



Published in final edited form as:

Obstet Gynecol. 2009 April ; 113(4): 825–832. doi:10.1097/AOG.0b013e31819bda7c.

The Tampon Test for Vulvodynia Treatment Outcomes Research: Reliability, Construct Validity, and Responsiveness

David C. Foster, MD MPH¹, Merrill Beth Kotok, BA, BSN, RNC¹, Li-Shan Huang, PhD², Arthur Watts, BS², David Oakes, PhD², Fred M. Howard, MD, MS¹, Chris J Stodgell, PhD¹, and Robert H. Dworkin, PhD³

¹Departments of Obstetrics and Gynecology, University of Rochester School of Medicine and Dentistry

²Department of Biostatistics, University of Rochester School of Medicine and Dentistry

³Department of Anesthesiology, University of Rochester School of Medicine and Dentistry

Abstract

Objective: A standardized tampon insertion and removal test, the “Tampon Test” provides an alternative to sexual intercourse pain as an outcome measure for vulvodynia research. We report upon the reliability, validity, and responsiveness to change of the “Tampon Test” as an outcome measure for vulvodynia clinical trials.

Methods: Outcome measures were assessed in women enrolled in the Vulvar Vestibulitis Clinical Trial, a randomized clinical trial of oral desipramine and topical lidocaine effectiveness. Reliability estimates of the Tampon Test using the Kappa statistic evaluated week to week measures at baseline. Tampon Test construct and discriminant validity were assessed through correlation to other outcome measures. Patients' ability to regularly perform the Tampon Test was compared to regularity of reporting intercourse pain.

Results: During the two-week baseline phase, vulvodynia-afflicted women reported stable mean Tampon Test scores 4.6 ± 2.6 (Week -2); 4.6 ± 2.7 (Week -1); and 4.7 ± 2.8 (Week 0) with moderate week-to week reliability, (weighted Kappa = 0.52). Over an 8 week phase of trial intervention, change in the Tampon Test measure significantly correlated to a number of outcome measures including: daily pain ($r=0.42$), intercourse pain ($r=0.35$), cotton swab vestibular pain ($r=0.38$), and the Brief Pain Inventory ($r=0.49$). Women with vulvodynia study subjects performed the Tampon Test 96.3% of the requested time, which was two-fold higher adherence than intercourse pain measurement (49.7%).

Conclusion: The Tampon Test reflects a “real life” experience that is reliable with good construct validity as shown by the breadth of correlated outcome measures. The Tampon Test is an appropriate outcome measure for vulvodynia research that can be considered for use as the primary efficacy endpoint in clinical trials of treatments for vulvodynia.

Clinical Trial Registration: ClinicalTrials.gov, www.clinicaltrials.gov, NCT00276068

Reprint requests: David C. Foster MD, MPH, Box 668, Strong Memorial Hospital 601 Elmwood Ave, Rochester, New York 14642
Phone: (585) 273-3233 Fax: (585) 276-1956 Email: david_foster@urmc.rochester.edu.

Financial Disclosure: The authors did not report any potential conflicts of interest.

Introduction

An estimated 7% of women will meet the diagnostic criteria for vulvodynia and those afflicted will commonly suffer significant psychosocial problems including sexual dysfunction, anxiety, infertility, and divorce.(1-4) Even though vulvodynia has been recognized to be a rather common affliction, evidence-based treatment options for vulvodynia are few, largely resulting from the dearth of randomized clinical trials (RCTs). As is evident from the clinicaltrials.gov website, research efforts to identify effective treatments for vulvodynia by RCTs have been limited, to date. The future expansion of RCTs for vulvodynia will require clear and widely accepted definitions of disease, inclusion / exclusion criteria, and outcome measures.(5) An expert panel, the Initiative on Methods, Measurements, and Pain Assessment in Clinical Trials (IMMPACT) group has defined what constitutes evidence of successful outcomes, known as “outcome domains”, for pain trials and has recommended standard measurement tools for these outcome domains.(6;7)

A standardized tampon insertion and removal test, the “Tampon Test” provides an alternative to sexual intercourse pain as an outcome measure for vulvodynia research. Although most vulvodynia-afflicted women seek treatment for a complaint of insertional dyspareunia, the assessment of “*intercourse pain*” as a primary outcome measure raises practical and methodologic difficulties. In severe cases, vulvodynia may be so intense that afflicted patients may completely abstain from intercourse. As a result, the use of intercourse pain as a primary outcome measure may be problematic for recruitment, data analyses, and generalization of results. Recent analysis of a large population-based sample found “pain with tampon insertion” to be one of the strongest risk factors for the development of vulvodynia.(8) The Tampon Test reflects a common, real-life experience well understood by patients and clinicians. Following IMMPACT recommendations, the Tampon Test incorporates important aspects of disease-specific, patient-reported outcomes using a numerical rating scale (NRS).(6;9) We examined the reliability, construct and discriminant validity, responsiveness to change, and feasibility of the Tampon Test as an outcome measure for vulvodynia clinical trials and we compared the Tampon Test to individual and composite measures of pain intensity/quality recommended by the IMMPACT group.

