Menopause: The Journal of The North American Menopause Society
Vol. 26, No. 6, pp. 659-664
DOI: 10.1097/GME.0000000000001287

Genitourinary syndrome of menopause symptom severity and impact outcome measures: are they reliable and correlated?

Joanie Mercier, PT, MSc,[1] Mélanie Morin, PT, PhD,[2] Barbara Reichetzer, MD, MSc,[3] Marie-Claude Lemieux, MD,[4] Samir Khalifé, MD,[5] and Chantale Dumoulin, PT, PhD[6]

Abstract

Objective: The Most Bothersome Symptom Approach (MBSA) assesses symptom severity of genitourinary syndrome of menopause (GSM), and identifies the symptom with the greatest burden. The Atrophy Symptoms Questionnaire (ASQ) assesses the impact of GSM symptoms on the activities of daily living. The psychometric properties of these outcomes remain to be completely assessed. This study aimed to assess the test-retest reliability of the MBSA and the ASQ and their convergent validity.

Method: One evaluator administered the outcomes twice 2 weeks apart to postmenopausal women with GSM and urinary incontinence. MBSA reliability was assessed with the graphical analysis of paired differences, Kappa coefficients and intraclass correlation coefficient (ICC), and the ASQ reliability, with paired *t* test and ICC. The convergent validity of both outcomes was investigated through their association with the Pearson correlation coefficient.

Results: Thirty-one women participated in the study. For the MBSA, the severity of the most bothersome symptom obtained a good reliability with 80% observed agreement between sessions, a substantial kappa (0.67 ± 0.12) and excellent ICC (0.88 [95% confidence interval 0.75-0.94]). For the ASQ, no significant difference was found between sessions ($P = 0.146$) and ICC indicated excellent reliability (0.85 [95% confidence interval 0.69-0.93]). There was a large, positive correlation between the severity of the most bothersome symptom selected by the participants in the MBSA and the ASQ total score for the two measurement sessions (T1: $r = 0.587$, $P = 0.001$ and T2: $r = 0.601$, $P < 0.001$).

Conclusions: The MBSA and the ASQ are reliable outcome measures in postmenopausal women with GSM and urinary incontinence. Our findings support good convergent validity of those two outcomes as they showed a significant positive correlation between the severity of GSM symptoms and their impact on activities of daily living.

Key Words: Genitourinary syndrome of menopause – Psychometric properties – Reliability – Validity.

Genitourinary syndrome of menopause (GSM), previously known as vulvovaginal atrophy, is defined as ''a collection of symptoms and signs associated with a decrease in estrogen and other sex steroids involving changes to the labia majora/minora, clitoris, vestibule/introitus, vagina, urethra, and bladder.''[1] This common condition, affecting approximately 50% of postmenopausal women,[2-4] leads to complaints such as vaginal dryness, dyspareunia, vulva pruritus, dysuria and urinary incontinence.[5-7] However, those symptoms will be experienced very differently among women with GSM with regard to the number of symptoms they have[2] and the severity of each symptom.[8-10] This heterogeneity in GSM symptoms makes their assessment difficult.

The most bothersome symptom approach

In the current literature, studies assessing GSM lack uniformity in symptoms evaluation.[11] For this reason, the Division of Drug Information of the US Food and Drug Administration (FDA) recommended the use of the Most Bothersome Symptom Approach (MBSA).[12] This approach makes women start by rating the severity of four main GSM symptoms. Then, the women have to identify the most bothersome among them.[12] A change in this symptom is the important element to evaluate after an intervention in this population.[12] The MBSA is therefore an interesting research and clinical outcome tool, as by choosing the most bothersome symptom, it addresses the issue of GSM symptom

Received September 4, 2018; revised and accepted November 12, 2018.
From the [1]Faculty of Medicine, School of Rehabilitation, University of Montreal, Research Centre of the Institut Universitaire de Gériatrie de Montréal, Montreal, Canada; [2]Faculty of Medicine, School of Rehabilitation, University of Sherbrooke, Research Centre of the Centre Hospitalier Universitaire de Sherbrooke, Sherbrooke, Canada; [3]Department of Obstetrics and Gynecology, Centre Hospitalier de l'Université de Montréal, Montreal, Canada; [4]Department of Obstetrics and Gynecology, Maisonneuve-Rosemont Hospital, Montreal, Canada; [5]Department of Obstetrics and Gynecology, Sir Mortimer B. Davis-Jewish General Hospital, McGill University, Montreal, Quebec, Canada; and [6]Faculty of Medicine, School of Rehabilitation, University of Montreal Research Centre of the Institut Universitaire de Gériatrie de Montréal, Montreal, Canada.
Funding/support: Regroupement Québécois de recherche sur le vieillissement.

