Arch Sex Behav (2018) 47:651–660
https://doi.org/10.1007/s10508-017-1076-4

ORIGINAL PAPER

Does Repeated Testing Impact Concordance Between Genital and Self-Reported Sexual Arousal in Women?

Julia Velten[1] · Meredith L. Chivers[2] · Lori A. Brotto[3]

Received: 22 November 2016 / Revised: 20 March 2017 / Accepted: 1 September 2017 / Published online: 19 September 2017

Abstract Women show a substantial variability in their genital and subjective responses to sexual stimuli. The level of agreement between these two aspects of response is termed sexual concordance and has been increasingly investigated because of its implications for understanding models of sexual response and as a potential endpoint in clinical trials of treatments to improve women's sexual dysfunction. However, interpreting changes in sexual concordance may be problematic because, to date, it still is unclear how repeated testing itself influences sexual concordance in women. We are aware of only one study that evaluated temporal stability of concordance in women, and it found no evidence of stability. However, time stability would be necessary for arguing that concordance is a stable individual difference. The main goal of this study was to investigate the test–retest reliability of sexual concordance in a sample of 30 women with sexual difficulties. Using hierarchical linear modeling, we found that sexual concordance was not influenced by repeated testing 12 weeks later, but showed test–retest reliability suggesting temporal stability. Our findings support the hypothesis that sexual concordance is a relatively stable individual difference and that changes in sexual concordance after treatment or experimental conditions could, therefore, be attributed to effects of those conditions.

✉ Lori A. Brotto
lori.brotto@vch.ca

1 Mental Health Research and Treatment Center, Ruhr-Universität Bochum, Bochum, Germany

2 Department of Psychology, Queen's University, Kingston, ON, Canada

3 Department of Obstetrics and Gynaecology, University of British Columbia, 6th Floor DHCC, 2775 Laurel Street, Vancouver, BC V5Z 1M9, Canada

Keywords Sexual concordance · Sexual response · Test–retest reliability · Low sexual desire · Vaginal photoplethysmography

Introduction

Female sexual arousal includes emotional, behavioral, and physiological responses to sexual stimulation that are interrelated, but also at least partly independent (Laan & Everaerd, 1995). The genital sexual response is a neurovascular process that includes increased blood flow to the genitals, resulting in a swelling of the clitoris, vagina, and vulva as well as vaginal lubrication (Levin & Wylie, 2008; Traish, Botchevar, & Kim, 2010). The subjective dimension of sexual arousal (SSA) reflects an individual's experience of being sexually aroused or "turned-on" (Chivers et al., 2010). The level of agreement between subjective and genital arousal is termed sexual concordance (Chivers et al., 2010), and scientists have grown increasingly interested in sexual concordance because of the implications for models of sexual response more broadly.

While, for instance, the linear model of sexual response by Masters and Johnson (1970) suggests that subjective and genital arousal run in parallel, the more recently described circular model of sexual response (Basson, 2001) emphasizes that both levels can be more or less disconnected in women. The substantial variation in female sexual concordance has elicited debate in the field. While some women exhibit very high sexual concordance, others show very low or even negative associations between subjective and genital arousal (Chivers et al., 2010). Though methodological and instrumentational artifacts have been suggested to account for this variability (Kukkonen, 2015), a large meta-analysis based on data from 2505 women has ruled out this assertion given that the large variability in concordance has also been found with other physiological measures, such as thermography. In addition, Bouchard, Chivers, and Pukall (2017) found positive correlations between different measurements of genital sexual arousal, namely vaginal vasocongestion and vulvar blood flow, suggesting convergence across physiological measures.

Compared to men, women, on average, demonstrate much lower, and more variable, sexual concordance (Chivers et al., 2010). Increasingly, researchers have been exploring moderators of sexual concordance such as sexual inhibition/excitation (Clifton, Seehuus, & Rellini, 2015; Velten, Scholten, Graham, Adolph, & Margraf, 2016), in order to better understand the multiple determinants of women's sexual response (Chivers et al., 2010). They have also been investigating the relevance of genital and subjective arousal as well as sexual concordance for women's sexual functioning (Brody, Laan, & Van Lunsen, 2003; Meston, Rellini, & McCall, 2010; Sarin, Amsel, & Binik, 2016). However, findings are equivocal in that others have not found concordance to be associated with, for instance, sexual dysfunctions in women (Brotto, Chivers, Millman, & Albert, 2016).

Assessment of Genital Sexual Arousal in Women

Though a variety of devices have been developed and tested, women's genital sexual response is most commonly assessed with a vaginal photoplethysmograph, a tampon-shaped device that measures blood flow in the vaginal walls. Vaginal pulse amplitude (VPA) is a reliable and valid measure of women's genital arousal (Laan & Everaerd, 1995; Suschinsky, Lalumière, & Chivers, 2009), although its use is not without problems. These include: the lack of an absolute VPA scale that can be compared across women and an inability to directly compare the magnitude of response to a measure of genital arousal in men (Kukkonen, Binik, Amsel, & Carrier, 2007). Nevertheless, the validity of the VPA measure has been shown across many studies (Chivers et al., 2010; Huberman & Chivers, 2015; Kukkonen, 2015). For example, one pharmacological study found women's VPA to predict their responses to sildenafil citrate on orgasm functioning (Basson & Brotto, 2003). Women's VPA response patterns are diverse: While most women respond with greater VPA to erotic stimuli (e.g., erotic fantasies or audiovisual media), some women show either no VPA response or even a decrease in response following erotic stimulus exposure (Chivers et al., 2010).

