Received: 15 December 2020 | Revised: 16 February 2021 | Accepted: 9 March 2021

DOI: 10.1111/nmo.14136

ORIGINAL ARTICLE

Neurogastroenterology & Motility NGM WILEY

Psychometric evaluation of an experience sampling method–based patient-reported outcome measure in functional dyspepsia

Tim Klaassen[1] | Fabienne G. M. Smeets[1] | Lisa Vork[1] | Jan Tack[2] | Nicholas J. Talley[3] | Magnus Simrén[4] | Qasim Aziz[5] | Alexander C. Ford[6,7,8] | Joanna W. Kruimel[1] | José M. Conchillo[1] | Carsten Leue[9] | Adrian A. M. Masclee[1] | Daniel Keszthelyi[1]

[1]Division of Gastroenterology-Hepatology, Department of Internal Medicine, School of Nutrition and Translational Research in Metabolism (NUTRIM), Maastricht University Medical Centre+, Maastricht, The Netherlands

[2]Department of Clinical and Experimental Medicine, Translational Research Centre for Gastrointestinal Disorders (TARGID), University of Leuven, Leuven, Belgium

[3]Faculty of Health and Medicine, University of Newcastle, Callaghan, NSW, Australia

[4]Department of Molecular and Clinical Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

[5]Centre for Neuroscience and Trauma, Wingate Institute of Neurogastroenterology, Blizard Institute, Barts and the London School of Medicine & Dentistry, Queen Mary University of London, London, UK

[6]Leeds Gastroenterology Institute, Leeds Teaching Hospitals NHS Trust, Leeds, UK

[7]Leeds Institute of Medical Research at St. James's, University of Leeds, Leeds, UK

[8]Leeds Institute of Biomedical and Clinical Sciences, University of Leeds, Leeds, UK

[9]Department of Psychiatry and Psychology, Maastricht University Medical Centre, Maastricht, The Netherlands

Correspondence
Tim Klaassen, Division of Gastroenterology-Hepatology, Department of Internal Medicine, School of Nutrition and Translational Research in Metabolism (NUTRIM), Maastricht University Medical Centre+, P.O. Box 5800, 6202 AZ Maastricht, The Netherlands.
Email: t.klaassen@maastrichtuniversity.nl

Abstract

Background: Due to important biases, conventional end-of-day and end-of-week assessment methods of gastrointestinal symptoms in functional dyspepsia (FD) are considered suboptimal. Real-time symptom assessment based on the experience sampling method (ESM) could be a more accurate measurement method. This study aimed to evaluate validity and reliability of an ESM-based patient-reported outcome measure (PROM) for symptom assessment in FD.

Methods: Thirty-five patients with FD (25 female, mean age 44.7 years) completed the ESM-based PROM (a maximum of 10 random moments per day) and an end-of-day symptom diary for 7 consecutive days. On day 7, end-of-week questionnaires were completed including the Nepean Dyspepsia Index (NDI) and Patient Assessment of Gastrointestinal Symptom Severity Index (PAGI-SYM).

Key Results: Experience sampling method and corresponding end-of-day scores for gastrointestinal symptoms were significantly associated (ICCs range 0.770–0.917).

Fabienne G. M. Smeets and Lisa Vork contributed equally.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2021 The Authors. Neurogastroenterology & Motility published by John Wiley & Sons Ltd.

However, end-of-day scores were significantly higher (Δ0.329–1.031) than mean ESM scores ($p < 0.05$). Comparing ESM with NDI and PAGI-SYM scores, correlations were weaker (Pearson's r range 0.467–0.846). Cronbach's α coefficient was good for upper gastrointestinal symptoms ($\alpha = 0.842$). First half-week and second half-week scores showed very good consistency (ICCs range 0.913–0.975).

Conclusion and Inferences: Good validity and reliability of a novel ESM-based PROM for assessing gastrointestinal symptoms in FD patients was demonstrated. Moreover, this novel PROM allows to evaluate individual symptom patterns and can evaluate interactions between symptoms and environmental/contextual factors. ESM has the potential to increase patients' disease insight, provide tools for self-management, and improve shared decision making. Hence, this novel tool may aid in the transition toward personalized health care for FD patients.

KEYWORDS
experience sampling method, functional dyspepsia, patient-reported outcome measure, questionnaires, symptom perception

Key points

- Accurate recording of symptoms is the cornerstone of the clinical evaluation of functional gastrointestinal disorders, including FD.
- Due to important biases, conventional end-of-day and end-of-week assessment methods of gastrointestinal symptoms are considered suboptimal.
- Our novel ESM-based PROM is valid and reliable to asses symptoms in FD.
- The use of ESM-based PROMs in patients with FD has the potential to aid in the shift towards personalised healthcare.

