Arch Sex Behav (2017) 46:671–684
DOI 10.1007/s10508-017-0952-2

ORIGINAL PAPER

Inhibiting Sexual Arousal to Children: Correlates and Its Influence on the Validity of Penile Plethysmography

Kelly M. Babchishin[1] · Susan D. Curry[1] · J. Paul Fedoroff[1] · John Bradford[2] · Michael C. Seto[1]

Received: 26 March 2015 / Revised: 27 January 2017 / Accepted: 28 January 2017 / Published online: 6 March 2017

Abstract The current study examined the extent to which 1136 men were able to inhibit their sexual arousal on a phallometric assessment, when instructed to do so. Although the observed changes between the two conditions (i.e., Normal and Suppression) were small to moderate in magnitude, the change was not more than what would be expected by measurement error for most participants (e.g., 83% of pedophilic sex offenders against children did not successfully inhibit their sexual arousal in the Suppression condition). There were very few variables that were associated with the ability to suppress. Higher Pedophilia Index scores in the Suppression condition predicted a greater likelihood of sexual recidivism among sex offenders (hazard ratio = 1.17, 95% CI [1.04, 1.32]), but the ability to suppress sexual arousal was not found to predict sexual recidivism. The current study highlights the importance of accounting for measurement error and found that, when doing so, most sex offenders against children are unable to successfully inhibit their sexual arousal to children when instructed to do so, and that the ability to suppress sexual arousal is not associated with recidivism.

Keywords Penile plethysmography · Phallometry · Sexual arousal · Pedophilia · Sex offenders · DSM-5

✉ Kelly M. Babchishin
Kelly.Babchishin@theroyal.ca

[1] The Royal's Institute of Mental Health Research, 1145 Carling Avenue, Ottawa, ON K1Z 7K4, Canada

[2] Waypoint Centre for Mental Health, Penetanguishene, ON, Canada

Introduction

Sexual offending against children can have far-reaching ramifications for the healthy development of victims (Beitchman et al., 1992; Browne & Finkelhor, 1986; Lindert et al., 2014; Paolucci, Genuis, & Violato, 2001). One factor influencing both the initiation and maintenance of sexual offending against children is pedophilia (Finkelhor, 1984; Hall & Hirchman, 1992; Seto, 2008), which is defined as a persistent and recurrent sexual interest in prepubescent children (American Psychiatric Association, 2013). Not all sex offenders against children have pedophilia, and not all pedophiles have committed sexual offenses against children; however, the prevalence of pedophilia among sex offenders is much higher than in the general population and other offender groups (Seto, 2008, 2013). Consequently, the identification of pedophilia is important in the comprehensive assessment and treatment of sex offenders against children; phallometry can be used to assess pedophilia.

Phallometry involves the physiological measurement of penile tumescence in response to various stimuli, such as audio descriptions of child and adult sexual interactions or pictures of nude or semi-clothed children and adults. Pedophilic sexual attraction is inferred from the amount of sexual arousal in response to child stimuli relative to adult stimuli. Phallometry is one of the best available measures for distinguishing sex offenders against children from non-sex offenders (Barsetti, Earls, Lalumière, & Bélanger, 1998; Freund, Watson, & Dickey, 1991; Marshall & Eccles, 1991; for a recent meta-analysis, see McPhail et al., 2017), as well as from sex offenders against adults and other men (Looman & Marshall, 2001; see also Lalumière, Harris, Quinsey, & Rice 2005).

In addition to pedophilia, the ability to suppress sexual arousal may be risk-relevant for sex offenders. Specifically, the inability to suppress sexual arousal has been associated with a greater like lihood of sexual risk taking among non-offending men (Bancroft, 2000; Bancroft et al., 2003; Janssen & Bancroft, 2007). Given

that sex offenders display greater sexual self-regulation problems than both non-sex offenders and non-offenders (Whitaker et al., 2008), it is likely that some would also have difficulty inhibiting their sexual arousal. In addition, sex offenders who are unable to suppress their arousal to child stimuli in a laboratory setting may present a heightened risk for sexual reoffending because these individuals may also have difficulty controlling their sexual arousal outside the laboratory (Marshall, 2007). Indeed, both the inability to control atypical sexual arousal (Clift, Rajlic, & Gretton, 2009) and problems with sexual self-regulation (Mann, Hanson, & Thornton, 2010) have been found to predict sexual recidivism. As such, it may be clinically useful to assess how well a sex offender is able to voluntarily suppress sexual arousal when instructed to do so. Relatedly, indicators of the ability to suppress sexual arousal might aid in identifying clients who could benefit from therapies designed to enhance voluntary control over sexual arousal to paraphilic and non-paraphilic stimuli (e.g., behavioral conditioning).

Ability to Suppress Sexual Arousal

Men differ in the ability to suppress their sexual arousal as assessed by phallometry (e.g., Abel, Blanchard, & Barlow, 1981; Adams, Mostinger, McAnulty, & Moore, 1992), with characteristics related to emotional detachment being related to a participant's ability to successfully inhibit sexual arousal (e.g., Mahoney & Strassberg, 1991). Specifically, men with greater emotional regulation have been found to be better able to inhibit emotional responses to sexual stimuli (Winters, Christoff, & Gorzalka, 2009). As such, correlates of emotional regulation, including older age (Gross et al., 1997), a greater number of friends (Lopes, Salovey, Côté, Beers, & Petty, 2005), higher intelligence (Singh & Singh, 2013), lower psychopathic traits (Malterer, Glass, & Newman, 2008), and absence of psychiatric disorders (Phillips, Ladouceur, & Drevets, 2008) might also be associated with a greater ability to suppress sexual arousal. In addition, men with a stronger sexual excitation system (i.e., high sex drive and greater sexual interest in the stimuli) have been found to have greater difficulty suppressing their sexual arousal than men with a weaker sexual excitation system (Bancroft & Janssen, 2000). As such, sex offenders against children with a high sex drive and greater evidence of pedophilic interest may have greater difficulty suppressing their sexual arousal to children during phallometric assessments than those who have lower sex drives and are not pedophilic.

Measurement of Change

Müller et al. (2014) examined the malleability of pedophilia as assessed by phallometry. The results showed a decrease in arousal to pedophilic stimuli and an increase in arousal to adult stimuli and so the authors concluded that pedophilia, as assessed by phallometry, was changeable. However, the measurement of change is inherently ambiguous. Observed change on a measure can be attributed to true change on the construct being assessed, to measurement error, or a mix of measurement error and true change. Mokros and Habermeyer (2016) presented statistical models that showed that regression to the mean (i.e., extreme scores becoming less extreme due to measurement error) and low reliability could explain Müller et al.'s (2014) findings.

The clinically significant change approach (further explained in the Method) offers one method of accounting for measurement errors in a two time-point design. Examining clinically meaningful change involves (a) defining a cutoff point to separate the dysfunctional from the functional population (in the current study, defined as the cutoff that separate pedophilic arousal and non-pedophilic arousal) and (b) evaluating the magnitude of the change to verify that the improvement exceeds the margin of measurement error. Such an approach provides a more conservative assessment of change than the common group-level approach, such as the one reported by Müller et al. (2014).

Current Study

The purpose of the current study was to: (1) examine differences in sexual arousal, as measured by phallometry, when sex offenders against children were asked to inhibit their sexual arousal (Suppression condition) compared to when the same offenders were asked to react normally to sexual stimuli (Normal condition), while accounting for measurement error; (2) determine the discriminative validity of sexual responses under the Suppression condition; (3) identify the correlates of ability to suppress sexual arousal; and (4) examine the extent to which ability to suppress sexual arousal predicts sexual and violent (including sexual) recidivism.

Method

Participants

Participants were 1612 men aged 18 years or older who underwent phallometric testing for child sexual interest between 1982 and 1999 (Mdn year = 1990) at a university teaching hospital in a Canadian city. Low responders ($n = 476$), defined as those who showed less than 2.5 mm change (Kuban, Barbaree, & Blanchard, 1999) to all of the sexual stimuli used in the Normal condition, irrespective of stimulus category, were removed from the analyses to increase the reliability and validity of phallometric data and change analyses, reducing the sample size to 1136. Of the remaining participants, there were also 125 low responders in the Suppression condition (less than 2.5 mm change in all the stimuli), but these men were kept in the analyses because removing these low responders would artificially reduce the number of participants who were successful in reducing their sexual arousal, as instructed.

