# Sexual Abuse: A Journal of Research and Treatment

**Discriminative and Predictive Validity of the Penile Plethysmograph in Adolescent Sex Offenders**

Robert J. W. Clift, Gordana Rajlic and Heather M. Gretton

# Discriminative and Predictive Validity of the Penile Plethysmograph in Adolescent Sex Offenders

Robert J. W. Clift
Gordana Rajlic
Heather M. Gretton
*Youth Forensic Psychiatric Services, Burnaby, British Columbia, Canada*

The current study addresses the discriminative and predictive validity of the penile plethysmograph (PPG) in a sample of 132 male adolescent sex offenders who were admitted to a sex offender treatment program. The subjects were assessed pretreatment and posttreatment, under two different instructions sets: an Arouse condition, where subjects allowed themselves to become aroused; and a Suppress condition, where subjects attempted to suppress or control their arousal. Under the pretreatment Arouse condition, the group with male child victims demonstrated significantly greater arousal to child stimuli than the group with female nonchild victims. However, the group with female child victims could not be differentiated from the group with female nonchild victims. Under the Suppress condition, there was no evidence of discriminative validity for any of the PPG indices. As for predictive validity, posttreatment arousal to male and female children was significantly related to sexual offense recidivism. Posttreatment inability to suppress deviant arousal to male and female children was also significantly related to sexual offense recidivism over the 6-year follow-up period. In light of these findings, the implications for adolescent sex offender research and treatment are discussed.

***Keywords:***   adolescents; discriminative validity; penile plethysmograph; phallometric
  assessment; predictive validity; sex offenders

In their meta-analysis of 95 sexual offender recidivism studies, Hanson and Morton-Bourgon (2004) identified a number of variables that predict an increased risk for sexual recidivism. The risk factors with the strongest predictive accuracy were those relating to sexual deviancy ($d = .30$) and antisocial orientation ($d = .23$). One method of determining sexual deviancy is through phallometric assessment with the penile plethysmograph (PPG). The PPG is used to measure changes in penile tumescence while males are presented with sexual stimuli that vary in content

(Seto, Lalumiere, & Blanchard, 2000). Relative increases in penile blood volume are interpreted as an index of the male's interest in that type of sexual scenario.

The PPG has been used extensively with adult sex offenders. Use of the PPG has been endorsed by the Association for the Treatment of Sexual Abusers (ATSA, 2005) and is supported by the empirical relationship between deviant sexual arousal and sexual recidivism (Hanson & Morton-Bourgon, 2004). However, relatively little research has been conducted using the PPG with adolescent sex offenders. We review the available research on the psychometric properties of the PPG here.

## Test–Retest Reliability

Test–retest reliability refers to the relationship between scores on a given test over more than one testing period. Becker, Hunter, Goodwin, Kaplan, and Martinez (1992) assessed the test–retest reliability of PPG stimuli in a small sample of 20 adolescent sex offenders. Their stimulus set was specifically designed for use with adolescents. It consisted of 19 two-minute audiotaped descriptions of various deviant sexual activities. Participants listened to the complete set of stimuli twice, on successive days. Several stimuli produced strong test–retest correlations. These included the stimuli depicting consensual sex with a male aged 9 to 12 years ($r = .83$), coercive sex with a female aged 13 to 18 years ($r = .82$), coercive sex with a female aged less than 8 years ($r = .80$), and voyeurism ($r = .80$). However, despite the short time period between tests, the test–retest correlations for four of the stimuli failed to reach significance. These included exhibitionism ($r = .21$), rape of an adult female ($r = .29$), incest with a female child ($r = .39$), and frottage ($r = .44$).

## Concurrent Validity

Concurrent validity refers to the relationship between scores on two measures, taken at approximately the same time (Groth-Marnat, 2003), which are thought to be measuring the same underlying construct (e.g., deviant arousal). Hunter, Becker, and Kaplan (1995) examined the concurrent validity of phallometrically measured deviant arousal and self-reported deviant arousal in a sample of 38 adolescents. Both measures were administered within the first 3 weeks of admission to a treatment program, although not necessarily on the same day. Hunter et al. (1995) used the same stimulus set as Becker et al. (1992) during the PPG assessments. Self-reported arousal was obtained using the Adolescent Sexual Interest Card Sort (ASIC). The ASIC consists of 64 sexual vignettes, which adolescents rate on a 5-point scale. Higher scores reflect greater arousal at the thought of engaging in that behavior.

Hunter et al. (1995) examined the correlation between the adolescents' mean responses on the ASIC and their mean maximum PPG responses for each of 14 categories of stimuli. Ten of the 14 comparisons failed to reach significance. The remaining four categories produced moderate correlations. These included aggressive

sex with a young female ($r = .44$), aggressive sex with a same-aged female ($r = .46$), aggressive sex with an adult female ($r = .35$), and consensual sex with a same-aged male ($r = .52$). Therefore, concurrent validity between the ASIC and the PPG was moderate for vignettes related to aggressive sex with females in different age groups and consensual sex with same-aged males. However, there was no evidence of concurrent validity between the ASIC and the PPG in assessing other types of sexual interests. The lack of concurrent validity between the PPG and the ASIC can be largely explained by the subjects' failure to self-report deviant arousal when completing the ASIC. As a group, the adolescents only reported positive mean arousal to one category of deviant stimuli (voyeurism). The authors concluded that self-report measures of deviant arousal, such as the ASIC, may not offer valid findings of deviant sexual interest (Hunter et al., 1995). That is, the validity of self-report measures may be compromised by youths' tendency to underreport deviant sexual arousal.

## Discriminative Validity

Seto et al. (2000) addressed the discriminative validity of the PPG with adolescent sex offenders. Discriminative validity refers to the ability of a measure to discriminate between groups based on an existing status or diagnosis (e.g., child molesters vs. nonoffenders). Seto et al. compared phallometric responses of 40 adolescent child molesters (aged 14-17), 75 young adult child molesters (aged 18-21), 23 young adult rapists (aged 18-21), and 16 nonoffenders (aged 18-21). The offenders with child victims were further divided by the sex of their victims. The subjects were presented with age-appropriate and deviant visual stimuli depicting nude male or female individuals walking toward the camera. The film clips were accompanied by audio vignettes that described individuals from the same age category engaging in nonsexual activities such as swimming. A pedophilic index was created by subtracting responses to adult stimuli from responses to very young children and prepubescent children.

Rapists and nonoffenders had negative pedophilic indices indicating that they preferred adult stimuli. As a group, offenders against children had significantly higher pedophilic indices than rapists or nonoffenders. Additionally, adolescent offenders against children and young adult offenders against children had similar pedophilic indices. However, further analysis revealed that adolescents with only female child victims did not differ significantly from rapists or nonoffenders in their pedophilic index, suggesting that youths with only female child victims were more similar to rapists and nonoffenders in their arousal patterns. Adolescents with any male victims and young adult child molesters had significantly higher mean pedophilic indices than comparison groups. Therefore, this study provides evidence of discriminative validity for the PPG with adolescent child molesters with male victims but not adolescent child molesters with female victims.

## Predictive Validity

Clift, Gretton, and Rajlic (2007) assessed the predictive validity of both pretreatment and posttreatment PPG responses in an adolescent sample ($n = 116$). That is, they assessed the PPG's ability to predict sexual recidivism over a 4-year follow-up period. The stimulus set consisted of pictures of clothed and unclothed children and adolescents, accompanied by audio descriptions of sexual activity. Clift et al. used the adolescents' PPG responses to form an array of deviance indices. This was done by subtracting the subject's $z$ score for appropriate stimuli from their mean $z$ score for different categories of sexually deviant stimuli.

The authors found no association between pretreatment deviance indices and sexual recidivism. This finding replicated findings from an earlier study conducted by Gretton, McBride, Hare, O'Shaughnessy, and Kumka (2001). However, Clift et al. (2007) did find a significant correlation between four of the posttreatment deviance indices and sexual recidivism. These four deviance indices represented arousal to stimuli that depicted forced and nonforced sexual interactions with male and female children. The findings from this study suggest that, in the adolescent population, posttreatment PPG indices have greater predictive validity than pretreatment PPG indices.

## Summary

To date, very little data have been published on the psychometric properties of the PPG with adolescent sexual offenders. Test–retest reliability for adolescent PPG stimuli has ranged from poor to strong (Becker et al., 1992), and concurrent validity of PPG scores with self-reported deviant arousal on the ASIC was low for the majority of categories that have been examined (Hunter et al., 1995). Seto et al. (2000) demonstrated good discriminative validity for PPG indices of deviant sexual interest for adolescent child molesters with male victims; however, PPG indices did not distinguish adolescent child molesters with female victims from rapists or nonoffenders. Finally, Clift et al. (2007) found mixed evidence for the predictive validity of the PPG over a 4-year follow-up period. There was no relationship between pretreatment deviance indices and sexual recidivism. However, there was a significant association between posttreatment deviant arousal to child stimuli and sexual recidivism. It is worth noting here that the bulk of the available research on the psychometric properties of the PPG—when used with adolescent sex offenders—has been conducted on small samples, which limited the statistical power of these studies.

