

Commentary: Phallometry in Court—Problems Outweigh Benefits

Roy O'Shaughnessy, MD

Phallometry evaluations are of great value in sex offender treatment programs, in assessing for a deviant drive related to criminal offenses and in monitoring treatment results, especially in connection with cognitive behavioral techniques that teach skills for reducing deviant sexual interests. The research on phallometry, however, is fraught with methodological problems that limit its utility in settings such as court procedures where there is a strong self-interest in producing results that suggest the absence of deviant sexual drives. The lack of consensus in methodology and scoring, the difficulty encountered in “nonadmitters,” the ability to dissimulate (“fake good”) on the assessments, and the lack of good specificity and sensitivity data limit the use of such procedures in any setting that could affect length of sentence or determination of civil commitment.

J Am Acad Psychiatry Law 43:154–8, 2015

Purcell *et al.*¹ have provided a welcome review of the use of a controversial test methodology in the court systems in Canadian jurisprudence. The article is both fascinating and disturbing, as they clearly document the gradually increasing use of a controversial test that has some utility in clinical settings with very questionable validity and reliability when used in legal settings. In their article, the authors note that the Supreme Court of Canada has rejected the use of phallometry evaluations in the guilt phase of proceedings but that courts have increasingly accepted it as part of the presentence hearings for convicted sex offenders. In particular, the courts have focused on the potential use of phallometry results in determining risk for reoffense and in prospects for treatment and rehabilitation. The authors note within Canadian jurisprudence that there is a strong emphasis on rehabilitation as part of the sentencing that may not be as firmly stressed in other jurisdictions. Even in jurisdictions that strongly emphasize rehabilitation and treatment of offenders, however, they document that the results of phallometry assessments may have significant implications for the length of sentence given for a sexual offense. Specifically, courts may see the offender's acceptance of testing as evidence of

either remorse or willingness to undergo therapy and, conversely, refusal for such testing as evidence of lack of remorse or treatability. The latter, of course, has significant implications for potentially longer sentences. These are particularly relevant in hearings regarding civil commitment proceedings or Dangerous Offender applications in which a convicted sexual offender may face indeterminate sentencing if prosecutors prove they continue to be at high risk of reoffending without effective treatment.

The authors raise considerable concerns regarding the use of phallometry evaluations. Partly, these reflect methodological deficiencies and in particular the great variations in different settings regarding how tests are conducted, the interpretation of the results, and the strength of any predictive value that evidence of deviant sexual arousal may have regarding future offending. They also raise significant questions about the adequacy of informed consent before the procedures and, in particular, comment on the degree of coercion or pressure placed on offenders to have such assessments, bearing in mind the previous comments.

The use of these evaluations, even in clinical settings, is controversial. Phallometry is a procedure that many have argued is invasive and, at the least, affronts dignity. The stimuli used generally depict visual, audiovisual, or auditory descriptions of overt sexual acts. In American jurisdictions, many of the images, especially of underaged children, that are used to measure pedophilic fantasies are considered

Dr. O'Shaughnessy is Clinical Professor of Psychiatry, Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada. Address correspondence to: Roy O'Shaughnessy, MD, Broadway Forensic Group, 101-1090 West Eighth Avenue, Vancouver, BC, Canada. E-mail: royjoshoughnessy@telus.net.

Disclosures of financial or other potential conflicts of interest: None.

illegal. The evaluation is all the more controversial when there is evidence that offenders are coerced into taking the procedure either to gain admittance to sexual offender treatment programs or as part of an evaluation, especially in a presentence stage of proceedings. The use of phallometry in such settings will only serve to increase the controversy and possibly to detract from the usefulness of phallometry within specialized treatment settings.