Materials and Methods

The Vulvar Vestibulitis Clinical Trial (VVCT) is an NIH (NICHD)--funded randomized, placebo-controlled, double-blinded clinical trial to study the clinical efficacy of four medical treatments for vulvar vestibulitis (localized vulvodynia): 1) topical lidocaine, 2) oral desipramine, 3) combined lidocaine and desipramine, and 4) placebo cream and tablets. The VVCT was conducted at Strong Memorial Hospital of the University of Rochester between August, 2002, and July, 2007 and the protocol was reviewed and approved by the University of Rochester Research Subjects Review Board (RSRB #8677). A blocked randomization scheme, utilizing a uniform random number generator and employing a block size of 8 ensured the four possible treatment combinations would occur equally or would not be greater than 2 assignments for any given treatment group. The duration of study drugs lasted 12 weeks with post-intervention follow-up at 16, 26, and 52 weeks. Clinical response from randomization to 12 weeks (the end of the randomized, blinded phase of the trial) was assessed by change in pain by Numeric Rating Scale (NRS) of a weekly Tampon Test compared to a number of measures with pre-existing reliability/validity data or prior published experience in vulvodynia clinical trials including: change in overall daily pain intensity (24 hour NRS)(7), the frequency of sexual intercourse (insertional attempts per week)(10), the change in intercourse pain NRS (10), vulvar algometer score(11) and the cotton swab test (CST) pain level by verbal reporting scale (VRS)(12). In addition, during each study visit subjects completed a battery of pain and health related quality-of- life measures recommended by IMMPACT including: the

Brief Pain Inventory, Short Form-McGill Pain Questionnaire (SF-MPQ), Profile of Mood States, and the Beck Depression Inventory.(7) For the primary outcome analysis of the clinical trial (to be published later), we hypothesized that the response rates would be 20% for the double placebo group, 50% for each treatment used alone and 80% when the two treatments are used together. Therapeutic response of desipramine / lidocaine was estimated from preliminary reported data from our group.(13) A Bonferroni--corrected 80% power level required a total of 104 subjects to complete the trial for a two-sided test with $\alpha = 0.05$. Assuming a 25% dropout rate we will therefore estimate 130 subjects were needed to be randomized into the trial.

Our present objective is to report data from pre-randomization (Baseline) through the first post-randomization visit (Week 8) in order to demonstrate the utility of the Tampon Test as an outcome measure for vulvodynia clinical trials. Baseline "cross-sectional" comparisons used the mean of the specific outcome variable over three pre-randomization time points as "Baseline" (Week -2, Week -1, Week 0). "Longitudinal" comparisons of outcome change over time used the mean of the specific outcome variable over three pre-randomization time points as "Baseline" (Week -2, Week -1, Week 0) and calculated the change in the respective mean of the outcome variable over three time points ending with Week 8 (Week 6, Week 7, Week 8).

Women were invited to participate if they reported greater than three continuous months' duration of vulvar symptoms of insertional dyspareunia and/or pain with tampon insertion, and were between 18 and 50 years of age. After informed consent, all study candidates completed a standard history and physical exam. To be included in the trial, participants needed to fulfill "Friedrich's Criteria" for the diagnosis of vulvodynia including tenderness localized within the vestibule confirmed by the Cotton Swab Test modified from the technique of Bergeron et al. (12) The Cotton Swab Test was performed on defined points of the labia majora, minora, and lower vagina. A "positive" Cotton Swab Test was operationally defined as follows. In four defined points (1:00, 5:00, 7:00, and 11:00) within the vulvar vestibule, the subjects should report mean score *equal to or greater than 4 out of 10* on a Verbal Rating Scale. This modified the criteria of Bergeron et al.(12) by excluding Cotton Swab Test testing at 12:00 and 6:00 of the vulvar vestibule as defined points. This modification was made with the intent of reducing the chance of inclusion of painful conditions such as Skenitis and vaginal forchette fissures that might evoke a pain response in those respective sites. The localized nature of pain was confirmed by finding all remaining Cotton Swab Test points tested in the lower vagina, labia majora, and labia minora to be non-painful, defined as a mean score *equal to or less than 2 out of 10* in pain on a Verbal Rating Scale. A second clinician-examiner would perform a second independent exam of the candidate and would need to concur with the diagnosis of vulvar vestibulitis. Additionally, eligible candidates did not demonstrate any other specific neuropathology, atrophic vaginitis, dermatitis such as vulvar dystrophy, or pathogens such as culture/smear-proven *Candida spp.* or *Herpes simplex*.