Considering that a substantial number of sexual offenses in the United States and Canada are committed by adolescent sex offenders (Juristat, 2003; U.S. Department of Justice, 2007), adolescent sexual offending is a serious concern for society. It is well established that sexual deviance is a risk factor for sexual recidivism (Hanson

& Morton-Bourgon, 2004). Yet in spite of our awareness that deviant sexual interest is a risk factor for sexual recidivism, the role of the PPG in the measurement of deviant sexual interests in adolescent males remains a neglected area of research. As the PPG is being used clinically with this population (e.g., McGrath, Cumming, & Burchard, 2003), there is a need for further research to inform clinicians about its valid application.

## Current Study

In the current study, we present three main findings. First, we examine the strength of the relationships between the age(s) and sex(es) of adolescent sex offenders' known victims and their PPG scores to stimuli depicting targets of a similar age and sex.

Second, we add to the literature on the psychometric properties of the PPG, when used with adolescents, by assessing the discriminative validity of the PPG using two sets of instructions: (1) an Arouse condition, where youths are instructed to allow themselves to become fully aroused by the stimuli; and (2) a Suppress condition, where youths are instructed to attempt to suppress or control their arousal to the stimuli. Both these sets of instructions are used in clinical practice (Howes, 1995).

Although the Arouse condition was the focus of Seto et al.'s (2000) study of discriminative validity, no study has examined discriminative validity under the Suppress condition. The Suppress condition has significance for clinical practice. The clinical assumption of the Suppress condition is that youths' ability to control their arousal during a PPG examination is indicative of their ability to control their arousal in real-life situations (e.g., Marshall, 2007). In other words, clinicians rationalize that, among sex offenders, the inability to suppress deviant arousal to a particular type of stimuli (e.g., young children) during a PPG examination may be linked to analogous patterns of victimization in their real-life histories. In the present study, we test this assumption.

Third, we examine the predictive validity of the PPG under both the Arouse and Suppress conditions. The predictive validity of the PPG was previously examined using the Arouse condition (Clift et al., 2007; Gretton et al., 2001). However, in clinical practice there is interest in the relationship between deviant arousal that youths are unable to inhibit during the Suppress condition (when the youths are actively trying to suppress deviant arousal) and risk for sexual recidivism. To date, there has been no study of the relationship between deviant arousal under the Suppress condition and risk for sexual recidivism in adolescents.

# Method

## Subjects

The subjects were 132 youths who were admitted to the Youth Sexual Offence Treatment Program (YSOTP) provided by Youth Forensic Psychiatric Services (YFPS) in Burnaby, British Columbia, or Prince George, British Columbia, and completed at least one valid PPG assessment. The YSOTP at YFPS (described below) is an outpatient program designed for adolescents who have been charged with, or convicted of a sexual offence, or have a history of sexual offending.

Adolescents admitted to the YSOTP at YFPS typically undergo both pretreatment and posttreatment PPG assessments as part of the treatment process. At the time of each PPG assessment, youths are informed that their PPG data will be used for research purposes. All subjects in the current study gave written informed consent to allow their data to be used. In addition to the youths' consent, we received permission to use these data from the YFPS Executive Committee. All aspects of the current study were conducted in compliance with the regulations of the YFPS internal research review board.

Pretreatment PPG data are available for 120 of the 132 subjects (90.9%), and posttreatment PPG data are available for 93 of the 132 subjects (70.5%). The typical reason for having only pretreatment data is that the subject did not complete treatment and, therefore, did not undergo posttreatment PPG assessment. The typical reason for having only posttreatment PPG data is that the subject underwent pretreatment PPG assessment before 1996. The PPG assessment procedure used at YFPS was modified in 1996, and therefore, assessments performed before 1996 are not directly comparable with the assessments described in the current study.

Some PPG assessments were not included in this analysis because they were deemed to be invalid by the PPG technologist. A PPG assessment was considered to be invalid if the subject showed no response to any of the stimuli. Seven pretreatment PPG assessments and nine posttreatment PPG assessments were excluded based on this criterion. One additional PPG assessment was excluded from the pretreatment analysis, as the subject displayed less than 1.5 mm of change to any of the test stimuli. Finally, some subjects had valid Arouse condition data, but their Suppress condition data were either invalid or incomplete (pretreatment, $n = 6$; posttreatment, $n = 8$).

The majority of subjects had been formally charged with or convicted of a sexual offence (88.6%), although some had a history of sexual offending that did not result in a formal charge (11.4%). Approximately one fifth (22.0%) of the subjects had been formally charged with multiple sexual offences. Ethnic information was reported in YFPS files for 114 subjects (86.4%). The majority of those subjects were Caucasian (65.8%) or First Nations (23.7%). The remaining 10.5% were mostly of Asian or South Asian heritage. The subjects' mean age was 16.5 years ($SD = 1.6$) at the beginning of treatment, 17.7 years ($SD = 1.6$) at discharge from treatment, and 23.6 years ($SD = 2.2$) at the end of the follow-up period.

## Youth Sexual Offence Treatment Program

As mentioned above, all subjects were admitted to the YSOTP at YFPS. The program followed a primarily cognitive–behavioral approach to treatment, consisting of both group and individual components (Ozeer, 2002). The group component consisted of 2-hour sessions, twice per week, and lasted approximately 6 to 8 months. The length of the individual component depended on the needs of the individual. The typical length of the complete YSOTP was about 1 to 2 years. Subjects attended treatment for a mean of 14.9 months ($SD$ = 10.5; range = 1-56 months).

Some of the modules offered during treatment included: (1) previous victimization, (2) human emotional needs, (3) problem cycles, (4) risk factors, (5) legal definitions of sexual offenses and the requirements for consent, (6) self-esteem, (7) cognitive distortions, (8) social skills enhancement, (9) communication skills, (10) empathy skills, (11) values and making moral choices, (12) cognitive–behavioral intervention strategies (e.g., covert sensitization, impulse control, problem solving, thought stopping, thought catching, emotional management, and positive self-talk), (13) intimate relations and sexuality, (14) relaxation, and (15) relapse prevention plans. The program was administered by a team of treatment providers that included psychologists, psychiatrists, social workers, and psychiatric nurses.

## Penile Plethysmograph Procedure

All PPG assessments were conducted as part of the youths' clinical assessment at an outpatient clinic. Assessments were conducted in a sound-attenuated room, in a laboratory setting, by a trained medical technologist. Subjects were tested using a RS3010 Medical Monitoring System with Stretchistor mercury strain gauges (i.e., a circumferential device was used in the current study). The recording equipment was contained in the technologist's room, separated from the subject's area by a wall containing a one-way mirror. The subjects communicated with the technologist over an intercom system.

The PPG procedure was discussed with subjects prior to each assessment. Once the technologist was satisfied that a subject understood the assessment procedure, the subject was asked to sign a consent form. He was then asked to sit in a chair with a sheet covering his lap. He was instructed to take a circumference measurement of his flaccid penis using a strip of paper, and the equipment was calibrated accordingly. The subject was then asked to place the mercury strain gauge on the midsection of his penis.

The assessment sessions were run in two parts: an Arouse condition followed by a Suppress condition. During the Arouse condition, the subject was instructed to allow himself to become aroused without trying to inhibit his responses. During the Suppress condition, the subject was asked to observe the stimuli but attempt to control or suppress his arousal. Each set began with a video clip from a sex education tape, which

depicted a man and woman having consensual intercourse. This was intended to stimulate full erection. If the technologist suspected that the subject had not achieved full erection, the subject was instructed to imagine an erotic scenario that would aid him in achieving a full erection.

Following the adult sex education tape, the subject was given time to detumesce below 20% of maximum erection achieved. He was then presented with a neutral stimulus, followed by the 10 test stimuli that comprise the Arouse condition, being allowed to detumesce below 20% of maximum erection achieved after each stimulus. The neutral stimulus and each of the test stimuli were 130 seconds in length. The stimuli were presented in VHS format. Each stimulus consisted of a still image briefly displayed on the screen, followed by a 90-second audio vignette read in a male voice, and then four more still images. The images used in the neutral stimulus depicted plant life, whereas the images used in the test stimuli depicted children or adolescents in varying states of undress. The vignette read during the neutral stimulus described nonsexual activity, whereas the vignette read during the test stimuli described nonforced, coercive, or forced sexual activity with a target in the same age range as the visual images. All sexual activity with child targets was defined as "coercive" or "forced". The term *coercive* was used for descriptions of nonforced sexual activity with children because children below the age of 12 cannot legally consent to sexual activity with partners of *any age* in Canada (Rodrigues, 2008). The term *forced* was used for descriptions of sexual activity that involved physical force. The stimulus set described in the current study was developed specifically for the use of YFPS, and it has been described previously by Clift et al. (2007).