Victim information was based on both self-report and official sources. Source of referral was available for 764 participants, with most (82%, $n = 625$) being referred from the criminal justice system (e.g., pre-trial or pre-sentence assessment), and the remainder

referred by family physicians (18%; $n = 139$). Some participants were referred by judges or defense lawyers as part of pre-trial evaluations; these offenders were classified based on their alleged offenses. Most participants self-reported or had an official charge for a sex offense against a child under age 12 ($n = 620$); the remaining participants self-reported a pedophilic sexual interest without a sex offense ($n = 45$), offended against a child between 12 and 16 years of age ($n = 246$), or offended against an adult (age 16 and older; $n = 100$). Eight solely had a non-contact sex offense, and 115 participants had no known sex offense, which included a group of non-offending community men who were recruited as a control group ($n = 83$). Offenders with non-contact sex offenses, self-reported pedophiles, or victims aged 12 to 16 were excluded from the discrimination and denial of offending analyses, but were included in the later analyses to determine the correlates of ability to suppress sexual arousal.

Participants were on average 36.1 years of age (SD = 11.8, $n = 1136$), with 11.2 years of education (SD = 3.4, $n = 973$). Less than half of the participants were single (41%, $n = 471$) at the time of the assessment. The remainder were married or in a common-law relationship (35%, $n = 397$), separated (12%, $n = 182$), divorced (8%, $n = 94$), or widowed (1%, $n = 9$).

Measures

Penile Plethysmography (PPG)

Phallometric testing in this study involved the physiological measurement of penile erection (i.e., penile plethysmography or PPG) in response to eight sexual activity vignettes recorded in a male monotone voice of approximately 2-min duration, with no visual stimuli (Abel et al., 1981). The child vignettes included two sets of stimuli, with three scenarios involving a man and a boy between the age of 8 and 10 (boy vignettes) and three scenarios involving a man and a girl between the age of 8 and 10 (girl vignettes). The three child scenarios differed in degree of participation: (1) the child initiates a sexual encounter, (2) the child appears to willingly participate in the initiation of a sexual encounter, and (3) a related child appears to agree to a sexual encounter. The two adult scenarios described a mutually consenting sexual encounter between a man and woman or a mutually consenting sexual encounter between two men. For the current study, we excluded the audiotapes that describe non-physical coercive, physical coercive, or sadistic sexual encounters with a child, as well as the non-sexual assault of a child. We also excluded a set of visual stimuli (i.e., 36 slides with four slides in each of the following categories: neutral, adult female, adult male, adolescent female, adolescent male, older female child, older male child, younger male child, and younger female child) typically presented in this setting's PPG assessment because the visual stimuli were not presented in both a Normal and a Suppression condition. Not all participants were given both child stimulus sets or both adult stimulus sets, because stimulus administration was determined by the attending psychiatrist and depended on the offenders' self-reported sexual interests and the gender of their victims based on self-report and official records; 302 were given both the female and male child set, 646 were only given the female child set, and 188 were only given the male child set.

Demographics

A number of demographic variables were transcribed from participants' assessment files, including age, level of education, marital status, victim characteristics, and number of sexual partners (defined as number of men or women with whom the participant had sexual intercourse since he was 16). Physician assessed level of denial (none vs. some) for each sex offense and degree of pedophilia (0–3, higher rating representing clients with a preference for physically immature individuals and established pattern of sexual behavior involving children) were also recorded.

Abel Becker Cognition Scale (ABCS; Abel, Becker, & Cunningham-Rathner, 1984)

The ABCS is a 29-item measure (5-point Likert scale), with scores ranging from 29 to 145 and higher scores indicating fewer attitudes tolerant of sexual offending toward children. The ABCS has good internal consistency (ranging from $\alpha = .74$ to .96 in subgroups; Tierney & McCabe, 2001). Data on the ABCS were not available for all participants because this measure was not used until April 1986.

Adversarial Sexual Beliefs (ASB; Burt, 1980)

The nine-item ASB (7-point Likert scale) assesses the belief that sexual relationships are exploitative or adversarial. Score ranged from 9 to 63, with higher scores indicating fewer adversarial sexual beliefs. It has been found to have adequate internal consistency ($\alpha = .80$; Burt, 1980). Data were not available for all participants because this measure was not used until November 1989.

Brief Psychiatric Rating Scale-Retardation Factor (BPRS; Overall & Gorham, 1962)

The BPRS is a widely used 18-item measure of psychiatric symptoms. BPRS Withdrawal-Retardation is comprised of the following subscales scored on a 7-point scale ranging from normal (0) to extremely ill (6): motor retardation, blunted affect, and emotional withdrawal (Overall & Pichot, 1967). Scores on the Withdrawal-Retardation factor can range from 0 to 18. BPRS Emotional Withdrawal assesses the individual ability to relate emotionally and includes withdrawal due to psychosis. Emotional Withdrawal scores can range from 0 to 6. The inter-rater reliability of these three items using interclass correlations ranged from .52 to .74, with Emotional Withdrawal being the lowest (Bell, Milstein, Beam-Goulet, Lysaker,

& Cicchetti, 1992). Correlations between these items and equivalent items on the Positive and Negative Syndrome Scale ranged from .43 to .89, and as with the inter-rater reliability, Emotional Withdrawal was the lowest correlation (Bell et al., 1992).

Derogatis Sexual Functional Inventory (DSFI; Derogatis & Melisaratos, 1979)

The DSFI assesses global sexual functioning and includes several subscales. Each subscale total score was transformed to *T* scores. The current study examined the following subscales: information (26, true/false items; higher scores indicating greater information regarding physiology, anatomy, and sexual relationship), experiences (24, yes/no items; higher scores indicated higher level of lifetime sexual experiences), sex drive (seven, 9-point Likert scale items; summary measure including sexual intercourse, masturbation, kissing and petting, sexual fantasy, and ideal frequency of intercourse, with higher scores indicating greater sex drive), attitudes (30, 5-point Likert scale items; a difference score representing liberalism–conservatism, higher scores representing greater liberalism), psychological symptoms (53, 5-point Likert items; higher scores indicating lower level of psychopathology), affects (40, 5-point Likert scale items measuring 20 positive affects and 20 negative affects; difference score representing ratio of positive–negative affect, with higher scores indicating greater positive affect), gender role definition (30, 5-point Likert scale adjectives; difference score representing masculinity–femininity, with higher scores reflecting understanding of gender stereotypes), sexual fantasy (20 items scored as present/absent; higher scores indicate a greater number of sexual fantasy themes), body image (15, 5-point Likert scale items; satisfaction with general and gender-specific attributes, with higher scores indicating a higher level of body appreciation), and sexual satisfaction (10 true/false items; higher scores represent higher level of sexual satisfaction). The DSFI subscales have been found to have adequate internal consistency (range from .61 [information] to .96 [attitude]) and test–retest reliability (range from .56 [information] to .97 [experiences] Derogatis, 1980). In the dataset used in the current study, internal consistency alpha ranged from .56 (sex drive) to .97 (psychological symptoms; Mdn = .69).

Screening Scale for Pedophilic Interest (SSPI; Seto & Lalumière, 2001)

The SSPI is a four-item scale constructed to assess pedophilia from child victim information commonly found in official files (Seto & Lalumière, 2001). Scores range from 0 to 5, with a higher score indicating greater likelihood of pedophilic sexual arousal. Items were coded as present or absent, with two points for having boy victims and one point for the presence of the remaining items: more than one child victim; any prepubescent victim (a child under the age of 12); and any extrafamilial victims. The SSPI has been found to accurately identify pedophilic sex offenders against children, as defined by phallometric responding, AUC = .70 (Seto & Lalumière, 2001). The SSPI also predicts sexual recidivism (Helmus, Ó Ciardha, & Seto, 2015; Seto, Harris, Rice, & Barbaree, 2004) and is significantly correlated with relative sexual arousal to children in both adult ($r = .28$; $p < .01$; Seto et al., 2004) and adolescent sex offenders against children ($r = .46$, $p < .005$; Seto, Murphy, Page, & Ennis, 2003). However, some studies (Canales, Olver & Wong, 2009; Moulden, Firestone, Kingston, & Bradford, 2009) have not found that the SSPI predicts sexual recidivism, with Moulden et al. (2009) using a sample that overlapped with the current study.