Subjects were provided with ORIGINAL REGULAR TAMPAX™ TAMPONS (Proctor & Gamble Corp., Cincinnati, OH) supplied in standard cardboard applicator for insertion. ORIGINAL REGULAR TAMPAX™ TAMPONS are 5.5 cm long and 1.5 cm. in diameter when contained in the cardboard applicator. The cardboard applicator length is 12.8 cm. ORIGINAL REGULAR TAMPAX™ TAMPONS are made of a combination of cotton and rayon, the exact fiber proportions are proprietary to Proctor & Gamble Corp.; the string is made of 100% cotton, and the applicator is made of cardboard.

Detailed instructions concerning the performance and documentation of the weekly Tampon Test, the daily 24 hour pain measure, and intercourse pain measure were given to each subject on the first pre-randomization Visit (Week -2) by the Research Nurse/ Coordinator. Each study

participant was verbally instructed to 1) deposit the tampon fully into the vagina above the level of the hymeneal ring via the cardboard applicator, 2) remove the applicator from the vagina, and 3) finally remove the tampon from the vagina via traction on the tampon string. The subject was instructed to remove tampon immediately following vaginal insertion. The subject was instructed not to lubricate the tampon prior to insertion and to insert / deposit the tampon using only the supplied cardboard applicator. On a weekly basis and in a consistent manner, the subject was instructed to insert and immediately remove the tampon and record the degree of pain during the entire insertion/removal experience on a 0 – 10 pain numeric rating scale – 0 meaning “no pain”; 10 meaning the worst possible pain. The subject would then record her level of pain by marking the corresponding number on a linear pain scale printed on the back of the first page of each week in her VVCT logbook. All information was reviewed and recorded during the weekly telephone call by the Research Nurse/Coordinator and later confirmed, following return of the VVCT Logbook on scheduled study visits. During the pre-randomization (Baseline) phase of the trial, eligible subjects were required to demonstrate an adequate baseline level of pain (average 4/10 or greater) on the Tampon Test to proceed to randomization. This criterion was used because lower baseline pain levels on the Tampon Test would limit the ability of the RCT to demonstrate greater improvement with treatment vs. placebo.

On a daily basis during the trial, subjects reported whether they experienced sexual intercourse in the last 24 hours. The possible responses were: #1—“No, too painful” would indicate the subject could not accept an approach to physical intimacy because of pain, #2 –“No, not interested” would indicate that the subject was not in the mood for sexual intimacy, #3—“No, no opportunity” would indicate that her partner was not available, #4—“Yes” would mean an attempt at sexual intercourse was made. If intercourse was attempted, the subject was asked to rate her level of pain during intercourse on a 0 – 10 pain scale – 0 meaning “no pain”; 10 meaning the worst possible pain. She would then record her level of pain by marking the corresponding number on a linear pain scale printed on the front of the daily diary page.

Other than the initial visit (Week -2) when two examiners confirmed the clinical diagnosis of localized vulvodynia, subjects were evaluated consistently during following visits by the same research clinician (DCF) with quantitative sensory tests (Cotton Swab Test and Algesiometer), selective palpation of pelvic muscles for pain, and a battery of psychometric tests. During each study visit of the trial, all components of the exam were performed by a single examiner in identical fashion to the first pre-randomization (Week -2) visit. The Algesiometer, generously supplied by Curnow and Morrison, Plymouth, UK, consisted of a mechanical pulse generator which drove a probe against the mucocutaneous surface of the vulva for a calibrated distance and force ranging from 176 mN to 1868 mN in 8 increments.(14) A standard 4- anatomic site test of the vestibule was routinely used as described by Eva et al.(11) We used a “method of limits” with the pain threshold determined as the first consistent verbal report of stimulus pain. (15) Subjects needed to demonstrate consistently positive responses for two consecutively increasing stimulus intensities. Algesiometer score resulted from the summation of the pain thresholds from the four anatomic sites (0 to 28 score range with higher score corresponding to less vestibular pain). During a pelvic exam conducted at each study visit, selective muscle palpation included digital palpation of the levator ani, obturator internus, and piriformis muscle groups. Notation was made for each muscle group, anatomic side, and pain level on a 0 to 4 scale corresponding to none, mild, moderate, and severe pain, respectively. In addition, the Brief Pain Inventory, Short Form-McGill Pain Questionnaire (SF-MPQ), Neuropathic Pain Scale, Profile of Mood States, Beck Depression Inventory, Sexual and Physical Abuse History, Multidimensional Pain Inventory, Dyadic Adjustment Scale, Communication Pattern Questionnaire, and Index of Sexual Satisfaction were administered and subjects were asked to answer psychometric questions according to their overall pain state.

This report focuses on the Brief Pain Inventory, SF-MPQ, the Neuropathic Pain Scale, the Profile of Mood States, and Beck Depression Inventory for the purpose of validating the Tampon Test based on psychometric measures recommended by IMMPACT for evaluating treatment efficacy and effectiveness.(6;7) Outcome domains (*in italics*) and recommended measures include: 1) *pain intensity*--pain over each 24 hr. period, pain with intercourse (if attempted), Cotton Swab Test, and Algesiometer score, 2) *pain quality*--SF-MPQ and Neuropathic Pain Scale, 3) *physical functioning*--Brief Pain Inventory Interference Scale score, and 4) *emotional functioning*—Beck Depression Inventory and Profile of Mood States.

