Original Paper

Real-time Symptom Assessment in Patients With Endometriosis: Psychometric Evaluation of an Electronic Patient-Reported Outcome Measure, Based on the Experience Sampling Method

Esther van Barneveld[1,2], MD, PhD; Arianne Lim[1,2], MD, PhD; Nehalennia van Hanegem[3], MD, PhD; Frits van Osch[4,5], PhD; Lisa Vork[4,6], MD, PhD; Joanna Kruimel[4,6], MD, PhD; Marlies Bongers[1,2,7], MD, PhD; Carsten Leue[8,9], MD, PhD

[1]GROW School of Oncology and Developmental Biology, Maastricht University, Maastricht, Netherlands

[2]Department of Gynaecology and Obstetrics, Maastricht University Medical Center+, Maastricht, Netherlands

[3]Department of Reproductive Medicine and Gynecology, University Medical Center Utrecht, Utrecht, Netherlands

[4]NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University, Maastricht, Netherlands

[5]Department of Clinical Epidemiology, VieCuri Medical Center, Venlo, Netherlands

[6]Department of Internal Medicine, Division of Gastroenterology and Hepatology, Maastricht University Medical Center+, Maastricht, Netherlands

[7]Department of Gynaecology and Obstetrics, Máxima Medical Center, Veldhoven, Netherlands

[8]School for Mental Health and Neuroscience, Maastricht University, Maastricht, Netherlands

[9]Department of Psychiatry and Psychology, Maastricht University Medical Center+, Maastricht, Netherlands

Corresponding Author:
Esther van Barneveld, MD, PhD
Department of Gynaecology and Obstetrics, Maastricht University Medical Center+
P. Debyelaan 25
Maastricht, 6229 HX
Netherlands
Phone: 31 613403725
Email: esthervb@gmail.com

Abstract

Background: The experience sampling method (ESM) holds advantages over traditional retrospective questionnaires including a high ecological validity, no recall bias, the ability to assess fluctuation of symptoms, and the ability to analyze the temporal relationship between variables.

Objective: This study aimed to evaluate the psychometric properties of an endometriosis-specific ESM tool.

Methods: This is a short-term follow-up prospective study, including patients with premenopausal endometriosis aged ≥18 years who reported dysmenorrhea, chronic pelvic pain, or dyspareunia between December 2019 and November 2020. An ESM-based questionnaire was sent out by a smartphone application 10 times a day during 1 week on randomly chosen moments. Additionally, patients completed questionnaires concerning demographics, end-of-day pain scores, and end-of-week symptom scores. The psychometric evaluation included compliance, concurrent validity, and internal consistency.

Results: Twenty-eight patients with endometriosis completed the study. Compliance for answering the ESM questions was as high as 52%. End-of-week pain scores were higher than ESM mean scores and showed peak reporting. ESM scores showed strong concurrent validity when compared with symptoms scored by the Gastrointestinal Symptom Rating Scale–Irritable Bowel Syndrome, 7-item Generalized Anxiety Disorders Scale, 9-question Patient Health Questionnaire, and the majority of questions of the 30-item Endometriosis Health Profile. Cronbach α coefficients demonstrated a good internal consistency for abdominal symptoms, general somatic symptoms, and positive affect, and an excellent internal consistency for negative affect.

Conclusions: This study supports the validity and reliability of a newly developed electronic instrument for the measurement of symptoms in women with endometriosis, based on momentary assessments. This ESM patient-reported outcome measure has the advantage of providing a more detailed view on individual symptom patterns and offers the possibility for patients to have insight in their symptomatology, leading to more individualized treatment strategies that can improve the quality of life of women with endometriosis.

XSL•FO
RenderX

Introduction

Endometriosis is defined as an estrogen-dependent condition with endometrium-like tissue outside the uterus [1]. The prevalence has been estimated at around 10% in women of reproductive age and up to 50% in women with chronic pelvic pain (CPP) or fertility problems [2,3]. Dysmenorrhea, CPP, dyspareunia, fatigue, and infertility are the leading symptoms [4,5]. Therefore, endometriosis is of significant social and psychological impact, decreasing patients' quality of life. Furthermore, the annual economic burden of women with endometriosis is high [6].

Endometriosis is currently managed by surgical or medical interventions; however, approximately 50% of women with endometriosis have recurrent symptoms over a period of 5 years, irrespective of the treatment approach [7]. The anatomical stage of endometriosis is not directly related to the degree of symptoms [8], which has led to the suggestion that the perception and recurrence of symptoms may be influenced by other factors, such as psychological distress [9-11]. Therefore, accurate symptom assessment, including patient-specific factors that could influence perceived endometriosis symptoms, is of utmost importance. Currently, there is no tool available that assesses symptoms and concurrently takes into account social and psychological factors, which could be of aid in the process toward a more individualized therapeutic approach. Furthermore, outcome measures for endometriosis symptoms concern retrospective questionnaires like the widely used 30-item Endometriosis Health Profile (EHP-30) [12]. A shorter recall period reduces recall bias, and the use of repeated measurements allows evaluating the variability and the effect of time‐varying factors between multiple measurements on the outcome.