Greenberg Sexual Preference Visual Analog Scale (GSPVAS; Greenberg, 1991)

The GVAS is a self-report measure designed to assess sexual preference. For the 6 items of this scale, individuals rate which of two stimuli they find more sexually attractive on a sliding scale ranging from 0–100: adult woman versus adult man, young boy versus young girl (10 years old or younger), adult man versus young boy, young girl versus adult woman, adult woman versus young boy, and young girl versus adult man. These are then combined to give a score of attraction to each of the categories (woman, man, girl, and boy). To calculate the score, the three values for each stimuli set (e.g., the three adult woman scores) are summed, divided by 2, and then 50 is subtracted from the score. As such, scores can range from 50 (if all three scores being added are zero) to 100 (if all three scores being added are 100), with higher scores representing a greater degree of sexual preference. In a study with data for 49 healthy controls and 36 admitted sexual offenders against children, controls scored significantly higher than the offenders on the Adult Woman and significantly lower than the offenders on both Young Girl and Young Boy (Fedoroff, Madrigrano, Curry, Bradford, & Ahmed, 2004). Scores on this scale were available from October 1991.

Wechsler Adult Intelligence Scale (WAIS-R; Welchsler, 1981)

The WAIS-R is a widely used standardized measure for assessing general intelligence (IQ). The WAIS-R has large normative samples and adequate reliability (Welchsler, 1981). The IQ score in the current sample ranged from 50 to 149 ($M = 94$, SD = 16.4; $n = 436$). IQ testing was not carried out with all participants.

Recidivism

Official recidivism outcomes were obtained from the Canadian Police Information Centre's (CPIC) national database of criminal arrests and convictions, maintained by the Royal Canadian Mounted Police. Any sexual recidivism was defined as any charge or con-

viction for a sexual offense (e.g., invitation to sexual touching), whether it involved contact or not. Violent (including contact sexual offense) recidivism was defined as any charge or conviction for a non-sexually violent and/or contact sexual offense (e.g., assault, assault causing bodily harm). Follow-up ranged from 0.3 months to 23.1 years for sexual recidivism ($M = 10.0$ years) and for violent recidivism ($M = 9.2$ years).

Procedure

This was a retrospective chart review study, and there are missing data for all measures, because not all measures were used throughout the study time period and because of clinical discretion (e.g., which stimulus sets were administered). All participants signed an informed consent form at the time of their assessment, which allowed the use of their data for research purposes. The Research Ethics Board of the hospital where the data were collected approved the current study.

As part of their clinical assessments, patients in this setting first completed interviews with a psychiatrist who has access to police reports and previous charts in order to assess diagnoses and record basic demographic information. Patients would then complete the phallometry assessments and a questionnaire battery. Participants completed two PPG sessions while being seated alone in a testing room, one in which they were instructed to allow normal arousal to occur (Normal condition) followed by instructions to attempt to suppress their arousal (Suppression condition). The Normal instructions were "You are about to hear a series of audiotape descriptions, each lasting about 2 min. There's nothing for you to do except listen to each tape, relax, and move as little as possible." The Suppression instructions were "For the next series of tapes, continue to listen to them as you have been doing, but from now on, try and suppress your sexual arousal. Use whatever mental means you wish to prevent yourself from feeling any sexual excitement." The order of sessions was always the same, with the Normal condition preceding the Suppression condition. This order was repeated for each PPG assessments, so in cases where both the female and male stimuli sets were presented, the order was: female stimuli set with Normal instructions, female stimuli set with Suppression instructions, male stimuli set with Normal instructions, and male stimuli set with Suppression instructions. This order was decided by clinical staff because the Suppression condition was considered to be most informative if it followed the Normal condition.

Data Analysis

We analyzed sexual arousal in response to child stimuli relative to adult stimuli. Raw data from each condition (Suppression and Normal) were first transformed into z-scores (within participants) to eliminate differences between trials (completed separately for the Normal and the Suppression conditions). These z scores were then used to calculate a *Pedophilia Index*, which was calculated by subtracting the subject's largest z score for adult stimuli from the largest z score for child stimuli. Standardized scores were used because phallometric data were scored as percent of estimated full erection from 1983 to 1994, and as millimeters of change in penile circumference since 1994. Using z scores allowed us to combine these data. Additionally, Harris, Rice, Quinsey, Chaplin, and Earls (1992) recommended a transformation of the individual raw scores into z scores to eliminate differences in overall responsivity among participants, as well as among the phallometric assessments.

Discrimination

The area under the receiver operating characteristic curve (AUC) was used to examine the extent to which the Pedophilia Index could discriminate between offender types. The Delong method for computing the standard error of the difference (Delong, Delong, & Clarke-Pearson, 1988), available in the pROC procedure in the R statistical package (Robin et al., 2011), was used to test the extent to which the two conditions differed in their level of group discrimination. If the 95% confidence interval of the AUC difference between conditions included zero, the difference in discrimination between the two conditions was not statistically significant at $p < .05$. Discrimination analyses only included sexual offenders against children (with at least one victim under 12) and sex offenders against adults or controls (non-offenders without sexual interest in children).

Change Across Conditions

Clinically Significant Change Approach Change on the Pedophilia Index between the Normal and Suppression conditions was examined using the clinically significant change approach (Jacobson, Follette, & Revenstorf, 1984). This approach considers measurement error (whether the change between the Normal and Suppression condition is statistically reliable, based on the Reliable Change Index) as well as the meaningfulness of the change (e.g., whether someone who shows a pedophilic index score in the Normal condition shows a non-pedophilic index score in the Suppression condition). Using Jacobson et al.'s (1984) formula, we calculated cutoff scores (Cutoff C method) to distinguish non-pedophilic from pedophilic sexual arousal patterns based on the Pedophilia Index. The non-pedophilic group was defined as those with a Pedophilia Index under $-.25$ in the Normal condition ($M = -1.88$, SD $= 0.90$, $n = 520$), and the pedophilic group was defined as those with a Pedophilia Index above .25 in the Normal condition ($M = 1.71$, SD $= 0.87$, $n = 517$). Those with Pedophilia Index scores between $-.25$ and .25 in the Normal condition ($n = 99$; 9%) were deemed to be less discriminating between children and adults and were not classified for this analysis. The cutoff C indicates a midpoint between the typical Pedophilia Index seen in non-pedophilic and pedophilic groups. Participants falling on the non-pe-

dophilic side of the cutoff score were classified as non-pedophilic. The cutoff C was calculated as follows:

$$\text{Cutoff } C = \frac{s_0\overline{X_1} + s_1\overline{X_0}}{s_0 + s_1}$$

where s_0 is the standard deviation of the non-pedophilic group, s_1 is the standard deviation of the pedophilic group, $\overline{X_1}$ is the mean of the pedophilic group, and $\overline{X_0}$ is the mean of the non-pedophilic group. Cutoff C was −0.0526 in the current study.

The Reliable Change Index (RC) was calculated as follows:

$$\text{RC} = \frac{x_2 - x_1}{S_{\text{diff}}}$$

where x_2 is the participant's Pedophilia Index during the Suppression condition, x_1 is the participant's Pedophilia Index during the Normal condition, and

$$S_{\text{diff}} = \sqrt{2(\text{SEM})^2}$$

in which $\text{SEM} = s_1\sqrt{1 - r_{xx}}$, with r_{xx} being test–retest reliability in the current sample ($r = .508$, $p < .001$). The S_{diff} was 1.222 in the current study.

The clinically significant change approach assesses the reliability of the change between the Normal and Suppression conditions to classify participants into four categories. The first category (*Recovered*, using the approach's terms) identifies individuals scoring below the cutoff (i.e., non-pedophilic) in the Suppression condition, with reliable change between the Suppression and Normal conditions (i.e., RC greater than ±1.64). The second category (*Improved*) identifies participants who do not fall within the non-pedophilic range in the Suppression condition, but who showed reliable change toward a non-pedophilic arousal pattern. The third category (*Unchanged*) identifies participants who did not show reliable change between the two conditions. Finally, the fourth category (*Deteriorated*) identifies participants who showed reliable change, but by producing more sexual arousal to child stimuli in the Suppression condition.

z scores computed across the two conditions. As a secondary measure of change, we also computed a z score across the two conditions. This allowed for a group-level analysis on general ability to suppress sexual arousal. The z scores are computed across the two conditions and thus represent change in arousal *without* accounting for measurement error.

Ability to Suppress

A series of bivariate correlations were used to identify correlates of ability to suppress sexual arousal, focusing on those we considered to be relevant to sexual self-regulation. The correlation coefficients and their 84% confidence intervals are provided; non-overlapping confidence intervals indicate that differences between correlations coefficients were statistically significant at $p < .05$ (Tryon, 2001).