Over the three pre-randomization (Baseline) Tampon Test assessments, test-retest reliability was assessed with a Kappa statistic, weighted Kappa statistic, and the Shrout-Fleiss intraclass correlation.(16) To evaluate construct validity, we performed Pearson and Spearman correlations examining associations between Tampon Test scores and the other outcome measures. The Tampon Test and the other outcome measures were analyzed in two ways: cross-sectional baseline values, and longitudinal change in values over time, without reference to treatment group allocation. Subject acceptance of the Tampon Test was evaluated by adherence to the measure compared to the intercourse pain measure. Correlations of the Tampon Test with Cotton Swab Test vaginal pain and with pelvic muscle pain to palpation were included to reflect specificity of the Tampon Test to pain localized to the vestibule compared to superficial vaginal and deep pelvic pain, respectively.

Results

Of the 150 women consented for the VVCT, 133 subjects were randomized and 118 subjects returned through the first post-randomization visit (Week 8). Table 1 summarizes characteristics of the 118 subjects who completed the trial from Baseline Week -2 to the first post-randomization visit, Week 8. Of the 17 consented candidates/subjects who were excluded or dropped out before drug randomization, 9 candidates decided not to participate in the trial, 5 candidates did not receive diagnostic agreement by examiners and 3 subjects did not demonstrate adequate levels of pain (4 out of 10 or greater) on initial Tampon Test. Of the 15 subjects randomized to study drug who did not complete the trial, there were 2 pregnancies, 4 removed by research staff because of concerning side effects (hypertension/tachycardia (1), elevated liver enzymes (1), symptomatic palpitations (1), poor record keeping (1)), and 9 elected to drop out of the study. Of subjects completing Week 8 (Table 1), mean age was 30.4 ± 7.6 , racial / ethnic mix was predominantly Caucasian -- non Hispanic, mean years of education was $16.0 \text{ years} \pm 3.0$, 69.5% reported being presently sexually active, 55.1% reported a history of pain with first sexual activity, and 63.6% reported a history of pain with first tampon insertion. Adherence to tampon insertion on a weekly basis was excellent, with 1136 tests completed out of 1180 subject weeks (96.3%) compared to intercourse pain measurement, for which only 586 tests were completed out of 1180 subject weeks (49.7%). Comparatively, the Tampon Test demonstrated a two-fold higher adherence compared to the intercourse pain measure, in spite of encouragement for both activities by the Research Nurse. Subjects were asked explain in the VVCT Logbook why they did not attempt intercourse. Subjects reported “no partner”-- 55.2% of un-attempted subject weeks, “too painful” -- 7.6% of un-attempted subject weeks, and “not interested” 37.2% of un-attempted subject weeks.

Test-retest reliability was estimated by examining week-to-week Tampon Test pain recorded by each subject during the pre-randomization (Baseline) Weeks -2, -1, and 0) phase of the trial. During the three weekly pre-randomization assessments, the Tampon Test means were 4.6 ± 2.6 (Week -2); 4.6 ± 2.7 (Week -1); and 4.7 ± 2.8 (Week 0), based on the 0 to 10 NRS. Weighted Kappa Tampon Test reliability was $K = 0.52$ for Weeks -2 and -1, $K = 0.52$ for Weeks -1 and 0 measures, and $K = 0.38$ for Weeks -2 and 0. Such Kappa values reflect moderate week-to-week agreement for Weeks -2 and -1 and Weeks -1 and 0 and fair week-to-week agreement

for Weeks -2 and 0. The Shrout-Fleiss intraclass correlation was 0.48 for the three baseline Tampon Test assessments and 0.74 for the average of the three baseline assessments.

For the cross-sectional assessment of construct validity, the Tampon Test significantly correlated with: “daily 24 hr. pain rating” $r = 0.38$ $P < 0.0001$, “intercourse pain” $r = 0.22$; $P = 0.04$, the Brief Pain Inventory $r = 0.34$; $P = 0.0001$, and the Neuropathic Pain Scale total score $r = 0.19$; $P = 0.03$. Spearman coefficients displayed similar results to Pearson coefficients for these correlations and scatterplot reviews for each of the correlations displayed a linear relationship pattern (scatterplot data not shown).

For the longitudinal assessment of construct validity and responsiveness to change, change in Tampon Test scores were significantly correlated with change in measures of: “daily 24 hr. pain” $r = 0.42$; $P < 0.0001$, “intercourse pain” $r = 0.35$; $P = 0.003$, Cotton Swab Test vestibule pain $r = 0.38$; $P < 0.0001$, Algesiometer scores $r = -0.33$; $P = 0.0004$, SF-MPQ sensory subscale scores $r = 0.30$; $P = 0.005$, Brief Pain Inventory Interference scale scores $r = 0.49$; $P < 0.0001$, and Neuropathic Pain Scale total scores $r = 0.33$; $P = 0.0005$. Spearman coefficients displayed similar results to Pearson coefficients for these correlations and scatterplot reviews for each of the correlations displayed a linear relationship pattern (scatterplot data not shown).