An endometriosis-specific patient-reported outcome measure (PROM) that uses the experience sampling method (ESM) has recently been developed at the Maastricht University Medical Center+ (MUMC+) [13]. This instrument provides a detailed insight into symptom patterns and helps in personalizing treatment strategies and, consequently, can aid in making meaningful changes to women's lives. The ESM is an electronic momentary assessment method characterized by randomly repeated self-reports in real-time moments that holds several advantages over traditionally used measurement tools, such as a high ecological validity and the ability to exclude recall bias as it is not relying on long-term memory [14]. The ESM offers the possibility to monitor clinically relevant experiences, such as pain, mood, and behavior, in the context of daily life and offers the ability to monitor fluctuation in symptoms to allow the assessment of the temporal relationship between variables [14-16]. It enables the assessment of various constructs and psychological mechanisms, for example, stress sensitivity and coping, which are difficult to assess using cross-sectional questionnaires [17]. It has frequently been used in psychiatric patients, and the evidence regarding the ESM in somatic illnesses is growing [18-24]. This study aimed to evaluate the psychometric properties of the above-mentioned endometriosis-specific ESM tool.

Methods

Ethics Approval

This prospective study was approved by the medical ethics committee of MUMC+ (Ref. No. 2019-1261).

Study Participants

Patients with premenopausal endometriosis aged ≥18 years, diagnosed by physical examination, imaging techniques, or laparoscopy, were recruited at the outpatient gynecology department at MUMC+, a tertiary referral center, between December 2019 and November 2020. Furthermore, patients were recruited by advertisement on the website of the Dutch Endometriosis Foundation. Subjects were eligible for inclusion if they reported one of the endometriosis-related pain symptoms (ie, dysmenorrhea, pelvic pain, or dyspareunia) on average at least 1 day per week during the last 3 months. Pregnant women and patients with any other organic explanation for CPP were not eligible for participation. Participants had to be able to understand written Dutch, and written informed consent was obtained from all participants before participation.

Data Collection

Overview

At baseline, patients were asked to fill in an electronic clinical case report form (eCRF) concerning demographic information and medical history in Castor electronic data capture (EDC) [25]. Then, during the 7-day study period, subjects completed repeated real-time questionnaires using the ESM tool and an end-of-day questionnaire. At the end of the study period, several retrospective questionnaires were completed.

ESM Questionnaire

The ESM questionnaire contains items concerning endometriosis symptoms, general somatic symptoms, psychological symptoms, and contextual/social information as well as questions concerning nutrition and medication. Depending on previous answers on trigger questions, this questionnaire comprises a minimum of 31 and a maximum of 42 items. Additionally, there is a morning questionnaire that includes questions concerning sleep and sexual behavior, which comprises a minimum of 4 and a maximum of 7 questions [13]. Most questions are scored on an 11-point Numeric Rating Scale (NRS; 0=*not at all* to 10=*very much so*). However, some questions were scored on a scale from –5 to +5 (0=*neutral*). The development of this ESM questionnaire for patients with endometriosis was previously

described [13]. The complete questionnaire was officially translated from Dutch to English by MediLingua Translations and is listed in Multimedia Appendix 1. The smartphone app MEASuRE (Maastricht Electronic Abdominal Symptom REporting), which in its current form was specifically developed for the use of the ESM in patients with endometriosis [13], was downloaded on participants' smartphones and activated for the 7-day course of the study period. Subjects were instructed to carry their smartphone with them during this week. The MEASuRE app was set to send out an auditory and written signal 10 times a day at random moments between 07:30 AM and 10:30 PM. The ESM questionnaire, also called "beep assessment" or "beep questionnaire," was available for 10 minutes after a signal, and after this time frame, questionnaires

were considered as missing data if not completed. Therefore, subjects were instructed to complete as many questionnaires as possible each day, but to skip the questionnaire if answering the questions was not appropriate, for example, when driving a car.

End-of-Day and End-of-Week Symptom Questionnaires

The Brief Pain Inventory (BPI Short Version—Dutch language) was used as an end-of-day pain diary. An average pain score experienced over the last 24 hours was given on an 11-point NRS (0=*not at all* to 10=*very much so*) at the end of each study day (on paper) [26]. At the end of the 7-day study period, patients completed validated symptom questionnaires using an eCRF system (Castor EDC) [25]. The questionnaires included are listed in Table 1.

Table 1. Self-reported questionnaires for retrospective symptom assessment, used for validation of the endometriosis-specific ESM[a] questionnaire.

Self-reported questionnaire	Content	Questions, n
Pain on NRS[b]	Abdominal pain, dysmenorrhea, and dyspareunia	3
EHP-30[c] [12]	Health-related quality of life for patients with endometriosis. Items assess symptoms on 5 domains: pain, control and powerlessness, social support, emotional well-being, and self-image	30
GSRS-IBS[d] [27]	Gastrointestinal symptoms. Items assess symptoms of pain, bloating, constipation, diarrhea, and satiety	13
PHQ-9[e] [28]	Screening for subclinical depressive symptoms	9
GAD-7[f] [29]	Screening for generalized anxiety disorder and assessment of severity	7

[a]ESM: experience sampling method.

[b]NRS: Numeric Rating Scale.

[c]EHP-30: 30-item Endometriosis Health Profile.

[d]GSRS-IBS: Gastrointestinal Symptom Rating Scale–Irritable Bowel Syndrome.

[e]PHQ-9: 9-question Patient Health Questionnaire.

[f]GAD-7: 7-item Generalized Anxiety Disorders Scale.

