PSYCHOMETRICS

Validation of the German Day-to-Day Impact of Vaginal Aging (DIVA) Questionnaire in Peri- and Postmenopausal Women

Check for updates

Michaela Gabes, PhD,[1,2] Petra Stute, MD,[3] and Christian Apfelbacher, PhD[1]

ABSTRACT

Background: The Day-to-Day Impact of Vaginal Aging (DIVA) questionnaire is a validated patient-reported outcome measure (PROM) capturing the impacts of vaginal symptoms in postmenopausal women.

Aim: We aimed to psychometrically validate the German version of the DIVA questionnaire.

Methods: Data was collected online and by paper-pencil. We ran confirmatory factor analyses to confirm the *a priori* four-factor structure of the DIVA. Internal consistency was calculated using Cronbach's alpha. Correlations with other outcome measures such as the Patient Health Questionnaire-4 (PHQ-4), the SF-12 SOEP (socio-economic panel) version and self-created anchor questions were calculated regarding convergent validity. Known groups regarding age, home country and disease severity were analyzed. Test-retest reliability after 1 week and responsiveness after 4 weeks were only descriptively assessed due to low sample sizes.

Main Outcome Measures: The DIVA questionnaire, the Menopause Rating Scale (MRS II), the PHQ-4 and the SF-12 SOEP version were the main outcome measures.

Results: 185 postmenopausal women reporting vaginal complaints participated in the survey. The mode of administration did not influence the severity of vaginal symptoms. The four-factor structure of the DIVA could be confirmed and the model fit indicated sufficient structural validity. Furthermore, strong internal consistency in all of the DIVA domains was found. Regarding convergent validity, no hypothesis has failed completely. The results regarding known-groups validity were mostly in line with our *a priori* hypotheses. Descriptive evidence for sufficient test-retest reliability and responsiveness was given, however, the sample size for the assessment of those two measurement properties was low.

Conclusion: This study supports the excellent structural validity, internal consistency and construct validity of the German version of the DIVA questionnaire. It can be recommended for the assessment of the impacts of vaginal symptoms in postmenopausal women in future clinical GSM trials. **Gabes M, Stute P, Apfelbacher C. Validation of the German Day-to-Day Impact of Vaginal Aging (DIVA) Questionnaire in Peri- and Postmenopausal Women. Sex Med 2021;9:100382**

Copyright © 2021 The Authors. Published by Elsevier Inc. on behalf of the International Society for Sexual Medicine. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Key words: Genitourinary Syndrome of Menopause; Patient-reported outcome measures; Measurement properties; Validity; Reliability; Responsiveness

https://doi.org/10.1016/j.esxm.2021.100382

INTRODUCTION

The term 'Genitourinary Syndrome of Menopause (GSM)' summarizes physiological changes due to menopausal sex hormone deficiency which often result in burdensome symptoms.[1] It affects up to 50% of all postmenopausal women worldwide.[2] Women usually report genital or rather vulvovaginal symptoms, such as vaginal dryness or burning, urinary symptoms of urgency and dysuria, and sexual symptoms, such as pain during intercourse or loss of sexual desire.[3,4] GSM was previously known as vulvovaginal atrophy (VVA) to which the urinary component was added to be more comprehensive.[5]

Genitourinary complaints during and after the menopausal transition affect the whole life of menopausal women: everyday activities, emotional well-being, sexual functioning, self-concept and body image, and interpersonal relations.[6,7] In clinical trials, these impacts are captured by patient-reported outcome measures (PROMs). PROMs are self-completed questionnaires reflecting the patient's perspective.[8]

A recent systematic review identified four PROMs for postmenopausal women with VVA or GSM. Two of those, the Vulvovaginal Symptoms Questionnaire (VSQ)[9] and the Day-to-Day Impact of Vaginal Aging (DIVA) questionnaire[10], showed good measurement properties and were recommended for use in future clinical trials.[11] According to the US Food and Drug Administration (FDA), for instance, evidence for patient input during PROM development is crucial.[12] Since patients were involved in the development of the DIVA, but not in the development of the VSQ, the DIVA should be preferred over the VSQ.[6,10,11]