Financial disclosure/conflicts of interest: None reported.

Ordre professionnel de la physiothérapie du Québec.

The abstract of this manuscript was presented at the International Continence Society's annual meeting in Florence, Italy, on 15 September.
Address correspondence to: Joanie Mercier, PT, MSc, Research Centre of the Institut Universitaire de Gériatrie de Montréal, 4565 ch. Queen-Mary, Montreal, Québec, Canada H3W 1W5. E-mail: Joanie.mercier.1@umontreal.ca

variability. It also allows a quick and easy overview of GSM symptoms. Sensitivity to change in this outcome measure has previously shown to be good in a study comparing a group treated with local hormonal therapy with a control group (significant difference in means between groups favoring local hormonal therapy, $P \leq 0.001$).[13] However, no study thus far has assessed other psychometric properties of this approach such as its test-retest reliability.

The atrophy symptom questionnaire

In large surveys, a negative impact of GSM symptoms on quality of life and sexual life has been reported in 25% to 80% and 59% to 76% of the participants, respectively.[4,8,10,14] However, little is known about the impact of GSM symptoms on women's activities of daily living (ADL). As for dyspareunia, it was widely shown to affect sexual activities in women with GSM.[14-16] In contrast, other GSM symptoms, such as vaginal dryness, are not well understood in terms of the impact on usual everyday activities. The Atrophy Symptom Questionnaire (ASQ) assesses the impact of different GSM symptoms on global ADL and dyspareunia on sexual function in sexually active women having intercourse.[17] This short questionnaire of five questions can be used in sexually and nonsexually active women as the total score can be adjusted according to their sexual status. The psychometric properties of the ASQ including its test-retest reliability have never been studied.

In the absence of a criterion standard, construct validity can be appraised by evaluating the convergent validity, correlating instrument scores with related concepts in which there is a hypothesized theoretical relationship.[18] A positive correlation is expected between the MBSA and the ASQ, given that a higher severity of GSM symptoms should have a higher impact on ADL. If this hypothesis is true, the association between the two outcome measures should support their convergent validity.

The primary objective of this study was to assess the test-retest reliability of the MBSA and the ASQ in postmenopausal women with GSM symptoms. The secondary objective was to investigate the convergent validity of the two outcomes measures by assessing their association in the same population.

METHOD

Between June 2015 and July 2017, community-dwelling women with GSM symptoms aged 55 years and older were recruited for this prospective cohort study from a randomized control trial on older women with urinary incontinence. To be included in this study, participants had to be postmenopausal (with their last menstrual period >12 months ago) and to present at least two of the following GSM signs assessed by a collaborating gynecologist: petechiae, absent rugae, decreased elasticity, and friability of the vaginal wall.[19] Women were excluded if any of the following applied: dermatological diseases of the vulva; significant prolapse (Pelvic Organ Prolapse Quantification system >2); radiation for gynecological cancer; vaginal or urinary infection within the previous 3 months; or taking antiestrogenic medication. The dosage of systemic and local hormonal therapy medication and the use of vaginal moisturizer had to be stable for at least 6 months before their participation in the study to ensure symptoms stability.

The study received ethical approval in May 2015 from the Institutional Review Board of the Institut Universitaire de Gériatrie de Montréal (Montreal, Canada) (approval number CER IUGM 12-13-002). Each volunteer provided written consent before her participation.