Following the Arouse condition the procedure was repeated with the same stimuli, using the instructions for the Suppress condition. To ensure that the subject attended to the stimuli, he was asked to give a verbal signal each time a yellow star appeared on the video screen. The yellow stars were imbedded randomly in the visual stimuli. If a subject did not appear to be attending to the stimuli, then their test would be deemed invalid.

The 10 test stimuli used in the Arouse and Suppress conditions were the following: (1) coercive sexual activity with a male infant, (2) coercive sexual activity with a very young male (aged 2-5), (3) coercive sexual activity with a prepubescent male (aged 6-11), (4) forced sexual activity with a prepubescent male (aged 6-11), (5) coercive sexual activity with a female infant, (6) coercive sexual activity with a very young female (aged 2-5), (7) coercive sexual activity with a prepubescent female (aged 6-11), (8) forced sexual activity with a prepubescent female (aged 6-11), (9) nonforced sexual activity with an adolescent female, and (10) forced sexual activity with an adolescent female. The infant stimuli were not analyzed in the current study because none of the subjects had committed offenses against children below the age of 2 years. The sexual activities described in the infant stimuli are fundamentally different from the sexual activities described in the other test stimuli, because of infants' physical and cognitive developmental immaturity.

**Table 1**
**Stimulus Categories Included in the Deviance Indices**

| Index | Stimulus Categories |
| --- | --- |
| Child Index | Very young male—coercive |
| | Prepubescent male—coercive |
| | Prepubescent male—with force |
| | Very young female—coercive |
| | Prepubescent female—coercive |
| | Prepubescent female—with force |
| Male Child Index | Very young male—coercive |
| | Prepubescent male—coercive |
| | Prepubescent male—with force |
| Female Child Index | Very young female—coercive |
| | Prepubescent female—coercive |
| | Prepubescent female—with force |
| Female Teen Force Index | Adolescent female—with force |

Changes in penile circumference were initially recorded as millimeters of change. These scores where then transformed into $z$ scores—a distribution with a mean of 0 and a standard deviation of 1. This was done separately for the Arouse condition and the Suppress condition. A positive $z$ score represents greater than average arousal for that individual across the stimuli that were presented in each set, and a negative $z$ score represents less than average arousal. There are two advantages to using $z$ scores. First, $z$ scores allow for direct comparison across subjects (Launay, 1994). Second, $z$ scores have been found to account for a higher percentage of variance than either raw scores (e.g., millimeters) or percentage of full erection achieved (Earls, Quinsey, & Castonguay, 1987). However, it should be noted that $z$ scores do not give a measure of the magnitude of an individual's penile responses. That is, subjects who respond strongly to all stimuli would produce similar $z$ scores to those subjects who respond weakly to all stimuli.

The $z$ scores were then used to calculate deviance indices, which are measures of arousal to deviant stimuli relative to arousal to age-appropriate stimuli. The four deviance indices were calculated by subtracting the subject's $z$ score for appropriate stimuli (i.e., nonforced sexual activity with an adolescent female) from his mean $z$ score for four categories of deviant stimuli. For example, the Child Index was created by subtracting the subject's $z$ score for appropriate stimuli from his mean $z$ score for all child stimuli (see Table 1 for a list of stimuli included in each index). The four deviance indices were the: (1) Child Index, (2) Male Child Index, (3) Female Child Index, and (4) Female Teen Force Index. A negative deviance index represents greater arousal to the age-appropriate nonforced stimulus than deviant stimuli, whereas a positive deviance index represents greater arousal to deviant stimuli than

**Table 2**
**Mean Arouse and Suppress Condition**
**Deviance Indices Before and After Treatment**

| Index | Arouse Condition | | Suppress Condition | |
|---|---|---|---|---|
| | M (SD) | Range | M (SD) | Range |
| Pretreatment | | | | |
| Child Index | −.15 (1.32) | −2.82 to 2.83 | −.20 (1.16) | −2.84 to 1.69 |
| Male Child Index | −.41 (1.44) | −2.91 to 2.83 | −.45 (1.25) | −2.91 to 2.11 |
| Female Child Index | .11 (1.31) | −2.83 to 2.83 | .06 (1.21) | −2.83 to 2.28 |
| Female Teen Force Index | .20 (1.45) | −3.04 to 3.10 | .10 (1.29) | −2.99 to 3.23 |
| Posttreatment | | | | |
| Child Index | −.22 (1.08) | −2.83 to 2.41 | −.43 (1.10) | −2.83 to 1.32 |
| Male Child Index | −.57 (1.20) | −2.83 to 2.41 | −.61 (1.14) | −2.92 to 1.96 |
| Female Child Index | .12 (1.08) | −2.83 to 2.52 | −.24 (1.21) | −2.83 to 1.71 |
| Female Teen Force Index | .17 (1.29) | −2.92 to 2.89 | .19 (1.61) | −3.08 to 2.90 |

Note: Arouse condition: pretreatment, $n = 112$; posttreatment, $n = 83$. Suppress condition: pretreatment, $n = 106$; posttreatment, $n = 75$.

to the age-appropriate nonforced stimulus. Sample means, standard deviations, and ranges for each of the four indices are reported in Table 2.

The stimulus set that is currently used at YFPS does not contain an age-appropriate nonforced homosexual stimulus. Therefore, it is important to take the subjects' sexual orientation into account when evaluating their PPG profiles. Three subjects had a self-reported homosexual orientation, as coded by YFPS file information. One of these subjects responded most strongly to the homosexual child stimuli. We decided that it was inappropriate to calculate deviance indices for this subject using heterosexual stimuli, and therefore, his data were removed from the PPG analyses. The two remaining subjects with a self-reported homosexual orientation responded as strongly—or more strongly—to stimuli that depicted female targets. Therefore, their data were not excluded.

## Data Coding and Analysis: Discriminative Validity

Historical (i.e., pretreatment) sexual offenses were coded from YFPS file information. YFPS files contain psychological assessment reports, treatment notes, police reports, and victim statements, as well as information from teachers, social workers, and other people who work directly with youths who have been convicted of criminal offenses. A youth was coded as having committed a sexual offense against a given type of victim if there was substantial information in their YFPS file to indicate that a sexual offense had occurred. Both contact and noncontact sexual offenses were coded as sexual offenses. Ninety-four percent of the sample had at least one contact sexual offence.

Complete information about victim age and sex was available for 116 subjects. The majority of our sample offended solely against prepubescent victims less than the age of 12 years (69.0%), 11.2% offended against victims less than the age of 12 *and* victims aged 12 or older, and 19.8% offended solely against victims aged 12 or older. Thirty-six percent of the sample had at least one male victim, whereas 64% offended exclusively against female victims.

Initially, we used correlational analyses to explore the relationships between victim age and gender characteristics and pretreatment and posttreatment deviance indices under the Arouse and Suppress conditions. In other words, we assessed the relationships between having *any* male child victims (present/absent), *any* female child victims (present/absent), *any* female nonchild victims (present/absent), and the deviance indices. Given that some subjects had victims in more than one age and sex category, some subjects were included in more than one comparison. This initial analysis was intended to determine if there was a relationship between offending against a certain type of victim and arousal to stimuli that depict that type of victim (e.g., male children) during the Arouse and Suppress conditions.

Following our preliminary analysis of the relationship between victim characteristics and deviant arousal, we formed three distinct groups of offenders based on victim age and sex characteristics. These groups were used for the analysis of discriminative validity. The three distinct groups were: (1) youths who had committed an offense against a male less than the age of 12 and had not committed an offense against anyone aged 12 or older, (2) youths who had committed an offense against a female less than the age of 12 and had not committed an offense against a male less than the age of 12 or anyone aged 12 or older, and (3) youths who had committed an offense against a female aged 12 or older and had not committed an offense against anyone less than the age of 12. For ease of communication, these three groups will be referred to as those with: (1) male child victims, (2) *only* female child victims, and (3) *only* female nonchild victims. It should be noted that the first group contains subjects with male child victims and subjects with both male and female child victims. These subjects were combined because of previous research, which demonstrated that adolescents with at least one male child victim respond more strongly to deviant sexual stimuli than adolescents with only female child victims or adolescents without child victims (Murphy, DiLillo, Haynes, & Steere, 2001; Seto et al., 2000). The relationship between deviant arousal and having any male child victim(s) was also supported by our initial correlational analysis (Table 3). Subjects who could not be classified into one of the three categories were excluded from the analysis of discriminative validity.

After the offenders were classified into three groups, their mean pretreatment and posttreatment deviance indices were compared using one-way analysis of variance (ANOVA), and significant differences were probed using post hoc Tukey–Kramer tests.