Table 3 displays a correlation matrix of pain intensity / quality measures and psychometric measures in addition to the Tampon Test correlations of Table 2. Of particular note, the highest correlation was found between the Baseline to Week 8 change in “24 hour pain” and change in the Brief Pain Inventory (BPI), $r = 0.55$. Additionally, there was a complete lack of correlation between changes in CST-evoked vestibular pain or algesiometer-evoked pain and changes in intercourse pain, $r = 0.01$ and $r = 0.00$, respectively. Comparing the correlation matrices of Table 3 to the corresponding correlations of the Tampon Test (Table 2) finds no single outcome measure surpasses the Tampon Test in breadth and strength of association.

We studied the potential impact of selected co-morbid conditions on Tampon Test pain. Unpaired sample *t* Tests were used assess possible co-morbid effects on the Tampon Test by selected historical categorical variables. No significant effect on Tampon Test pain was found in the presence of endometriosis, irritable bowel syndrome, interstitial cystitis, history of rape / sexual abuse, or a report of “never using tampons before”. A significant difference in tampon test pain was found when fibromyalgia was present ($t = 2.30$; $P = 0.02$). A linear model was developed incorporating fibromyalgia and “overall 24 hour pain” as independent variables regressed against tampon test pain. “Overall 24 hour pain” remained highly predictive of tampon test pain, adjusting for the presence of fibromyalgia, ($t = 3.76$; $P < 0.001$). On the other hand, fibromyalgia no longer significantly predicted tampon test pain, adjusting for “overall 24 hour pain” ($t = 1.68$; ns). When Tampon Test was done within 7 days of onset of menses, Tampon Test pain was not significantly different: 5.3 ± 1.7 within 7 days of menses, 4.2 ± 2.2 outside of time period, ($t = 1.47$; ns). As is evident in Table 2, Tampon Test scores were not significantly correlated with measures of levator, obturator, and piriformis muscle pain to palpation nor with the Cotton Swab (Q Tip) assessment of vaginal pain. The Tampon Test scores also did not significantly correlate with variation in mood or affect as reflected by the Beck Depression Inventory, SF-MPQ affective subscale, and the Profile of Mood States.

Discussion

A consensus group (IMMPACT) has published recommendations for the conduct of clinical trials in chronic pain and describes the “ideal” primary outcome measure to include qualities of: appropriateness of content, reliability, validity, responsiveness, and limited participant burden.⁽⁷⁾ The Tampon Test is a readily understandable “real life” outcome measure that demonstrated good week-to-week reliability using weighted Kappa and intra-class correlation

coefficients. The consistent means and variability for the Tampon Test over Weeks -2, -1, and 0 indicate the absence of change in pain intensity, secondary to a practice effect. The Tampon Test was significantly associated with a number of the IMMPACT core outcome dimensions and specifically recommended measures (6;7) including: 1) *pain intensity*--pain over each 24 hr. period, pain with intercourse (if attempted), Cotton Swab Test and Algesiometer scores, 2) *pain quality*--McGill Pain Questionnaire--Short form (MPQ-SF), and 3) *physical functioning*--Brief Pain Inventory (BPI) scores. Pearson and Spearman correlations were consistent demonstrating statistical robustness of the findings. Comparing the correlations of all outcome measures examined in the present analyses, no other measure correlated as highly or as frequently with other outcome measures as the Tampon Test.

With respect to construct validity, we evaluated two dimensions: first, the Tampon Test baseline values were compared cross-sectionally to baseline values of other outcome measures and second Tampon Test change longitudinally over time to changes in other outcome measures. Our intent was to evaluate the ability of the Tampon Test to measure the severity of pain through the cross-sectional comparisons and to evaluate the ability of the Tampon Test to measure response to treatment through the longitudinal comparisons. Table 2 indicates that the Tampon Test displays broader and stronger associations with changes in outcome measures over time compared to cross-sectional associations at baseline. This would suggest that the Tampon Test has stronger validity in measuring response to treatment over time compared to measuring pain severity at a single time point. The ability of an outcome measure to measure change over time exemplifies the quality of responsiveness, a critical requirement for clinical trial outcome measures, which must reflect improvement (or worsening) and successfully distinguish efficacy among different treatment groups.