Assessment of Subjective Sexual Arousal in Women

The most convenient method to assess SSA is to ask participants to report the level of sexual arousal that they are experiencing before and after viewing a sexual stimulus (e.g., To what degree do you feel sexually aroused right now?). The advantage of this method is ease of administration and applicability for both questionnaire and laboratory studies. Another means of capturing self-reported sexual arousal is with the use of a continuous measure of SSA during erotic stimulus presentation. This approach allows one to elucidate the time course of sexual arousal and reduces response biases (Laan & Everaerd, 1995). Additionally, continuous ratings may be less susceptible to impression management biases than post-stimulus ratings alone (Huberman, Suschinsky, Lalumière, & Chivers, 2013). Interpretation of continuously measured SSA is, however, not without problems. Most devices that are used to continuously measure SSA, such as the arousometer (Brotto, Seal, & Rellini, 2012; Meston et al., 2010), allow participants to indicate their level of mental arousal on a scale from, for example, −2 (*turned off*), to 0 (neutral) to 7 (*maximum level of arousal*). As women differ in their awareness of sexual arousal cues (Brotto et al. 2016), some women might experience subtle changes in their SSA that cannot be adequately captured by the relatively low resolution of the scale. Indeed, most experimental designs do not involve prompting women to "check in" with their level of mental arousal and make adjustments to their ratings, but rather, they are instructed only at the outset to indicate any perceived level of mental sexual arousal. Women may, however, differ in their interpretation of this task or may forget to indicate their arousal for certain periods of time, which may also impact their SSA values. Though some researchers have addressed this by reminding research participants to report on their level of subjective sexual arousal during the testing period (Boyer, Pukall, & Chamberlain, 2013; Kukkonen, Binik, Amsel, & Carrier, 2010), this may elicit an experimental demand on participants and impact their SSA ratings.

Operationalization of Sexual Concordance

As noted earlier, sexual concordance is operationalized as the correlation between VPA and SSA. Because of heterogeneous VPA response patterns (i.e., decreased, unchanged, or increased VPA in sexual arousal conditions), experimental designs that calculate sexual concordance using between-subject correlations of VPA and SSA have to be interpreted with caution (Prause & Janssen, 2006). Alternatively, within-subject correlations can be calculated using contiguously measured SSA and VPA (Meston et al., 2010; Rellini, McCall, Randall, & Meston, 2005). With respect to these within-subject correlations, methodological problems arising from unstandardized use of the arousometer may still contribute to spurious concordance estimates. Often, no specific instructions are given on how often the arousometer should be adjusted. There must be at least one change in values for both SSA and VPA to allow one to calculate within-subject concordance estimates, such that, if a participant does not indicate any change in their SSA with the erotic stimulus (i.e., there is no increase or decrease in their level of perceived sexual arousal), a within-subject correlation cannot be produced. In this instance, the researcher may label this individual as a "non-responder" and exclude them from analyses. However, this, too, is potentially problematic because her "non-response" may be important information reflecting her experience or an inadequate stimulus. Unfortunately, many researchers do not report on the number of such excluded subjects.

Another statistical method for estimating sexual concordance using contiguously measured subjective and genital sexual arousal is hierarchical linear modeling (HLM), also called multilevel modeling. HLM can be used to analyze repeated measurement data nested within individuals. The advantages of HLM over within-subjects analyses include: (1) It allows the analysis of repeated measurements taken within subjects, which are neither conceptually nor typically statistically independent, (2) it simultaneously estimates variance and covariance components for within- and between-subject levels of analysis, (3) it improves estimate of effects within subjects, and (4) it has lower Type 1 error rates (Gelman & Hill, 2007; Raudenbush & Bryk, 2002). HLM overcomes the issue of individual differences in VPA response during an erotic stimulus condition because the analysis is conducted within participants and the between-subject comparisons are based on the strength of the within-subject relationships between VPA and continuously measured SSA. However, one shortcoming in the application of HLM for the assessment of sexual arousal concordance is that most researchers only use data that were acquired during an erotic stimulus presentation, and not those collected during a baseline period or during a neutral film (Clifton et al., 2015; Velten et al., 2016), which may lead to biased concordance estimates.

To our knowledge, the effect of repeated measurement (i.e., different assessment points) on sexual concordance has only been reported in one study including sexually healthy participants (Suschinsky & Lalumière, 2011). In that study, 20 men and 18 women were tested 1 month apart. Among women ($M_{age} = 21.6$), there was no significant correlation among concordance scores between sessions 1 and 2 ($r = -.20$, $p = .44$).

Demonstration of the temporal stability of sexual concordance is crucial for study designs that measure concordance before and after an experimental manipulation or a treatment in order to infer whether changes in concordance reflect the impact of a study manipulation or whether changes reflect typical variation in concordance. Demonstrating stability of sexual concordance would also support the hypothesis that concordance may be an individual difference in sexual response, potentially related to sexual functioning.

Additionally, the patterns of sexual concordance during an erotic stimulus are yet to be thoroughly examined and reported. Though the meta-analysis concluded that greater stimulus length may be associated with a larger concordance estimate (Chivers et al., 2010), it may also be that sexual concordance decreases over time, given the dynamics of vaginal vasocongestion, and the finding that VPA tends to increase sharply at the beginning of an erotic stimulus, reaches a peak after only 21 s of an erotic stimulus (Huberman, Dawson & Chivers, in press), and then stabilizes. On the other hand, levels of SSA show greater fluctuation over the course of a stimulus and may take longer to reach maximum levels (Huberman & Chivers, 2015). While these findings do not directly contradict the Chivers et al. (2010) meta-analysis, they call into question whether longer stimulus presentations often lead to higher concordance estimates.

Current Study

The main goal of the current study was to assess the test–retest reliability of sexual arousal concordance in a sample of women with sexual difficulties. Additional goals included exploring three potential methodological moderators of concordance: (1) frequency of reporting change in subjective sexual arousal; (2) the magnitude of change in VPA from baseline genital response; and (3) the stimulus length. These moderators were selected for analysis given that different experimental designs differ on these factors, yet it is unknown whether they directly impact concordance. For example, some studies provide instruction to participants to report on their SSA, resulting in more frequent changes in the levels of SSA, whereas other designs do not provide instruction to participants during the testing. Whether the frequency of reporting changes in SSA impacts sexual concordance is unknown. Furthermore, the duration of erotic film exposure varies widely between 90 s and 15 min, and there is evidence that the impact of stimulus duration on concordance may be gender specific (Chivers et al., 2010). A better understanding of methodological moderators may inform best practices in using sexual concordance as an outcome variable.