As its name says, the DIVA is capturing the impact of vulvovaginal complaints in aging, postmenopausal women. It is composed of 23 items on four scales, everyday activities, emotional well-being, sexual functioning and self-concept and body image. Four items of the sexual functioning scale can only be answered by recently sexually active women, and thus there is a long-form with 23 items only for women with recent sexual activity and a short-form with 19 items which can be answered by all women regardless of their sexual activity status. Each item can be rated on a scale from 0 to 4. The domain scores are determined by the mean of the corresponding items, the higher the mean scores, the greater the symptom impact.[10] The DIVA is currently available in English, Spanish and Italian[10,13-15] and has recently been also linguistically validated in German using cognitive debriefing interviews with the target population to ensure aspects of the comprehensibility of the translated DIVA. However, there is a lack of psychometric validity evidence of the German DIVA. This study aimed to psychometrically validate the German DIVA in a German population of postmenopausal women with vulvovaginal complaints.

MATERIALS AND METHODS

Participants and Procedure

We conducted an online survey that only included postmenopausal women reporting vaginal GSM symptoms related to menopause, such as dryness, burning, pruritus or vaginal pain. The inclusion criteria were assured by several filter questions. The online survey was created with soscisurvey.de and contained general questions (about age, home country, mother tongue, hormone intake), the German version of the DIVA questionnaire and several other PROMs, such as the Menopause Rating Scale (MRS II)[16-19], five self-created anchor questions (one general (anchor 1) and one specific for each of the four DIVA domains (anchor 2 − 5)), the Patient Health Questionnaire-4 (PHQ-4)[20] and the SOEP (socio-economic panel) version of the short-form

health survey (SF-12).[21] The German version of the DIVA has not been available until now, however, the German DIVA has been recently created by a translation office and linguistically validated using cognitive interviews with the target population. Participating women were asked to leave their e-mail address at the end of the survey to be contacted for two follow-up assessments after seven days and 4 weeks.

The link to the survey was distributed in several Facebook groups for postmenopausal women, flyers with the link were sent to gynecologic practices and clinics or directly handed out to the target population and the e-mail distribution list of a cooperating gynecologist (PS) was used to spread the survey. Furthermore, a paper-pencil version of the survey was given to women participating at the screening for a clinical GSM trial by one of our research partners.

The study was approved by the Ethics Committee of the University of Regensburg (file number: 19-1554-101).

Analysis

Before the validation analyses started, it was assessed using Mann-Whitney-U-tests whether the two samples with a different mode of administration (online vs paper-pencil) significantly differed in the main symptom of GSM, vaginal dryness (assessed by item 10 of the MRS II), on the one hand, and in GSM symptom severity (assessed by the five anchor questions), on the other.

Structural Validity. Structural validity refers to the degree to which the scores of a PROM are an adequate reflection of the dimensionality of the construct to be measured.[22,23] Confirmatory factor analyses were carried out to confirm the four-factor structure of both, the short- and the long-version of the DIVA in the current study population. The following goodness-of-fit indices and cut-off levels were used: comparative fit index (CFI)/ Tucker-Lewis Index (TLI) $\geq$ 0.95 adequate fit, $\geq$ 0.90 acceptable fit; root mean square error of approximation (RMSEA) < 0.05 good fit, < 0.08 acceptable fit; standardized root mean square residual (SRMR) < 0.05 adequate fit, < 0.8 still acceptable fit and Chi square/df ration (χ^2/df) < 2 adequate fit, < 3 acceptable fit.[24-26] According to the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative, the structural validity rating of an instrument is considered to be sufficient if either the CFI or TLI is > 0.95 or the RMSEA is < 0.06 or the SRMR is < 0.08.[27]

Internal Consistency. Internal consistency refers to the intercorrelation of the single items, that is, the interrelatedness among the items.[23] Cronbach's alpha was calculated for the four domains of the DIVA. For the third domain, internal consistency was determined for each version separately. According to the COSMIN group, a Cronbach's alpha $\geq$ 0.7 is considered to be appropriate.[27]

Test-Retest Reliability. Test-retest reliability reflects the level of agreement between two measurements in patients which are stable on another criterion.[22] All women who voluntarily left their e-mail address at the end of the survey were contacted again after one week to participate in a first follow-up assessment. Only women showing no change in the anchor question(s) were analyzed. The calculation of intraclass correlation coefficients (ICCs) with 95% confidence intervals was initially planned, however, the sample size was low and for this reason only descriptive statistics were determined for baseline and the first follow-up.