Discriminant validity was evident through several observations: first, the Tampon Test was not influenced by co-morbid conditions such as endometriosis and interstitial cystitis, second, the Tampon Test did not correlate with evoked pain measures outside of the vestibule, and third, the Tampon Test did not correlate with psychometric measures of affect. To elaborate, the Tampon Test showed good correlation with cotton swab and algesiometer assessments of vestibular pain--change over time, a pivotal assessment for future studies of response to treatment. In contrast, testing of other anatomic regions of the genital tract including cotton swab-evoked vaginal pain and palpation-evoked levator muscle pain failed to correlate with the Tampon Test. With respect to the "*emotional functioning*" outcome dimension, the Tampon Test did not significantly correlate with the Beck Depression Inventory (BDI), Profile of Mood States (POMS), and McGill Affective scores. As a result, the Tampon Test may be less influenced by the subject's affect or short-term emotional variation thereby strengthening discriminant validity over the duration a clinical trial.

In vulvodynia outcomes research to date, primary outcome variables have fallen into three major categories, characterized by both strengths and weaknesses: 1) composite pain scores (commonly a combination of psychometric tests, personal pain assessment, and practitioner pain assessment), 2) individually designed subject questionnaires and clinical assessment instruments, and 3) quantitative sensory testing (QST). (10;17-19) Composite pain scores commonly consist of one or several psychometric tests combined with other measures of patient-reported outcomes and clinician-reported outcomes, which may have variable reliability and validity. Many composite measures lack specificity with regards to vulvodynia and could be influenced by co-morbid pain and mood disorders. The complexity of composite scores may also hinder interpretation. Individually designed questionnaires and exam assessments used in single studies may be quite specific to vulvodynia but commonly overlooks reliability and validity testing. In contrast, to composite scores and individually designed assessment tools, QST provides reliable measures that can be specifically designed for vulvar pain assessment. Several research groups have developed algesiometers for QST assessments

that produce calibrated mechanical stimuli for pain testing of the vulva.(14: 20) Unfortunately, QST assessments are instrument dependent making replication difficult when the instruments are not commercially available. QST measures also lack direct clinical relevance which may limit their use as a primary endpoint, although QST may be quite valuable as surrogate or secondary endpoint.

There are several limitations of the present study and of the Tampon Test as an outcome measure for vulvodynia clinical trials. Some subjects may report Tampon Test baseline data that may be too low to permit effective analysis of outcomes in an RCT. The proportion excluded from the present study was small (3%) but does dilute the potential pool of subjects. Some vulvodynia afflicted individuals reported a much higher level of intercourse pain in contrast to tampon insertion pain, highlighting the fact that the Tampon Test does not fully replace “intercourse pain” as an outcome measure. The Tampon Test, by nature, evokes a “self-inflicted” pain compared to “partner-inflicted” intercourse pain, and this difference may lead to a distinctly different experience and perception of pain. Intercourse pain also carries a psychosexual dimension that cannot be equated to pain associated with tampon insertion nor can intercourse pain be equated to evoked pain of the cotton swab or algesiometer to the vestibule, as is evident in Table 3. Intercourse pain outcome measures will therefore remain a major, albeit problematic focus of vulvodynia RCT's

A crucial facet of pain outcomes research is the development of well defined, understandable, reliable, and valid outcome variables. Pain with tampon insertion is common symptom in women with vulvodynia that in many cases precedes the development of intercourse pain.(8) Among the criteria developed for evaluating the quality of chronic pain outcome measures, the greatest weight has been given to appropriateness of the measure's content, reliability, validity, responsiveness, and limited respondent burden.(7) In addition, for assessments of pain intensity, study endpoints with patient-reported outcomes on a numeric rating scale are preferred. The Tampon Test therefore fulfills key attributes of a core outcome measure for vulvodynia pain. Rather than simply being a surrogate for “intercourse pain,” the Tampon Test reflects another “real life” behavior. We have shown in this report that the Tampon Test is a reliable and valid outcome measure, one which is associated with a wide range of other pain outcome measures in both cross-sectional and longitudinal assessments. Importantly, its excellent adherence rate of over 95% indicates that patients find it an acceptable and feasible approach to evaluating their pain.

Acknowledgements

Supported by a grant from the National Institutes of Health: RO-1 HD040123-05.

Reference List

1. Goetsch MF. Vulvar vestibulitis: prevalence and historic features in a general gynecologic practice population. *American Journal of Obstetrics & Gynecology* Jun;1991 164(6 Pt 1):1609–14. [PubMed: 1646569]discussion
2. Stewart DE, Reicher AE, Gerulath AH, Boydell KM. Vulvodynia and psychological distress. *Obstetrics & Gynecology* Oct;1994 84(4):587–90. [PubMed: 8090398]
3. Nunns D, Mandal D. Psychological and psychosexual aspects of vulvar vestibulitis. *Genitourinary Medicine* Dec;1997 73(6):541–4. [PubMed: 9582478]
4. Meana M, Binik YM, Khalife S, Cohen D. Psychosocial correlates of pain attributions in women with dyspareunia. *Psychosomatics* Nov;1999 40(6):497–502. [PubMed: 10581978]
5. Landry, A, Bergeron, S, Dupuis, M, Desrochers, G. The Treatment of Provoked Vestibulodynia, a critical review. *Clinical Journal of Pain* Jan 2;2008 24(2):155–171. [PubMed: 18209522]