Method

Participants

Participants were part of a larger study evaluating outcomes of group mindfulness-based sex therapy on sexual desire, sexual response, and affect (Brotto & Basson, 2014). Women seeking treatment for sexual desire and/or arousal concerns and who lived in a large metropolitan Canadian city were eligible to participate. Inclusion criteria were: age between 19 and 65 years, fluent in English, willingness to participate in four group sessions and to complete daily homework, and ability to participate in pre- and post-treatment assessments (consisting of both self-report questionnaires and in-laboratory psychophysiological sexual arousal assessments) at three time points. Women with difficulties in achieving orgasm were also included as long as those were not experienced as more distressing than the desire and/or arousal concerns. At the time of the phone screen, we excluded any woman with dyspareunia (chronic genital pain not resolved with a personal lubricant).

The present set of analyses focused only on a subset of 30 women who participated in two pre-treatment psychophysiological testing sessions, separated by approximately 3 months. Their mean age was 42.0 years (SD = 11.2; range 28–63). A total of 83.3% participants were in committed relationships, 10.0% were casually dating, and 6.6% were single or separated. Mean relationship length was 14.7 years (SD = 12.4). Most participants were of Euro-Canadian descent (80%) followed by East-Asian descent (16.7%). This was a highly educated group in that

76.7% had some post-secondary education, and 13.3% had an advanced graduate degree. Seventeen participants met the criteria for the DSM-IV-TR (American Psychiatric Association, 2000) diagnosis of Hypoactive Sexual Desire Disorder, five met the criteria for Female Sexual Arousal Disorder, and eight met the criteria for both disorders.

Measures

Subjective Sexual Arousal (SSA)

Subjective sexual arousal was measured in two ways: both with a discrete and a continuous measure. Discrete SSA was measured by having the participants rate their response to the erotic stimulus with a 34-item version of the Film Scale (Heiman & Rowland, 1983). The items listed different emotional or physiological responses to film stimuli and were answered on a seven-point Likert scale from 1 (*not at all*) to 7 (*intensely*). A score of mental or subjective arousal was calculated using the following six items of the scale: sensuous, a desire to be close to someone, sexy, loving, sexually attractive, and easy to arouse.

Continuous SSA was measured during stimulus presentation with an arousometer that was constructed by a local engineer modeled after the one described by Rellini et al. (2005). The device consisted of a computer optic mouse mounted on a plastic track with ten intervals, affixed to the armrest of a reclining chair. Women were instructed to use the arousometer continuously to indicate changes in mental sexual arousal on a 10-point scale from −2 (*sexually turned off*), 0 (*neutral/no sexual arousal*), and 7 (*highest level of arousal*) during the entire duration of the erotic films. Similar devices have been used to assess subjective sexual arousal in previous laboratory studies (Clifton et al., 2015; Rellini et al., 2005).

Physiological Sexual Arousal

Vaginal pulse amplitude (VPA) was used as a measure of genital sexual response using a vaginal photoplethysmograph equipped with an orange-red spectrum light source (Behavioral Technology Inc., Salt Lake City, UT) during the experimental procedure. The signal was sampled at 200 Hz, band-pass-filtered (0.5–30 Hz), and recorded continuously during the stimulus presentation. Data were acquired and processed using a data acquisition unit Model MP150 and AcqKnowledge version 3.8.1 (BIOPAC Systems, Inc., Santa Barbara, CA). A trained research assistant visually inspected the data and performed smoothing of movement artifacts prior to data reduction and analysis.

Procedure

An experienced research assistant with specialized training in the assessment and diagnosis of sexual dysfunctions conducted telephone screening to assess the eligibility of potential participants. Following the telephone screen, she mailed a consent form to eligible women interested in participating. Women were also mailed a package of questionnaires and asked to return them completed at the time of their sexual arousal assessment. The sexual arousal assessment took place in a sexual psychophysiology laboratory, located at a university hospital. Following signed consent, participants were tested by a female researcher. Participants were first shown the vaginal photoplethysmograph and encouraged to ask any questions about how to insert it. Participants were also reminded to use the arousometer to capture their subjective sexual arousal throughout the erotic film presentation. The researcher instructed participants to monitor their subjective feelings of sexual arousal by using this device. She also explained that "subjective feelings of sexual arousal" means how mentally sexually aroused the participant feels while watching the film.

The female researcher then left the room while participants inserted the probe and informed the researcher via intercom of their readiness. The researcher then initiated the video sequence. Women watched a 3-min documentary about Hawaii followed by a 7-min erotic film that depicted a male–female couple engaging in foreplay, cunnilingus, fellatio, and penile–vaginal intercourse. They were then instructed to remove the probe and meet the researcher in a separate room. After a debriefing period, the researcher promptly disinfected the probe in a solution of Cidex OPA (ortho-phthalaldehyde 0.55%), a high-level disinfectant (Advanced Sterilization Products, Irvine, CA, USA), following each session. Using different video stimuli, the in-laboratory testing procedure was repeated 12 weeks later at the end of the wait-list period. The sequence of films was randomly chosen for each participant. All procedures were approved by the Clinical Research Ethics Board at the University of British Columbia. All procedures were carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki (2013).

Data Reduction and Analyses

Off-line, plethysmography data were band-pass-filtered (0.5–20 Hz). Then, in agreement with standardized procedures, movement artifacts, defined by sudden and drastic changes in pulse amplitude, were visually identified and deleted by being marked as missing for data analysis (Prause & Janssen, 2006). Data from the arousometer and the vaginal photoplethysmograph were averaged across 30-s intervals, resulting in a total of 26 data points; 13 per session for each condition, per participant.[1]

Stimulus duration was operationalized as the number of 30-s bins. To estimate the frequency of reporting continuous sexual arousal, the number of arousometer movements throughout the stimulus was operationalized as the number of different arousometer values throughout the stimulus, using change of SSA relative to the previous 30-s time interval. To increase comparability

[1] The 14th bin was not analyzed due to missing data.

of our results to previous studies, we calculated the mean subjective and genital response for each stimulus category. Mean genital response, calculated using the average VPA of the 13, 30-s bins, is reported as mean VPA during the erotic stimulus (in mV). The mental arousal subscale of the Film Scale and the mean of the continuously measured SSA are also reported. Additionally, within-subject correlations were calculated. General linear modeling for repeated measurements was used to investigate differences between participants and within-subject changes over time. SPSS 23 (IBM, 2012) was used for these analyses.