Statistical Analysis

As the identical ESM questionnaire is repeatedly answered several times a day, we expected each patient to complete on average 5 of 10 assessments [13]. Power calculation resulted in a minimum of 25 patients who needed to complete the study protocol to validate this tool [30]. All analyses were performed using SPSS (SPSS Statistics for Windows, Version 25.0, Released 2017; IBM Corp.).

The compliance was calculated as the percentage of ESM questionnaires that were completed by all participants during the study period, per day, and measurement moment. Concurrent validity was assessed by comparing ESM scores with end-of-day pain scores (ie, on day level; no repeated measures within the day) and with end-of-week pain scores (ie, on week level; no repeated measures within the week). To compare ESM scores with end-of-day pain scores, the mean and maximum scores for the ESM were calculated for each of the 7 days, that is, combining all repeated measurements of each subject for the concerning day. Associations between end-of-day scores and ESM scores were tested using linear mixed effects models with the end-of-day score as the dependent and the mean ESM score

as the independent variable, a random intercept, corrected for repeated measures using a first-order autoregressive (AR1) correlation structure. The level of agreement between end-of-day and ESM scores was evaluated by calculating intraclass correlation coefficients (ICCs), based on a single-rating consistency 2-way model. The ICCs that were calculated in this study were used to measure agreement between different scores that aimed to measure the same construct, such as end-of-week scores compared with mean ESM scores. To compare ESM scores with end-of-week questionnaire scores, all ESM measures were averaged to 1 score per subject. Differences between the measurement methods were tested using paired sample *t* tests, and Pearson correlation coefficients (*r*) were calculated. A value of *r* greater than 0.7 is considered a strong correlation, between 0.4 and 0.7 a moderate correlation, and anything less than 0.4 is considered a weak or no correlation [31]. For this matter, however, we notice that the paired *t* test cannot cope with the multilevel structure of the ESM pain scores. Nevertheless, to the best our knowledge, there is no alternative analysis that could cope with this while also comparing the multiple ESM scores with a day/week mean score. Results are displayed for creating insight; no conclusions will be made from significance testing.

The internal consistency of the ESM-PROM was evaluated by dividing the items into 4 symptom domains and calculating the Cronbach coefficient per domain.

Results

Study Population

In total, 64 women were recruited. Thirty-six patients were excluded as they did not meet inclusion criteria (n=6), declined participation (n=9), did not respond to e-mail or phone calls after providing the patient information leaflet (n=17), or had to quit the study because of technical phone problems (n=4). Of 28 patients who completed the study, 15 were recruited at the outpatient clinic and 13 were recruited by advertisement. Baseline characteristics are listed in Table 2.

Table 2. Baseline characteristics (N=28).

Characteristics	Values
Recruited by MUMC+,[a] n (%)	15 (54)
Recruited by advertisement, n (%)	13 (46)
Age, mean (SD)	35.46 (9.39)
BMI, mean (SD)	25.97 (5.09)
Educational level, n (%)	
High school	2 (7)
College or university	26 (93)
Work status, n (%)	
Student	2 (7)
Unemployed	7 (25)
Employed	19 (68)
Relationship status, n (%)	
Single	7 (25)
In relationship	21 (75)
Parity, n (%)	
0	13 (46)
≥1	15 (54)
Having a menstrual cycle, n (%)	17 (61)
Use of hormones, n (%)	22 (79)
Oral contraceptives	5 (18)
Levonorgestrel-releasing intrauterine system (Mirena)	9 (32)
Progestins	8 (29)
GnRH[b]	1 (4)
Regular use of pain medication, n (%)	20 (71)
Smoking, n (%)	6 (21)
Alcohol, n (%)	21 (75)
History of abdominal surgery for endometriosis, n (%)	20 (71)
Fertility treatment, n (%)	7 (25)
Current child wish, n (%)	6 (21)
Unplanned childlessness, n (%)	4 (14)
Traumatic life event, n (%)	3 (11)

[a]MUMC+: Maastricht University Medical Center+.

[b]GnRH: gonadotropin-releasing hormone.

Compliance

Of all ESM morning questionnaires, 87.2% were completed and those took on average 18 seconds to complete. The completion rate of the ESM beep assessments was 52.1%, which corresponds to a mean number of completed assessments of 37 out of 70 per individual over the 7 days. It took participants on average 2 minutes 11 seconds to complete a beep questionnaire. The lowest number of completed measurements per subject was 19; the highest number was 63. Table 3 shows the mean percentage of completed ESM assessments per study day. The response rate was highest on the first 2 study days, with on average 65% of completed assessments. During the study period,

the compliance decreased with the lowest number of completed assessments on the last study day (34%). Table 4 shows the mean percentage of completed ESM assessments per measurement moment during the day, with 1 indicating the first assessment of the day (morning) and 10 the last assessment of the day (evening). Response rates fluctuated between 39% and 56% during the day.

In total, 1.8% of questionnaires were started but not completed. The end-of-day pain questionnaire (on paper) was on average completed by 67.6%, and all participants completed the end-of-week questionnaire.

Table 3. Mean % completed ESMa assessments per study day.

Study days	Mean completion, %
1	64.6
2	65.4
3	59.6
4	56.1
5	45.7
6	38.9
7	34.3

[a]ESM: experience sampling method.

Table 4. Mean % of completed ESMa assessments per measurement moment during the day, averaged for all subjects and all 7 days during the study period.

Measurement moment	Mean completion, %
1	39.29
2	48.98
3	52.55
4	55.61
5	51.02
6	52.04
7	56.63
8	52.55
9	58.16
10	54.08

[a]ESM: experience sampling method.