6. Turk DC, Dworkin RH, Burke LB, Gershon R, Rothman M, Scott J, et al. Developing patient-reported outcome measures for pain clinical trials: IMMPACT recommendations. *Pain* Dec 6;2005 125(3):208–15. [PubMed: 17069973]
7. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain* Jan;2005 113(12):9–19. [PubMed: 15621359][see comment]. [Review] [88 refs]
8. Harlow BL, Wise LA, Stewart EG. Prevalence and predictors of chronic lower genital tract discomfort. *American Journal of Obstetrics & Gynecology* Sep;2001 185(3):545–50. [PubMed: 11568775]
9. Dworkin RH, Backonja M, Rowbotham MC, Allen RR, Argoff CR, Bennett GJ, et al. Advances in neuropathic pain: diagnosis, mechanisms, and treatment recommendations. *Archives of Neurology* Nov;2003 60(11):1524–34. [PubMed: 14623723][see comment]. [Review] [70 refs]
10. Bergeron S, Binik YM, Khalife S, Pagidas K, Glazer HI, Meana M, et al. A randomized comparison of group cognitive-behavioral therapy, surface electromyographic biofeedback, and vestibulectomy in the treatment of dyspareunia resulting from vulvar vestibulitis. *Pain* Apr;2001 91(3):297–306. [PubMed: 11275387]
11. Eva LJ, Reid WM, MacLean AB, Morrison GD. Assessment of response to treatment in vulvar vestibulitis syndrome by means of the vulvar algometer. *American Journal of Obstetrics & Gynecology* Jul;1999 181(1):99–102. [PubMed: 10411802]
12. Bergeron S, Binik YM, Khalife S, Pagidas K, Glazer HI. Vulvar vestibulitis syndrome: reliability of diagnosis and evaluation of current diagnostic criteria. *Obstetrics & Gynecology* Jul;2001 98(1):45–51. [PubMed: 11430955]
13. Foster, DC.; Duguid, KM. Open label study of oral desipramine and topical lidocaine for the treatment of vulvar vestibulitis; Abstract, Int. Conf. on Mechanism and Treatment of Neuropathic Pain; Rochester, NY. 1998.
14. Curnow JS, Barron L, Morrison G, Sergeant P. Vulval algometer. *Medical & Biological Engineering & Computing* May;1996 34(3):266–9. [PubMed: 8762838]
15. Dotson RM. Clinical neurophysiology laboratory tests to assess the nociceptive system in humans. *Journal of Clinical Neurophysiology* Jan;1997 14(1):32–45. [PubMed: 9013358][Review] [74 refs]
16. Shrout P, Fliess JL. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* 1979;86:420–8. [PubMed: 18839484]
17. Danielsson I, Torstensson T, Brodda-Jansen G, Bohm-Starke N. EMG biofeedback versus topical lidocaine gel: a randomized study for the treatment of women with vulvar vestibulitis. *Acta Obstetrica et Gynecologica Scandinavica* 2006;85(11):1360–7. [PubMed: 17091418]
18. Bornstein J, Livnat G, Stolar Z, Abramovici H. Pure versus complicated vulvar vestibulitis: a randomized trial of fluconazole treatment. *Gynecologic & Obstetric Investigation* 2000;50(3):194–7. [PubMed: 11014954]
19. Nyirjesy P, Sobel JD, Weitz MV, Leaman DJ, Small MJ, Gelone SP. Cromolyn cream for recalcitrant idiopathic vulvar vestibulitis: results of a placebo controlled study. *Sexually Transmitted Infections* Feb;2001 77(1):53–7. [PubMed: 11158692]
20. Pukall CF, Binik YM, Khalife S. A new instrument for pain assessment in vulvar vestibulitis syndrome. *Journal of Sex & Marital Therapy* Mar;2004 30(2):69–78. [PubMed: 15043051]

Table 1
Demographic and clinical characteristics of the participants

Personal History	N	
Age (yrs. \pm s.d.)	118	30.4 \pm 7.6
-- White, not Hispanic	110	93.2%
-- White, Hispanic	4	3.4%
-- Asian or Pacific Islander	3	2.5%
-- Black, Hispanic	1	0.8%
-- Other	1	0.8%
Years of education	118	16.0 \pm 3.0
Age of onset of vulvodynia pain	118	24.3 + 7.4
Sexually active now	82/118	69.5%
History of tampon pain with insertion	75/118	63.6%
First sexual experience "moderate to severe pain"	65/118	55.1%
Tampon Test (adhered to protocol per subject week)	1136 / 1180	96.3%
Intercourse at least once per subject week	586 / 1180	49.7%

Table 2

Outcome measure Baseline values, Pearson correlation coefficients, and change in outcome measures from Baseline to Week 8. (95% confidence intervals)