Hierarchical linear modeling (HLM) analyses were conducted in the R environment (R Development Core Team, 2010) using the package nlme (Pinheiro, Bates, DebRoy, & Sarkar, 2015). For this study, the major advantage of HLM was that it conducts a within-subject analysis of the relationship between VPA and continuously measured SSA and uses the coefficients that describe this relationship (i.e., slope and intercept) as outcome variables to test differences between participants. Thus, the predictive qualities of Level 2 predictors (i.e., inter-individual differences such as number of arousometer movements or VPA change from baseline) as well as Level 1 predictors (i.e., intra-individual changes such as VPA, SSA, stimulus duration or assessment point) could be tested. To assess the relationship between SSA and VPA, the intercepts and slopes of the linear regressions of SSA predicted by VPA, and vice versa, were computed for each participant and then used as outcome variables for the linear model to assess the impact of the different predictor variables (e.g., arousometer movements or assessment point). Six different models were constructed to test our hypotheses. For example, the model used to investigate the impact of arousometer movements and VPA on SSA used the following formula:

$$\begin{aligned} y(\mathrm{SSA})_{ij} = {} & \beta_{0i} + \beta_1(\mathrm{VPA})_{ij} + \beta_2(\text{Arousometer movements})_i \\ & + \beta_3(\text{Arousometer movements})_i * (\mathrm{VPA})_{ij} \\ & + r_{0i} + r_{1i} * (\mathrm{Time})_{ij} + r_{2i} * (\text{Assessment point})_i + \varepsilon_{ij} \end{aligned}$$

where SSA $_{ij}$ is the ith individual's SSA at the jth time point.

In this model, β_{0i} is the individual-specific intercept, β_1 is the individual-specific slope, r_{0i} describes the random intercept, r_{1i} and r_{2i} describe the random slope, and ε_{ij} indicates the residuals. Level 2 predictors were grand-mean-centered, and Level 1 predictors were group-mean-centered before data analysis. Data were estimated using maximum likelihood estimation. The slopes and intercepts for each model were allowed to vary across participants in order to take into account within-subject differences in the baseline levels and time courses of the continuously measured VPA or SSA. We specified the covariance matrices of all tested models as first-order autoregressive structures in order to fit our model to the correlation between the repeated measures within participants (for an overview, see Singer & Willett, 2003). When significant two-way interaction effects were identified, post hoc simple slope analyses were conducted to determine whether the slopes of the two predicted lines differed from zero (Cohen, Cohen, West, & Aiken, 2013).

Results

Descriptive Analyses

Table 1 provides an overview of the descriptive values for all relevant predictor and outcome variables. For both genital and subjective arousal, skewness was below a value of 2 and kurtosis below a value of 7. Hence, an indication for substantial non-normality was not found (Kim, 2013). Average SSA did not significantly differ between the first and second assessment points. The same was true for average VPA during the erotic film condition. In addition, within-subject correlations ($r[25] = .34$ [range $-.85 < r < .86$]) for the first and $r(23) = .30$ (range $-.90 < r < .91$) for the second assessment point) did not differ significantly between assessments. The mean difference in within-subject concordance between first and second assessment was 0.05 (SD = 0.58). Percent increase in VPA between neutral and erotic film condition was significantly greater at the first assessment point (75.61 vs. 58.0%, respectively). There was no significant difference in the number of arousometer movements between the two assessment sessions. All sexual arousal measurements significantly differed between participants.

A number of different VPA response patterns between neutral and erotic conditions were observed: Across all participants and both assessment points, during six sessions (10.0%), VPA actually decreased from the neutral to the erotic film and in four sessions (6.7%) did not change between the conditions. In 14 sessions (23.3%), VPA increased by at least 10% compared to the baseline, in four sessions (6.7%) increased by at least 30%, and in 32 sessions (53.5%) increased at least 50% compared to baseline. Complete arousometer data were available for 56 sessions across all participants and both assessment points. There was substantial variability with respect to the number of arousometer movements during the erotic stimulus presentation: during four (7.1%) sessions, the arousometer was not moved at all, in 14 sessions (25%) was moved once or twice, in 12 sessions (21.4%) three to four times, in nine sessions (16.1%) five times, and in 17 sessions (30.4%) moved six times or more.

Correlational Analyses

Bivariate Pearson correlation coefficients between predictor and outcome variables were calculated to enable comparisons with studies using similar statistical methods (Table 2).

Average levels of SSA were highly correlated with the number of arousometer movements ($.43 < r < .63$, $p < .01$) and a discrete measure of SSA assessed using the Film Scale (.58 <

Table 1 Sexual arousal measurements during erotic stimulus presentation at two different assessment points

Assessment point		Min	Max	*M*	SD	Within subjects *p*	Between subjects *p*
Mental sexual arousal, (arousometer, mean)	1	−0.62	3.06	1.49	1.39	.709	<.001
Range = −2 to 7	2	0.00	4.68	1.52	1.45		
Arousometer movements	1	0.00	6.00	3.59	2.03	.517	<.001
Range = 0–6	2	0.00	6.00	3.93	2.16		
Mental sexual arousal (film scale)	1	1.00	5.50	2.36	1.37	.873	<.001
Range = 1–7	2	1.00	6.83	2.33	1.40		
Genital response (VPA in mV, mean)	1	.012	.120	0.051	0.023	.723	<.001
	2	.015	.089	0.050	0.020		
Change in genital response (percentage increase from baseline)	1	−44.38	295.13	75.61	74.61	.014	<.001
	2	−85.50	211.20	58.00	67.22		
Within-subject correlations	1	−.85	.86	0.34	0.51	.560	<.001
	2	−.90	.91	0.30	0.61		