**Table 3**
**Correlations Between Victim Characteristics and**
**Pretreatment Arousal and Suppression Indices (*p* Value)**

| Index | AMCV, Present/Absent | AFCV, Present/Absent | AFNCV, Present/Absent |
|---|---|---|---|
| Pretreatment—Arousal | | | |
| Child Index | .27 (.007)* | −.03 (.770) | −.21 (.038) |
| Male Child Index | .27 (.007)* | −.05 (.645) | −.19 (.068) |
| Female Child Index | .25 (.015)* | −.01 (.932) | −.22 (.030) |
| Female Teen Force Index | .13 (.197) | −.02 (.856) | −.11 (.290) |
| Pretreatment—Suppression | | | |
| Child Index | .12 (.248) | .04 (.728) | −.12 (.245) |
| Male Child Index | .19 (.075) | −.07 (.540) | −.07 (.509) |
| Female Child Index | .04 (.704) | .14 (.199) | −.16 (.125) |
| Female Teen Force Index | .00 (.996) | −.01 (925) | .04 (.697) |

Note: AMCV = *any* male child victims; AFCV = *any* female child victims; AFNCV = *any* female nonchild victims. Arouse condition, *n* = 97; Suppress condition, *n* = 91.
*$p < .025$, the accepted α level after the Bonferroni adjustment.

   Finally, we determined the sensitivity and specificity of the deviance indices under two instruction sets. In this context, sensitivity refers to the proportion of subjects who have a certain type of victim and are identified by the deviance indices as having interest in that type of victim. Specificity refers to the proportion of subjects who *do not* have a particular type of victim and are identified by the deviance indices as *not* having interest in that type of victim. We initially determined the sensitivity and specificity of the deviance indices using a cutoff score of zero. Zero was chosen as the cutoff score because this represents an equal amount of arousal to deviant stimuli and age-appropriate stimuli. We also determined the sensitivity of deviance index scores at predetermined specificity levels of 90%. The sensitivity and specificity of the Male Child Index were calculated for subjects with and without male child victims; the sensitivity and specificity of the Female Child Index were calculated for subjects with and without *only* female child victims; and the sensitivity and specificity of the Female Teen Force Index were calculated for subjects with and without *only* female nonchild victims (i.e., we used the same categorization for this analysis that we used for the ANOVA).

   Although we examined discriminative validity using both pretreatment and posttreatment deviance indices, it should be noted that the subjects were classified into groups based on *pretreatment* victim type (i.e., victim age and sex). That is, subjects were not reclassified based on new offences that might have occurred during the treatment period. This was not possible, as there was limited file information about offences that occurred during the treatment period, and British Columbia Corrections records do not contain information about victim age and sex.

## Data Coding and Analysis: Predictive Validity

New charges and convictions were coded from British Columbia Corrections records. Recidivism was defined as a formal charge and/or conviction during the posttreatment follow-up period and was divided into three categories: violent offenses, sexual offenses, and nonviolent offenses. Violent offenses included robbery, assault (including threats, intimidation, and criminal harassment), homicide (including attempted homicide), weapons offenses, kidnapping, and arson with intent to injure. Sexual offenses included both contact and noncontact offenses. All other offenses were coded as nonviolent, including breaches of recognizance.

The follow-up period began when the subject was discharged from treatment. The length of the follow-up period ranged from 13 to 104 months, with a mean of 71.2 months ($SD = 21.2$) or 5.9 years ($SD = 1.8$). At the end of the treatment period, 127 of the 132 subjects (96.2%) were free in the community. The remaining 5 subjects were released within 10 months. In addition to the 5 subjects who were incarcerated at the beginning of the follow-up period, some subjects received new sentences during part of the follow-up period, which could have substantially reduced their opportunity to reoffend. For the analysis of predictive validity we required a minimum follow-up time of 1 year *free in the community*. Two subjects were excluded because of this criterion. After the exclusions, the subjects' actual time free in the community ranged from 12 to 104 months, with a mean of 67.4 months ($SD = 20.9$) or 5.6 years ($SD = 1.7$).

The predictive validity of the deviance indices was assessed using Cox proportional hazards regression. This method had the advantage of taking into account the variable times at risk among subjects (Luke & Homan, 1998). For this analysis we used time *free in the community* as the measure of time at risk. This allowed us to control for the possibility that some subjects were incarcerated (for nonsexual offenses) during a portion of the follow-up period. The outcome variable was a dichotomous measure of sexual recidivism.

Cox proportional hazards model allows one to estimate the "hazard" (or risk) of an event of interest given the subjects' scores on explanatory variables (e.g., the risk of reoffence based on deviance indices). Hazard ratios can be interpreted as a measure of effect size and as a measure of relative risk. For example, a hazard ratio of 1.5 would mean that every increase of one unit on the independent variable (e.g., an increase of one on a deviance index) results in an increase of 1.5 times in the hazard rate.

Predictive validity was also assessed using area under the curve (AUC) statistics. AUC statistics are regularly used in the risk assessment literature because they are not constrained by base rates or selection ratios. In this context, an AUC statistic represents the probability that a randomly selected recidivist would demonstrate greater deviant arousal than a randomly selected nonrecidivist (Mossman & Somoza, 1991). Possible scores on this statistic range from 0 to 1, with 0.5 indicating prediction at chance levels and 1.0 representing an entirely accurate measure (i.e., no overlap between recidivists and nonrecidivists).

Because of the large number of analyses that were conducted in the current study, the Bonferroni method was used to control the familywise (FW) error rate to .10 for each set of analyses.

# Results

## Victim Characteristics and Deviant Arousal

We first analyzed the relationships between victim characteristics and pretreatment deviant arousal. These findings are presented in Table 3. There was a significant positive relationship between having *any* male child victims and the pretreatment Arouse condition Male Child Index, the pretreatment Arouse condition Female Child Index, and the pretreatment Arouse condition Child Index. However, there was no relationship between having *any* female child victims and any of the pretreatment Arouse condition deviance indices.

After using the Bonferroni method to control the FW error rate to .10 (i.e., $\alpha = .025$, as four comparisons were being made), there were marginally significant negative relationships between having *any* female nonchild victims and both the pretreatment Arouse condition Child Index and the pretreatment Arouse condition Female Child index. However, there was no relationship between having *any* female nonchild victims and the pretreatment Arouse condition Female Teen Force Index. Additionally, there were no significant relationships between having *any* male child victims, *any* female child victims or *any* female nonchild victims, and the pretreatment Suppress condition deviance indices (Table 3).

None of the posttreatment deviance indices were significantly correlated with pretreatment victim characteristics (Arouse condition: $r$s ranged from $-.14$ to .16, all $p$s ≥ .178; Suppress condition: $r$s ranged from $-.07$ to .11, all $p$s ≥ .374).

## Discriminative Validity: Arouse Condition

Under the Arouse condition, mean pretreatment deviance indices were compared for the group with male child victims (as described in the Methods section), the group with *only* female child victims, and the group with *only* female nonchild victims, using one-way ANOVA (see Table 4). After using the Bonferroni method to control the FW error rate to .10 (i.e., $\alpha = .025$, as four comparisons were being made), the three groups had significantly different means on the pretreatment Child Index, $F(2, 82) = 4.19$, $p = .019$, and the pretreatment Male Child Index, $F(2, 82) = 4.49$, $p = .014$. Additionally, the three groups had marginally significantly different means on the pretreatment Female Child Index, $F(2, 82) = 3.38$, $p = .039$. The groups did not have significantly different means on the pretreatment Female Teen Force Index, $F(2, 82) = 1.13$, $p > .30$. This finding was expected, as none of the groups had offended exclusively using force against female adolescents.

**Table 4**
**Comparison of the Three Offender Types**
**on the Pretreatment Arousal and Suppression Indices**

| Index | MCV, Mean (SD) | FCV, Mean (SD) | FNCV, Mean (SD) | F | p Value |
|---|---|---|---|---|---|
| Pretreatment—Arousal | | | | | |
| Child Index | .54 (1.42) | −.29 (1.12) | −.47 (1.55) | 4.19 | .019* |
| Male Child Index | .37 (1.55) | −.60 (1.21) | −.65 (1.61) | 4.49 | .014* |
| Female Child Index | .70 (1.39) | .03 (1.14) | −.29 (1.60) | 3.38 | .039 |
| Female Teen Force Index | .55 (1.68) | .19 (1.42) | −.14 (1.31) | 1.13 | .329 |
| Pretreatment—Suppression | | | | | |
| Child Index | .07 (1.00) | −.20 (1.14) | −.19 (1.17) | 0.52 | .600 |
| Male Child Index | −.04 (1.08) | −.59 (1.18) | −.34 (1.30) | 1.67 | .194 |
| Female Child Index | .19 (1.13) | .20 (1.23) | −.04 (1.10) | 0.27 | .764 |
| Female Teen Force Index | .04 (1.00) | .05 (1.29) | .11 (1.60) | 0.02 | .982 |

Note: MCV = male child victims (as described in the Methods section); FCV = *only* female child victims; FNCV = *only* female nonchild victims. Arouse condition: $ns$ = 26, 42, and 17, respectively; Suppress condition: $ns$ = 24, 41, and 17, respectively.
*$p < .025$, the accepted α level after the Bonferroni adjustment.

