *Psychological Bulletin*, 113, 164-180.
- Kendler, K., Bulik, C., Silberg, J., Hettema, J., Myers, J., & Prescott, C. (2000). Childhood sexual abuse and adult psychiatric and substance use disorders in women. *Archives of General Psychiatry*, 57, 953-959.
- Ketterlinus, R. D., Lamb, M. E., Nitz, K., & Elster, A. B. (1992). Adolescent nonsexual and sex-related problem behaviors. *Journal of Adolescent Research*, 7, 431-456.
- Kilpatrick, A. C. (1987). Childhood sexual experiences: Problems and issues in studying long-range effects. *Journal of Sex Research*, 23, 173-196.
- Kinsey, A., Pomeroy, W. B., & Martin, C. E. (1948). *Sexual behavior in the human male*. Philadelphia: W. B. Saunders.
- Landis, J. (1956). Experiences of 500 children with adult sexual deviation. *Psychiatric Quarterly Supplement*, 30, 91-109.
- Laumann, E. O., Gagnon, J. H., Michael, R. T., & Michaels, S. (1994). *The social organization of sexuality: Sexual practices in the United States*. Chicago: University of Chicago Press.
- Letourneau, E. J., & Lewis, T. C. (1999). The portrayal of child sexual assault in introductory textbooks. *Teaching of Psychology*, 26, 253-258.
- Lilienfeld, S. O. (in press). When worlds collide: Social science, politics, and the Rind et al. (1998) child sexual abuse meta-analysis. *American Psychologist*.
- Lodico, M., Gruber, E., & DiClemente, R. (1996). Childhood sexual abuse and coercive sex among school-based adolescents in a midwestern state. *Journal of Adolescent Health*, 18, 211-217.
- Long, P., & Jackson, J. (1993). Initial emotional response to childhood sexual abuse: Emotion profiles of victims and relationship to later adjustment. *Journal of Family Violence*, 8, 167-181.
- López, F., Carpintero, E., Hernández, M., & Fuertes, A. (1995). Prevalencia y consecuencias del abuso sexual al menor en España [Prevalence and consequences of child sexual abuse in Spain.] *Child Abuse & Neglect*, 19, 1039-1050.
- Masson, J. M. (1984). *The assault on truth: Freud's suppression of seduction theory*. New York: Penguin Books.
- McCarty, R. (1999, November/December). A brief comment by APA executive director for science. *Psychological Science Agenda*, 12(6), 2-3.
- Money, J., & Weinrich, J. D. (1983). Juvenile, pedophile, heterophile: Hermeneutics of science, medicine and law in two outcome studies. *International Journal of Medicine and Law*, 2, 39-54.
- Mullen, P. E., Martin, J. L., Anderson, J. C., Romans, S. E., & Herbison, G. P. (1993). Childhood sexual abuse and mental health in adult life. *British Journal of Psychiatry*, 163, 721-732.
- Nagy, S., DiClemente, R., & Adcock, A. G. (1995). Adverse factors associated with forced sex among southern adolescent girls. *Pediatrics*, 96, 944-946.
- Nash, C., & West, D. (1985). Sexual molestation of young girls: A retrospective survey. In D. J. West (Ed.), *Sexual victimization* (pp. 1-92). Brookfield, VT: Gower.