Convergent and Divergent Validity. Construct validity including convergent/divergent validity and known-groups validity refers to the degree to which the scores of a PROM are consistent with hypotheses based on the assumption that the PROM validly measures the construct to be measured.[23] Convergent validity is assessed by the Spearman rank correlation with other outcome measures. We formulated *a priori* hypotheses about correlations with other PROMs (the five anchor questions, the PHQ and the SF-12 SOEP-version) to test whether the domains of the DIVA validly measure the constructs to be measured. We expected Spearman rank correlations ≥ 0.5 with instruments measuring similar constructs and correlations between 0.3 and 0.5 with instruments measuring related, but dissimilar constructs (see Table 1).

Known-Groups Validity. We formulated *a priori* hypotheses about differences in known groups. We did not expect any significant differences in the DIVA domains between different age categories $(40 - 49, 50 - 59, 60 - 69, \geq 70$ years) and the three different home countries (Germany, Austria, Switzerland) of the

participants. However, regarding different levels of disease severity according to item 10 of the MRS II and the first anchor question, we did expect significant differences in the DIVA domains. Kruskal-Wallis-tests were carried out to assess our hypotheses.

Responsiveness. Responsiveness refers to the ability of a PROM to detect change over time in the construct to be measured.[23] All women who left their e-mail address at the end of the survey were contacted again after four weeks for the participation in a second follow-up assessment. Spearman rank correlations between the domain change scores and the five anchor question change scores were calculated to see if the anchors could be considered as appropriate (> 0.3).[28] Since the sample size of the second follow-up was low, only descriptive statistics were calculated.

All analyses were carried out using IBM SPSS Statistics 25 and Mplus 8.4 software (*Muthen & Muthen, Los Angeles, CA*).

RESULTS

Participants

In the online survey, 144 postmenopausal women reported vaginal symptoms related to menopause and therefore fulfilled the inclusion criteria. The mean age of the online sample was 53.7 years (SD = 5.7, median: 53, range: $41 - 85$). Missing values increased towards the end of the survey. The paper-pencil version was completed by 41 women with a mean age of 59.5 years (SD = 6.1, Median: 58, Range: $50 - 71$).

We neither found a significant difference in item 10 of the MRS II about vaginal dryness ($P = .702$) nor in any of the five anchor questions about symptom severity ($P \geq 0.065$) regarding the mode of administration. The two samples were therefore

Table 1. Expected correlations of the DIVA domains with other items and PROMs

	Anchors					PHQ			SF-12 SOEP*	
	1	2	3	4	5	Depr.	Anx.	Total	PCS	MCS
Domain 1										
Domain 2										
Domain 3 (short)										
Domain 3 (long)										
Domain 4										

*negative correlations expected (the higher the quality of life, the lower the symptom impact).
Note: Green: high correlations (≥ 0.5) expected
Yellow: medium correlations ($0.3 - 0.5$) expected
Grey: no hypotheses
Domain 1: activities of daily living,
Domain 2: emotional well-being,
Domain 3: sexual functioning,
Domain 4: self-concept and body image.
Anx. = Anxiety; Depr. = Depression; MCS = Mental Health; PCS = Physical Health; PHQ = Patient Health Questionnaire ;
PROMs = patient-reported outcome measures; SOEP = Socio-economic panel.