One evaluator administered the outcome measures twice (at T1 and T2), 2 weeks apart. For both outcomes, participants were asked to refer to the last 4 weeks. The MBSA proposed by the FDA assesses four GSM symptoms: (1) vaginal dryness, (2) vaginal itching/irritation, (3) dysuria, and (4) dyspareunia, where participants have to rate the severity of each symptom on a 4-point scale (0 = not present, 1 = mild, 2 = moderate, or 3 = severe).[12] Thereafter, a single symptom is selected by the women as the most bothersome symptom. There is no total score for this outcome measure, each symptom being assessed separately. Regarding the ASQ, it is composed of four items to assess the impact of different GSM symptoms (1-vaginal dryness, 2-vaginal soreness, 3-vulvovaginal irritation, and 4-vaginal discharge) on ADL and one item (5-dyspareunia) to assess the impact on sexual function and satisfaction.[17] Only sexually active women having intercourse are requested to answer this last question. Each of the five items is rated on a 4-point scale (0 = none, 1 = mild, 2 = moderate, and 3 = severe). For the total score, the individual item scores are summed and divided by five for the women having sexual intercourse or by four for those not having sexual intercourse. Scores range from 0 to 3. Higher scores indicate higher impact of GSM symptoms on ADL.

Statistical analysis

The test-retest reliability of the MBSA was assessed in two different ways, because this outcome measure's scales have been used in the literature as both categorical variables[20-22] and as continuous variables.[13,21,23-33] Therefore, agreement between the severity of symptoms included in the MBSA taken at T1 and T2 was first observed by graphical analysis of paired differences and by the weighted Kappa (K) statistic. Kappa values less than 0 were considered to be a poor strength of agreement, 0 to 0.20 slight, 0.21 to 0.40 fair, 0.41 to 0.60 moderate, 0.61 to 0.80 substantial, and 0.81 to 1.00 almost perfect.[34] Second, intraclass correlation coefficients (ICCs) were computed. ICC values less than 0.4 were considered poor, 0.40 to 0.75 fair to good and >0.75 to 1.00 excellent.[35] To assess the test-retest reliability of the ASQ, paired *t* test and ICC were used. Finally, as data from the severity of the most bothersome symptom selected by participants in the MBSA and from the ASQ total score were normally distributed at T1 and T2, the correlation between those two outcome measures was assessed with Pearson correlation coefficients. The strength of the association was considered small for *r* values between 0.10 and 0.29, medium for *r* values ranging from 0.30 to 0.49, and large for *r* values between 0.5 and 1.0.[36]

RESULTS

A total of 31 women aged between 57 and 82 years old (67.9 ± 6.9 years old) were recruited. They had a mean parity of 1.9 ± 1.1 and a mean body mass index of 26.2 ± 4.3 kg/m^2. Five participants (16%) had stress urinary incontinence and 26 had urgency urinary incontinence (84%). Twenty participants (64%) had had vaginal intercourse in the past 4 weeks. Twelve participants (39%) were receiving treatment for GSM symptoms: seven were having local hormonal therapy (HT), two were using a nonhormonal vaginal moisturizer, one was taking a systemic HT, one was on local HT and a systemic HT, and one was using local HT and a nonhormonal vaginal moisturizer. Table 1 summarizes the study participants' demographics.

Most bothersome symptom approach

A description of the MBSA results at T1 and T2 is detailed in Table 2. A large variability of symptoms was found as participants reported one to four symptoms, and the severity of each symptom differed for each of them. At T1 and T2, the most common most bothersome symptom chosen by the participants was vaginal dryness, followed by dyspareunia and vaginal itching/irritation. None of the participants chose dysuria as their most bothersome symptom. Participants rated the severity of their most bothersome symptom from mild to severe, but it was mostly moderate or severe. The mean severity of the most bothersome symptom for all participants was 2.40/3 ± 0.66 at T1 and 2.35/3 ± 0.71 at T2.

Table 3 summarizes the results of the reliability of each symptom of the MBSA including observed agreement between T1 and T2, Kappa values and ICC values. Vaginal dryness and dyspareunia symptoms obtained an observed agreement of 74% between T1 and T2, substantial Kappa's strength of agreement ($K = 0.63 \pm 0.11$ and 0.64 ± 0.10, respectively) and excellent ICC values (ICC = 0.92 [95% confidence interval (CI) 0.80-0.96] and 0.95 [95% CI 0.90-0.98], respectively). Vaginal itching/irritation symptom had an observed agreement of 58%, a moderate Kappa's strength of agreement ($K = 0.41 \pm 0.12$), and an excellent ICC value (ICC = 0.85 [95% CI 0.69-0.93]). Although the dysuria symptom showed an observed agreement between measurement sessions of 87% and a moderate Kappa strength of agreement ($K = 0.44 \pm 0.23$), this symptom is the only one of the MBSA that obtained a poor ICC value (ICC = 0.33 [95% CI −0.43-0.69]).