1 | INTRODUCTION

Functional dyspepsia (FD) is one of the most common functional gastrointestinal disorders, recently renamed in disorders of gut-brain interaction, with an estimated prevalence of 10%–15% in the general population.[1] According to the Rome IV criteria for functional gastroduodenal disorders, it is defined by the presence of various symptoms in the absence of organic, systemic, or metabolic diseases that could explain complaints.[2] Among the heterogeneous presentation of patients with FD, four core symptoms have been defined: early satiation, postprandial fullness, epigastric burning, and epigastric pain.[2,3] Quality of life and work productivity are impaired in patients with FD.[4] Moreover, up to 40% of patients will consult a physician, having substantial financial implications.[5] The diagnosis of FD largely relies on symptoms, since underlying pathophysiologic mechanisms remain unclear and specific biological markers are currently lacking. Hence, symptom assessment is warranted to evaluate treatment efficacy. The Food and Drug Administration (FDA) recommends the use of well-defined patient-reported outcome measures (PROMs) for evaluation of treatment outcomes in clinical trials.[6] A recent systematic review identified 20 PROMs for assessment of dyspeptic symptoms. However, no single instrument has undergone all the development steps recommended by the FDA. Therefore, no consensus has yet been reached with regard to the most relevant outcome measure in patients with FD.[7]

The currently used assessment methods to evaluate dyspeptic symptoms and response to treatment are mainly retrospective, self-reported questionnaires, based on daily or weekly monitoring. This has important limitations. Firstly, retrospective questionnaires are prone to recall bias.[8] Secondly, symptom variability in functional disorders can occur due to external triggers, such as intake of food or psychological factors,[9–11] which cannot be accurately captured by retrospective assessments, thereby resulting in ecological bias. Thirdly, lack of patient adherence or fake adherence is common problems that arise with the use of paper questionnaires.[12] These limitations underline the need for a reliable method for symptom assessment to evaluate treatment efficacy in patients with FD.

The experience sampling method (ESM) might overcome these limitations. ESM is characterized by random, repeated assessments in a patient's current state and environment. Assessment has to be completed within a short time after an auditory signal, and questions always relate to current symptoms, contextual factors, and psychological factors. Therefore, ESM might be able to reduce the risk of recall and ecological bias and capture symptom variability

over time.[13,14] Current use of ESM in gastrointestinal disorders is limited. Several studies evaluated the use of ESM in patients with irritable bowel syndrome (IBS) and found good correlation between symptom scores on ESM and retrospective questionnaires. However, abdominal pain scores were significantly higher in retrospective questionnaires compared with mean scores derived from ESM. Interestingly, the scores for abdominal pain on retrospective questionnaires seemed to represent the peak scores measured by ESM.[15–17] Currently, there are no data available on the use of ESM in patients with FD. Recently, an ESM-based PROM was developed for patients with FD.[18] Therefore, the present study aimed to assess the validity and reliability of this novel FD-specific ESM-based PROM.

2 | METHODS

The study protocol was approved by the Medical Ethics Committee of the Maastricht University Medical Centre+ (MUMC+), Maastricht, the Netherlands (ID METC19-077), and performed in full accordance with the Declaration of Helsinki (latest amendment by the World Medic Association in 2013) and Dutch Regulations of Medical Research involving Human Subjects (WMO, 1998). This prospective observational study was performed at the MUMC+ from May 29, 2020, until October 1, 2020. This study was registered in the US National Library of Medicine (http://www.clinicaltrials.gov, ID NCT04204421).

2.1 | Subjects

Recruitment of patients with FD, aged between 18 and 75 years, took place at the outpatient clinic of Gastroenterology and Hepatology of the MUMC+, a secondary/tertiary academic hospital. Additionally, patients with FD that participated in other studies of the MUMC+ were contacted to participate in the current study (NCT02522000, NCT03652571). Functional dyspepsia, including subtype assessment, was diagnosed according to the Rome IV criteria, which were evaluated by a trained clinical researcher in a face-to-face interview. For all subgroups of FD, subjects needed to have symptoms for at least 6 months. As per Rome IV definitions, in order to fulfill criteria for postprandial distress syndrome (PDS), subjects needed to experience either (1) an uncomfortably full feeling after meal of a normal portion size for at least 3 days or more per week in the past 3 months that restricted their normal activities or (2) have an uncomfortable feeling of early satiation that resulted in inability to finish a normal-sized meal for at least 3 days or more per week in the past 3 months. In order to fulfill criteria for epigastric pain syndrome (EPS), subjects needed to experience either (1) chest pain or (2) pyrosis for at least 1–2 days per week that restricted their normal activities. In order to fulfill criteria for overlap syndrome (OS), subjects needed to fulfill both PDS and EPS criteria. Apart from the Rome IV criteria, no specific minimum symptom frequency/intensity on a weekly basis was used as an entry criterion. Exclusion criteria were the initiation of regularly used medication from 1 month before inclusion until the end of the study period, a history of upper gastrointestinal surgery, history of radiation therapy to the abdomen, and pregnancy. Subjects could only participate if they understood the Dutch language and were able to use the smartphone application. Moreover, as studies in the general population show patients fulfilling criteria for both FD and IBS are common,[19–23] fulfilling criteria for IBS was not deemed an exclusion criterion, as to adequately reflect FD patients in the general population.