Post hoc Tukey–Kramer tests were performed on the deviance indices that were obtained during the pretreatment Arouse condition. These post hoc tests revealed that the group with male child victims had a significantly higher mean Child Index than the group with *only* female nonchild victims ($p = .040$; $d = .69$). The group with male child victims also had a significantly higher Child Index than the group with *only* female child victims ($p = .040$; $d = .65$). The group with *only* female child victims could not be differentiated from the group with *only* female nonchild victims using the pretreatment Child Index ($p = .871$; $d = .13$).

Post hoc tests revealed that the group with male child victims had a significantly higher mean pretreatment Male Child Index than the group with *only* female child victims ($p = .018$; $d = .70$). The group with male child victims also had a higher pretreatment Male Child Index than the group with *only* female nonchild victims; however, that finding was only marginally significant ($p = .055$; $d = .65$). As expected, the group with *only* female child victims and the group with *only* female nonchild victims had similar means on the pretreatment Male Child Index ($p = .991$; $d = 0.04$).

The pretreatment Female Child Index did not differentiate the group with *only* female child victims from the group with *only* female nonchild victims ($p = .670$; $d = .23$). Additionally, the group with *only* female child victims and the group with male child victims did not score significantly different on this index ($p = .111$; $d = .53$). However, the group with male child victims had a significantly higher pretreatment Female Child Index than the group with *only* female nonchild victims ($p = .047$; $d = .66$).

**Table 5**
**Proportion of Participants Identified With Pretreatment Deviant
Arousal at Cutoff of 0 and the Cutoff Corresponding to Specificity of 90%**

| Arouse Condition | | | Suppress Condition | | |
|---|---|---|---|---|---|
| Index Cutoff | Sensitivity | Specificity | Index Cutoff | Sensitivity | Specificity |
| Pretreatment Male Child Index | | | | | |
| 0.00 | 61.5% | 67.8% | 0.00 | 54.2% | 65.5% |
| 1.19[a] | 42.3% | 90% | 0.98 | 12.5% | 90% |
| Pretreatment Female Child Index | | | | | |
| 0.00 | 45.2% | 44.2% | 0.00 | 58.5% | 51.2% |
| 2.24 | 2.4% | 90% | 1.36 | 19.5% | 90% |
| Pretreatment Female Teen Force Index | | | | | |
| 0.00 | 41.2% | 48.5% | 0.00 | 47.1% | 52.3% |
| 2.45 | 5.9% | 90% | 1.53 | 23.5% | 90% |

Note: Arouse condition, $n = 85$; Suppress condition, $n = 82$.
a. A cutoff score of 1.19 would represent mean arousal to male child stimuli that is 1.19 standard deviations greater than the subject's arousal to age-appropriate nonforced stimuli.

The mean Arouse condition posttreatment deviance indices were not significantly different for the group with male child victims, the group with *only* female child victims, and the group with *only* female nonchild victims (*F*s ranged from .04 to 1.33; all *p*s ≥ .272).

## Discriminative Validity: Suppress Condition

Under the Suppress condition, the group with male child victims, the group with *only* female child victims, and the group with *only* female nonchild victims did not have significantly different means on any of the pretreatment deviance indices (all *p*s ≥ .194; see Table 4) or any of the posttreatment deviance indices (*F*s ranged from .37 to .86; all *p*s ≥ .427).

## Sensitivity of the Deviance Indices: Arouse Condition

At a cutoff score of 0, the pretreatment Arouse condition Male Child Index had a sensitivity of 61.5%, meaning that 61.5% of youths with male child victims were identified as having deviant arousal to male children (Table 5). The specificity of the pretreatment Arouse condition Male Child Index was 67.8% at a cutoff score of 0, meaning that more than two thirds of subjects who did not have a male child victim did not demonstrate deviant arousal on the Male Child Index. At a predetermined specificity level of 90% (i.e., cutoff score of 1.19), the pretreatment Male Child Index had a sensitivity of 42.3%, meaning that more than 40% of the subjects with male child victims were identified as having deviant arousal on the Male Child Index.

**Table 6**
**Proportion of Participants Identified with Posttreatment Deviant Arousal**
**at Cutoff of 0 and the Cutoff Corresponding to Specificity of 90%**

| Arouse Condition | | | Suppress Condition | | |
|---|---|---|---|---|---|
| Index Cutoff | Sensitivity | Specificity | Index Cutoff | Sensitivity | Specificity |
| Posttreatment Male Child Index | | | | | |
| 0.00 | 40.9% | 72.1% | 0.00 | 38.1% | 69.2% |
| 1.04[a] | 22.7% | 90% | 0.81 | 14.3% | 90% |
| Posttreatment Female Child Index | | | | | |
| 0.00 | 53.3% | 45.7% | 0.00 | 37% | 51.5% |
| 1.83 | 6.7% | 90% | 1.40 | 3.7% | 90.0% |
| Posttreatment Female Teen Force Index | | | | | |
| 0.00 | 53.8% | 46.2% | 0.00 | 50% | 50% |
| 1.82 | 7.7% | 90% | 2.35 | 8.3% | 90% |

Note: Arouse condition, $n = 65$; Suppress condition, $n = 60$.
a. A cutoff score of 1.04 represents mean arousal to male child stimuli that is 1.04 standard deviations greater than the subject's arousal to age-appropriate nonforced stimuli.

During the Arouse condition, the pretreatment Female Child Index and the pretreatment Female Teen Force Index had low sensitivity and specificity at classifying subjects by the presence or absence of *only* female child victims and *only* female nonchild victims, respectively (Table 5). Although the subjects were grouped based on pretreatment victim characteristics, the sensitivity and specificity of the posttreatment Arouse condition deviance indices are presented in Table 6 for comparison.

## Sensitivity of the Deviance Indices: Suppress Condition

At a cutoff score of 0, the pretreatment Suppress condition Male Child Index had a sensitivity of 54.2% and a specificity of 65.5% (Table 5). When 90% specificity was required (i.e., using a cutoff score of .98), the pretreatment Suppress condition Male Child Index had a sensitivity of 12.5%. In general, as with the Arouse condition, the other pretreatment deviance indices did not perform as well as the Male Child Index at classifying subjects by the presence or absence of particular types of victims (Table 5). For comparison, the sensitivity and specificity of the posttreatment Suppress condition deviance indices are presented in Table 6.

## Outcome: Recidivism Rates

Over the approximately 6-year follow-up period, 53 of the 132 subjects (40.2%) were charged with a new crime.[1] Fifty-one subjects (38.6%) were charged with a nonviolent crime, 28 (21.2%) were charged with a violent crime, and 10 (7.6%) were charged with a sexual crime.

The 10 sexual recidivists were charged with 22 new sexual offenses. One subject had seven new sexual offenses, two subjects had four, and seven subjects had one. The majority of new sexual offenses were sexual assaults ($n = 14$). Other new sexual offenses included sexual interference ($n = 2$), committing indecent acts in public ($n = 2$), indecent exposure ($n = 1$), invitation to sexual touching ($n = 1$), communication via a computer to lure a child ($n = 1$), and inciting fear of a sexual offence involving a child less than the age of 14 ($n = 1$). Nine of the 10 subjects who recidivated sexually also committed other types of offences during the follow-up period. One committed a violent crime, three committed nonviolent crimes, and five committed both violent and nonviolent crimes in addition to their sexual offense(s).

For comparison, we also calculated recidivism rates for those subjects who had a valid posttreatment PPG assessment (i.e., had completed treatment) and had a minimum follow-up time of 12 months in the community. Twenty-seven of those 83 subjects (32.5%) were charged with a new crime during the follow-up period. Twenty-six subjects (31.3%) were charged with a nonviolent crime, 14 (16.9%) were charged with a violent crime, and 7 (8.4%) were charged with a sexual crime.

## Predictive Validity: Arouse Condition

After using the Bonferroni method to control the FW error rate to .10 (i.e., $\alpha = .025$, as four comparisons were being made), three posttreatment Arouse condition deviance indices were significantly related to sexual recidivism: The Child Index, Exp($b$) = 2.98; Male Child Index, Exp($b$) = 2.30; and Female Child Index, Exp($b$) = 3.31. These findings indicate that for every one-point increase on the Child Index, Male Child Index, and Female Child Index, subjects were 2.98 times, 2.30 times, and 3.31 times more likely to commit a new sexual offense, respectively (Table 7). The associated AUC statistics (Table 8) for these indices were medium to large (cf., Rice & Harris, 2005); however, they did not reach statistical significance: Child Index (AUC = .71, $p = .067$), Male Child Index (AUC = .70, $p = .080$), and Female Child Index (AUC = .71, $p = .064$).

None of the pretreatment deviance indices calculated from Arouse condition data were significantly related to sexual recidivism (Table 7). Additionally, their associated AUC statistics were low (range = .34-.47; all $p$s > .10; Table 8). In fact, the AUC statistics were below .5, suggesting that nonrecidivists scored higher on these indices than recidivists.