Concurrent Validity

End-of-Day Pain Score Compared With ESM Score

In this analysis, 19 patients were included, as 9 patients did not complete or return the end-of-day questionnaire. The average *mean* and *maximum* daily abdominal pain scores, as scored by the ESM, were 3.29 (SD 2.76) and 4.80 (SD 2.96), respectively. The average abdominal pain score in the end-of-day questionnaire was 4.31 (SD 2.32). Mean differences show a

tendency to peak reporting as the *maximum* ESM scores show less difference to the end-of-day score than the *mean* ESM scores. When correcting for repeated measures (AR1 covariate structure), associations between end-of-day scores and corresponding *mean* and *maximum* ESM scores were significant. Furthermore, both ICCs between end-of-day diary scores and *mean* and *maximum* ESM scores demonstrate a good agreement between the 2 assessment methods, indicating that both measurement methods assess the same constructs (Table 5).

XSL•FO

RenderX

Table 5. Mean difference, association, and intraclass correlation coefficients (ICCs) between end-of-day diary scores and experience sampling method (ESM) mean and maximum scores for abdominal pain in 19 patients.[a]

	Mean difference		Association		ICCs	
	Difference	SE	Estimate	SE	ICC	95% CI
End-of-day vs ESM *mean* score	0.94	2.08	0.53[b]	0.08	0.80[b]	0.71-0.86
End-of-day vs ESM *maximum* score	–0.53	0.19	0.47[b]	0.06	0.81[b]	0.73–0.87

[a]Mean differences tested using the paired sample *t* test. A positive difference indicates a higher end-of-day score than the ESM. Significance of associations tested using mixed linear models with end-of-day diary scores as the dependent variable and ESM mean or maximum scores as the independent variable, corrected for repeated measures (first-order autoregressive covariate structure). The estimate indicates the strength and direction of the association.

[b]$P<.001$.

End-of-Week Scores Compared With ESM

The comparison and correlation between ESM and end-of-week pain scores scored on the NRS are shown in Table 6. End-of-week scores for abdominal pain and dysmenorrhea were significantly higher than mean ESM scores. However, the end-of-week score for abdominal pain did not differ significantly from ESM maximum scores, indicating peak level reporting at the end of the week. Dyspareunia scored by the ESM and at the end of the week were not significantly different. Pearson correlation showed strong and significant correlations for all pain scores (ie, abdominal pain, dysmenorrhea, and dyspareunia). The dysmenorrhea score for the ESM was calculated as the average abdominal pain score on days that patients experienced vaginal blood loss.

Table 6. End-of-week pain scores compared with experience sampling method (ESM) scores (mean scores on subject level).[a]

End-of-week vs ESM	End-of-week score, mean (SD)	ESM score, mean (SD)	Mean difference	SE	Pearson correlation coefficient (*r*)
Abdominal pain	4.96 (2.36)	3.21 (2.99); 4.71 (2.95)[b]	1.807; 0.371[b]	0.394; 0.524[b]	0.821; 0.830[b,c]
Dysmenorrhea	5.50 (2.59)	4.59 (3.18)	1.765	1.641	0.813[d]
Dyspareunia	2.69 (2.36)	2.63 (2.36)	–0.167	0.749	0.945[c]

[a]Differences were tested using the paired sample *t* test.

[b]Maximum (SD) values.

[c]$P<.001$.

[d]$P<.05$.

To visualize the comparison between the ESM and end-of-day as well as end-of-week scores in more detail, average abdominal pain scores for the ESM per measurement moment and the end-of-day and end-of-week scores are shown in Figure 1. This figure depicts a highly fluctuating pattern of abdominal pain during the 7-day study period when assessed using the ESM. The end-of-day scores are extrapolated to the entire day and the end-of-week scores to the entire week. The discrepancy between the assessment methods is highlighted here.

Figure 1. Abdominal pain scores (on an 11-point Numeric Rating scale) for ESM, end-of-day, and end-of-week scores over the 7-day study period, averaged for all participants. Each day, 10 assessments were available for the ESM; 1 assessment was completed at the end of each day and 1 assessment was completed at the end of the week. ESM: experience sampling method.
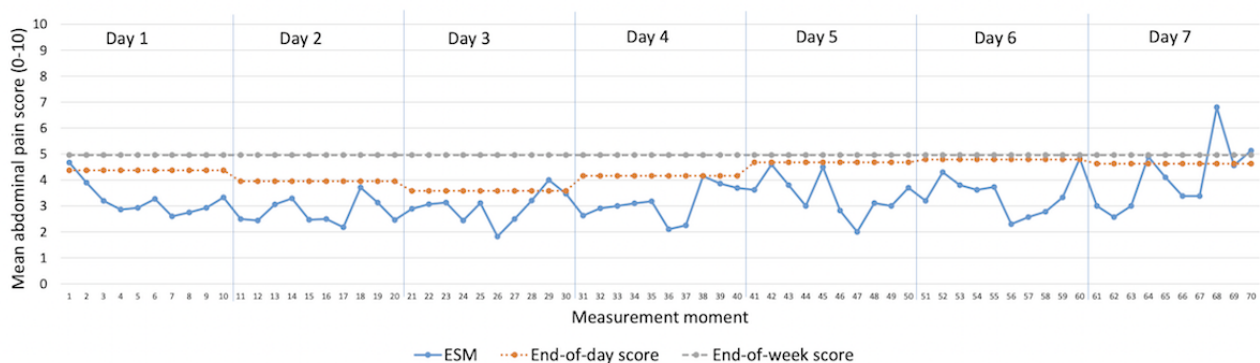
The correlation between ESM and end-of-week scores is shown in Table 7 and assessed by the EHP-30, Gastrointestinal Symptom Rating Scale–Irritable Bowel Syndrome (GSRS-IBS), 9-question Patient Health Questionnaire (PHQ-9), and 7-item Generalized Anxiety Disorders Scale (GAD-7), which do not allow the comparison between mean scores due to differences in scoring systems. Strong and significant correlations were found between the ESM and all symptoms scored by the GSRS-IBS, GAD-7, and PHQ-9, except for "feeling down," which showed a significant but moderate correlation. Strong and significant correlations between the ESM and EHP-30 were found in 11 of 18 possible compared items (61%). Four items (22%) showed a moderately significant correlation, and for 3 items (17%), a small, nonsignificant correlation was found.