Recidivism

Cox regression analyses were used to examine the predictive validity of pedophilia indices calculated separately for the Normal and Suppression conditions and the clinically significant change classification for sexual and violent (including sexual) recidivism, taking time at risk into account. The dependent variable was time at risk (for non-recidivists) or "survival time" (for recidivists). Cox regression calculates Exp(B), which represents a hazard ratio that indicates a potential predictor's relationship with recidivism. For example, a hazard ratio of 1.00 indicates no predictive relationship, whereas a hazard ratio of 1.10 indicates that each one-point increase on the scale increases the hazard by 1.10, or 10%.

Results

Discrimination

Table 1 shows the ability of the Pedophilia Index to discriminate sex offenders against children from non-offenders and sex offenders against children in the two study conditions, as well as denial status of the sex offenses (admitters vs. deniers). The Pedophilia Index produced slightly lower discrimination in the Suppression condition (AUC = .62, 95% CI [.56, .67]) than in the Normal condition (AUC = .64, 95% CI [.59, .69]) when comparing sex offenders against children and non-offending men, but this difference did not reach statistical significance (difference = .02, 95% CI [−.04, .08]). The Suppression condition also provided similar discrimination between sex offenders against children who denied their offense and non-offending men than the Normal condition (difference = .01, 95% CI [−.06, .08]). The analyses were also conducted on a male PPG score (average of the boy categories minus adult male category) and a female PPG score (average of the girl categories minus adult female category) and produced similar findings (see Appendix 1).

The Pedophilia Index from the Suppression condition did not discriminate between sex offenders against children and sex offenders against adults, and the two conditions (Suppression and Normal) provided similar levels of discrimination, as indicated by nonsignificant Delong tests (see Table 1). Discrimination in both the Normal and Suppression conditions was higher for sex offenders against children who admitted their offense compared to those who denied some aspect of their sexual offense.

Ability to Suppress Sexual Arousal

Table 2 shows the (lack of) change in the Pedophilia Index between the two conditions for sex offenders against children, after accounting for fluctuation in scores that can be expected due to mea-

Table 1 Discriminative validity of the Pedophilia Index as indicated by AUC values

	AUC [95% CI]		Difference [95% CI]
	Suppression condition	Normal condition	Difference across conditions
Sex offenders against children ($n = 620$) versus non-offenders ($n = 115$)[a]	**.62 [.56, .67]**	**.64 [.59, .69]**	.022 [−.039, .084]
Deniers only ($n = 286$ vs. 115)	**.58 [.52, .64]**	**.59 [.53, .65]**	.008 [−.057, .075]
Admitters only ($n = 330$ vs. 115)	**.65 [.60, .71]**	**.68 [.63, .74]**	.032 [−.034, .098]
Sex offenders against children ($n = 604$) versus sex offenders against adults ($n = 100$)[a,b]	.52 [.46, .58]	.52 [.46, .58]	.003 [−.058, .064]
Deniers only ($n = 279$ vs. 32)	.47 [.37, .57]	.51 [.41, .60]	.039 [−.063, .140]
Admitters only ($n = 321$ vs. 67)	.56 [.48, .63]	.55 [.47, .62]	−.008 [−.086, .070]

Bolded AUC values are statistically significant at $p < .05$. Positive difference scores indicate that discrimination was higher in the Normal conditions than in the Suppression conditions. If the 95% confidence interval of the difference score does not include zero, then the difference in discrimination between the two conditions is statistically significant. Four sexual offenders against children and 1 sex offender against adults did not have a denial score

[a] Forty-five non-offending self-reported pedophiles, 246 hebephilic offenders, and 8 non-contact offenders were excluded from these analyses

[b] Sixteen offenders with both a victim under 12 and over 16 years of age were excluded from these analyses

surement error rather than true change. Based on the clinically significant change approach (Jacobson et al., 1984), most pedophilic sex offenders against children were unchanged; 83% did not show a change greater than what is expected based on measurement error). Change was also asymmetric: participants classified as pedophilic in the Normal conditions were more likely to show less pedophilic arousal in the Suppression condition (17%) as opposed to greater pedophilic arousal (0.3%), after accounting for measurement error. Specifically, 15% were found to be recovered (i.e., changed to a non-pedophilic score and displayed enough change to be considered reliable), 2% were improved (i.e., changed to a less pedophilic score, displayed enough change to be considered reliable, but did not reach a non-pedophilic score), and 0.3% deteriorated (i.e., changed to a more pedophilic score and displayed enough change to be considered reliable).

As expected, the ability to suppress sexual arousal differed between sex offenders against children who were pedophilic in the Normal condition and those classified as non-pedophilic in the Normal condition, $\chi^2(3) = 100.37, p < .001, n = 582$. A larger number of non-pedophilic sex offenders against children (20.4%) deteriorated in the Suppression condition (i.e., reliably changed to a more pedophilic score), and none of the non-pedophilic sex offenders against children were improved in the Suppression condition. Those who denied their offense showed similar change on the Pedophilia Index as those who admitted their offenses against children.

Table 3 presents participants' relative change on their pedophilia index scores between the Normal and Suppression conditions. These analyses, which do not account for measurement error, found a consistent reduction in PPG indices from the Normal to the Suppression condition for all groups (overall, deniers, and admitters), with small to moderate effect sizes.

Correlates of Ability to Suppress Sexual Arousal

The ability to suppress was defined as change on the Pedophilia Index between the Normal and Suppression condition, calculated so that positive scores indicated greater change toward a non-pedophilic phallometric index in the Suppression condition. The indicators for ability to suppress were classified into four domains: Markers of emotional regulation, markers of pedophilia, sexual history variables, and markers of high sexual excitation (cf. sex drive; see Table 4). Non-overlapping 84% confidence intervals indicated that differences between variables were statistically significant at $p < .05$. We ran the analyses separately for all participants and those who were classified as pedophilic during the Normal condition.

Very few indicators were found to predict ability to inhibit sexual arousal in the current sample. Among pedophilic offenders, greater self-reported sexual arousal to women and lower self-reported sexual arousal toward boys were significantly associated with a greater ability to suppress sexual arousal. However, the SSPI, emotional withdrawal as assessed by the BPRS, as well as other markers of pedophilia (e.g., victim age, number of child victims), were not associated with ability to suppress sexual arousal. Among pedophilic offenders, scores indicative of normal sexual functioning and greater sexual experience were also related to a greater ability to suppress sexual arousal.

Ability to Suppress and Recidivism

We also examine the extent to which the pedophilia indices and the ability to suppress sexual arousal was related to sexual and violent (including sexual) recidivism. The Pedophilia Index in the

Table 2 Individual change in Pedophilia Index as assessed by phallometric assessments

	Pedophilic in the Normal condition		Non-pedophilic in the Normal condition		Differences in classification between pedophilic and non-pedophilic offenders (χ^2)
	Status	No. of participants (%)	Status	No. of participants (%)	
Sex offenders against children					
	Recovered	47 (15.1%)	Recovered	2 (0.7%)	$\chi^2(3) = 100.37$, $p < .001$, $n = 582$
	Improved	5 (1.6%)	Improved	0 (0.0%)	
	No change	259 (83.0%)	No change	213 (78.9%)	
	Deteriorated	1 (0.3%)	Deteriorated	55 (20.4%)	
	$n = 312$		$n = 270$		
Admitters	Recovered	25 (13.9%)	Recovered	1 (0.8%)	$\chi^2(3) = 63.94$, $p < .001$, $n = 308$
	Improved	4 (2.2%)	Improved	0 (0.0%)	
	No change	151 (83.9%)	No change	95 (74.2%)	
	Deteriorated	0 (0.0%)	Deteriorated	32 (25.0%)	
	$n = 180$		$n = 128$		
Deniers	Recovered	21 (16.3%)	Recovered	1 (0.7%)	$\chi^2(3) = 39.44$, $p < .001$, $n = 270$
	Improved	1 (0.8%)	Improved	0 (0.0%)	
	No change	106 (82.2%)	No change	117 (83.0%)	
	Deteriorated	1 (0.8%)	Deteriorated	23 (16.3%)	
	$n = 129$		$n = 141$		
Differences in classification between admitters and deniers (χ^2)	$\chi^2(3) = 2.68$, $p = .443$, $n = 309$		$\chi^2(2) = 3.14$, $p = .209$, $n = 269$		