Good reliability in the selection of the most bothersome symptom among the four symptoms of the MBSA was obtained with an observed agreement between T1 and T2 of 90% and an almost perfect Kappa strength of agreement ($K = 0.84 \pm 0.09$). For the severity of the most bothersome symptom selected by participants, an observed agreement between T1 and T2 of 80% was reached with a substantial Kappa's strength of agreement ($K = 0.67 \pm 0.12$; $P < 0.001$) and an excellent ICC value (ICC = 0.88 [95% CI 0.75-0.94].

TABLE 1. *Participants demographics (n = 31)*

Parameters	
Mean age ± SD, y	67.9 ± 6.9
Parity ± SD (deliveries)	1.9 ± 1.1
Mean body mass index (BMI), kg/m^2	26.2 ± 4.3
Continence status (%)	
Stress urinary incontinence	5 (16%)
Urgency urinary incontinence	26 (84%)
Sexual status (%)	
Having intercourse	20 (64%)
No intercourse	11 (35%)
Treatment for GSM (%)	
None	19 (61%)
Following treatment	12 (39%)
Local hormonal therapy	9 (29%)
Systemic hormonal therapy	2 (6%)
Nonhormonal vaginal moisturizer	3 (10%)

GSM, genitourinary syndrome of menopause; SD, standard deviation.

TABLE 2. *Description of the most bothersome symptom approach results at T1 and T2*

Description	T1 (n = 31)	T2 (n = 31)
Number of symptoms (n, %)		
1	5 (16%)	4 (13%)
2	16 (52%)	13 (42%)
3	7 (23%)	11 (35%)
4	3 (10%)	3 (10%)
Mean ± SD	2.26 ± 0.85	2.42 ± 0.85
Presence of symptoms (n, %)		
Vaginal dryness	26 (84%)	28 (90%)
Vaginal itching/irritation	19 (61%)	22 (71%)
Dysuria	4 (13%)	4 (13%)
Dyspareunia	21 (68%)	21 (68%)
Severity of symptoms (mean ± SD)		
Vaginal dryness (/3)	1.78 ± 1.04	2.00 ± 0.93
Vaginal itching/irritation (/3)	0.97 ± 1.0	1.16 ± 0.93
Dysuria (/3)	0.16 ± 0.45	0.13 ± 0.34
Dyspareunia (/3)	1.59 ± 1.2	1.58 ± 1.3
Symptoms selected as the most bothersome symptom (n, %)		
Vaginal dryness	15 (48%)	18 (58%)
Vaginal itching/irritation	6 (19%)	6 (19%)
Dysuria	0 (0%)	0 (0%)
Dyspareunia	10 (32%)	7 (23%)
Severity of the most bothersome symptom (n, %)		
Mild	3 (10%)	4 (13%)
Moderate	12 (39%)	12 (39%)
Severe	16 (52%)	15 (48%
Mean (/3) (mean ± SD)	2.40 ± 0.66	2.35 ± 0.71

SD, standard deviation.

Atrophy symptom questionnaire

A description of the ASQ results at T1 and T2 is detailed in Table 4. A large variability was also found in this questionnaire, because participants reported one to five GSM symptoms having a negative impact on their ADL and the severity differed for each item. Vaginal dryness was the most common symptom impacting ADL in our population at T1 and T2, followed by dyspareunia, vulvovaginal irritation, vaginal soreness, and vaginal discharge. The mean total score of the ASQ was 0.82/3 ± 0.32 at T1 and 0.89/3 ± 0.39 at T2.