## Predictive Validity: Suppress Condition

After using the Bonferroni method to control the FW error rate to .10 (i.e., $\alpha = .025$, as four comparisons were being made), two of the posttreatment Suppress condition deviance indices were *marginally* related to sexual recidivism (Table 8). These were the Child Index, Exp($b$) = 3.70, and Female Child Index, Exp($b$) = 4.33. That is, for every one-point increase on the Child Index, subjects were 3.70 times more likely to

**Table 7**
**Cox Proportional Hazards Regression Statistics for Pretreatment**
**and Posttreatment Deviance Indices and Sexual Recidivism**

| Index | b | SE | Wald | p Value | Exp(b) |
|---|---|---|---|---|---|
| Pretreatment Arouse | | | | | |
| Child Index | −.10 | .249 | .16 | .694 | .91 |
| Male Child Index | −.04 | .232 | .03 | .873 | .96 |
| Female Child Index | −.15 | .246 | .37 | .541 | .86 |
| Female Teen Force Index | −.23 | .230 | 1.03 | .311 | .79 |
| Pretreatment Suppress | | | | | |
| Child Index | −.20 | .305 | .43 | .512 | .82 |
| Male Child Index | −.07 | .283 | .06 | .803 | .93 |
| Female Child Index | −.30 | .297 | 1.03 | .310 | .74 |
| Female Teen Force Index | −.32 | .275 | 1.32 | .251 | .73 |
| Posttreatment Arouse | | | | | |
| Child Index | 1.09 | .431 | 6.43 | .011* | 2.98 |
| Male Child Index | .84 | .340 | 6.03 | .014* | 2.30 |
| Female Child Index | 1.20 | .496 | 5.83 | .016* | 3.31 |
| Female Teen Force Index | .16 | .327 | .24 | .622 | 1.18 |
| Posttreatment Suppress | | | | | |
| Child Index | 1.31 | .603 | 4.71 | .030 | 3.70 |
| Male Child Index | .68 | .398 | 2.93 | .087 | 1.98 |
| Female Child Index | 1.47 | .671 | 4.78 | .029 | 4.33 |
| Female Teen Force Index | .37 | .307 | 1.42 | .234 | 1.44 |

Note: Pretreatment: Arouse condition, $n = 110$; Suppress condition, $n = 104$. Posttreatment: Arouse condition, $n = 82$; Suppress condition, $n = 74$.
*$p < .025$, the accepted α level after the Bonferroni adjustment.

commit a new sexual offense. Similarly, for every one-point increase on the Female Child Index, subjects were 4.33 times more likely to commit a new sexual offense. After using the Bonferroni method to control the FW error rate to .10, the AUC statistics for the Child Index and Female Child Index were significant (AUC = .80, $p = .01$ and AUC = .84, $p = .004$, respectively; Table 8). The AUC statistic for the Male Child Index was large (AUC = .72); however, it did not reach significance ($p = .062$).

None of the pretreatment deviance indices calculated from the Suppress condition data were significantly related to sexual recidivism (Table 8). Additionally, their associated AUC statistics were low (range = .43-.48; all $ps > .5$; Table 8).

## Discussion

We examined the strength of the relationship between the age(s) and sex(es) of adolescent sex offenders' known victims under two different conditions: the pretreatment

### Table 8
### Area Under the Curve Statistics for Pretreatment
### and Posttreatment Deviance Indices and Sexual Recidivism

| Index | Arouse Condition | | | | Suppress Condition | | | |
|---|---|---|---|---|---|---|---|---|
| | AUC | SE | 95% CI | p Value | AUC | SE | 95% CI | p Value |
| Pretreatment | | | | | | | | |
| Child Index | .45 | .13 | .20-.69 | .604 | .47 | .13 | .21-.74 | .798 |
| Male Child Index | .47 | .12 | .23-.71 | .764 | .48 | .14 | .22-.75 | .855 |
| Female Child Index | .41 | .13 | .16-.66 | .357 | .43 | .11 | .21-.66 | .530 |
| Female Teen Force Index | .34 | .06 | .21-.46 | .108 | .43 | .10 | .23-.63 | .506 |
| Posttreatment | | | | | | | | |
| Child Index | .71 | .11 | .50-.92 | .067 | .80* | .11 | .59-1.00 | .011 |
| Male Child Index | .70 | .11 | .48-.92 | .080 | .72 | .11 | .51-.92 | .062 |
| Female Child Index | .71 | .11 | .51-.92 | .064 | .84* | .11 | .63-1.00 | .004 |
| Female Teen Force Index | .58 | .12 | .35-.81 | .491 | .64 | .11 | .42-.86 | .219 |

Note: AUC = area under the curve; 95% CI = 95% confidence interval. Pretreatment: Arouse condition, $n = 110$; Suppress condition, $n = 104$. Posttreatment: Arouse condition, $n = 82$; Suppress condition, $n = 74$.
*$p < .025$, the accepted α level after the Bonferroni adjustment.

Arouse condition and the pretreatment Suppress condition. Overall, subjects with *any* male child victims displayed the greatest levels of deviant arousal during the pretreatment Arouse condition. As would be expected, there was a significant positive relationship between having *any* male child victims and pretreatment deviant arousal to all child stimuli. Additionally, as a group, subjects with *any* female nonchild victims displayed more arousal to nonforced interactions with adolescent females than to child stimuli during the Arouse condition—although these findings were only marginally significant. Somewhat surprisingly, there was no relationship between having *any* female child victims and pretreatment deviant arousal to child stimuli. There were also no significant relationships between having *any* male child victims, *any* female child victims, or *any* female nonchild victims and the pretreatment Suppress condition deviance indices.

After examining the relationship between the age(s) and sex(es) of the subjects' victims, we examined the pretreatment discriminative validity of the PPG. Again, this was done under both the Arouse condition and the Suppress condition. In this context, the clinical assumption underlying the Suppress condition is that failure to suppress deviant arousal (e.g., deviant arousal to children) in a controlled situation is related to a propensity to offend against the analogous victim type in the real world. Results from the present study did not support this connection.

The present study did find evidence for the discriminative validity of some PPG indices under the pretreatment Arouse condition. The strongest evidence for discriminative validity was found for the Arouse condition Male Child Index, which

measures deviant sexual interest in male children. Using group means, the Male Child Index significantly discriminated between subjects with male child victims and subjects without male child victims. Using a cutoff score of 1.19 on the Male Child Index, more than 40% of subjects who had male child victims would be identified as having deviant sexual interest in male children, whereas only 10% of subjects without male child victims would be identified as having such interests.

Using group means, the pretreatment Child Index, which measures deviant sexual interest in both male and female children, also discriminated between subjects with male child victims and subjects with female nonchild victims. However, none of the deviance indices differentiated between the group with female child victims and the group with female nonchild victims.

These findings are consistent with previous research on the discriminative validity of the PPG in adolescent offenders. Using a pedophilic index, Seto et al. (2000) were able to discriminate adolescents (aged 14-17) with male child victims from young adult rapists (aged 18-21) and nonoffenders (aged 18-21). The authors were also able to discriminate young adults with male or female child victims from young adult rapists and nonoffenders. However, they were unable to differentiate adolescents with female child victims from rapists and nonoffenders.

It is not clear why the PPG has demonstrated low discriminative validity for adolescents with female child victims. However, Seto et al. (2000) offered several possible explanations. One possibility is that adolescents with female child victims have pedophilic interests that cannot be detected using current PPG methodology. A second possibility is that adolescents with female child victims may be better able to simulate nonpedophilic interests than other sex offenders. A third possibility is that some adolescents with female child victims do not have pedophilic interests. That is, some adolescent child molesters may victimize female children for reasons other than pedophilic interest. For example, one study found that adolescents who had been sexually abused by men offended against young females to reduce tension and anxiety surrounding fears that they might be homosexual (Becker, Hunter, Stein, & Kaplan, 1989). It has also been suggested that some adolescents may offend against female children because of curiosity, low self-esteem, and a lack of social skills (Fehrenbach, Smith, Monastersky, & Deisher, 1986; Hunter, Goodwin, & Becker, 1994; Ryan, Lane, Davis, & Isaac, 1987).

Hunter et al. (1994) hypothesized that adolescents may not yet have developed a fixed pattern of sexual arousal that would lead to consistent patterns of sexual behavior. Results from the present study, as well as the study by Seto et al. (2000), partially support this hypothesis. Although there appears to be a link between sexual deviance and male child victimization in adolescence, the connection between female child victimization and deviant arousal to children (male, female, or both) is less clear. Research on the etiology of pedophilia may shed some light on these findings. It has been suggested that homosexual pedophilia develops at a younger age than heterosexual pedophilia (Hunter et al., 1994). Retrospective self-report data on deviant

sexual arousal of adult sex offenders offer some support for this suggestion (e.g., Marshall, Barbaree, & Eccles, 1991). Data from the current study are also consistent with this suggestion, given the link between homosexual child victimization in adolescence and measurable sexual deviance. However, more research is clearly needed in this area.