Recovered: Participants had reliable change toward more non-pedophilic responses in the Suppression condition and reached the cutoff for a non-pedophilic score in the Suppression condition. *Improved*: Participants did not pass the non-pedophilic cutoff in the Suppression condition, but showed reliable change toward more non-pedophilic responses in the Suppression condition. *Unchanged*: participants did not show reliable change between the two conditions. *Deteriorated*: participants showed reliable change, but toward more pedophilic responses in the Suppression condition

Table 3 Means and standard deviations of PPG z scores computed across the two conditions for sexual offenders against children

	Mean (SD)		Difference (SD)	Cohen $d_{\text{repeated measure}}$ [95% CI]
	Normal	Suppression		
Pedophilic in the Normal condition ($n = 312$)	1.60 (1.12)	0.71 (1.39)	0.88 (1.62)	**0.38 [0.26, 0.50]**
Deniers ($n = 129$)	1.52 (1.10)	0.62 (1.31)	0.90 (1.66)	**0.38 [0.18, 0.58]**
Admitters ($n = 180$)	1.66 (1.13)	0.80 (1.43)	0.86 (1.60)	**0.38 [0.53, 0.22]**
Non-pedophilic in the Normal condition ($n = 270$)	−1.86 (1.14)	−0.98 (1.76)	−0.88 (1.88)	**−0.33 [−0.20, −0.45]**
Deniers ($n = 140$)	−1.96 (1.12)	−1.17 (1.77)	−0.80 (1.90)	**−0.30 [−0.12, −0.46]**
Admitters ($n = 128$)	−1.74 (1.15)	−0.77 (1.73)	−0.97 (1.90)	**−0.36 [−0.18, −0.54]**

Five sexual offenders against children were missing denial status. Repeated measure Cohen's d corrected for correlation ($r = .508$) between the two assessments (Borenstein, Hedges, Higgins, & Rothstein, 2009; Cohen, 1988). Bolded values are statistically significant at $p < .05$

Suppression condition was found to significantly predict subsequent sexual recidivism (hazard ratio = 1.17, 95% CI [1.04, 1.32], $p = .010$, $n = 419$) and violent (including sexual) recidivism (hazard ratio = 1.12, 95% CI [1.02, 1.24], $p = .024$, $n = 419$; see Table 5). The predictive accuracy of the Pedophilia Index derived from the Normal condition only reached statistical significance for the pedophilic offender group (sexual recidivism hazard ratio = 1.49, 95% CI [1.04, 2.13], $p = .028$). Lastly, despite the clinically significant change classification (recovered, improved, unchanged, or deteriorated) having a positive beta for the complete sample and the pedophilic group, indicating that a greater difficulty of suppressing sexual arousal was associated with an

Table 4 Correlates of the ability to suppress

Variable ($n_{\text{complete sample}}/n_{\text{pedophilic}}$)	*r* [84% CI]	
	Complete sample	Pedophilic at the Normal condition
Markers of emotional regulation		
Age at assessment ($n = 1134/589$)	.02 [−.02, .06]	−.01 [.05, −.07]
IQ ($n = 436/214$)	−.04 [.06, −.14]	.02 [−.08, 12]
BPRS-retardation factor		
Motor retardation ($n = 696/391$)	.03 [−.02, .08]	.03 [−.04, .10]
Blunted affect ($n = 695/390$)	−.01 [.04, −.06]	.01 [−.06, .08]
Emotional withdrawal ($n = 701/393$)	.01 [−.04, .06]	.04 [−.03, .10]
Education ($n = 971/508$)	−.02 [.02, −.06]	.04 [−.06, .08]
Markers of pedophilia		
Degree of pedophilia ($n = 714/347$)	.05 [−.003, .10]	−.10 [−.02, −.17]†
SSPI scores ($n = 578/322$)	−.02 [.04, −.08]	−.04 [.04, −.12]
Greenberg Visual Analog Scale—Women ($n = 548/278$)	**.13 [.04, .22]**	**.13 [.05, .21]**
Greenberg Visual Analog Scale—Men ($n = 548/278$)	**−.11 [−.02, −.20]**	.04 [−.04, 12]
Greenberg Visual Analog Scale—Girls ($n = 548/278$)	.04 [−.05, .13]	−.07 [.02, −.15,]
Greenberg Visual Analog Scale—Boys ($n = 548/278$)	−.07 [.02, −.16]†	**−.12 [−.04, −.20]**
Youngest child victim age ($n = 673/365$)	.004 [−.05, .06]	.01 [−.06, .08]
Number of child victims ($n = 931/505$)	−.04 [.01, −.09]	−.08 [.11, −.14]
Sexual domain variables		
Abel Cognition scale ($n = 941/487$)	−.02 [.03, −.07]	.05 [−.01, .11]
Adversarial sexual beliefs ($n = 670/344$)	−.06 [−.01, −.44]	−.05 [.03, −.12]
DSFI		
Information ($n = 1097/577$)	−.03 [.01, −.07]	.03 [−.03, .09]
Attitudes ($n = 1095/576$)	−.01 [.03. −.05]	.02 [−.04, .08]
Psychological symptoms ($n = 1096/577$)	−.03 [.01, −.07]	−.01 [.05, −.07]
Affects ($n = 1096/577$)	−.06 [−.02, −.10]†	.01 [−.05, .07]
Gender role definition ($n = 1095/576$)	−.02 [.02, −.06]	−.04 [.02, −.10]
Body image ($n = 1091/573$)	.04 [−.003, .08]	.06 [.001, .12]
Sexual satisfaction ($n = 1093/571$)	−.03 [.01, −.07]	.03 [−.03, .09]
Markers of high sexual excitation system		
Number of past female sexual partners ($n = 817/416$)	.04 [−.01, .09]	.07 [.001, .14]
Frequency of masturbation in the past year ($n = 847/433$)	.03 [−.02, .08]	−.02 [.05, −.09]
DSFI		
Experiences ($n = 1097/577$)	.04 [−.002, .08]	**.12 [.06, .18]**
Sex drive ($n = 1096/577$)	.02 [−.02, .06]	.08 [.02, .14]†
Sexual fantasy ($n = 1096/577$)	.02 [−.02, .06]	.04 [−.02, .10]
Sexual functioning ($n = 1088/571$)	.004 [−.04, .05]	**.10 [.04, .16]**

Bolded correlations are statistically significant at $p < .05$. Negative correlations indicate that lower scores on the characteristics are associated with a greater ability to suppress sexual arousal. Non-overlapping 84% confidence intervals indicate that the two correlations significantly differed from one another at $p < .05$; $^{\dagger}p < .10$

BPRS Brief Psychiatric Rating Scale, *SSPI* Screening Scale for Pedophilic Interest, *DSFI* Derogatis Sexual Functioning Inventory

increased risk of sexual and violent recidivism, this association did not reach statistical significance. We also computed the PPG indices using the average arousal (rather than maximum) and using *z* scores computed across to the two conditions (rather than sep arately for the Suppression and Normal conditions); both resulted in similar findings (see Appendix 2).

Discussion

The aim of the current study was to examine the extent to which sex offenders against children could voluntarily suppress their sexual arousal, assess the influence of the ability to suppress on the discriminative and predictive validity of phallometric assess

Table 5 Predictive validity of the Pedophilia Index and ability to suppress

	Sexual recidivism			Violent (including sexual) recidivism		
	β [95% CI]	p	Reoffend/total N	β [95% CI]	p	Reoffend/total N
Complete sample						
Pedophilia Index (Normal condition)	1.06 [0.95, 1.19]	.274	89/419	1.04 [0.95, 1.14]	.450	129/419
Pedophilia Index (Suppression condition)	**1.17 [1.04, 1.32]**	**.010**	**89/419**	**1.12 [1.02, 1.24]**	**.024**	**129/419**
Clinically significant change classification[a]	1.11 [0.82, 1.49]	.513	89/419	1.08 [0.85, 1.38]	.539	129/419
Pedophilic group						
Pedophilia Index (Normal condition)	**1.49 [1.04, 2.13]**	**.028**	**35/179**	**1.34 [1.01, 1.78]**	**.046**	**53/179**
Pedophilia Index (Suppression condition)	**1.27 [1.06, 1.52]**	**.008**	**35/179**	1.15 [0.99, 1.32]	.068	53/179
Clinically significant change classification[a]	1.10 [0.59, 2.05]	.773	35/179	0.96 [0.59, 1.57]	.877	53/179
Child victim group						
Pedophilia Index (Normal condition)	1.09 [0.94, 1.26]	.253	54/251	1.03 [0.92, 1.16]	.603	83/251
Pedophilia Index (Suppression condition)	1.17 [1.00, 1.36]	.056	54/251	1.08 [0.96, 1.23]	.211	83/251
Clinically significant change classification[a]	0.92 [0.64, 1.33]	.668	54/251	1.01 [0.74, 1.37]	.952	83/251

Bolded values are statistically significant at $p < .05$. Higher change score indicates greater change toward a non-pedophilic PPG index in the Suppression condition

[a] Higher scores on the clinically significant change classification indicate greater difficulty at suppressing sexual arousal

ments, and identify the correlates of ability to suppress. We found that most (5 out of 6) sex offenders against children did not display reliable change after being instructed to suppress their sexual arousal during a penile plethysmography assessment. The clinically significant change approach used in the current study allows for the classification of participants into recovered, deteriorated, improved, or unchanged groups. Eighty-three percent of sex offenders against children identified as having pedophilic arousal in the Normal condition and 79% of sex offenders against children identified as having non-pedophilic arousal in the Normal condition were classified as unchanged in the Suppression condition. The ability to suppress sexual arousal was similar for sex offenders against children who denied their offenses compared to those who admitted their offenses. Lastly, although the trend was toward greater difficulty of suppressing sexual arousal to be associated with an increased risk of sexual recidivism, the association was not statistically significant.