Table 2. Demographics of the sample

Variable	n = 185
Mean age (SD)	54.9 (± 6.2)
Home country	
- Germany	148 (80.0%)
- Austria	12 (6.5%)
- Switzerland	22 (11.9%)
- Other*	3 (1.6%)
Mother tongue	
- German	182 (98.4%)
- Other†	3 (1.6%)
Cycle	
- No period for at least 2 mo	23 (12.5%)
- No period for at least 1 y	124 (67.4%)
- No womb anymore	37 (20.1%)
Hormone intake	
- No hormone intake	124 (67.0%)
- Hormone intake for birth control‡	4 (2.2%)
- Hormone intake for the treatment of menopausal complaints	50 (27.0%)
- Other	7 (3.8%)

*Other home countries were the USA, Italy and Spain.
†Other mother tongues were Russian, Polish and French.
‡It was not asked what contraception was used.

summarized and considered as one sample (Ø age = 54.9 years, SD = 6.2, Median: 54, Range: 41 − 85). Further demographics about the sample are presented in Table 2.

The first domain of the DIVA was completed by 174 women, the second domain by 176 women, the short-version of the third domain by 158 women, the long-version by 132 women with recent sexual activity and the fourth domain by 159 women. The PHQ-4 and the SOEP-version of the SF-12 were completed by 158 women.

Analysis

Structural Validity. Goodness-of-fit indices indicated an acceptable to adequate model fit. According to the COSMIN criteria, the structural validity of the DIVA would be rated as sufficient since a SRMR < 0.08 was found (see Table 3).

Internal consistency. We found a Cronbach's alpha of 0.84 for the first domain, 0.91 for the second domain, 0.84 for the short-version of the third domain and 0.93 for the long-version of the third domain, and a Cronbach's alpha of 0.93 for the fourth domain. Thus, the internal consistency of all domains of the DIVA can be considered as sufficient according to the COSMIN criteria.[27]

Test-Retest Reliability. Sixty-four women left their e-mail address at the end of the survey and were contacted after 1 week (and once reminded if necessary). Thirty-one women participated a second time, however only 26 completed the whole follow-up survey after 8.84 days on average. Most of the women (n = 15 − 18) showed no change in any of the five anchor questions (see Appendix A).

The sample size of women reporting stability in any of the five anchor questions was too low to determine ICCs. Therefore, only means (and standard deviations) of the DIVA domains at baseline and 1-week follow-up were contrasted. It was shown descriptively that means were very similar to identical in both measurements (see Appendix B). Thus, we assumed a stable measurement of the DIVA domains.

Convergent and Divergent Validity. We could confirm 13 of the 31 formulated hypotheses (see Table 4). Some correlations were slightly underestimated (e.g. with the PHQ), some were slightly overestimated (e.g. domain 3 with the anchor questions), however, no hypothesis has failed completely.

Known-Groups Validity. As expected, we did not find any significant differences in the DIVA domains between the different age categories. However, against our expectation, domain 1 and 2 differed significantly in the home country of the participants (see Table 5). When interpreting this result, it should be taken into account that we only had 9 − 12 women from Austria and 20 − 22 women from Switzerland, but 101 − 138 women from Germany. Regarding disease severity measured by item 10 of the MRS II and the first anchor questions, all domains were able to significantly differ between different severity groups. Only the third domain failed to significantly differ between different severity groups measured by

Table 3. Fit-indices for the short-version (n = 152) and the long-version (n = 128) of the DIVA to confirm the four-factor structure

Fit indices	Adequate (acceptable) model fit	DIVA Short-version	DIVA Long-version
CFI	≥ 0.95 (≥ 0.90)	0.942	0.932
TLI	≥ 0.95 (≥ 0.90)	0.932	0.923
RMSEA	< 0.05 (< 0.08)	0.070 (90% CI: 0.055 − 0.084)	0.072 (90% CI: 0.058 − 0.084)
SRMR	< 0.05 (< 0.08)	0.053*	0.061*
$\chi 2$/df	< 2 (< 3)	1.734	1.657

*sufficient according to the COSMIN (= COnsensus-based Standards for the selection of health Measurement INstruments) criteria. CFI = comparative fit index; df = degrees of freedom; DIVA = Day-to-Day Impact of Vaginal Aging questionnaire; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; TLI = Tucker-Lewis Index