- Nathan, D. (1990, June 12). What McMartin started: The ritual sex abuse hoax. *Village Voice*, 35, 36–44.
- Nathan, D. & Snedeker, M. (1995). *Satan's silence: Ritual abuse and the making of a modern American witchhunt*. New York: Basic Books.
- Nelson, J. A. (1989). Intergenerational sexual contact: A continuum model of participants and experiences. *Journal of Sex Education & Therapy*, 15, 3–12.
- Nelson, A. & Oliver, P. (1998). Gender and the construction of consent in child-adult sexual contact: Beyond gender neutrality and male monopoly. *Gender & Society*, 12, 554–577.
- Neumann, D. A., Houskamp, B. M., Pollock, V. E., & Briere, J. (1996). The long-term sequelae of childhood sexual abuse in women: A meta-analytic review. *Child Maltreatment*, 1, 6–16.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Nunnally, J. C., & Bernstein, I. A. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Okami, P. (1990). Sociopolitical biases in the contemporary scientific literature on adult human sexual behavior with children and adolescents. In J. R. Feierman (Ed.), *Pedophilia: Biosocial dimensions* (pp. 91–121). New York: Springer-Verlag.
- Olafson, E., Corwin, D., & Summit, R. (1993). Modern history of child sexual abuse awareness: Cycles of discovery and suppression. *Child Abuse & Neglect*, 17, 7–24.
- Ondersma, S. J., Chaffin, M., & Berliner, L. (1999). Comments on Rind et al. meta-analysis controversy. *APSAC Advisor*, 12, 2–5.
- Ondersma, S. J., Chaffin, M., Berliner, L., Cordon, I., Goodman, G., & Barnett, D. (2001). Sex with children is abuse: Comment on Rind, Tromovitch, and Bauserman (1998). *Psychological Bulletin*, 127, 707–714.
- Pedhazur, E. (1997). *Multiple regression in behavioral research: Explanation and prediction*. New York: Harcourt Brace.
- Percy, W. (1996). *Pederasty and pedagogy in archaic Greece*. Chicago: University of Illinois Press.
- Pope, H. G. & Hudson, J. I. (1995, Fall). Does childhood sexual abuse cause adult psychiatric disorders? Essentials of methodology. *Journal of Psychiatry and Law*, 363–381.
- Popper, K. (1961). *The logic of scientific inquiry*. New York: Basic Books.
- Resnick, M., & Blum, R. (1994). The association of consensual sexual intercourse during childhood with adolescent health risk and behaviors. *Pediatrics*, 94, 907–913.
- Rind, B. (2001). Gay and bisexual adolescent boys' sexual experiences with men: An empirical examination of psychological correlates in a nonclinical sample. *Archives of Sexual Behavior*, 30, 345–368.
- Rind, B., Bauserman, R., & Tromovitch, P. (2000). Science versus orthodoxy: Anatomy of the congressional condemnation of a scientific article and reflections on remedies for future ideological attacks. *Applied and Preventive Psychology*, 9, 211–225.
- Rind, B., Bauserman, R., & Tromovitch, P. (2001a, July/August). The condemned meta-analysis on child sexual abuse: Good science and long-overdue skepticism. *Skeptical Inquirer*, 25, 68–72.
- Rind, B., Bauserman, R., & Tromovitch, P. (2001b). Moralistic psychiatry, Procrustes' bed, and the science of child sexual abuse: A reply to Spiegel. *Sexuality & Culture*, 5(1), 75–85.
- Rind, B., & Tromovitch, P. (1997). A meta-analytic review of findings from national samples on psychological correlates of child sexual abuse. *Journal of Sex Research*, 34, 237–255.
- Rind, B., Tromovitch, P., & Bauserman, R. (1998). A meta-analytic examination of assumed properties of child sexual abuse using college samples. *Psychological Bulletin*, 124, 22–53.
- Rind, B., Tromovitch, P., & Bauserman, R. (2000a). Condemnation of a scientific article: A chronology and refutation of the attacks and a discussion of threats to the integrity of science. *Sexuality & Culture*, 4(2), 1–62.
- Rind, B., Tromovitch, P., & Bauserman, R. (2000b). Debunking the false allegation of "statistical abuse": A reply to Spiegel. *Sexuality & Culture*, 4(2), 101–111.
- Roland, B., Zelhart, P., & Dubes, R. (1989). MMPI correlates of college women who reported experiencing child/adult sexual contact with father, stepfather, or with other persons. *Psychological Reports*, 64, 1159–1162.
- Rosenthal, D., Smith, A., & de Visser, R. (1999). Personal and social factors influencing age at first intercourse. *Archives of Sexual Behavior*, 28, 319–333.
- Rosenthal, R. (1984). *Meta-analytic procedures for social research*. London: Sage.
- Rosenthal, R. (1994). Science and ethics in conducting, analyzing, and reporting psychological research. *Psychological Science*, 5, 127–134.
- Rosenthal, R., & Rosnow, R. (1969). *Artifact in behavioral research*. New York: Academic Press.
- Rosenthal, R., & Rubin, D. B. (1982). A simple general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, 74, 166–169.
- Rosnow, R., & Rosenthal, R. (1997). *People studying people: Artifacts and ethics in behavioral research*. New York: Freeman.
- Sagan, C. (1995). *The demon-haunted world: Science as a candle in the dark*. New York: Random House.
- Sandfort, T. G. (1992). The argument for adult-child sexual contact: A critical appraisal and new data. In W. O'Donohue & J. H. Geer (Eds.), *The sexual abuse of children: Theory and research* (Vol. 1, pp. 38–48). Hillsdale, NJ: Erlbaum.
- Sarbo, A. (1985). Childhood molestations: Adult personality correlates and predictors of trauma (Doctoral dissertation, Georgia State University, 1984). *Dissertation Abstracts International*, 45, 3960.
- Sarnoff, S. (2001). *Sanctified snake oil: How government-supported junk science endangers our economy, safety and freedom*. Westport, CT: Praeger.
- Saunders, B. E., Kilpatrick, D. G., Hanson, R. F., Resnick, H. S., & Walker, M. E. (1999). Prevalence, case characteristics, and long-term psychological correlates of child rape among women: A national survey. *Child Maltreatment*, 4, 187–200.
- Schalow, P. G. (1989). Male love in early modern Japan: A literary depiction of the "youth." In M. B. Duberman, M. Vicinus, & G. Chauncey, Jr. (Eds.), *Hidden from history* (pp. 118–128). New York: New American Press.
- Schlessinger, L. (1999, March 23). The Dr. Laura show [Radio broadcast]. Los Angeles, CA: Premiere Radio Networks.
- Schultz, L. G. (1980). *The sexual victimology of youth*. Springfield, IL: Charles C Thomas.
- Schultz, L. G., & Jones, P. (1983). Sexual abuse of children: Issues for social science and health professionals. *Child Welfare*, 62, 99–109.
- Sedney, M., & Brooks, B. (1984). Factors associated with a history of childhood sexual experience in a nonclinical female population. *Journal of the American Academy of Child Psychiatry*, 23, 215–218.
- Seligman, M. E. P. (1994). *What you can change and what you can't*. New York: Knopf.
- Sheskin, D. J. (1997). *Handbook of parametric and nonparametric statistical procedures*. New York: CRC Press.
- Sidran Foundation. (1994). *Traumatic memories* [Brochure]. Lutherville, MD: Author.
- Silliman, M. (1993). Self-esteem and locus of control of women who report sexual abuse during childhood. *Psychological Reports*, 72, 1294.
- Sommers, C. H. (1995). *Who stole feminism? How women have betrayed women*. New York: Touchstone.
- Sommers, C. H. (2001). *The war against boys: How misguided feminism is harming our young men*. New York: Simon & Schuster.
- Spiegel, D. (2000a). The price of abusing children and numbers. *Sexuality & Culture*, 4(2), 63–66.