Of note, there was no evidence for discriminative validity based on posttreatment deviance indices. One possible explanation for this finding relates to the way that the subjects were categorized. That is, subjects were categorized based on the age(s) and sex(es) of their pretreatment victims. Some subjects may have had new victims during the treatment period, and we were not able to reliably code the age(s) and sex(es) of new victims during the treatment period. There was limited YFPS file information about new victims acquired during the treatment period, and although we had access to British Columbia Corrections records, they do not contain information about the age(s) and sex(es) of victims. For the same reason it was not possible to assess the relationship between posttreatment deviant arousal and the age(s) and sex(es) of victims during the posttreatment period. It is possible that posttreatment deviance indices would be more strongly related to the age(s) and sex(es) of posttreatment victims than to pretreatment victims. In other words, it is possible that some subjects' victim age and sex preferences changed during treatment, because of maturation (Hunter et al., 1994), or the effects of treatment itself, and that their posttreatment deviances indices would therefore be related to *posttreatment* offending patterns and not to past offending patterns.

We also examined the predictive validity of pre- and posttreatment PPG-measured deviant arousal and sexual recidivism under the Arouse and Suppress conditions. We found that none of the pretreatment measures of deviant arousal were associated with sexual recidivism. However, there was evidence of a relationship between posttreatment measures of arousal to all child stimuli (stimuli that depict forced and nonforced sexual interactions with male and female children) and sexual recidivism under the Arouse condition. Based on Cox regression analyses, all three child indices were significantly related to sexual recidivism, and although their associated AUC statistics did not reach significance, their effect sizes were medium to large (cf., Rice & Harris, 2005). Under the Suppress condition, the Child and Female Child indices were marginally related to sexual recidivism, based on Cox regression analysis, and significantly related to sexual recidivism, based on AUC statistics. These findings suggest that posttreatment arousal to child stimuli and failure to suppress that arousal, as measured by the PPG, predict risk for sexual recidivism with accuracy in the moderate to high range (cf., Rice & Harris, 2005). Overall, these findings are consistent with Hanson and Morton-Bourgon's (2004) meta-analysis, which found that sexual interest in children, as measured by phallometric assessment, is a strong predictor of sexual recidivism.

Quinsey, Harris, Rice, and Cormier (2006) have suggested, based on PPG studies conducted with adults, that posttreatment deviance indices are not effective at predicting sexual recidivism. However, among this adolescent sample, the strongest

findings relate to the predictive validity of posttreatment deviance indices. Youth who were still aroused to child stimuli following treatment, or did not suppress their deviant sexual arousal, were at greater risk for sexual recidivism.

These findings make sense from a clinical perspective. If a youth has been through an average of 14.9 months of treatment, he has had an opportunity to learn about deviant arousal patterns and to learn and practice suppression of deviant arousal. The current data suggest that if a youth has deviant arousal after treatment, and does not demonstrate an ability to suppress that arousal, then this residual deviant arousal may well be a potent indicator of risk. This risk may be more pronounced because post-treatment deviant interest and lack of suppression suggests that: (a) deviant arousal and failure to suppress remains in spite of treatment; and (b) it may have stabilized through the maturity of the youth.

Another possible explanation for this discrepancy may be related to the way the PPG procedure is used with adolescents in this setting and the way the PPG has been presented to research subjects. The methodology and instructions used during PPG assessments differ by laboratory (Howes, 1995). Depending on the instructions that are used, it may be that some sex offenders view the PPG assessment procedure as a test of sexual deviance (i.e., a test with potentially negative consequences; Laws & Holmen, 1978). This interpretation of the PPG procedure may lead some subjects who are involved with the judicial system to attempt to inhibit or distort their responses (Farkas, 1978). The subjects in the current study were informed that the results of their PPG assessments would be used for treatment and research proposes *only* and would not be used for any court proceedings. As a part of the treatment process, it was explained that the PPG assessments would help the subjects and their therapists understand the subjects' deviant arousal patterns. Given that the PPG was presented as a part of the treatment process—and not a risk assessment measure—and that by the end of treatment the youths would have had on average 14.9 months of treatment, it may be that these youths were less likely to attempt to distort their arousal during their posttreatment PPG assessments. This may have added to the validity of the posttreat-ment PPG assessments analyzed in the current study and added to their predictive validity. However, whether posttreatment predictive validity is influenced by instruc-tion set is not known and could not be determined by the present study.

In terms of predictive validity, the deviance indices calculated from the Suppress condition data have implications beyond those calculated from the Arouse condition data. During sex offender treatment programs, adolescent sexual offenders are often taught techniques to help them reduce or control their deviant sexual arousal (Aylwin, Reddon, & Burke, 2005; Bourke & Donohue, 1996; Hunter & Becker, 1994). The clinical assumption is that teaching sex offenders to suppress their deviant arousal will reduce their probability of reoffending (e.g., Marshall, 2007). We tested the assumption that failure to suppress deviant arousal is related to risk for sexual recidi-vism under the Suppress condition. During the Suppress condition, subjects were told to attempt to control or suppress their arousal to each of the stimuli. Therefore, the suppression indices represent both deviant arousal *and* the adolescents' ability/

willingness to suppress their deviant arousal. In other words, a youth who was not aroused by a particular category of stimuli would score low on its associated index, in the same way that they did during the Arouse condition. A youth who was sexually aroused by a particular category of stimuli but was willing and able to control his arousal to that category of stimuli during the Suppress condition would also score low on its associated Suppress index.

As would be predicted from clinical assumptions, deviance indices calculated from the posttreatment Suppress condition data performed similarly to the deviance indices calculated from the posttreatment Arouse condition data. That is, the data from the suppression indices suggest that an inability to control deviant arousal to stimuli that depict forced and nonforced sexual interactions with male and female children is related to sexual recidivism. This finding gives provisional support to the rationale for teaching adolescent sex offenders techniques to control or reduce deviant sexual arousal. To the extent that adolescent sex offenders can be taught to control their deviant arousal patterns, it may be possible to reduce their risk of sexual recidivism. Future research should directly examine the effects of treatment on changes in arousal patterns and on changes in suppression abilities between pretreatment and posttreatment.

## Limitations and Future Directions

There are some limitations of the current study that should be noted. One limitation is the lack of a peer-aged consenting homosexual stimulus. We attempted to control for subjects' homosexual orientation by eliminating subjects who self-identified as homosexual. However, it is possible that some subjects had a predominantly homosexual orientation and did not report this. It is also possible that some subjects were unaware that they had a predominantly homosexual orientation (e.g., because of their developmental stage; Smith, Dermer, & Astramovich, 2005). Future studies could control for this potential confound by including a peer-aged homosexual stimulus. This was not possible in the current study as we were using real-world data collected at laboratories that did not use peer-aged homosexual stimuli at the time.

The sample used in the current study was relatively large in comparison to other samples that have been used to study the PPG in adolescent sex offenders (e.g., Hunter et al., 1994; Hunter et al., 1995; Murphy et al., 2001). However, an increased sample size would have increased the statistical power of the comparisons that were made. A related issue was sample attrition (i.e., the loss of subjects between pretreatment and posttreatment). In the current study, the relationships between the posttreatment Arouse condition child indices and sexual recidivism produced moderate to strong AUC statistics; however, they did not reach statistical significance. It is possible that these comparisons would have reached significance with a larger sample size.

## General Issues Related to PPG Assessment

There are two issues related to PPG assessments that merit further discussion. The first issue relates to the stimuli and procedures that have been used. As noted above, different PPG laboratories have used different stimulus sets, different instruction sets, and presented the stimuli for different periods of time. Because of this variability, several authors have called for standardization in PPG assessment procedures (e.g., Launay, 1999; Marshall & Fernandez, 2000). To date, this goal has not been accomplished. To some degree, the variability in stimulus sets and procedures limits the generalizability of findings from a given study. However, the current study provides some findings that are consistent with previous findings in the field, despite the fact that different stimulus sets, procedures, and deviance indices were used. For example, the findings related to discriminative validity in the current study were consistent with those of Seto et al. (2000). This would suggest that PPG assessment procedures have discriminative validity for adolescents with male child victims but not for adolescents with female child victims. Although it would be ideal to achieve a consensus on PPG assessment procedures, consistent findings from multiple studies increase our confidence in the generalizability of the results.

Second, several authors have noted ethical concerns related to the PPG. These concerns include the invasiveness of the procedure and the necessity of exposing subjects to sexually explicit materials (Hunter & Lexier, 1998; Seto, 2001). These concerns may be particularly pronounced when dealing with younger adolescents or adolescents who have been victims of sexual abuse themselves and may be suffering from posttraumatic stress disorder. Given these concerns, individual therapists should weigh the potential benefits of phallometric assessment (e.g., obtaining a more complete picture of the client's sexual interests or examining suppression abilities) with the potential negative effects of the examination. All subjects in the current study had an opportunity to discuss the PPG procedure with their therapist and express any concerns that they may have had. In addition, all PPG assessments were conducted by trained medical technicians, who had extensive experience conducting PPG assessments. In addition to these precautions, therapists working with sexually abusive adolescents should consult the Practice Standards and Guidelines for Members of the Association for the Treatment of Sexual Abusers (ATSA, 2005).