Table 4. Correlations of the DIVA domains with other PROMs

	Anchors					PHQ			SF-12 SOEP*	
	1	2	3	4	5	Depr.	Anx.	Total	PCS	MCS
Domain 1	0.414	0.465	0.510	0.340	0.556				-0.452	
n	175	175	175	175	175				158	
Domain 2	0.365	0.537	0.698	0.487	0.727	0.533	0.533	0.579		-0.483
n	176	176	176	175	176	158	158	158		157
Domain 3 (short)	0.233	0.242	0.225	0.350	0.334					
n	158	158	158	157	158					
Domain 3 (long)	0.372	0.275	0.286	0.417	0.376					
n	132	132	132	132	132					
Domain 4	0.447	0.595	0.603	0.673	0.685					-0.386
n	159	159	159	158	159					155

*negative correlations expected (the higher the quality of life, the lower the symptom impact).
Note: Green: high correlations (≥ 0.5) expected
Yellow: medium correlations ($0.3 - 0.5$) expected
Grey: no hypotheses
Domain 1: activities of daily living
Domain 2: emotional well-being
Domain 3: sexual functioning
Domain 4: self-concept and body image
Anx. = Anxiety; Depr. = Depression; PCS = Physical Health; PHQ = Patient Health Questionnaire; PROMs = patient-reported outcome measures; SOEP = Socio-economic panel; MCS = Mental Health.

item 10 of the MRS II. This is not surprising since domain 3 was only correlating low with this item (short-version: $r = 0.198$, long-version: $r = 0.232$). In total, we could confirm 16 of our 20 *a priori* formulated hypotheses.

Responsiveness. All of the 64 women who left their e-mail address at the end of the survey were contacted again after 4 weeks (and once reminded if necessary). Twenty-eight women participated again, however, only 26 women completed the

Table 5. *P*-values of the single Kruskal-Wallis-tests regarding known-groups validity

	p-values			
	Age	Home country	Disease severity	
			Item 10 (MRS II)	Anchor 1
Domain 1	.215	< .001*	.001*	< .001*
Domain 2	.755	.002*	.010*	< .001*
Domain 3 (short)	.294	.143	.107	.028*
Domain 3 (long)	.462	.102	.096	.001*
Domain 4	.747	.256	< .001*	< .001*

*significant (α = 0.05).
Note: Grey: no significant differences expected
Green: significant group differences expected
Bold: Hypothesis confirmed
Domain 1: activities of daily living
Domain 2: emotional well-being
Domain 3: sexual functioning,
Domain 4: self-concept and body image
MRS II = Menopause Rating Scale.

whole follow-up survey again after 30.71 days on average. Correlations of the DIVA domain change scores with the anchor change scores were between -0.032 and 0.529. The appropriateness of the anchor questions could only be confirmed in 10 of 20 cases. Most of the participants did not report any change in one of the five anchor questions (see Appendix C).

All women reporting any improvement in the anchor questions after 4 weeks were summarized and their means (and standard deviations) were descriptively contrasted in Appendix D. Apparently, an improvement in all domains of the DIVA could be detected.

DISCUSSION

This study confirmed several excellent measurement properties of the DIVA questionnaire in German language. We could confirm the four-factor structure of the DIVA with an acceptable model fit. All domains showed strong internal consistency. Our results regarding construct validity were mostly in line with our *a priori* hypotheses. The high correlation of the second domain with the PHQ-4 showed that both are measuring a similar construct that may be emotional well-being. The medium correlation of the first, the second and the fourth domain with the SOEP version of the SF-12, an instrument measuring general health-related quality of life, indicated that those domains are measuring something more specific, but also something general such as the overall impact on the quality of life of affected women. Our self-created anchor questions correlated pretty well, even better than expected, with the first, second and fourth domain, however those questions did not appropriately correlate with the third domain about sexual functioning. It would have been worthwhile including a reference measure about sexual functioning such as the Female Sexual Function Index (FSFI)[29] to confirm the construct of the third domain. We could only descriptively show that stable women after 1 week had stable DIVA scores and women reporting an improvement after 4 weeks showed improved DIVA scores. Furthermore, the results of the cognitive interviews in the context of the linguistic validation supports the content validity of the DIVA questionnaire. Those validation results are consistent with all prior validation results reported in the recent systematic review.[11]