Table 7. Correlation between end-of-week scores and experience sampling method (ESM) scores (concurrent validity).

Symptoms	Pearson correlation coefficient (r)
Correlation between EHP-30[a] and ESM	
EHP-30: Been unable to go to social events because of the pain? ESM: My symptoms are getting in the way of my social activities	0.520[b]
EHP-30: Been unable to do jobs around the home because of the pain? ESM: My symptoms are getting in the way of my household chores	0.222[c]
EHP-30: Found it difficult to exercise or do the leisure activities you would like to do because of the pain? ESM: My symptoms are getting in the way of sports/hobbies	0.505[b]
EHP-30: Found it difficult to stand because of the pain? ESM: I feel pain when I am standing	0.145[c]
EHP-30: Found it difficult to sit because of the pain? ESM: I feel pain when I am sitting	0.554[b]
EHP-30: Found it difficult to walk because of the pain? ESM: I feel pain when I am walking	0.446[d]
EHP-30: Been unable to sleep properly because of the pain ESM: I slept well	0.606[b]
EHP-30: Had to go to bed/lie down because of the pain? ESM: I feel that I need to rest	0.254[c]
EHP-30: Been unable to do the things you want to do because of the pain? ESM: My symptoms are getting in the way of my activities	0.527[b]
EHP-30: Generally felt unwell? ESM: I feel well physically	–0.502[b]
EHP-30: Felt depressed? ESM: I feel dispirited (down)	0.729[e]
EHP-30: Felt weepy/tearful? ESM: I feel emotional	0.748[b]
EHP-30: Felt bad-tempered or short-tempered? ESM: I feel irritable	0.729[b]
EHP-30: Couldn't do certain tasks at work because of the pain? ESM: My symptoms are getting in the way of work	0.466[d]
EHP-30: Experienced pain during or after sexual intercourse? ESM: I suffered from pain during or after sexual intercourse	0.688[d]
EHP-30: Has avoided intercourse because of the pain? ESM: I avoided sexual intercourse because of pain symptoms	0.466[d]
Correlation between GSRS-IBS[f] and ESM	
GSRS: Have you been bothered by nausea? ESM: I feel nauseous	0.803[e]
GSRS: Has your stomach felt bloated? ESM: I feel discomfort due to being bloated	0.841[e]
GSRS: Have you suffered from a visibly distended abdomen? ESM: My stomach is distended (swollen)	0.859[e]
GSRS: Have you been bothered by abdominal pain? ESM: I suffer from abdominal pain	0.691[e]
Correlation between PHQ-9/GAD-7[g] and ESM	
PHQ-9: Feeling down, depressed, or hopeless ESM: I feel dispirited (down)	0.380[d]

Symptoms	Pearson correlation coefficient (*r*)
PHQ-9: Feeling tired or having little energy	–0.778[e]
ESM: I feel… (very tired – full of energy)	
GAD-7: feeling nervous, anxious, or on edge	0.726[e]
ESM: I feel stressed	
GAD-7: Worrying too much about different things	0.738[e]
ESM: I am worried	
GAD-7: trouble relaxing	–0.496[b]
ESM: I feel relaxed	
GAD-7: Becoming easily annoyed or irritable	0.533[b]
ESM: I feel irritable	

[a]EHP-30: 30-item Endometriosis Health Profile. The questionnaire starts with "During the last 4 weeks, how often, because of your endometriosis, have you…." Scale = 5-point Likert (Never – Always).

[b]*P*<.01 when differences were assessed using a paired samples *t* test.

[c]Not significant.

[d]*P*<.05 when differences were assessed using a paired samples *t* test.

[e]*P*<.001 when differences were assessed using a paired samples *t* test.

fGSRS-IBS: Gastrointestinal Symptom Rating Scale–Irritable Bowel Syndrome. Questions are asked over the past week. The scale used was a 7-point Likert scale (from no discomfort at all to Very severe discomfort).

gPHQ-9: 9-question Patient Health Questionnaire; GAD-7: 7-item Generalized Anxiety Disorder Scale. The questionnaires start with "Over the last 2 weeks, how often have you been bothered by any of the following problems?" The scale is a 4-point Likert scale (from Not at all to Nearly every day).