A group-level approach, similar to Müller et al. (2014), found a small to moderate size effect in the ability to suppress sexual arousal, whereas the clinically significant change approach found only 15% of pedophilic offenders and 1% non-pedophilic offenders showed reliable change toward a less pedophilic arousal response. We also found that the observed change was asymmetrical. Participants classified as pedophilic in the Normal conditions were more likely to show less pedophilic arousal in the Suppression condition (17%) as opposed to greater pedophilic arousal (0.3%), after accounting for measurement error. In contrast, participants with non-pedophilic arousal patterns in the Normal condition were less likely to be classified as recovered or improved (1%) and more likely to score more pedophilic scores in the Suppression condition (20%; i.e., deteriorated), after accounting for measurement error. This trend could suggest a regression toward the mean, even after attempting to account for measurement error using the clinically significant change approach.

The divergence between the group-level and individual clinically significant change findings is unsurprising given that the group-level approach does not account for imperfect reliability or use meaningful cutoffs. The divergence in findings has been replicated in other studies. For example, whereas the group-level analyses in Nunes, Babchishin, and Cortoni's (2011) treatment outcome study revealed significant change in medium magnitude for the majority of the measures examined, the individual-level analyses revealed that only about one-third of participants showed reliable change and reached functional levels post-treatment.

Given that Mokros and Habermeyer (2016) found that low reliability and regression toward the mean can explain observed changes in phallometrically assessed sexual arousal in a sample drawn from the same setting as the current study (Müller et al., 2014) and that the current study found only a moderate ability for the PPG index to discriminate sexual offenders against children from non-offenders (e.g., AUC = .62 for the Suppression condition and .64 for Normal condition), it is important that change analyses account for measurement reliability. Of course, the clinically significant change approach (Jacobson et al., 1984) used in the current study is only as accurate as the reliability estimate from which the calculations are based. As summarized by Mokros and Habermeyer (2016), the reliability estimates of PPG assessments vary across settings. A better approach would be to use the observed data to model the reliability of phallometric assessment in that setting. Such an approach, however, requires more than two time-points (see Bryk & Raudenbush, 1987).

Most participants in the current study were referred by the criminal justice system (82%), and as such, it is possible if not likely that participants actively attempted to suppress their sexual arousal toward children, even during the Normal condition. It is also possible that participants did not try to suppress their arousal when instructed to. No systematic methods to detect faking other than camera observation of participants—to check whether they were paying attention to the visual stimuli—were used during these phallometric assessments. It is possible that testing in a less adversarial setting would increase observed differences between the Normal and Suppression conditions, because fewer participants might be motivated to suppress their sexual arousal in the Normal condition. It is also possible that changes between the Normal and Suppression conditions were not due to participants' ability to suppress, but an attenuation of response caused by a practice effect. We used clinical assessment data that were not counterbalanced. An experiment in which participants were assured their results would not affect their legal status, and the two conditions are counterbalanced, would provide a stronger test of these alternative explanations.

The current study did not find strong correlates of ability to suppress sexual arousal, with most correlations being small in magnitude. Greater sexual experience, scores indicative of normal sexual functioning, and less self-reported sexual interest in boys were positively associated with the ability to inhibit sexual arousal in pedophilic offenders. A higher degree of pedophilia as assessed by psychiatrists and higher sex drive approached statistical significance, with high sex drive associated with higher ability to suppress and a higher degree of pedophilia being associated with less ability to suppress.

Unlike Winters et al. (2009), the current study found some support for sexual experience and sexual drive being positively related to ability to suppress sexual arousal. Bancroft and Janssen (2000) found that men with a stronger sexual excitation system have greater difficulty suppressing their sexual arousal than men with a weaker sexual excitation system. We found inconsistent results: Although higher self-reported sexual interest in boys was associated with a lower ability of suppressing sexual arousal, higher sexual drive was associated with a greater rather than lower ability to suppress. Despite research suggesting that sexual self-regulation is related to emotional regulation and can affect physiological sexual response (e.g., Mahoney & Strassberg, 1991; Winters et al., 2009), emotional withdrawal and age were not related to the ability to suppress sexual arousal in the current study. Of note, the emotional withdrawal measure used in the current study was narrow in scope and further studies with alternative emotional regulation measures would be useful. In addition, our current sample only included a minority of men who could successful inhibit sexual arousal (e.g., only 15% of pedophilic sex offenders against children). The current study also did not include treatment information, and it is possible that treatment participation may have influence phallometrically assessed sexual arousal to children.

Conclusion

We found that the Suppression instructions provided similar group discrimination compared to the Normal instructions and that, when accounting for measurement error, most sexual offenders were not able (or did not attempt) to voluntarily suppress their sexual arousal during the Suppression condition. Additional research is needed to discover the factors underlying ability to suppress, which may assist identifying men who are able to voluntarily alter their sexual arousal. One prior study found that the inability to control atypical sexual arousal post-treatment was related to sexual recidivism among adolescent sex offenders (Clift et al., 2009). Although lower ability to suppress sexual arousal was positively associated to sexual recidivism in the current study, this association did not reach statistical significance. The Pedophilia Index obtained in the Suppression condition, however, predicted sexual recidivism (the Pedophilia Index obtained from the Normal condition only predicted recidivism for the pedophilic group of offenders). As such, it is possible that the Suppression condition may be useful in assessing recidivism risk. Replication of this effect is required.

Acknowledgements We thank Lisa Murphy and Aharon Zohar for their assistance collecting these data. Funding for this project was provided in part by the University of Ottawa Medical Research Fund and the Canadian Institutes of Health Research (CIHR Banting Fellowship).

Compliance with Ethical Standards

Conflict of interest All authors declare that they have no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Appendix 1

See Table 6.

Table 6 Discrimination of the Pedophilia Index as indicated by AUC values, index computed as average of stimuli

	Female PPG Index		Male PPG Index	
	Suppression condition	Normal condition	Suppression condition	Normal condition
	AUC [95% CI]			
Sex offenders against children versus non-offenders	**.58 [.51, .65]**	**.61 [.55, .67]**	**.69 [.62, .76]**	.56 [.48, .63]
Deniers only	.54 [.46, .62]	.55 [.48, .62]	**.67 [.59, .76]**	.54 [.46, .63]
Admitters only	**.61 [.54, .69]**	**.66 [.60, .73]**	**.71 [.63, .78]**	.56 [.48, .65]
Sex offenders against children versus rapists	.49 [.42, .55]	.51 [.45, .57]	**.63 [.54, .72]**	.51 [.42, .60]
Deniers only	.42 [.29, .54]	.44 [.32, .55]	.56 [.40, .71]	.54 [.37, .70]
Admitters only	.54 [.46, .61]	.57 [.49, .64]	**.67 [.56, .78]**	.50 [.39, .62]

Bolded values are statistically significant, $p < .05$. PPG female: $n_{\text{sex offenders against children}} = 561$ ($n_{\text{deniers}} = 262$; $n_{\text{admitters}} = 296$; 3 missing denial status), $n_{\text{nonsex offenders}} = 61$, $n_{\text{rapists}} = 86$ ($n_{\text{deniers}} = 24$; $n_{\text{admitters}} = 62$). Fourteen sex offenders against children (SOC) removed from SOC versus rapists analyses. PPG male: $n_{\text{sex offenders against children}} = 244$ ($n_{\text{deniers}} = 109$; $n_{\text{admitters}} = 133$; 2 missing denial status), $n_{\text{nonsex offenders}} = 60$, $n_{\text{rapists}} = 43$ ($n_{\text{deniers}} = 15$; $n_{\text{admitters}} = 27$; 1 missing denial status). Eight sex offenders against children (SOC) removed from SOC versus rapists analyses. Not all of our participants had the two sets of child categories: 302 were given both the female and male child set, 646 were only given the female child set, and 188 were only given the male child set. The female PPG index is the average of the female child stimuli minus the one adult female stimulus; the male PPG index is the average of the male child stimuli minus the one adult male stimulus, if provided

Appendix 2

See Tables 7 and 8.