This study faced a number of limitations. For the test-retest and the responsiveness assessment, only a few participants from baseline left their e-mail address and of those, only very few completed the consecutive follow-up assessments. Thus, only a descriptive comparison was possible. The information about the cycle may have had only limited informative value since many women had a hormone replacement therapy or a hormonal contraception in parallel. Furthermore, since the online data collection was undertaken electronically, there were no clinic visits and therefore the participants did not undergo clinical examination and/or laboratory testing to evaluate the etiology of their symptoms. Other potential causes of the symptoms such as vaginal infections, contact allergies, female sexual dysfunction etc. could not be excluded. Thus, the study sample included both clinic and non-clinic women which is more representative of the GSM patient population. However, the patients who completed the paper-pencil version had physician-confirmed diagnoses. Another issue regarding the online data collection is the fact that online activity presumes computer or smartphone literacy and internet access what automatically excludes those without, especially since our target population contains many mid-aged and older women. Nevertheless, social media offer an alternative recruitment source, without geographical limitations.

Future validation studies should focus on the estimation of minimal important difference (MID) values since those have only been determined once so far in a randomized controlled clinical trial after 12 weeks.[30] Having a MID values for every scale helps to give an interpretation of a change in score after specific time intervals, since MID values provide information on whether a change is perceived by patients as beneficial and it therefore considered to be clinically meaningful or simply due to measurement error.[31,32]

CONCLUSIONS

In conclusion, this study showed several excellent measurement properties of the DIVA questionnaire in terms of validity and reliability. Our findings were consistent with the evidence generated in the original development and validation[6,10] and further validation studies of the DIVA.[14,33] We confirmed its recommendation for the assessment of the impacts of postmenopausal vulvovaginal complaints in future clinical GSM trials according to the COSMIN criteria since we did not find any insufficient measurement properties.

ETHICAL APPROVAL

The study was approved by the Ethics Committee of the University of Regensburg (file number: 19-1554-101).

CONSENT TO PARTICIPATE

Informed consent to participate was given by the sample population.

AVAILABILITY OF DATA AND MATERIAL

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Corresponding Author: Michaela Gabes, PhD, Dr.-Gessler-Str. 17, 93051 Regensburg, Germany, Tel: +49 941 944-5227; Fax: +49 941 944-5252; E-mail: michaela.gabes@med.ogu.de

STATEMENT OF AUTHORSHIP

Michaela Gabes: Conceptualization, methodology, analysis and interpretation of the data, writing-original draft, final approval of the submitted manuscript; Petra Stute: Supervision, critical revision for important intellectual content, final approval of the submitted manuscript; Christian Apfelbacher: Project administration, supervision, critical revision for important intellectual content, final approval of the submitted manuscript.

REFERENCES

1. Portman DJ, Gass ML. 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. **Menopause 2014;21:1063–1068.**

2. 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 Womens Health 2013;5:437–447.**

3. Gandhi J, Chen A, Dagur G, et al. Genitourinary syndrome of menopause: an overview of clinical manifestations, pathophysiology, etiology, evaluation, and management. **Am J Obstet Gynecol 2016;215:704–711.**

4. Farrell AM E. Genitourinary syndrome of menopause. **Aust Fam Physician 2017;46:481–484.**

5. Shifren JL, Zincavage R, Cho EL, et al. Women's experience of vulvovaginal symptoms associated with menopause. **Menopause 2019;26:341–349.**

6. Huang AJ, Luft J, Grady D, et al. The day-to-day impact of urogenital aging: Perspectives from racially/ethnically diverse women. **J Gen Intern Med 2010;25:45–51.**

7. Simon JA, Kokot-Kierepa M, Goldstein J, et al. Vaginal health in the United States: Results from the Vaginal Health: Insights, views & attitudes survey. **Menopause 2013;20:1043–1048.**

8. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: A structured review. **J Eval Clin Pract 2006;12:559–568.**

9. Erekson EA, Yip SO, Wedderburn TS, et al. The vulvovaginal symptoms questionnaire: A questionnaire for measuring vulvovaginal symptoms in postmenopausal women. **Menopause 2013;20:973–979.**