Columns 5 and 6 reflect the results of general linear modeling for repeated measurements to assess differences between assessment points (within subjects) and participants (between subjects)

Table 2 Bivariate correlations between different sexual arousal measurements at both assessment points

	Assessment point		1	2	3	4	5	6	7	8	9	10	11	12
1	Mental sexual arousal (arousometer, mean)	1	1	.49*	.63**	.58**	.58**	.50**	−.17	−.12	−.05	.18	.20	−.12
2		2		1	.18	.43*	.44*	.78**	−.32	−.15	−.31	−.09	.09	.10
3	Arousometer movements	1			1	.47*	.35	.31	.09	.08	−.18	.14	−.02	.04
4		2				1	.34	.33	.07	.26	.11	.15	.25	.06
5	Mental sexual arousal (film scale)	1					1	.68**	.00	−.05	−.28	−.28	.03	−.06
6		2						1	−.02	−.10	−.25	−.23	.21	.11
7	Genital response (VPA in mV, mean)	1							1	.42*	.45*	−.07	−.03	−.35
8		2								1	.36	.31	.10	.31
9	Genital response (% increase from baseline)	1									1	.55**	.61**	.14
10		2										1	.51**	.48*
11	Within-subject correlations	1											1	.48*
12		2												1

The first assessment point reflects the first testing session and the next one reflects the second testing session, for all variables presented

* $p < .05$; ** $p < .01$

$r < .78, p < .01$) in each of the assessments. None of the measures of SSA were significantly correlated with measures of genital sexual arousal or within-subjects concordance. The only aspect of sexual response that was correlated with within-subjects concordance was VPA increase measured in percent change ($.48 < r < .68$, $p < .01$). Within-subjects concordance estimates at both assessment points were positively correlated ($r[22] = .48$, $p < .01$).

Hierarchical Linear Modeling

In Model 1 and Model 2, we investigated the influence of repeated testing and stimulus duration on sexual arousal and concordance. In Model 1, SSA was a significant predictor of VPA. Additionally, an interaction term of duration and SSA was negatively predictive of VPA, indicating that the predictive value of SSA on VPA diminished with increasing duration of the stimulus presentation. Session and any interaction term including this variable were not predictive of VPA. In Model 2, VPA was a significant predictor of SSA. Moreover, duration was a positive predictor, indicating that SSA increased over the duration of the erotic film presentation. Session was not a significant predictor of SSA (Table 3).

In Model 3 and Model 4, we investigated the impact of the percentage of change of VPA compared to baseline VPA on sexual response and concordance. In both models, VPA and SSA were significantly predictive of each other. Not surprisingly, the percentage increase in VPA compared to the baseline predicted greater VPA during the erotic stimulus presentation. The significant interac-

Table 3 Prediction of genital/subjective sexual arousal by time course of stimulus presentation, subjective/genital sexual arousal, and assessment point

	Model 1				Model 2			
Outcome	Genital sexual arousal (VPA)				Subjective sexual arousal (SSA)			
Fixed effects	*b*	SE (*b*)	*t* (*df*)	*p*	*b*	SE (*b*)	*t* (*df*)	*p*
Duration	6.65^{-04}	3.52^{-04}	1.89 (691)	.059	0.13	0.02	5.19 (691)	<.001
SSA/VPA	14.75^{-04}	4.78^{-04}	3.08 (691)	.002	9.71	3.31	2.93 (691)	.004
Assessment point	0.83^{-04}	21.50^{-04}	0.04 (691)	.969	0.03	0.09	0.29 (691)	.772
Duration * SSA/VPA	-4.88^{-04}	1.16^{-04}	−4.19 (691)	<.001	−0.36	0.73	−0.50 (691)	.618
Duration * assessment point	-0.99^{-04}	2.28^{-04}	−0.43 (691)	.666	0.00	0.02	−0.01 (691)	.993
SSA/VPA * assessment point	0.04^{-04}	4.66^{-04}	0.01 (691)	.992	−2.73	3.25	−0.84 (691)	.401
Duration * SSA/VPA * assessment point	0.59^{-04}	1.12^{-04}	0.52 (691)	.600	0.34	0.70	0.48 (691)	.631

Table 4 Prediction of genital/subjective sexual arousal by subjective/genital sexual arousal and VPA increase

	Model 3				Model 4			
Outcome	Genital sexual arousal (VPA)				Subjective sexual arousal (SSA)			
Fixed effects	*b*	SE (*b*)	*t* (*df*)	*p*	*b*	SE (*b*)	*t* (*df*)	*p*
SSA/VPA	20.12^{-04}	47.73^{-05}	4.22 (695)	<.001	11.88	3.43	3.47 (695)	<.001
VPA increase[a]	0.96^{-04}	3.69^{-05}	2.60 (695)	.009	0.00	0.00	0.39 (695)	.695
SSA/VPA * VPA increase[a]	0.24^{-04}	0.70^{-05}	3.38 (695)	.001	−0.01	0.04	−0.22 (695)	.827

[a] VPA increase in percent compared to baseline level

tion term in Model 3 indicated that sexual concordance was greater in women with a greater VPA increase compared to the baseline level. A post hoc simple slope analysis showed that, in women with a VPA decrease or a smaller increase compared to baseline (one SD below average), SSA did not predict VPA ($b = -0.0010$, SE = 0.0011, $t = -0.97$, $p = .330$), whereas in women with a greater VPA increase compared to baseline (one SD above average) SSA was a positive predictor of VPA ($b = 0.0075$, SE = 0.0012, $t = 6.09$, $p < .001$). VPA increases from baseline, and the interaction term was not predictive of SSA in Model 4 (Table 4).