The test-retest reliability analysis showed no significant difference between T1 and T2 for the ASQ total score

TABLE 3. *Test-retest reliability of the most bothersome symptom approach (n = 31 participants)*

Symptoms	Observed agreement between T1 and T2 n (%)	Observed nonagreement between T1 and T2: Higher severity observed at T2 n (%)	Observed nonagreement between T1 and T2: Lower severity observed at T2 n (%)	Kappa (κ) ± SE	*P*	ICC (95% CI)	*P*
Vaginal dryness	23 (74%)	7 (22%)	1 (3%)	0.63 ± 0.11[a]	<0.001	0.92[b] (0.80-0.96)	<0.001
Vaginal itching/irritation	18 (58%)	10 (32%)	3 (10%)	0.41 ± 0.12[c]	<0.001	0.85[b] (0.69-0.93)	<0.001
Dysuria	27 (87%)	2 (6%)	2 (6%)	0.44 ± 0.23[c]	0.006	0.33[d] (-0.43-0.69)	0.147
Dyspareunia	23 (74%)	4 (13%)	4 (13%)	0.64 ± 0.10[a]	<0.001	0.95[b] (0.90-0.98)	<0.001
Selection of the most bothersome symptom	28 (90%)			0.84 ± 0.09[e]	<0.001		
Most bothersome Symptom	25 (80%)	2 (6%)	4 (13%)	0.67 ± 0.12[a]	<0.001	0.88[b] (0.75-0.94)	<0.001

Kappa strength of agreement.
CI, confidence interval; ICC, intraclass correlation coefficients; SE, standard error.
[a]Substantial.
[b]Excellent.
[c]Moderate.
[d]Poor.
[e]Almost perfect ICC value.

($P = 0.146$). Based on the ICC results, excellent reliability was obtained (0.85 [95% CI 0.69-0.93]; $P < 0.001$).

Correlation between the MBS severity and the ASQ

There was a large, positive correlation between the severity of the most bothersome symptom selected by participants in the MBSA and the ASQ total score for the two measurement sessions (T1: $r = 0.587$; $P = 0.001$, T2: $r = 0.601$; $P < 0.001$).

DISCUSSION

The findings from this study showed good test-retest reliability of the severity of the most bothersome symptom using the MBSA and of the ASQ total score. Moreover, the strong positive correlation found between the two outcome measures provides support for their convergent validity.

TABLE 4. *Description of the Atrophy Symptom Questionnaire results at T1 and T2*

Description	T1 (n = 31)	T2 (n = 31)
Number of symptoms with impact on activities of daily living (n, %)		
1	6 (19%)	5 (16%)
2	15 (48%)	14 (45%)
3	9 (29%)	7 (22%)
4	1 (3%)	3 (10%)
5	0 (0%)	2 (6%)
Mean ± SD	2.16 ± 0.77	2.45 ± 1.1
Symptoms with impact on activities of daily living (n, %)		
Vaginal dryness	26 (84%)	28 (90%)
Vaginal soreness	4 (13%)	5 (16%)
Vulvo-vaginal irritation	14 (45%)	16 (52%)
Dyspareunia	21 (68%)	19 (61%)
Vaginal discharge	2 (6%)	6 (19%)
Symptoms impact's severity (mean ± SD)		
Vaginal dryness (/3)	1.68 ± 1.08	1.77 ± 0.84
Vaginal soreness (/3)	0.16 ± 0.45	0.19 ± 0.48
Vulvo-vaginal irritation (/3)	0.65 ± 0.88	0.71 ± 0.82
Dyspareunia (/3)	1.95 ± 0.84	1.95 ± 1.0
Vaginal discharge (/3)	0.60 ± 0.32	0.23 ± 0.50
Total score (/3) (mean ± SD)	0.82 ± 0.32	0.89 ± 0.39

SD, standard deviation.

For each individual symptom of the MBSA, good test-retest reliability was observed with the graphical analysis, the Kappa strength of agreement and the ICC value for the symptoms of vaginal dryness and dyspareunia. For the vaginal itching/irritation symptoms, moderate test-retest reliability was obtained, as the agreement between sessions was 58%, with a moderate Kappa's strength of agreement. This symptom showed an excellent ICC but the 95% confidence intervals were wider. Only the dysuria symptoms obtained a nonsignificant value (ICC = 0.33 [95% CI −0.43-0.69], $P = 0.147$). This result could be explained by either the instability of the measurements or the rare presence and the low severity of this symptom among our population (four participants, 13%).