10. Huang AJ, Gregorich SE, Kuppermann M, et al. Day-to-Day impact of vaginal aging questionnaire: A multidimensional measure of the impact of vaginal symptoms on functioning and well-being in postmenopausal women. **Menopause 2015;22:144–154.**

11. Gabes M, Knuttel H, Stute P, Apfelbacher CJ. Measurement properties of patient-reported outcome measures (PROMs) for women with genitourinary syndrome of menopause: A systematic review. **Menopause 2019;26:1342–1353.**

12. US Food and Drug Administration. Guidance for industry: Patient-Reported Outcome Measures: Use in medical product development to support labeling claims. 2009. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims. Accessed Decemder 12, 2020.

13. Moral E, Delgado JL, Carmona F, et al. The impact of genitourinary syndrome of menopause on well-being, functioning, and quality of life in postmenopausal women. **Menopause 2018;25:1418–1423.**

14. Nappi RE, Palacios S, Bruyniks N, et al. The burden of vulvovaginal atrophy on women's daily living: Implications on quality of life from a face-to-face real-life survey. **Menopause 2019;26:485–491.**

15. Palacios S, Nappi RE, Bruyniks N, et al. The European Vulvovaginal Epidemiological Survey (EVES): Prevalence, symptoms and impact of vulvovaginal atrophy of menopause. **Climacteric 2018;21:286–291.**

16. Heinemann K, Assmann A, Mohner S, et al. [Reliability of the Menopause Rating Scale (MRS): Investigation in the German population]. **Zentralbl Gynecol 2002;124:161–163.**

17. Heinemann LA, DoMinh T, Strelow F, et al. The Menopause Rating Scale (MRS) as outcome measure for hormone treatment? A validation study. **Health Qual Life Outcomes 2004;2:67.**

18. Potthoff P, Heinemann LA, Schneider HP, et al. [The Menopause Rating Scale (MRS II): Methodological standardization in the German population]. **Zentralbl Gynecol 2000;122:280–286.**

19. Schneider HP, Heinemann LA, Rosemeier HP, et al. The Menopause Rating Scale (MRS): Comparison with Kupperman index and quality-of-life scale SF-36. **Climacteric 2000; 3:50–58.**

20. Kroenke K, Spitzer RL, Williams JB, et al. An ultra-brief screening scale for anxiety and depression: The PHQ-4. **Psychosomatics 2009;50:613–621.**

21. Nübling M, Andersen HH, Mühlbacher A. Entwicklung eines Verfahrens zur Berechnung der körperlichen und psychischen Summenskalen auf Basis der SOEP-Version des SF 12 (Algorithmus). DIW Berlin, German Institute for Economic Research; 2006.

22. Mokkink LB, Prinsen CA, Patrick DL, et al. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) - user manual. 2018. Available at: https://www.cosmin.nl/tools/guideline-conducting-systematic-review-outcome-measures/. Accessed Decemder 12, 2020.

23. Prinsen CA, Vohra S, Rose MR, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. **Trials 2016;17:449.**

24. Brown TA. Confirmatory Factor Analysis for Applied Research. New York: Guilford Press; 2015.

25. Moosbrugger H, Schermelleh-Engel K. Exploratorische (EFA) und Konfirmatorische Faktorenanalyse (CFA). In: Moosbrugger H, Kelava A, editors. Testtheorie und Fragebogenkonstruktion. Berlin Heidelberg: Springer; 2012. p. 326–343.

26. Schmitt TA. Current methodological considerations in exploratory and confirmatory factor analysis. J Psychoeduc Assess 2011;29:304–321.

27. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018;27:1147–1157.

28. Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008;61:102–109.

29. Isidori AM, Pozza C, Esposito K, et al. Development and validation of a 6-item version of the female sexual function index (FSFI) as a diagnostic tool for female sexual dysfunction. J Sex Med 2010;7:1139–1146.

30. Gibson CJ, Huang AJ, Larson JC, et al. Patient-centered change in the day-to-day impact of postmenopausal vaginal symptoms: Results from a multicenter randomized trial. Am J Obstet Gynecol 2020;223:99.e91–99.e99.