In Model 5 and Model 6, we investigated how the number of times the arousometer was moved throughout the stimulus presentation predicted sexual arousal or moderated concordance. In Model 5, SSA was again a significant predictor of VPA. Additionally, the number of arousometer movements also predicted VPA. The interaction term between arousometer movements and SSA showed a trend toward predicting VPA. On average, 3.75 (SD = 2.08) arousometer movements per session could be identified. The post hoc simple slope analysis showed that in women with no or few arousometer movements (one SD below average) SSA did not predict VPA ($b = -.0032$, SE = .0024, $t = 1.33$, $p = .185$), while in women who used the arousometer more often (one SD above average) SSA was a positive predictor of VPA ($b = .0026$, SE = .0010, $t = 2.50$, $p = .007$). In Model 6, VPA was again a significant predictor of SSA. Moreover, SSA was significantly greater in participants that moved the arousometer more often during stimulus presentation. No significant interaction effect was found (Table 5).

Discussion

The main aim of this study was to investigate the impact of repeated testing on the agreement of genital and subjective sexual arousal in a sample of women with sexual difficulties. Another objective was to evaluate the relevance of other factors related to study setting and method, namely the frequency of reporting continuous sexual arousal, the magnitude of change in VPA from baseline genital response, and stimulus length.

Our study provides evidence of the test–retest reliability of sexual concordance and indicates that repeated testing did not increase or decrease sexual concordance in women, at least among women with sexual dysfunctions. Within-subject correlations between contiguously measured VPA and SSA were comparable to those measured in other studies (Chivers et al., 2010) and did not differ significantly between the first and second assessment point. In addition, the multilevel analysis showed that assessment point did not predict either VPA or SSA, nor moderate sexual concordance estimates.

In contrast, a previous study investigating the time stability of sexual concordance in women found no significant stability other

Table 5 Prediction of genital/subjective sexual arousal by time, subjective/genital sexual arousal, and arousometer use

	Model 5				Model 6			
Outcome	Genital sexual arousal (VPA)				Subjective sexual arousal (SSA)			
Fixed effects	*b*	SE (*b*)	*t* (*df*)	*p*	*b*	SE (*b*)	*t* (*df*)	*p*
SSA/VPA	3.09^{-03}	8.63^{-04}	3.58 (695)	<.001	11.18	3.11	3.67 (695)	<.001
Arousometer use[a]	2.90^{-03}	12.26^{-04}	2.38 (695)	.018	0.12	0.04	2.58 (695)	.010
SSA/VPA * arousometer use[a]	-0.77^{-03}	4.31^{-04}	−1.78 (695)	.076	1.04	1.18	0.73 (695)	.467

[a] Number of arousometer movements across 30-s intervals

the course of 1 month (Suschinsky & Lalumière, 2011). However, this study included only sexually functional participants and differed from ours in important methodological aspects, such as participant age, erotic stimuli (erotic narratives vs. videos), and statistical analysis used. Studies including both sexually functional and dysfunctional participants may help to clarify which factors (i.e., sexual function, age, methods) impact the time stability of sexual concordance in women.

Women's self-reported and genital sexual response, as well as their concordance estimates, ranged considerably across the participants. In line with other studies, our study therefore supports the notion of great variability in women's sexual responses to erotic stimuli as well as in their concordance levels (Chivers et al., 2010).

A secondary goal of the study was to explore three moderators of concordance, the first of which was frequency of reporting change in subjective sexual arousal. We found that women not only differed in the level of subjective arousal that they were experiencing, but also that there was substantial variability in the frequency of usage of the arousal measurement device throughout a testing session. Greater number of arousometer movements was not only associated with greater levels of SSA (Model 6), but was also predictive of VPA (Model 5). The number of arousometer movements may be influenced by how often women perceived changes in their subjective arousal and how they interpreted the task of adjusting their arousal continuously during an erotic condition. The number of movements may also reflect third variable confounds, such as impression management (e.g., Huberman et al., 2013), which has been found to significantly predict distinct measures of subjective arousal before and after an erotic stimulus presentation. As arousometer usage is an important predictor of arousal, the impact of different arousal measurement devices (e.g., computer mouse, potentiometer), scale resolution, visual feedback or on-screen instructions should be investigated further, particularly since there are differences across experimental laboratories in the instructions given to research participants about how often to monitor their SSA.

The magnitude of change in VPA was also explored as a moderator of concordance. In Model 3 and Model 4, we investigated the impact of different VPA response patterns on sexual arousal and concordance. These analyses were conducted to evaluate how VPA differences between baseline and erotic condition influenced SSA and VPA during erotic condition as well as the agreement between those two measures. As expected, greater increase in VPA from baseline was predictive of greater VPA during the erotic stimulus presentation. Interestingly, greater VPA response compared to baseline was also predictive of sexual concordance. This finding could also be explained by a range restriction of VPA in women with a less pronounced VPA response, which would attenuate the bivariate relationship to SSA. Among women who showed a greater VPA response, SSA was more strongly predictive of VPA. This finding implies that sexual concordance may be underestimated among women who do not show pronounced increases or decreases in their VPA responses to an erotic stimulus. On the other hand, studies that exclude women with VPA "non-response" may overestimate sexual concordance in women.

Erotic stimulus length was also explored as a potential moderator of concordance, and we found evidence for this to be relevant for women's sexual response and concordance. This stands in contrast to the findings of the concordance meta-analysis (Chivers et al., 2010) which found no significant relationship between stimulus length and concordance for women (but a positive association for men). We found that SSA significantly increased with the duration of the stimulus (Model 2).

This finding can be explained by at least two different factors: First, women may need time to tune in with their bodily sensations and to gradually become aware of physical sexual arousal (Huberman et al., in press). Second, sexual stimuli that were presented in this study became increasingly arousing and sexually explicit over time. Therefore, women may experience lower levels of arousal during the depiction of foreplay and kissing compared to the parts of the videos that showed oral or manual genital stimulation or penile-vaginal intercourse. In contrast, genital sexual arousal did not significantly increase with stimulus duration (Model 1). This finding is in line with previous studies which have shown that time between stimulus presentation and peak genital response measured with a vaginal photoplethysmograph may be as short as 21 s (Huberman et al., 2017). After this steep increase, VPA tends to stabilize and does not show significant fluctuations over time (Huberman & Chivers, 2015).