## Summary and Conclusions

In summary, this article adds to the literature on the psychometric properties of the PPG when used with adolescent sex offenders. Consistent with previous research, the PPG demonstrated discriminative validity for adolescent sex offenders with male child victims. This is an important finding, as adult sex offenders with an interest in male children may offend against a disproportionately high number of victims (Abel, Becker, Cunningham-Rathner, Mittelman, & Rouleau, 1988; Abel et al., 1987). If we

are able to detect this group of offenders early enough and provide them with adequate treatment, this could greatly reduce the number of child molestations that occur.

In terms of predictive validity, both posttreatment arousal to children and an inability to suppress that arousal were related to sexual offense recidivism over the 6-year follow-up period. This is the first study to examine whether failure to suppress deviant arousal in a controlled clinical laboratory setting is related to risk for sexual recidivism in adolescent sex offenders. Given that a relationship was found between posttreatment failure to suppress deviant arousal to child stimuli and sexual recidivism, it would suggest that teaching adolescent sex offenders the ability to suppress deviant arousal within treatment may have relevance to reducing recidivism risk. Based on the current findings, further research on the PPG with adolescent sex offenders is warranted—including research that further examines the role of suppression of deviant sexual arousal in reducing risk for sexual recidivism.

## Acknowledgments

## Note

1. For ease of communication, both charges and convictions are referred to as charges.

## References

Abel, G. G., Becker, J. V., Cunningham-Rathner, J., Mittelman, M. S., & Rouleau, J. L. (1988). Multiple paraphilic diagnoses among sex offenders. *Bulletin of the American Academy of Psychiatry and the Law*, *16*, 153-168.

Abel, G. G., Becker, J. V., Mittelman, M. S., Cunningham-Rathner, J., Rouleau, J. L., & Murphy, W. D. (1987). Self-reported sex crimes of nonincarcerated paraphiliacs. *Journal of Interpersonal Violence*, *2*, 3-25.

Association for the Treatment of Sexual Abusers. (2005). *Practice standards and guidelines for members of the Association for the Treatment of Sexual Abusers*. Beaverton, OR: Author.

Aylwin, A. S., Reddon, J. R., & Burke, A. R. (2005). Sexual fantasies of adolescent male sex offenders in residential treatment: A descriptive study. *Archives of Sexual Behavior*, *34*, 231-239.

Becker, J. V., Hunter, J. A., Goodwin, D., Kaplan, M. S., & Martinez, D. (1992). Test-retest reliability of audio-taped phallometric stimuli with adolescent sex offenders. *Annals of Sex Research*, *5*, 45-51.

Becker, J. V., Hunter, J. A., Stein, R. M., & Kaplan, M. S. (1989). Factors associated with erection in adolescent sex offenders. *Journal of Psychopathology and Behavioral Assessment*, *11*, 353-362.

Bourke, M. L., & Donohue, B. (1996). Assessment and treatment of juvenile sex offenders: An empirical review. *Journal of Child Sexual Abuse*, *5*, 47-70.

Clift, R. J. W., Gretton, H. M., & Rajlic, G. (2007). The relationship between deviant arousal and sexual recidivism in adolescent sex offenders. In M. C. Calder (Ed.), *Children and young people who sexually abuse* (pp. 96-102). Dorset, UK: Russell House.

Earls, C. M., Quinsey, V. L., & Castonguay, L. G. (1987). A comparison of three methods of scoring penile circumference changes. *Archives of Sexual Behavior*, *16*, 493-500.

Farkas, G. M. (1978). Comments on Levin et al. and Rosen and Kopel: Internal and external validity issues. *Journal of Consulting and Clinical Psychology*, *46*, 1515-1516.

Fehrenbach, P. A., Smith, W., Monastersky, C., & Deisher, R. W. (1986). Adolescent sexual offenders: Offender and offence characteristics. *American Journal of Orthopsychiatry*, *56*, 225-233.

Gretton, H. M., McBride, M., Hare, R. D., O'Shaughnessy, R., & Kumka, G. (2001). Psychopathy and recidivism in adolescent sex offenders. *Criminal Justice and Behavior*, *28*, 427-449.

Groth-Marnat, G. (2003). *Handbook of psychological assessment*. Hoboken, NJ: John Wiley.

Hanson, R. K., & Morton-Bourgon, K. (2004). *Predictors of sexual recidivism: An updated meta-analysis* (User Rep. No. 2004-02). Ottawa, Ontario, Canada: Public Safety and Emergency Preparedness Canada.

Howes, R. J. (1995). A survey of plethysmographic assessment in North America. *Sexual Abuse: A Journal of Research and Treatment*, *7*, 9-24.

Hunter, J. A., Jr., & Becker, J. V. (1994). The role of deviant sexual arousal in juvenile sexual offending. Etiology, evaluation, and treatment. *Criminal Justice and Behavior*, *21*, 132-149.

Hunter, J. A., Becker, J. V., & Kaplan, M. S. (1995). The adolescent sexual interest card sort: Test-retest reliability and concurrent validity in relation to phallometric assessment. *Archives of Sexual Behavior*, *24*, 555-561.

Hunter, J. A., Goodwin, D. W., & Becker, J. V. (1994). The relationship between phallometrically measured deviant sexual arousal and clinical characteristics in juvenile sexual offenders. *Behavior Research and Therapy*, *32*, 533-538.

Hunter, J. A., & Lexier, L. J. (1998). Ethical and legal issues in the assessment and treatment of juvenile sex offenders. *Child Maltreatment*, *3*, 339-348.

Juristat. (2003). *Sexual offences in Canada* (Catalogue No. 85-002-XIE, Vol. 23, No. 6). Ottawa, Ontario, Canada: Statistics Canada.

Launay, G. (1994). The phallometric assessment of sex offenders: Some professional and research issues. *Criminal Behavior and Mental Health*, *4*, 48-70.

Launay, G. (1999). The phallometric assessment of sex offenders: An update. *Criminal Behaviour and Mental Health*, *9*, 254-274.

Laws, D. R., & Holmen, M. L. (1978). Sexual response faking by pedophiles. *Criminal Justice and Behavior*, *5*, 343-356.

Luke, D. A., & Homan, S. M. (1998). Time and change: Using survival analysis in clinical assessment and treatment evaluation. *Psychological Assessment*, *10*, 306-378.

Marshall, W. L. (2007). Covert association: A case demonstration with a child molester. *Clinical Case Studies*, *6*, 218-231.

Marshall, W., Barbaree, H., & Eccles, A. (1991). Early onset and deviant sexuality in child molesters. *Journal of Interpersonal Violence*, *6*, 323-336.

Marshall, W. L., & Fernandez, Y. M. (2000). Phallometric testing with sexual offenders: Limits to its value. *Clinical Psychology Review*, *20*, 807-822.

McGrath, R. J., Cumming, G. F., & Burchard, B. L. (2003). *Current practices and trends in sexual abuser management: The Safer Society 2002 Nationwide Survey*. Brandon, VT: Safer Society Foundation.

Mossman, D., & Somoza, E. (1991). ROC curves, test accuracy, and the description of diagnostic tests. *Journal of Neuropsychiatry & Clinical Neurosciences*, *3*, 330-333.

Murphy, W. D., DiLillo, D., Haynes, M. R., & Steere, E. (2001). An exploration of factors related to deviant sexual arousal among juvenile sex offenders. *Journal of Research and Treatment*, *13*, 91-103.

Ozeer, S. (2002). *Standards and guidelines for the assessment, treatment and management of youth who have sexually offended*. Unpublished document, Youth Forensic Psychiatric Services.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk* (2nd ed.). Washington, DC: American Psychological Association.

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law and Human Behavior*, *29*, 615-620.

Rodrigues, G. P. (2008). *Pocket criminal code 2009*. Toronto, Ontario, Canada: Thomson-Carswell.

Ryan, G., Lane, S., Davis, J., & Isaac, C. (1987). Juvenile sex offenders: Development and correction. *Child Abuse and Neglect*, *11*, 385-395.

Seto, M. C. (2001). The value of phallometry in the assessment of male sex offenders. *Journal of Forensic Psychology Practice*, *1*, 65-75.

Seto, M. C., Lalumiere, M. L., & Blanchard, R. (2000). The discriminative validity of a phallometric test for pedophilic interests among adolescent sex offenders against children. *Psychological Assessment*, *12*, 319-327.

Smith, S. D., Dermer, S. B., & Astramovich, R. (2005). Working with nonheterosexual youth to understand sexual identity development, at-risk behaviors, and implications for health care professionals. *Psychological Reports*, *96*, 651-654.

U.S. Department of Justice, Federal Bureau of Investigation. (2007, September). *Crime in the United States, 2006*. Retrieved November 23, 2007, from http://www.fbi.gov/ucr/cius2006/