Table 7 Predictive accuracy of the average PPG indices

	Sexual recidivism			Violent (including sexual) recidivism		
	β [95% CI]	p	Reoffend/total N	β [95% CI]	p	Reoffend/total N
Complete sample						
PPG mean female (Normal condition)	1.09 [0.93, 1.26]	.286	77/359	1.10 [0.97, 1.24]	.149	112/359
PPG mean female (Suppression condition)	**1.24 [1.06, 1.45]**	**.008**	**77/359**	**1.19 [1.05, 1.36]**	**.009**	**112/359**
PPG mean male (Normal condition)	1.18 [0.92, 1.50]	.195	45/188	1.08 [0.88, 1.33]	.452	61/188
PPG mean male (Suppression condition)	1.00 [0.79, 1.27]	.990	45/188	0.97 [0.79, 1.19]	.773	61/188
Pedophilic group						
PPG mean female (Normal condition)	**1.49 [1.05, 2.11]**	**.025**	**30/150**	**1.42 [1.06, 1.89]**	**.017**	**44/150**
PPG mean female (Suppression condition)	**1.38 [1.08, 1.77]**	**.010**	**30/150**	**1.26 [1.03, 1.55]**	**.025**	**44/150**
PPG mean male (Normal condition)	**1.52 [1.03, 2.24]**	**.036**	**17/84**	1.32 [0.95, 1.84]	.103	25/84
PPG mean male (Suppression condition)	1.38 [0.86, 2.22]	.181	17/84	1.16 [0.80, 1.67]	.439	25/84
Child victim group						
PPG mean female (Normal condition)	1.12 [0.93, 1.35]	.252	49/226	1.10 [0.95, 1.28]	.210	73/226
PPG mean female (Suppression condition)	**1.26 [1.03, 1.54]**	**.027**	**49/226**	1.17 [0.99, 1.38]	.060	73/226
PPG mean male (Normal condition)	1.17 [0.87, 1.57]	.295	29/105	1.03 [0.81, 1.31]	.825	43/105
PPG mean male (Suppression condition)	1.01 [0.75, 1.35]	.957	29/105	0.95 [0.75, 1.20]	.666	43/105

Bolded values are statistically significant, $p < .05$

Table 8 Recidivism analyses based on *z* scores computed across the two conditions

	Sexual recidivism			Violent (including sexual) recidivism		
	β [95% CI]	*p*	Reoffend/total *N*	β [95% CI]	*p*	Reoffend/total *N*
Complete sample						
PPG average across (Normal condition)	1.05 [0.95, 1.17]	.309	89/419	1.03 [0.95, 1.12]	.430	129/419
PPG average across (Suppression condition)	**1.25 [1.09, 1.42]**	**.001**	**89/419**	**1.16 [1.04, 1.29]**	**.008**	**129/419**
Pedophilic group						
PPG average across (Normal condition)	1.28 [0.96, 1.70]	.096	35/179	1.25 [0.99, 1.57]	.064	53/179
PPG average across (Suppression condition)	**1.41 [1.14, 1.73]**	**.001**	**35/179**	**1.19 [1.01, 1.40]**	**.034**	**53/179**
Child victim group						
PPG average across (Normal condition)	1.07 [0.94, 1.22]	.318	54/251	1.03 [0.93, 1.15]	.590	83/251
PPG average across (Suppression condition)	**1.20 [1.02, 1.42]**	**.031**	**54/251**	1.09 [0.96, 1.25]	.175	83/251

Bolded values are statistically significant, $p < .05$

References

Abel, G. G., Becker, J., & Cunningham-Rathner, J. (1984). Complications, consent, and cognitions in sex between children and adults. *International Journal of Law and Psychiatry, 7,* 89–103. doi:10.1016/0160-2527(84)90008-6.

Abel, G. G., Blanchard, E. B., & Barlow, D. H. (1981). Measurement of sexual arousal in several paraphilias: The effect of stimuli modality, instructional set and stimulus content on the objective. *Behaviour Research and Therapy, 19,* 25–33. doi:10.1016/0005-7967(81)90109-1.

Adams, H. E., Motsinger, P., McAnulty, R. D., & Moore, A. L. (1992). Voluntary control of penile tumescence among homosexual and heterosexual subjects. *Archives of Sexual Behavior, 21,* 17–31. doi:10.1007/bf01542714.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Press.

Bancroft, J. (2000). Individual differences in sexual risk taking by men: A psychosociobiological approach. In J. Bancroft (Ed.), *The role of theory in sex research* (pp. 177–212). Bloomington: Indiana University Press.

Bancroft, J., & Janssen, E. (2000). The dual control model of male sexual response: A theoretical approach to centrally mediated erectile dysfunction. *Neuroscience and Biobehavioral Reviews, 24,* 571–579. doi:10.1016/s0149-7634(00)00024-5.

Bancroft, J., Janssen, E., Strong, D., Carnes, L., Vukadinovic, Z., & Long, J. S. (2003). Sexual risk taking in gay men: The relevance of sexual arousability, mood and sensation seeking. *Archives of Sexual Behavior, 32,* 555–572. doi:10.1080/00224490409552226.

Barsetti, I., Earls, C. M., Lalumière, M. L., & Bélanger, N. (1998). The differentiation of intrafamilial and extrafamilial heterosexual child molesters. *Journal of Interpersonal Violence, 13,* 275–286. doi:10.1177/088626098013002007.

Beitchman, J. H., Zucker, K. J., Hood, J. E., DaCosta, G. A., Akman, D., & Cassavia, E. (1992). A review of the long-term effects of child sexual abuse. *Child Abuse and Neglect, 16,* 101–118. doi:10.1016/0145-2134(92)90011-f.

Bell, M., Milstein, R., Beam-Goulet, J., Lysaker, P., & Cicchetti, D. (1992). The Positive and Negative Syndrome Scale and the Brief Psychiatric Rating Scale: Reliability, comparability, and predictive validity. *Journal of Nervous and Mental Disease, 180*(11), 723–728. doi:10.1097/00005053-199211000-00007.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, England: John Wiley & Sons Ltd.

Browne, A., & Finkelhor, D. (1986). Impact of child sexual abuse: A review of the research. *Psychological Bulletin, 99,* 66–77. doi:10.1037/0033-2909.99.1.66.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101,* 147–158. doi:10.1037/0033-2909.101.1.147.

Burt, M. (1980). Cultural myths and supports for rape. *Journal of Personality and Social Psychology, 38,* 217–230. doi:10.1037/0022-3514.38.2.217.

Canales, D. D., Olver, M. E., & Wong, S. C. P. (2009). Construct validity of the Violence Risk Scale—Sexual Offender Version for measuring sexual deviance. *Sexual Abuse: A Journal of Research and Treatment, 21,* 474–492. doi:10.1177/1079063209344990.

Clift, R. J., Rajlic, G., & Gretton, H. M. (2009). Discriminative and predictive validity of the penile plethysmograph in adolescent sex offenders. *Sexual Abuse: A Journal of Research and Treatment, 21,* 335–362. doi:10.1177/1079063209338491.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics, 44,* 837–845. doi:10.2307/2531595.

Derogatis, L. R. (1980). Psychological assessment of psychosexual functioning. *Psychiatric Clinics of North America, 3,* 113–131.

Derogatis, L. R., & Melisaratos, N. (1979). The DSFI: A multidimensional measure of sexual functioning. *Journal of Sex and Marital Therapy, 5,* 244–281. doi:10.1080/00926237908403732.

Fedoroff, J. P., Madrigrano, G., Curry, S., Bradford, J., & Ahmed, A. G. (2004). *Evaluation of the Sexual Behaviour Clinic's assessment protocol for child molesters*. Unpublished raw data.

Finkelhor, D. (1984). *Child sexual abuse: New theory and research*. New York: Free Press.