Sexual concordance—when operationalized as SSA predicting VPA—was moderated by stimulus duration with greater concordance levels at the beginning of the stimulus presentation (Model 1). This result can be explained by the differential effect of stimulus duration on SSA and VPA. In contrast, Chivers et al. (2010) reported that greater stimulus length was related to greater concordance in women, when concordance was assessed using within-subjects correlations. This finding has important implications for the interpretation of existing concordance data as well as for the planning of future studies. To validly assess sexual concordance, it may be useful to choose sexual stimuli that are long enough for women to perceive sexual arousal cues, which may be between 2 and 5 min (Huberman et al., 2017). In addition, continuously measured arousal data should not be aggregated over several minutes, so that the time course of both genital and subjective sexual arousal can be captured.

Future studies should also investigate how women differ in their stability or variability of sexual concordance over time. It is possible that, for some women, concordance is a more trait-like phenomenon, whereas for others, it may be more malleable, making it a potential target for sex therapy. One factor of potential relevance is interoception, which refers to an individual's ability to recognize internal physiological states, such as heart rate or respiration rate (Mehling et al., 2012). While some studies suggest that interoceptive awareness is relevant for sexual concordance in women, it remains an open question whether greater interoception is also related to greater stability of sexual concordance over time.

Limitations

Several limitations challenge the internal validity and generalizability of our findings. Our sample consisted of a relatively small number of women with different sexual concerns. Even though our sample was diverse with respect to age, most of our participants indicated Euro-Canadian ethnicity and a high level of education. Therefore, we cannot extend our findings to other ethnic groups or women with lower education. Furthermore, all participants met DSM-IV criteria for a sexual dysfunction. It is unknown if the finding of test–retest reliability in concordance would be found among a sample of women free of sexual concerns. The finding that women with and without sexual difficulties showed different effect sizes for concordance—even though those group differences were not statistically significant (Chivers et al., 2010)—leaves open the possibility that concordance reliability may be different for women without sexual concerns. Another limitation is that we only tested concordance at two assessment points, separated by 12 weeks. The impact of additional assessment points on concordance should be the focus of future studies.

Implications

There are clinical implications of the findings that can be taken into consideration. Given that concordance has been proposed as an endpoint in some clinical trials of an experimental treatment (e.g., Basson & Brotto, 2003; Brotto et al., 2016; Chivers & Rosen, 2010), the current findings suggest that changes in concordance associated with treatment are unlikely to be attributable to systematic measurement error produced by repeated measurements of subjective arousal. In other words, increases in concordance with a particular treatment may be more likely attributed to the treatment than to methodological artifacts. Treatments designed to improve a woman's interoceptive ability, or to align the self-reported with the physiological responding, may be especially appropriate to employ concordance as an endpoint.

Taken together, our results suggest temporal stability of concordance between genital and self-reported sexual arousal when women were tested in a controlled laboratory setting separated by 12 weeks. We also found evidence that stimulus length and VPA response patterns moderated concordance values, and frequency of monitoring subjective arousal showed a trend toward predicting concordance. We strongly support the use of HLM as a statistical alternative to both between-subjects and within-subjects analyses in studies of sexual arousal concordance, as they allow the investigator to explore the role of these and other moderators.

Acknowledgements Funding was provided by Women and Children's Health Research Institute (CA) and Canadian Institutes of Health Research.

Compliance with Ethical Standards

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

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

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: American Psychiatric Association.

Basson, R. (2001). Human sex-response cycles. *Journal of Sex and Marital Therapy, 27,* 33–43. doi:10.1080/00926230152035831.

Basson, R., & Brotto, L. A. (2003). Sexual psychophysiology and effects of sildenafil citrate in oestrogenised women with acquired genital arousal disorder and impaired orgasm: A randomised controlled trial. *BJOG: An International Journal of Obstetrics & Gynaecology, 110,* 1014–1024. doi:10.1111/j.1471-0528.2003.02438.x.

Bouchard, K. N., Chivers, M. L., & Pukall, C. (2017). Effects of genital response measurement device and stimulus characteristics on sexual concordance in women. *Journal of Sex Research.* doi:10.1080/00224499.2016.1265641.

Boyer, S. C., Pukall, C. F., & Chamberlain, S. M. (2013). Sexual arousal in women with provoked vestibulodynia: The application of laser Doppler imaging to sexual pain. *Journal of Sexual Medicine, 10,* 1052–1064. doi:10.1111/j.1743-6109.2012.02855.x.

Brody, S., Laan, E., & Van Lunsen, R. H. W. (2003). Concordance between women's physiological and subjective sexual arousal is associated with consistency of orgasm during intercourse but not other sexual behavior. *Journal of Sex and Marital Therapy, 29,* 15–23.

Brotto, L. A., & Basson, R. (2014). Group mindfulness-based therapy significantly improves sexual desire in women. *Behaviour Research and Therapy, 57,* 43–54. doi:10.1016/j.brat.2014.04.001.

Brotto, L. A., Chivers, M. L., Millman, R. D., & Albert, A. (2016). Mindfulness-based sex therapy improves genital-subjective arousal concordance in women with sexual desire/arousal difficulties. *Archives of Sexual Behavior, 45,* 1907–1921. doi:10.1007/s10508-015-0689-8.

Brotto, L. A., Seal, B. N., & Rellini, A. (2012). Pilot study of a brief cognitive behavioral versus mindfulness-based intervention for women with sexual distress and a history of childhood sexual abuse. *Journal of Sex and Marital Therapy, 38,* 1–27. doi:10.1080/0092623X.2011.569636.

Chivers, M. L., & Rosen, R. C. (2010). Phosphodiesterase type 5 inhibitors and female sexual response: Faulty protocols or paradigms? *Journal of Sexual Medicine, 7,* 858–872. doi:10.1111/j.1743-6109.2009.01599.x.