Utility of Phallometry in Treatment Settings

Since phallometry was initiated by Freund² in Toronto, it has been well accepted by psychiatrists specializing in the assessment and treatment of sex offenders. It is a tool to assist in the assessment of sexual drive and in particular deviant drives toward children and sexually aggressive behavior. Similar to any test, phallometry is at its greatest utility when applied to a cooperative patient in a collaborative relationship with a treatment team focused on the goal of helping the offender control or reduce any deviant drive that in turn is likely to contribute to the person's ability to refrain from further offending. As a result, both the offender and general society receive considerable benefits. In such settings, there is often substantial pressure on offenders to partake in the assessment and the treatment process. Generally speaking, it is not overtly coercive, and consent is voluntary. As part of a pretreatment assessment, phallometry may demonstrate the presence of deviant sexual arousal that can then become a target for behavioral interventions commonly offered in sex offender treatment programs. In addition, a thorough assessment may reveal the presence of other paraphilic interests that have not been disclosed or known, which may be the foci of treatment. The graphic demonstration of deviant arousal is also very helpful in challenging individuals who maintain some degree of denial as to whether they have underlying deviant sexual interests and in turn can lead to improvement in collaboration. Phallometry also has value in enabling repeat assessments during or at the completion of portions of the treatment program. As part of many sex offender treatment programs, the offenders are taught various cognitive techniques to control or reduce deviant drive. Their abilities to implement these can be evaluated through subsequent phallometry examinations. For those offenders who are unable to suppress a

deviant drive, despite appropriate cognitive behavioral therapy, it raises the opportunity to discuss further interventions such as medication.

Unfortunately, many sex offenders, even those willing to participate in sex offender treatment programs, are not as fully cooperative and collaborative with the treatment team as the somewhat idealized description noted above. Nonetheless, phallometry is still useful, especially with "admitters" of sexual offenses and deviant arousal patterns. As noted below, however, the problem of interpretation of the results, especially in "nonadmitters," is problematic, even in postsentencing treatment settings. It is a much greater problem at the presentence level of assessment because of the motivation to appear less deviant.

Methodologic Concerns in Phallometry Assessment

Fedoroff *et al.*² compiled a thorough review of the methodological limitations in the use of phallometry evaluations. They noted multiple areas of concern that are relevant to whether phallometry should be used in a forensic setting, as opposed to a treatment setting. They noted from the outset that there has been inconsistency across various laboratories, in terms of stimulus sets used, how data are interpreted, and how scoring is conducted. In some laboratories the minimum response requirements are very low, raising significant questions regarding their validity and reliability. Laboratories score the results differently (e.g., measuring area under the curve versus *z*-scores). Some laboratories place differential weight on absolute scores versus relative scores for penile responses. Relative scoring evaluates the amount of arousal in response to the deviant stimuli compared with the response to consensual adult stimuli in the same individual, whereas absolute scores simply look at response levels to any stimuli. Laboratories also varied on the interpretation of levels of significance in response values. It is not clear what percentage of full erection to a deviant sexual response should be considered clinically significant. In conservative laboratories, it is considered significant if the deviant response is equal to or greater than the nondeviant response, whereas others opine that scores that are still below response to normal stimuli may be pathological.

Fedoroff *et al.*² reported that various laboratories use different stimulus sets that in turn generate different

degrees of response. Videos generally produced the greatest erectile response, but is that meaningful in terms of clinical significance compared with less robust responses to stimuli such as audiotapes or slides?

Complicating the concerns about validity further is the fact that sex offenders are by nature heterogeneous. Not all child molesters have deviant arousal responses to children. Likewise, individuals who commit rape are even more heterogeneous, such that some admitted rapists would not show any deviant arousal. On other studies many nonrapists or normal individuals responded to rape stimuli with increased sexual arousal. There are particular problems in interpreting normal responses or nonresponses where the offender does not respond with significant tumescence to any stimuli.

More recently, there has been laboratory research on individuals with sexually aggressive behavior that has somewhat improved the ability to discriminate among sadists, rapists, and nonrapists. Harris *et al.*³ introduced a new stimulus set in comparing 12 rapists and 14 control subjects. They found that the best discrimination was the response to stimuli focusing on nonconsent or resistance. Seto *et al.*⁴ studied three groups of men in nonforensic and noncriminal settings, including 18 self-identified sexual sadists, 23 nonsadists, and 22 men with some sadistic interests. They used new stimuli in an attempt to separate arousal responses to violence from those to nonconsensual sexual activity. They noted that there was no reason for any of the volunteers to “fake good,” and they were all quite cooperative. There was close association between their subjective ratings toward their sexually sadistic interests with the phallometry evaluation showing arousal to sadistic themes. Sadists showed a significant difference on the violence index but not on the nonconsensual index.