Freund, K., Watson, R. J., & Dickey, R. (1991). Sex offenses against female children perpetrated by men who are not pedophiles. *Journal of Sex Research, 28,* 409–423. doi:10.1080/00224499109551616.

Greenberg, D. M. (1991). *Greenberg's Sexual Preference Visual Analogue Scale (GSPVAS)*. Ottawa, ON: Forensic Program, Royal Ottawa Hospital.

Gross, J. J., Carstensen, L. L., Pasupathi, M., Tsai, J., Götestam Skorpen, C., & Hsu, A. Y. (1997). Emotion and aging: Experience, expression, and control. *Psychology and Aging, 12,* 590–599. doi:10.1037/0882-7974.12.4.590.

Hall, G. C. N., & Hirschman, R. (1992). Sexual aggression against children: A conceptual perspective of etiology. *Criminal Justice and Behavior, 19,* 8–23. doi:10.1177/0093854892019001003.

Harris, G. T., Rice, M. E., Quinsey, V. L., Chaplin, T. C., & Earls, C. (1992). Maximizing the discriminant validity of phallometric assessment data. *Psychological Assessment, 4,* 502–511. doi:10.1037/1040-3590.4.4.502.

Helmus, L., Ó Ciardha, C., & Seto, M. C. (2015). The Screening Scale for Pedophilic Interests (SSPI): Construct, predictive, and incremental

validity. *Law and Human Behavior, 39,* 35–43. doi:10.1037/lhb0000099.

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15,* 336–352. doi:10.1016/s0005-7894(84)80002-7.

Janssen, E., & Bancroft, J. (2007). The dual control model: The role of sexual inhibition and excitation in sexual arousal and behavior. In E. Janssen (Ed.), *The psychophysiology of sex* (pp. 197–222). Bloomington, IN: Indiana University Press.

Kuban, M., Barbaree, H. E., & Blanchard, R. (1999). A comparison of volume and circumference phallometry: Response magnitude and method agreement. *Archives of Sexual Behavior, 28,* 345–359. doi:10.1023/A:1018700813140.

Lalumière, M. L., Harris, G. T., Quinsey, V. L., & Rice, M. E. (2005). *The causes of rape: Understanding individual differences in male propensity for sexual aggression*. Washington, DC: American Psychological Association.

Lindert, J., von Ehrenstein, O. S., Grashow, R., Gal, G., Braehler, E., & Weisskopf, M. G. (2014). Sexual and physical abuse in childhood is associated with depression and anxiety over the life course: Systematic review and meta-analysis. *International Journal of Public Health, 59,* 359–372. doi:10.1007/s00038-013-0519-5.

Looman, J., & Marshall, W. L. (2001). Phallometric assessments designed to detect arousal to children: The responses of rapists and child molesters. *Sexual Abuse: A Journal of Research and Treatment, 13,* 3–13. doi:10.1177/107906320101300102.

Lopes, P. N., Salovey, P., Côté, S., Beers, M., & Petty, R. E. (2005). Emotion regulation abilities and the quality of social interaction. *Emotion, 5,* 113–118. doi:10.1037/1528-3542.5.1.113.

Mahoney, J. M., & Strassberg, D. S. (1991). Voluntary control of male sexual arousal. *Archives of Sexual Behavior, 20,* 1–16. doi:10.1007/bf01543003.

Malterer, M. B., Glass, S. J., & Newman, J. P. (2008). Psychopathy and trait emotional intelligence. *Personality and Individual Differences, 44,* 735–745. doi:10.1016/j.paid.2007.10.007.

Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse: A Journal of Research and Treatment, 22,* 191–217. doi:10.1177/1079063210366039.

Marshall, W. L. (2007). Covert association: A case demonstration with a child molester. *Clinical Case Studies, 6,* 218–231. doi:10.1177/1534650105280329.

Marshall, W. L., & Eccles, A. (1991). Issues in clinical practice with sex offenders. *Journal of Interpersonal Violence, 6,* 68–93. doi:10.1177/088626091006001006.

McPhail, I. V., Hermann, C. A., Fernane, S., Fernandez, Y. M., Nunes, K. L., & Cantor, J. M. (2017). *Validity in phallometric tests for pedohebephilic interests: A meta-analytic review*. Manuscript submitted for publication.

Mokros, A., & Habermeyer, E. (2016). Regression to the mean mimicking changes in sexual arousal to child stimuli in pedophiles. *Archives of Sexual Behavior, 45,* 1863–1867. doi:10.1007/s10508-015-0652-8.

Moulden, H. M., Firestone, P., Kingston, D., & Bradford, J. (2009). Recidivism in pedophiles: An investigation using different diagnostic methods. *Journal of Forensic Psychiatry & Psychology, 20,* 680–701. doi:10.1080/14789940903174055.

Müller, K., Curry, S., Ranger, R., Briken, P., Bradford, J., & Fedoroff, J. (2014). Changes in sexual arousal as measured by penile plethysmography in men with pedophilic sexual interest. *Journal of Sexual Medicine, 11,* 1221–1229. doi:10.1111/jsm.12488.

Nunes, K. L., Babchishin, K. M., & Cortoni, F. (2011). Measuring treatment change in sex offenders: Clinical and statistical significance. *Criminal Justice and Behavior, 38,* 157–173. doi:10.1177/0093854810391054.

Overall, J. E., & Gorham, D. R. (1962). The Brief Psychiatric Rating Scale. *Psychological Reports, 10,* 799–812. doi:10.1037/t01554-000.

Overall, J. E., & Pichot, P. (1967). Major psychiatric disorders: A four-dimensional model. *Archives of General Psychiatry, 16,* 146–151. doi:10.1001/archpsyc.1967.01730200014003.

Paolucci, E. O., Genuis, M. L., & Violato, C. (2001). A meta-analysis of the published research on the effects of child sexual abuse. *Journal of Psychology, 135,* 17–36. doi:10.1080/00223980109603677.

Phillips, M. L., Ladouceur, C. D., & Drevets, W. C. (2008). A neural model of voluntary and automatic emotion regulation: Implications for understanding the pathophysiology and neurodevelopment of bipolar disorder. *Molecular Psychiatry, 13,* 833–857. doi:10.1038/mp.2008.82.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics, 12,* 77. doi:10.1186/1471-2105-12-77.

Seto, M. C. (2008). *Pedophilia and sexual offending against children: Theory, assessment, and intervention*. Washington, DC: American Psychological Association.

Seto, M. C. (2013). *Internet sex offenders*. Washington, DC: American Psychological Association.

Seto, M. C., Harris, G. T., Rice, M. E., & Barbaree, H. E. (2004). The Screening Scale for Pedophilic Interests predicts recidivism among adult sex offenders with child victims. *Archives of Sexual Behavior, 33,* 455–466. doi:10.1023/b:aseb.0000037426.55935.9c.

Seto, M. C., & Lalumière, M. L. (2001). A brief screening scale to identify pedophilic interests among child molesters. *Sexual Abuse: A Journal of Research and Treatment, 13,* 15–25. doi:10.1177/107906320101300103.

Seto, M. C., Murphy, W. D., Page, J., & Ennis, L. (2003). Detecting anomalous sexual interests in juvenile sex offenders. *Annals of the New York Academy of Sciences, 989,* 118–130. doi:10.1111/j.1749-6632.2003.tb07298.x.

Singh, P., & Singh, N. (2013). Difficulties in emotion regulation: A barrier to academic motivation and performance. *Journal of the Indian Academy of Applied Psychology, 39,* 289–297.

Tierney, D. W., & McCabe, M. P. (2001). An evaluation of self-report measures of cognitive distortions and empathy among Australian sex offenders. *Archives of Sexual Behavior, 30,* 495–519. doi:10.1177/1524838001002003004.

Tryon, W. W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: An integrated alternative method of conducting null hypothesis statistical tests. *Psychological Methods, 6,* 371–386. doi:10.1037/1082-989x.6.4.371.

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale—Revised*. New York: Psychological Corporation.

Whitaker, D. J., Le, B., Hanson, R. K., Baker, C. K., McMahon, P. M., Ryan, G., … Rice, D. D. (2008). Risk factors for the perpetration of child sexual abuse: A review and meta-analysis. *Child Abuse and Neglect, 32,* 529–548. doi:10.1016/j.chiabu.2007.08.005.

Winters, J., Christoff, K., & Gorzalka, B. B. (2009). Conscious regulation of sexual arousal in men. *Journal of Sex Research, 46,* 330–343. doi:10.1080/00224490902754103.