Selected outcome measures	Baseline (mean ± SE)	TT Baseline correlated* to Baseline outcome values	Change Baseline to Week 8 (mean ± SE)	TT change correlated* to change from Baseline to Week 8
Tampon test	4.67 ± 0.19	1.00	-1.53 ± 0.19	1.00
Overall pain (24 hrs.)	1.90 ± 0.16	0.38 (0.23, 0.52); P < 0.0001	-0.60 ± 1.19	0.42 (0.24, 0.55); P < 0.0001
Intercourse pain	5.82 ± 0.24	0.22 (0.05, 0.38); P = 0.04	-1.60 ± 0.23	0.35 (0.17, 0.50); P = 0.003
CST (vestibule)	22.12 ± 0.46	0.11 (-0.06, 0.28); P = 0.21	-9.13 ± 0.94	0.38 (0.21, 0.53); P < 0.0001
CST (vagina)	2.39 ± 0.27	0.08 (-0.10, 0.25); P = 0.40	-0.76 ± 0.33	0.12 (-0.07, 0.30); P = 0.23
Algesiometer	10.70 ± 0.63	-0.16 (-0.33, 0.00); P = 0.06	5.68 ± 1.14	-0.33 (-0.49, -0.15); P = 0.0004
Pelvic muscle pain (mean)	0.64 ± 0.05	-0.09 (-0.25, 0.08); P = 0.34	-0.17 ± 0.05	0.12 (-0.06, 0.30); P = 0.19
SF-MPQ total score	13.56 ± 0.76	0.15 (-0.03, 0.31); P = 0.11	-4.92 ± 0.74	0.23 (0.04, 0.40); P = 0.02
SF-MPQ affective subscale	2.50 ± 0.27	0.13 (-0.04, 0.30); P = 0.16	-1.27 ± 0.26	0.16 (-0.03, 0.33); P = 0.12
SF-MPQ sensory subscale	10.96 ± 0.60	0.19 (0.02, 0.35); P = 0.054	-3.73 ± 0.58	0.30 (0.12, 0.46); P = 0.0052
Brief Pain Inventory	20.46 ± 1.60	0.34 (0.17, 0.48); P < 0.0001	-7.91 ± 1.60	0.49 (0.34, 0.62); P < 0.0001
Neuropathic Pain Scale	43.69 ± 1.37	0.19 (0.01, 0.35); P = 0.03	-12.7 ± 1.52	0.33 (0.16, 0.49); P = 0.0005
Beck Depression Inventory	9.58 ± 0.79	0.03 (-0.14, 0.21); P = 0.71	**	**
Profile of Mood States	86.68 ± 2.75	0.10 (-0.07, 0.27); P = 0.29	**	**

SF-MPQ = Short-form McGill Pain Questionnaire

CST = Cotton Swab Test

TT = Tampon Test

* Pearson Correlation Coefficient (95% C.I.)

** not available at Week—8

Pearson correlation matrix (95% confidence intervals) for selected outcome variables based upon changes in outcome measures from Baseline to Week 8.

Table 3

	24hr Pain	Intercourse Pain	CST-Vestibule	Algesiometer	Pelvic Muscle Pain
24hr Pain	1.00	0.33* (0.15, 0.48)	0.22 (0.03, 0.39)	-0.18 (-0.36, 0.00)	-0.03 (-0.12, 0.15)
Intercourse Pain	0.33* (0.15, 0.48)	1.00	-0.01 (-0.35, 0.22)	0.00 (-0.24, 0.24)	-0.22 (-0.44, 0.01)
CST-Vestibule	0.22 (0.03, 0.39)	-0.01 (-0.35, 0.22)	1.00	-0.35* (-0.50, -0.17)	0.28* (0.10, 0.44)
Algesiometer	-0.18 (-0.36, 0.00)	0.00 (-0.24, 0.24)	-0.35* (-0.50, -0.17)	1.00	0.02 (-0.16, 0.20)
Pelvic Muscle Pain	-0.03 (-0.12, 0.15)	-0.22 (-0.44, 0.01)	0.28* (0.10, 0.44)	0.02 (-0.16, 0.20)	1.00

	SF-McGill Total	Brief Pain Inventory	Neuropathic Pain Scale
24hr Pain	0.12 (-0.06, 0.30)	0.55* (0.40, 0.66)	0.29* (0.11, 0.45)
Intercourse Pain	0.40* (0.18, 0.58)	0.38* (0.16, 0.57)	0.40* (0.17, 0.58)
CST-Vestibule	0.10 (-0.08, 0.30)	0.16 (-0.02, 0.33)	0.17 (-0.01, 0.34)
Algesiometer	0.00 (-0.19, 0.18)	-0.16 (-0.34, 0.02)	-0.16 (-0.33, 0.03)
Pelvic Muscle Pain	0.14 (-0.04, 0.32)	0.01 (-0.17, 0.19)	0.01 (-0.20, 0.17)

* P<0.01