Chivers, M. L., Seto, M. C., Teresa, L., Lalumie, M. L., Lalumière, M. L., Laan, E., & Grimbos, T. (2010). Agreement of self-reported and genital measures of sexual arousal in men and women: A meta-analysis. *Archives of Sexual Behavior, 39,* 5–56. doi:10.1007/s10508-009-9556-9.

Clifton, J., Seehuus, M., & Rellini, A. H. (2015). Testing cognitive predictors of individual differences in the sexual psychophysiological responses of sexually functional women. *Psychophysiology, 52,* 957–968. doi:10.1111/psyp.12423.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. London: Routledge.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and hierarchical/multilevel models*. Cambridge: Cambridge University Press.

Heiman, J. R., & Rowland, D. L. (1983). Affective and physiological sexual response patterns: The effects of instructions on sexually functional and dysfunctional men. *Journal of Psychosomatic Research, 27,* 105–116. doi:10.1016/0022-3999(83)90086-7.

Huberman, J. S., & Chivers, M. L. (2015). Examining gender specificity of sexual response with concurrent thermography and plethysmography. *Psychophysiology, 52,* 1382–1395. doi:10.1111/psyp.12466.

Huberman, J. S., Dawson, S. J., & Chivers, M. L. (in press). Examining the time course of genital and subjective sexual responses in women and men with concurrent plethysmography and thermography. *Biological Psychology*.

Huberman, J. S., Suschinsky, K. D., Lalumière, M. L., & Chivers, M. L. (2013). Relationship between impression management and three measures of women's self-reported sexual arousal. *Canadian Journal of Behavioural Science, 45,* 259–273. doi:10.1037/a0033397.

IBM. (2012). *IBM SPSS statistics for Macintosh*. Armonk, NY: Author.

Kim, H.-Y. (2013). Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics, 38,* 52–54. doi:10.5395/rde.2013.38.1.52.

Kukkonen, T. M. (2015). Devices and methods to measure female sexual arousal. *Sexual Medicine Reviews, 3,* 225–244. doi:10.1002/smrj.58.

Kukkonen, T. M., Binik, Y. M., Amsel, R., & Carrier, S. (2007). Thermography as a physiological measure of sexual arousal in both men and women. *Journal of Sexual Medicine, 4,* 93–105. doi:10.1111/j.1743-6109.2006.00399.x.

Kukkonen, T. M., Binik, Y. M., Amsel, R., & Carrier, S. (2010). An evaluation of the validity of thermography as a physiological measure of sexual arousal in a non-university adult sample. *Archives of Sexual Behavior, 39,* 861–873. doi:10.1007/s10508-009-9496-4.

Laan, E., & Everaerd, W. (1995). Determinants of female sexual arousal: Psychophysiological theory and data. *Annual Review of Sex Research, 6,* 32–76.

Levin, R. J., & Wylie, K. (2008). Vaginal vasomotion—Its appearance, measurement, and usefulness in assessing the mechanisms of vasodilatation. *Journal of Sexual Medicine, 5,* 377–386. doi:10.1111/j.1743-6109.2007.00669.x.

Masters, W. H., & Johnson, V. E. (1970). *Human sexual inadequacy*. Boston: Little, Brown, and Company.

Mehling, W. E., Price, C., Daubenmier, J. J., Acree, M., Bartmess, E., & Stewart, A. (2012). The Multidimensional Assessment of Interoceptive Awareness (MAIA). *PLoS ONE, 7,* e48230. doi:10.1371/journal.pone.0048230.

Meston, C. M., Rellini, A. H., & McCall, K. (2010). The sensitivity of continuous laboratory measures of physiological and subjective sexual arousal for diagnosing women with sexual arousal disorder. *Journal of Sexual Medicine, 7,* 938–950. doi:10.1111/j.1743-6109.2009.01548.x.

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2015). *nlme: Linear and nonlinear mixed effects models*. Retrieved from http://cran.r-project.org/package=nlme.

Prause, N., & Janssen, E. (2006). Blood flow: Vaginal photoplethysmography. In I. Goldstein, C. M. Meston, S. R. Davis, & A. M. Traish (Eds.), *Women's sexual function and dysfunction: Study, diagnosis and treatment* (pp. 359–367). Boca Raton, FL: Taylor & Francis.

R Development Core Team. (2010). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Rellini, A. H., McCall, K. M., Randall, P. K., & Meston, C. M. (2005). The relationship between women's subjective and physiological sexual arousal. *Psychophysiology, 42,* 116–124. doi:10.1111/j.1469-8986.2005.00259.x.

Sarin, S., Amsel, R., & Binik, Y. M. (2016). A streetcar named "derousal"? A psychophysiological examination of the desire–arousal distinction in sexually functional and dysfunctional women. *Journal of Sex Research, 53,* 711–729. doi:10.1080/00224499.2015.1052360.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford: Oxford University Press.

Suschinsky, K. D., & Lalumière, M. L. (2011). Category-specificity and sexual concordance: The stability of sex differences in sexual arousal patterns. *The Canadian Journal of Human Sexuality, 20,* 93–109.

Suschinsky, K. D., Lalumière, M. L., & Chivers, M. L. (2009). Sex differences in patterns of genital sexual arousal: Measurement artifacts or true phenomena? *Archives of Sexual Behavior, 38,* 559–573. doi:10.1007/s10508-008-9339-8.

Traish, A. M., Botchevar, E., & Kim, N. N. (2010). Biochemical factors modulating female genital sexual arousal physiology. *Journal of Sexual Medicine, 7,* 2925–2946. doi:10.1111/j.1743-6109.2010.01903.x.

Velten, J., Scholten, S., Graham, C. A., Adolph, D., & Margraf, J. (2016). Investigating female sexual concordance: Do sexual excitation and sexual inhibition moderate the agreement of genital and subjective sexual arousal in women? *Archives of Sexual Behavior, 45,* 1957–1971. doi:10.1007/s10508-016-0774-7.

World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *Journal of the American Medical Association, 310,* 2191–2194. doi:10.1001/jama.2013.281053.