To our knowledge, this is the first study assessing the test-retest reliability and the convergent validity of the MBSA and the ASQ.[12] Evaluation of the reliability of an outcome is important in order to assess the stability of individuals' responses over time, and it is therefore an essential prerequisite to investigate. Results of the MBSA have been reported in different ways in the literature: using the score of the severity of the most bothersome symptom selected by participants,[11] using the score of each symptom's severity[11] or with a total score of several symptoms.[11] Because there is a wide variability in GSM symptoms among women and our results showed a good reliability for the severity of the most bothersome symptom selected by participants, this symptom should be favored to evaluate change in GSM symptoms after an intervention. Its use allows the inclusion of every woman, sexually active or not, in the assessment of GSM symptoms.[12] Furthermore, the MBS has shown to be sensitive to changes in studies, comparing a group treatment (12-week local hormonal therapy) to a control group in women with GSM (placebo treatment group [significant mean difference between groups, $P \leq 0.001$]).[13]

Although MBSA scales have been used in the literature as categorical variables[20-22] and as continuous variables,[13,21,23-33] the results of the MBSA should be presented as continuous

variables. In the description of the MBSA, the US FDA recommends the assessment of the mean change from the baseline in the most bothersome symptom selected by participants in their guidance document to industry for studies of estrogen and estrogen/progestin drug products for the treatment of GSM symptoms.[12]

For the ASQ, the excellent reliability of the total score obtained in this study also supports its use in future studies investigating the impact of GSM symptoms on ADL.

As hypothesized, our results showed that more severe GSM symptoms (as measured by the severity of the most bothersome symptom in the MBSA) have a higher negative impact on ADL (as measured by the ASQ). The fact that the constructs of these two outcome measures are related supports their convergent validity.

Regarding the other psychometric parameters, their internal consistency was not investigated for the MBSA and the ASQ because it was not applicable. For example, in a questionnaire like the International Consultation on Incontinence Questionnaire-urinary incontinence short form (ICIQ-UI SF), it is expected that the three questions included in the total score (about the frequency, the quantity, and the impact on everyday life of urinary incontinence) will correlate. In the MBSA and the ASQ, it is not expected that the results of each item will correlate as women with GSM show variability in GSM symptoms, regarding the number and severity. Moreover, change in only one symptom (the most bothersome symptom selected by participants) should be considered in the MBSA.

The importance of assessing ADL in different populations or conditions has been highlighted by the ''International Classification of Functioning, Disability and Health'' of the World Health Organization.[37] Defined as ''the individual's ability to execute tasks or actions,''[37] ADL is influenced by the body function, state of health, and contextual factors (ie, environmental and personal factors). The ASQ provides an overview of the impact of GSM symptoms on global ADL. To our knowledge, no other study exists in the literature assessing GSM impact on ADL specifically, or the association between the severity of GSM symptoms and their impact on ADL.

The present study has some limitations. First, the reliability results of the MBSA and ASQ are closely linked to the population under study, that is, postmenopausal women with mild to severe GSM symptoms, aged between 57 and 82. Moreover, all women included in this study had urinary incontinence. Even if urinary incontinence is part of the symptoms of the GSM and present in approximately 30% to 52% of women with this syndrome,[2,8] it is not possible to conclude that the results would have been different in a population of women with GSM with a lower prevalence of urinary incontinence. However, the two outcome measures investigated in our study focus on vulvovaginal symptoms and no questions are related to urinary incontinence. Therefore, the interference of urinary symptoms with the results of those two outcome measures is less likely. Nevertheless, reliability and validity of MBSA and ASQ should be undertaken in a population of women with GSM with and without urinary incontinence and in other populations with similar symptoms such as premenopausal women, postpartum women and in those with vulvovaginal atrophy following treatment for a gynecological cancer. Second, to obtain the same reliability as our results, the time frame of the outcomes must refer to the last four weeks. This information was not specified in the description of either outcome measures but was defined by our team to enhance clarity. Thirdly, the convergent validity of the two outcome measures was determined by assessing their correlation. The investigation of their association with other GSM validated questionnaires could be useful to further support their validity. Finally, the ASQ includes two aspects in the wording of the questions: (1) the impact of the GSM symptom on ADL and (2) the frequency of the GSM symptom (eg, sensation of dryness all the time, dryness interferes with ADL). To be sure that the impact of GSM symptoms on ADL was taken into account, we asked participants to consider the whole question (ie, both aspects). When a few participants hesitated between the two aspects, we asked them to answer according to the impact on ADL. This methodology must be considered to obtain the same reliability as in our study.