- Spiegel, D. (2000b). Real effects of real child sexual abuse. *Sexuality & Culture*, 4(4), 99–105.
- Spiegel, D. (in press). Who is in Procrustes' bed? *Sexuality & Culture*.
- Stein, J. A., Golding, J. N., Siegel, J. M., Burnam, M. A., & Sorenson, S. B. (1988). Long-term psychological sequelae of child sexual abuse. In G. E. Wyatt & G. J. Powell (Eds.), *Lasting effects of child sexual abuse* (pp. 135–154). Newbury Park, CA: Sage.
- Stock, J. L., Bell, M. A., Boyer, D. K., & Connell, F. A. (1997). Adolescent pregnancy and sexual risk-taking among sexually abused girls. *Family Planning Perspective*, 29, 200–203, 227.
- Summit, R. (1983). The child sexual abuse accommodation syndrome. *Child Abuse & Neglect*, 7, 177–193.
- Szasz, T. (1990). *Sex by prescription: The startling truth about today's sex therapy*. New York: Syracuse University Press.
- Trumbach, R. (1977). London's sodomites: Homosexual behavior and western culture in the 18th century. *Journal of Social History*, 11, 1–33.
- Vasey, P. L. (1995). Homosexual behavior in primates: A review of evidence and theory. *International Journal of Primatology*, 16, 173–203.
- Webster's Third New International Dictionary (1981). Springfield, MA: Merriam-Webster.
- West, D. J. (1998). Boys and sexual abuse: An English opinion. *Archives of Sexual Behavior*, 27, 539–559.
- West, D. J., & Woodhouse, T. (1993). Sexual encounters between boys and adults. In C. Li, D. West, & T. Woodhouse (Eds.), *Children's sexual encounters with adults* (pp. 3–137). New York: Prometheus.
- Williams, W. L. (1996). Two-spirit persons: Gender nonconformity among Native American and Native Hawaiian youths. In R. C. Savin-Williams & K. M. Cohen (Eds.), *The lives of lesbians, gays, and bisexuals* (pp. 416–435). New York: Harcourt Brace.
- Wisniewski, N. (1990). A path analytic model of the aftereffects of childhood sexual victimization. (Doctoral dissertation, Kent State University, 1989). *Dissertation Abstracts International*, 51, 2079–2080.

Received May 24, 2001

Revision received July 3, 2001

Accepted July 3, 2001

Members of Underrepresented Groups: Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write to Demarie Jackson at the address below. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.
- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.
- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In your letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, "social psychology" is not sufficient—you would need to specify "social cognition" or "attitude change" as well.
- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

Write to Demarie Jackson, Journals Office, American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242.