Internal Consistency

To determine the internal consistency of the ESM-PROM for endometriosis, items were categorized into 4 constructs: abdominal symptoms, general somatic symptoms, positive affect, and negative affect. Cronbach coefficients for each of these domains are shown in Textbox 1. Results demonstrate a good internal consistency for abdominal symptoms, general somatic symptoms, and positive affect, and an excellent internal consistency for negative affect.

Textbox 1. Internal consistency for 4 symptom domains within the experience sampling method patient-reported outcome measure, reflected by the Cronbach coefficient.

Abdominal symptoms (Cronbach =.793)

Abdominal pain, bloating, distended abdomen, urge to defecate, vaginal blood loss

General somatic symptoms (Cronbach =.767)

Dizziness, nausea, headache, muscle pain, shortness of breath

Mental—positive affect (Cronbach =.716)

Cheerful, relaxed

Mental—negative affect (Cronbach =.935)

Dispirited/down, emotional, stressed, worried, irritable

Discussion

Principal Findings

This study evaluated the concurrent validity and internal consistency of a previously developed PROM based on the ESM principle for momentary symptom assessment in endometriosis [13]. In this study, we demonstrated that ESM abdominal pain scores in women with endometriosis are strongly correlated with the corresponding end-of-day and end-of-week pain scores. However, end-of-week scores were significantly higher than *mean* ESM scores but did not differ significantly from the *maximum* ESM level during the day. This indicates peak reporting of pain scores after a period of recall. Although we notice that the paired *t* test cannot cope with the multilevel structure of the ESM pain scores and results should be

interpreted with caution, the phenomenon of peak reporting was also described earlier in an irritable bowel syndrome population [16].

ESM scores showed strong concurrent validity when compared with symptoms scored by the GSRS-IBS, GAD-7, and PHQ-9, except for a significant moderate correlation for the symptom "feeling down." This is probably due to the use of different words to express a depressed mood by the ESM ("feeling dispirited, down") versus PHQ ("feeling down, depressed, hopeless"). Furthermore, the present mental state of a person filling in a retrospective questionnaire might impact the result (psychological bias). The correlation for "feeling down" as assessed by ESM and EHP-30 was strong. Concerning the comparison of the ESM with EHP-30, social and behavioral items were less strongly correlated than the occurrence and

severity of physical and psychological symptoms. For example, nonsignificant correlations between the ESM and EHP-30 were found for the experience of being unable to do household chores, difficulty standing, and the need to rest because of symptoms. This could be explained by the fact that social or behavioral factors are less suitable for psychometric evaluation as they are not scored on a fixed scale but either occur or do not. In addition, behavioral and social items might be impacted more strongly by ecological and recall biases. This strengthens the advantages of the ESM tool, as social and behavioral responses to symptoms are important to estimate disease burden. Cronbach coefficients demonstrated a good internal consistency for abdominal symptoms, general somatic symptoms and positive affect, and an excellent internal consistency for negative affect within the ESM-PROM. As the ESM is meant to measure fluctuations within symptoms and experiences, a test-retest reliability is not a good measure of reproducibility of ESM data. Therefore, no test-retest reliability analyses were performed.

A strength of this study is that the evaluated tool was previously designed based on specific recommendations for PROM development [32]. Only 1.8% of questionnaires were started but not completed, indicating that the questionnaire was not too time-consuming. A large number of repeated measures significantly increase the power of the analyses, and the criteria for power calculation, as advised by Schuster et al [30], were met. As the ESM questionnaire is made available by a smartphone app, a potential limitation is that the compliance depends on technical issues, such as signal blocking and network failure by the data company. However, the app does not need an internet connection as it functions offline. Smartphone holders using a specific type of smartphone experienced blocking of ESM messages that could not be addressed, and thus, they could not complete the study (n=4). Another drawback of the ESM is the patient burden, as the repeated assessments during the day are more time-consuming than most conventional assessment methods. This may result in low compliance.

However, in this study, 52% of the total assessments were completed, which is considerably higher than the generally accepted completion rates of 33% in other studies [33,34]. Although 2 of 28 individual patients with endometriosis completed only 19 measurements out of 70 (=27.1% and therefore did not meet the 33% criterion), they were not excluded from the study. Furthermore, a variable number of questions of the ESM assessment due to trigger questions could result in giving negative answers to prevent follow-up questions and therefore win time. As opposed to frequently used static measures, which describe health status at a specific time point, the ESM provides dynamic assessments of symptomatology, allowing for the detection of symptom fluctuation and symptom formation, that is, how symptoms impact symptoms. Given that endometriosis is a heterogeneous disorder often presenting with psychological disturbances or general somatic complaints, which is also reflected by lower levels of quality of life [9], items were included on mental state and frequently reported physical complaints in addition to endometriosis-specific symptoms. Furthermore, environmental and behavioral items were added as they represent and influence disease burden. The assessment of patients' mental states using the ESM has been thoroughly described in psychiatric and psychosomatic research, also concerning transdiagnostic issues [14,17-24,35,36]. Accordingly, the ESM yields detailed data of the dynamics of endometriosis symptoms in daily life, which could be used to support self-insight, shared decision-making, and prediction of treatment satisfaction in clinical practice. This may also lead to individualized treatment strategies that can improve quality of life and decrease costs [14,17]. The present endometriosis-specific ESM questionnaire was developed in Dutch and was officially translated by following the translation guidelines [37]. Official English language validation is currently planned.