CONCLUSION

Our findings suggest that the MBSA and the ASQ are reliable outcome measures to use with postmenopausal women with GSM and urinary incontinence. Moreover, the most bothersome item selected by participants in the MBSA should be favored to evaluate changes in severity of GSM symptoms after an intervention. Finally, our results show a strong positive correlation between the severity of GSM symptoms and their impact on ADL, providing support for their convergent validity.

REFERENCES

1. Portman DJ, Gass MLS; Vulvovaginal Atrophy Terminology Consensus Conference Panel. Genitourinary syndrome of menopause: new terminology for vulvovaginal atrophy from the International Society for the Study of Women's Sexual Health and The North American Menopause Society. *J Sex Med* 2014;11:2865-2872.
2. Erekson EA, Li F-Y, Martin DK, Fried TR. Vulvovaginal symptoms prevalence in postmenopausal women and relationship to other menopausal symptoms and pelvic floor disorders. *Menopause* 2016;23:368-375.
3. DiBonaventura M, Luo X, Moffatt M, Bushmakin AG, Kumar M, Bobula J. The association between vulvovaginal atrophy symptoms and quality of life among postmenopausal women in the United States and Western Europe. *J Womens Health (Larchmt)* 2015;24:713-722.
4. Kingsberg SA, Wysocki S, Magnus L, Krychman ML. Vulvar and vaginal atrophy in postmenopausal women: findings from the REVIVE (REal Women's VIews of Treatment Options for Menopausal Vaginal ChangEs) survey. *J Sex Med* 2013;10:1790-1799.
5. Parish SJ, Nappi RE, Krychman ML, et al. Impact of vulvovaginal health on postmenopausal women: a review of surveys on symptoms of vulvovaginal atrophy. *Int J Women Health* 2013;5:437.
6. Stika CS. Atrophic vaginitis. *Dermatol Ther* 2010;23:514-522.
7. Management of symptomatic vulvovaginal atrophy: 2013 position statement of The North American Menopause Society. *Menopause* 2013;20:888-902.
8. Nappi R, Kokot-Kierepa M. Vaginal health: insights, views & attitudes (VIVA)—results from an international survey. *Climacteric* 2012;15:36-44.