Harris *et al.* concluded that these two studies did in fact provide increasing evidence that phallometry could be very effective in separating sexually aggressive (rapists), from sexually sadistic, from normal males in volunteer populations of admitters. Their work again supports the utility of phallometry evaluations in a treatment context where a collaborative relationship exists.

Validity: Sensitivity versus Specificity

Federoff *et al.*,² in a review of existing studies on the validity of phallometry evaluation, noted appropriately that most studies in fact did not provide data

regarding sensitivity and specificity. Further, most studies eliminate “nonresponders” (i.e., those who do not show significant levels of erectile response to any stimuli), raising even further questions regarding the sensitivity of phallometry evaluations. Blanchard *et al.*⁵ reported sensitivity and specificity for their results, noting that phallometry had a sensitivity of only 0.46 but a specificity of 0.92. They argued that it was likely that the sensitivity was in fact higher, because not all sex offenders would have deviant arousal in any event. Other studies have shown similar results, indicating that sensitivity is consistently less than specificity in phallometry evaluations. Given such low sensitivity, it is challenging for the examiner to know what to make of a negative phallometry evaluation (i.e., one in which there is no clear evidence of deviant sexual arousal).

Admitters Versus Nonadmitters

Lanyon and Thomas⁶ reviewed the question of deception in sex offender assessments, including phallometry evaluations. They appropriately note the discrepancy between admitters, those who acknowledge that they have committed a sexual offense, have a deviant drive, or both, who are substantially different from nonadmitters, who may deny one or both aspects. Admitters provide the necessary information for appropriate risk assessment, treatment planning, and monitoring, whereas nonadmitters provide little information that can be seen as reliable. Lanyon and Thomas argued that as a result we need much more carefully validated assessments of nonadmitters. Unfortunately, they note that, in the existing research in phallometry assessments, there has been little actual validation of assessment methods in nonadmitters.

Lanyon and Thomas ask two key questions as part of the focus of the textbook, *Clinical Assessment of Malingering and Deception*.⁷ These are:

How accurate is the instrument for its intended assessment purposes?

Can it be deceived?

They note methodological deficiencies similar to those raised by Federoff *et al.*⁶ They also argue that there is no study that can clearly affirm the presence or absence of deviant interest in nonadmitters given the inherent limitations in sensitivity and specificity in research of nonadmitters. Given the heterogeneity

of sex offender populations, you cannot tell if a non-admitter who does not show deviant arousal on phallometry is actually free of deviant arousal or simply "faking good."

The Faking Good Problem

Fedoroff *et al.*² document the multiple physical and psychological methods of obfuscation used by offenders who want to fake their responses on phallometry assessments. In the literature, there have been descriptions of offenders attempting to create arousal to normal stimuli or to diminish their arousal to deviant stimuli. Some of the attempts are quite crude (e.g., attempting to manipulate the gauge, which can generally be determined by the technician looking at the graph). Other techniques, such as "pumping" in which the examinee contracts perineal muscles with the hope of becoming tumescent to normal stimuli can often be picked up as little spikes on the graphic representation. Some offenders use aversion of eyes from the stimuli, although this can generally be evaluated by having the technician monitor the offender or by using more sophisticated visual measurement technology. In a recent study, Trottier *et al.*⁸ used eye tracking to identify individuals trying to falsify responses on phallometry evaluations, with some success. Techniques to evaluate deception are challenging and increase the complexity of the examination procedure. They also are of questionable value, as many offenders are simply able to control arousal responses by using fantasy during the stimuli. Hall *et al.*⁹ asked 122 sex offenders in treatment to inhibit their sexual arousal on phallometry evaluation. Eighty per cent were able to do so without resorting to demanding physical activities. Simple techniques, such as having an aversive fantasy while watching deviant imagery, are effective in many individuals in reducing their arousal. Likewise, if an offender has little, if any, arousal to adult consensual sexual activities, he may resort to his own deviant fantasies during such stimuli to produce arousal for the desired stimulus.