Conclusions

This study shows the validity and reliability of the newly developed electronic instrument for the momentary measurement of symptoms in women with endometriosis, based on the ESM. Compliance for answering the ESM questions was as high as 52%. We demonstrated peak reporting of pain scores after a period of recall since end-of-week pain scores were significantly higher than *mean* ESM scores but did not differ significantly from the *maximum* ESM level during the day. This ESM-PROM has the advantage of providing a more detailed view on individual symptom patterns, with the option to analyze interactions between symptoms, the psychological state, and environmental factors, to aid disease monitoring in both clinical practice and research settings. The ESM could offer the possibility of self-insight of patients in their symptomatology, leading to more individualized treatment strategies and shared decision-making in treatment choices, which can considerably improve the quality of life of women suffering from endometriosis.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Table S1. Set of questions (q) for the endometriosis specific ESM-PROM [13].
[PDF File (Adobe PDF File), 258 KB-Multimedia Appendix 1]

References

1. Kennedy S, Bergqvist A, Chapron C, D'Hooghe T, Dunselman G, Greb R, ESHRE Special Interest Group for EndometriosisEndometrium Guideline Development Group. ESHRE guideline for the diagnosis and treatment of endometriosis. Hum Reprod 2005 Oct;20(10):2698-2704. [doi: 10.1093/humrep/dei135] [Medline: 15980014]

2. Dunselman GA, Vermeulen N, Becker C, Calhaz-Jorge C, D'Hooghe T, De Bie B, et al. ESHRE guideline: management of women with endometriosis. Hum Reprod 2014;29(3):400-412. [doi: 10.1093/humrep/det457]

3. Shafrir AL, Farland LV, Shah DK, Harris HR, Kvaskoff M, Zondervan K, et al. Risk for and consequences of endometriosis: A critical epidemiologic review. Best Pract Res Clin Obstet Gynaecol 2018 Aug;51:1-15. [doi: 10.1016/j.bpobgyn.2018.06.001] [Medline: 30017581]

4. Bellelis P, Dias JAJ, Podgaec S, Gonzales M, Baracat EC, Abrão MS. Epidemiological and clinical aspects of pelvic endometriosis: a case series. Rev Assoc Med Bras 2010;56(4):467-471.

5. Davis GD, Thillet E, Lindemann J. Clinical characteristics of adolescent endometriosis. J Adolesc Health 1993;14(5):362-368. [doi: 10.1016/s1054-139x(08)80008-0]

6. Nnoaham KE, Hummelshoj L, Webster P, d'Hooghe T, de Cicco Nardone F, de Cicco Nardone C, et al. Impact of endometriosis on quality of life and work productivity: a multicenter study across ten countries. Fertil Steril 2011;96(2):366-373. [doi: 10.1016/j.fertnstert.2011.05.090]

7. Becker CM, Gattrell WT, Gude K, Singh SS. Reevaluating response and failure of medical treatment of endometriosis: a systematic review. Fertil Steril 2017;108(1):125-136. [doi: 10.1016/j.fertnstert.2017.05.004]

8. Vercellini P, Fedele L, Aimi G, Pietropaolo G, Consonni D, Crosignani PG. Association between endometriosis stage, lesion type, patient characteristics and severity of pelvic pain symptoms: a multivariate analysis of over 1000 patients. Hum Reprod 2007;22(1):266-271. [doi: 10.1093/humrep/del339]

9. Vitale SG, La Rosa VL, Rapisarda AMC, Laganà AS. Impact of endometriosis on quality of life and psychological well-being. J Psychosom Obstet Gynaecol 2017;38(4):317-319. [doi: 10.1080/0167482x.2016.1244185]

10. Laganà AS, La Rosa VL, Rapisarda AMC, Valenti G, Sapia F, Chiofalo B, et al. Anxiety and depression in patients with endometriosis: impact and management challenges. Int J Womens Health 2017;9:323-330. [doi: 10.2147/ijwh.s119729]

11. Vitale SG, Petrosino B, La Rosa VL, Rapisarda AMC, Laganà AS. A systematic review of the association between psychiatric disturbances and endometriosis. J Obstet Gynaecol Can 2016;38(12):1079-1080. [doi: 10.1016/j.jogc.2016.09.008]

12. Jones G, Kennedy S, Barnard A, Wong J, Jenkinson C. Development of an endometriosis quality-of-life instrument: the endometriosis health profile-30. Obstet Gynecol 2001;98(2):258-264. [doi: 10.1097/00006250-200108000-00014]

13. van Barneveld E, Lim A, van Hanegem N, Vork L, Herrewegh A, van Poll M, et al. Patient-reported outcome measure for real-time symptom assessment in women with endometriosis: focus group study. JMIR Form Res 2021;5(12):e28782. [doi: 10.2196/28782]

14. van Os J, Verhagen S, Marsman A, Peeters F, Bak M, Marcelis M, ESM-MERGE Investigators. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. Depress Anxiety 2017;34(6):481-493. [doi: 10.1002/da.22647]

15. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415]

16. Mujagic Z, Leue C, Vork L, Lousberg R, Jonkers DMAE, Keszthelyi D, et al. The experience sampling method---a new digital tool for momentary symptom assessment in IBS: an exploratory study. Neurogastroenterol Motil 2015;27(9):1295-1302. [doi: 10.1111/nmo.12624]

17. Verhagen SJW, Hasmi L, Drukker M, van Os J, Delespaul PAEG. Use of the experience sampling method in the context of clinical trials. Evid Based Ment Health 2016;19(3):86-89. [doi: 10.1136/ebmental-2016-102418]