9. Kingsberg SA, Krychman M, Graham S, Bernick B, Mirkin S. The women's EMPOWER survey: identifying women's perceptions on vulvar and vaginal atrophy and its treatment. *J Sex Med* 2017;14:413-424.
10. Santoro N, Komi J. Prevalence and impact of vaginal symptoms among postmenopausal women. *J Sex Med* 2009;6:2133-2142.
11. Weber MA, Limpens J, Roovers JPWR. Assessment of vaginal atrophy: a review. *Int Urogynecol J* 2015;26:15-28.
12. US Department of Health and Human Services FaDA, Center for Drug Evaluation and Research (CDER). Guidance for Industry. Estrogen and estrogen/progestin drug products to treat vasomotor symptoms and vulvar and vaginal atrophy symptoms—recommendations for clinical evaluation. 2003.
13. Ettinger B, Hait H, Reape KZ, Shu H. Measuring symptom relief in studies of vaginal and vulvar atrophy: the most bothersome symptom approach. *Menopause* 2008;15:885-889.
14. Simon JA, Kokot-Kierepa M, Goldstein J, Nappi RE. Vaginal health in the United States: results from the vaginal health: insights, views & attitudes survey. *Menopause* 2013;20:1043-1048.
15. Nappi R, Palacios S. Impact of vulvovaginal atrophy on sexual health and quality of life at postmenopause. *Climacteric* 2014;17:3-9.
16. Avis NE, Brockwell S, Randolph JF Jr, et al. Longitudinal changes in sexual functioning as women transition through menopause: results from the Study of Women's Health Across the Nation (SWAN). *Menopause* 2009;16:442-452.
17. Davila GW, Singh A, Karapanagiotou I, et al. Are women with urogenital atrophy symptomatic? *Am J Obstet Gynecol* 2003;188:382-388.
18. OECD. OECD Guidelines on Measuring Subjective Well-being, OECD Publishing. 2013. Available at: http://dx.doi.org/10.1787/9789264191655-en.
19. Greendale G, Zibecchi L, Petersen L, Ouslander J, Kahn B, Ganz P. Development and validation of a physical examination scale to assess vaginal atrophy and inflammation. *Climacteric* 1999;2:197-204.
20. Kagan R, Williams RS, Pan K, Mirkin S, Pickar JH. A randomized, placebo-and active-controlled trial of bazedoxifene/conjugated estrogens for treatment of moderate to severe vulvar/vaginal atrophy in postmenopausal women. *Menopause* 2010;17:281-289.
21. Labrie F, Archer D, Bouchard C, et al. Intravaginal dehydroepiandrosterone (Prasterone), a physiological and highly efficient treatment of vaginal atrophy. *Menopause* 2009;16:907-922.
22. Bachmann GA, Schaefers M, Uddin A, Utian WH. Microdose transdermal estrogen therapy for relief of vulvovaginal symptoms in postmenopausal women. *Menopause* 2009;16:877-882.
23. Bachmann GA, Komi JO, Group OS. Ospemifene effectively treats vulvovaginal atrophy in postmenopausal women: results from a pivotal phase 3 study. *Menopause* 2010;17:480-486.
24. Simon J, Nachtigall L, Gut R, Lang E, Archer DF, Utian W. Effective treatment of vaginal atrophy with an ultra–low-dose estradiol vaginal tablet. *Obstet Gynecol* 2008;112:1053-1060.
25. Simon JA, Reape KZ, Wininger S, Hait H. Randomized, multicenter, double-blind, placebo-controlled trial to evaluate the efficacy and safety of synthetic conjugated estrogens B for the treatment of vulvovaginal atrophy in healthy postmenopausal women. *Fertil Steril* 2008;90:1132-1138.
26. Simon JA, Bouchard C, Waldbaum A, Utian W, Zborowski J, Snabes MC. Low dose of transdermal estradiol gel for treatment of symptomatic postmenopausal women: a randomized controlled trial. *Obstet Gynecol* 2007;109:588-596.
27. Bachmann G, Bouchard C, Hoppe D, et al. Efficacy and safety of low-dose regimens of conjugated estrogens cream administered vaginally. *Menopause* 2009;16:719-727.
28. Freedman M, Kaunitz AM, Reape KZ, Hait H, Shu H. Twice-weekly synthetic conjugated estrogens vaginal cream for the treatment of vaginal atrophy. *Menopause* 2009;16:735-741.
29. Labrie F, Archer D, Bouchard C, et al. High internal consistency and efficacy of intravaginal DHEA for vaginal atrophy. *Gynecol Endocrinol* 2010;26:524-532.
30. Labrie F, Archer D, Bouchard C, et al. Intravaginal dehydroepiandrosterone (prasterone), a highly efficient treatment of dyspareunia. *Climacteric* 2011;14:282-288.
31. Archer DF, Kimble TD, Lin FY, Battucci S, Sniukiene V, Liu JH. A randomized, multicenter, double-blind, study to evaluate the safety and efficacy of estradiol vaginal cream 0.003% in postmenopausal women with vaginal dryness as the most bothersome symptom. *J Womens Health (Larchmt)* 2018;27:231-237.
32. Gaspar A, Brandi H, Gomez V, Luque D. Efficacy of erbium: YAG laser treatment compared to topical estriol treatment for symptoms of genitourinary syndrome of menopause. *Lasers Surg Med* 2017;49:160-168.
33. Simon JA, Archer DF, Kagan R, et al. Visual improvements in vaginal mucosa correlate with symptoms of VVA: data from a double-blind, placebo-controlled trial. *Menopause* 2017;24:1003-1010.
34. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 1977;33:363-374.
35. Fleiss JL. *The Desing and Analysis of Clinical Experiments*. New York, NY: John Wiley & Sons; 1986.
36. Cohen J. Differences between correlation coefficients. Chapter 4.*Statistical Power Analysis for the Behavioral Sciences*, (Revised Edition). New York, NY: Academic Press; 1977. pp. 109-143.
37. World Health Organization. *International Classification of Functioning, Disability and Health: ICF*. Geneva: World Health Organization; 2001.