Summary and Conclusions

Phallometry is a useful tool in treatment settings, especially with voluntary patients who have been able to forge a collaborative relationship with the treatment team. It is helpful in both the assessment and the follow-up phases of treatment to measure poten-

tial deviant arousal that may be present and the ability of the patient to suppress such arousal after a treatment program. It is particularly relevant in those who admit to the offensive behavior and to deviant arousal but is considerably less effective in those who are nonadmitters. There are substantial methodological limitations inherent in the technology, such that there is very limited information as to the validity, particularly the sensitivity and specificity of the test, especially in nonadmitters. It is relatively easy to dissimulate on the phallometry evaluations. A negative result offers very little in the way of meaningful information, as we simply cannot tell whether it is a true negative, meaning that the person does not have deviant arousal, or that he has dissimulated.

There are justifiable concerns about using phallometry evaluation results in any court proceeding, even at a presentence level. Although all courts have rejected the use of phallometry at the trial phase, there is substantial concern regarding its utility, even at a presentence level. The authors note that some courts have perceived the willingness of the offender to undergo a test as evidence of his remorse and potential treatability and likewise have viewed an individual's reluctance to take a phallometry examination as evidence of potential higher risk or lack of remorse. In fairness, the limitations of phallometry evaluations are such that the conclusions are likely overvalued or possibly inappropriate.

It is easy to conceive of situations where the court may be given an erroneous impression by phallometry results. For example, consider two offenders, each of whom has been convicted of sexually assaulting prepubescent children. The first admits to the offense and cooperates with a phallometry evaluation that demonstrates deviant sexual arousal. He agrees to treatment. As part of the standard risk assessment process, the court is informed that offenders who have deviant arousal are at greater risk of reoffending than those without evidence of paraphilia, and accordingly he is given a lengthy sentence with a condition for treatment. In contrast is the pedophile who commits sexual offenses against children but denies deviant arousal and is able to control his responses to a phallometry evaluation such that he does not have any laboratory affirmation of pedophilic behavior. The court is led to believe he does not have deviant arousal and therefore is less of a risk, and accordingly he receives a more lenient sentence. In review of the two similar cases, many would agree that the pedo-

phile able to fake responses on a phallometry evaluation is probably a greater risk than the pedophile taking the test willingly and cooperatively. In such instances, phallometry may in fact lead us to incorrect conclusions, because it lacks the sensitivity and specificity that should be required of any physiological test that is used for legal purposes. One could easily argue that there is no real way to pass a phallometry examination, but there are ways to fail.

At this point, I do not think phallometry evaluations are at the level of validity that can be reliably used in court proceedings. Much work is needed to standardize stimulus sets and reach agreement on methodology, including scoring and administration, as well as more research to clarify sensitivity and specificity before it reaches the threshold, where it can be reliably used in court proceedings.

References

1. Purcell MS, Chandler JA, Federoff JP: The use of phallometric evidence in Canadian criminal law. *J Am Acad Psychiatry Law* 43:141–53, 2015
2. Federoff JP, Kuban M, Bradford J: Laboratory measurement of penile response in the assessment of sexual interests, in *Sex Offenders: Identification, Risk Assessment, Treatment, and Legal Issues*. Edited by Saleh F, Grudzinskas AJ, Brodsky DJ, *et al*. New York: Oxford University Press, 2009
3. Harris GT, Lalumiere M, Seto MC, *et al*: Explaining the erectile responses of rapists to rape stories: the contribution of sexual activity, non-consent, and violence with injury. *Arch Sex Behav* 41:221–9, 2012
4. Seto MC, Lalumiere M, Harris G, *et al*: The sexual response of sexual sadists. *J Abnorm Psychol* 121:739–53, 2012
5. Blanchard R, Kuban ME, Blak T, *et al*: Phallometric comparison of pedophilic interest in nonadmitting sexual offenders against stepdaughters, biological daughters, other biologically related girls, and unrelated girls. *Sex Abuse* 18:1–14, 2006
6. Lanyon R, Thomas M: Detecting deception in sex offender assessment, in *Clinical Assessment of Malingering and Deception*. Edited by Rogers R. New York: The Guilford Press, 2008, pp 285–300
7. Rogers R: *Clinical Assessment of Malingering and Deception*. New York: The Guilford Press, 2008
8. Trotter D, Rouleau JL, Renaud P, *et al*: Using eye tracking to identify faking attempts during penile plethysmography assessment. *J Sex Res* 51:946–55, 2014
9. Hall GC, Proctor WC, Nelson GM: Validity of physiological measures of pedophilic sexual arousal in a sexual offender population. *J Consul Clin Psychol* 56:118–22, 1988