18. Vork L, Keszthelyi D, Mujagic Z, Kruimel JW, Leue C, Pontén I, et al. Development, content validity, and cross-cultural adaptation of a patient-reported outcome measure for real-time symptom assessment in irritable bowel syndrome. Neurogastroenterol Motil 2018;30(3):e13244. [doi: 10.1111/nmo.13244]

19. Vork L, Mujagic Z, Drukker M, Keszthelyi D, Conchillo JM, Hesselink MAM, et al. The experience sampling method: evaluation of treatment effect of escitalopram in IBS with comorbid panic disorder. Neurogastroenterol Motil 2019;31(1):e13515. [doi: 10.1111/nmo.13515]

20. Herrewegh A, Vork L, Eurelings E, Leue C, Kruimel J, van Koeveringe G, et al. The development of a patient-reported outcome measure for real-time symptom assessment in a population with functional urologic complaints: a focus group study. Neurourol Urodyn 2018;37(8):2893-2903. [doi: 10.1002/nau.23808]

21. Martin EC, Leue C, Delespaul P, Peeters F, Janssen AML, Lousberg R, et al. Introducing the DizzyQuest: an app-based diary for vestibular disorders. J Neurol 2020;267(suppl 1):3-14. [doi: 10.1007/s00415-020-10092-2]

22. Vork L, Keszthelyi D, van Kuijk SMJ, Quetglas EG, Törnblom H, Simrén M, et al. Patient-specific stress-abdominal pain interaction in irritable bowel syndrome: an exploratory experience sampling method study. Clin Transl Gastroenterol 2020;11(7):e00209. [doi: 10.14309/ctg.0000000000000209]

23. Drukker M, Peters JCH, Vork L, Mujagic Z, Rutten BPF, van Os J, et al. Network approach of mood and functional gastrointestinal symptom dynamics in relation to childhood trauma in patients with irritable bowel syndrome and comorbid panic disorder. J Psychosom Res 2020;139:110261. [doi: 10.1016/j.jpsychores.2020.110261]

24. Kreiter D, Drukker M, Mujagic Z, Vork L, Rutten B, van Os J, et al. Symptom-network dynamics in irritable bowel syndrome with comorbid panic disorder using electronic momentary assessment: a randomized controlled trial of escitalopram vs. placebo. J Psychosom Res 2021;141:110351. [doi: 10.1016/j.jpsychores.2020.110351]

25. Castor Electronic Data Capture. 2019. URL: https://castoredc.com [accessed 2022-12-21]

26. Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. Ann Acad Med Singap 1994;23(2):129-138. [doi: 10.1097/01893697-199614020-00035]

27. Wiklund IK, Fullerton S, Hawkey CJ, Jones RH, Longstreth GF, Mayer EA, et al. An irritable bowel syndrome-specific symptom questionnaire: development and validation. Scand J Gastroenterol 2003;38(9):947-954. [doi: 10.1080/00365520310004209]

28. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x]

29. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092]

30. Schuster R, Schreyer ML, Kaiser T, Berger T, Klein JP, Moritz S, et al. Effects of intense assessment on statistical power in randomized controlled trials: simulation study on depression. Internet Interv 2020;20:100313. [doi: 10.1016/j.invent.2020.100313]

31. Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med 2018;18(3):91-93. [doi: 10.1016/j.tjem.2018.08.001]

32. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes 2006;4:79. [doi: 10.1186/1477-7525-4-79]

33. Palmier-Claus JE, Myin-Germeys I, Barkus E, Bentley L, Udachina A, Delespaul PA, et al. Experience sampling research in individuals with mental illness: reflections and guidance. Acta Psychiatr Scand 2011;123(1):12-20. [doi: 10.1111/j.1600-0447.2010.01596.x]

34. Delespaul P. Assessing Schizophrenia in Daily Life: The Experience Sampling Method. Maastricht: Datawyse/Universitaire Pers Maastricht; 1995.

35. Myin-Germeys I, Peeters F, Havermans R, Nicolson NA, DeVries MW, Delespaul P, et al. Emotional reactivity to daily life stress in psychosis and affective disorder: an experience sampling study. Acta Psychiatr Scand 2003;107(2):124-131. [doi: 10.1034/j.1600-0447.2003.02025.x]

36. Verhagen SJW, Berben JA, Leue C, Marsman A, Delespaul PAEG, van Os J, et al. Demonstrating the reliability of transdiagnostic mHealth routine outcome monitoring in mental health services using experience sampling technology. PLoS ONE 2017;12(10):e0186294. [doi: 10.1371/journal.pone.0186294]

37. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 2000;25(24):3186-3191. [doi: 10.1097/00007632-200012150-00014]

Abbreviations

AR1: first-order autoregressive
CPP: chronic pelvic pain
eCRF: electronic clinical case report form
EDC: electronic data capture
EHP-30: 30-item Endometriosis Health Profile
ESM: experience sampling method
GAD-7: 7-item Generalized Anxiety Disorders Scale
GSRS-IBS: Gastrointestinal Symptom Rating Scale–Irritable Bowel Syndrome
ICC: intraclass correlation coefficient
MEASuRE: Maastricht Electronic Abdominal Symptom REporting
MUMC+: Maastricht University Medical Center+
NRS: Numeric Rating Scale
PHQ-9: 9-question Patient Health Questionnaire
PROM: patient-reported outcome measure

XSL•FO
RenderX