31. Brunner HI, Higgins GC, Klein-Gitelman MS, et al. Minimal clinically important differences of disease activity indices in childhood-onset systemic lupus erythematosus. Arth Care Res 2010;62:950–959.

32. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Cont Clini Trials 1989;10:407–415.

33. Hunter MM, Nakagawa S, Van Den Eeden SK, Kuppermann M, Huang AJ. Predictors of impact of vaginal symptoms in postmenopausal women. Menopause 2016;23 (1):40–46.

APPENDIX A

Number of women with either a stable or changed disease severity after 1 week based on the five anchor questions

Anchor	Change after 1 week					
	A good deal better	Somewhat better	Slightly better	**No change**	Slightly worse	Somewhat worse
	-3	-2	-1	**0**	1	2
1	1	1	7	17	5	-
2	-	2	7	17	3	2
3	-	3	10	15	2	1
4	-	1	6	18	6	-
5	-	-	9	18	4	-

Grey: Women reporting no change.

APPENDIX B

Descriptive comparison of the means (and standard deviations) of the DIVA domains in women reporting no change (change in the corresponding anchor question = 0) at baseline and 1-week follow-up

		Anchors				
		1	2	3	4	5
Domain 1	T0	0.647 (0.64) n = 17	0.918 (0.90) n = 17			
	T1	0.647 (0.83) n = 17	0.929 (1.03) n = 17			
Domain 2	T0	0.721 (0.93) n = 17		1.133 (1.24) n = 15		
	T1	0.721 (0.99) n = 17		1.150 (1.11) n = 15		
Domain 3 (short)	T0	2.075 (0.77) n = 16			2.000 (0.93) n = 18	
	T1	1.707 (0.89) n = 15			2.294 (1.05) n = 17	
Domain 3 (long)	T0	2.082 (0.74) n = 15			2.153 (0.90) n = 16	
	T1	1.769 (0.83) n = 13			2.341 (1.06) n = 15	
Domain 4	T0	1.506 (0.94) n = 17				1.906 (1.12) n = 17
	T1	1.520 (1.15) n = 15				1.824 (1.18) n = 17

Domain 1: activities of daily living
Domain 2: emotional well-being
Domain 3: sexual functioning
Domain 4: self-concept and body image
T0 = baseline,
T1 = 1-week follow-up.

APPENDIX C

Number of women with either a changed or stable disease severity after four weeks based on the five anchor questions

Anchor	Change after 4 weeks					
	A good deal better -3	**Somewhat better** -2	**Slightly better** -1	No change 0	Slightly worse 1	Somewhat worse 2
1	1	1	11	14	1	-
2	1	5	5	15	2	-
3	-	4	5	17	2	-
4	-	3	7	15	3	-
5	-	1	10	12	5	-

Grey: Women reporting an improvement.

APPENDIX D

Descriptive comparison of the means (and standard deviations) of the DIVA domains in women reporting an improve- ment (change in the corresponding anchor question < 0) at baseline and 4-week follow-up

		Anchors				
		1	2	3	4	5
Domain 1	T0	1.185 (0.90) n = 13	1.346 (1.01) n = 11			
	T2	0.917 (0.85) n = 12	1.146 (1.01) n = 11			
Domain 2	T0	0.981 (1.07) n = 13		1.694 (1.07) n = 9		
	T2	0.667 (0.76) n = 12		1.222 (0.94) n = 9		
Domain 3 (short)	T0	1.954 (0.70) n = 13			2.044 (0.73) n = 9	
	T2	1.667 (1.13) n = 12			1.467 (1.22) n = 9	
Domain 3 (long)	T0	2.000 (0.75) n = 12			2.250 (0.90) n = 8	
	T2	1.354 (0.85) n = 11			1.476 (0.83) n = 7	
Domain 4	T0	1.985 (1.14) n = 13				1.891 (1.07) n = 11
	T2	1.850 (1.16) n = 12				1.740 (0.99) n = 10

Domain 1: activities of daily living,
Domain 2: emotional well-being
Domain 3: sexual functioning
Domain 4: self-concept and body image
T0 = baseline
T2 = 4-week follow-up