2.2 | Data collection

Experience sampling method and an end-of-day symptom diary were collected during seven consecutive days. On day 7, subjects completed validated symptom questionnaires using an electronic case report form (eCRF) system (CastorEDC).

2.3 | ESM

The MEASuRE-D application was developed for the use of ESM in patients with FD.[18] Figure S1 displays the home screen of the application. The subjects downloaded this application on their smartphones. During their regular daily life, subjects completed ESM for 7 consecutive days. In order to complete the real-time questionnaires as often as possible, subjects were instructed to carry their smartphone with them during the week. The MEASuRE-D application sent out a haptic, auditory, and written signal 10 times per day between 07:00 and 22:00 at randomly chosen moments, with a time interval of at least 15 min between consecutive signals. Following a signal, the ESM-questionnaire (called 'Beep vragenlijst', Figure S1) was available for 10 min. On all measurement moments, the questions were repeated in the same order, and scored on an 11-point numeric rating scale (NRS) (0 = not at all to 10 = very severely). The development of this ESM-based questionnaire has been described previously.[18]

2.4 | End-of-day diary

A 7-day end-of-day symptom diary was used to evaluate symptom severity on a daily basis. Gastrointestinal symptoms (i.e., upper abdominal fullness, upper abdominal heaviness, bloating, upper abdominal pain, upper abdominal burning sensation, lower abdominal pain, nausea, belching, heartburn, regurgitation, ability to eat normal portion sizes, vomiting, and urge to defecate) were scored using an 11-point NRS (0 = not at all to 10 = very severely) at the end of each test day. This symptom diary was built into the MEASuRE-D application and made available between 19:00 and 0:00. Subjects were instructed to manually open the application to fulfill this diary, as no signal was sent to the smartphone to indicate availability of this questionnaire. In the application, this list was called 'Avondvragenlijst' (Figure S1).

2.5 | End-of-week questionnaires

At the end of the study period, validated questionnaires were completed using an eCRF, assessing upper gastrointestinal symptoms and mental health status. Regarding upper gastrointestinal symptom severity, the Nepean Dyspepsia Index (NDI; 0–4 scale for occurrence of core complaints, 0–5 scale for severity of core complaints, 0–4 scale for hinderance due to core complaints, recall period 14 days)[24–26] and Patient Assessment of Gastrointestinal Symptom Severity Index (PAGI-SYM; 1–6 scale; composes subscores for postprandial fullness, nausea/vomiting, bloating, upper abdominal pain, lower abdominal pain, heartburn/regurgitation, recall period 14 days)[27] were completed.

The Generalized Anxiety Disorder Scale-7 (GAD-7; 0–3 scale; total composite score for severity of anxiety symptoms; recall period of 14 days),[28] the Hospital Anxiety and Depression Scale (HADS; 0–3 scale, total composite scores for severity of anxiety and depression, recall period 1 week),[29] and the Patient Health Questionnaire-9 (PHQ-9; 0–3 scale; total composite score for severity of depressive symptoms, recall period 14 days)[30] were collected regarding anxiety and depression.

2.6 | Statistical analyses

Sample size was based on previous studies using ESM data that have shown sample sizes between 20 and 30 subjects to be sufficient for analyses.[31,32] Moreover, in IBS patients, sample sizes of 26–37 were used to evaluate a novel ESM-based PROM.[16,33,34] The present study was an exploratory study on the use of an ESM-based PROM in patients with FD. Therefore, we aimed to include at least 30 valid cases with a maximum of 36. Subjects were included in the analyses only when at least 1/3 of the total number of ESM assessments (i.e., 23 out of 70) were completed.[35,36]

All analyses were performed using R version 3.6.3. Continuous outcomes are presented as mean ± standard deviation (SD) and tested using paired or independent samples *t*-test. Proportions for categorical variables were tested using the χ^2-test. For all analyses, $p < 0.05$ was considered statistically significant.

Experience sampling method scores were compared with end-of-day diary scores and with end-of-week questionnaire scores to assess concurrent validity. In order to compare ESM scores with end-of-day diary scores, mean and maximum scores for ESM were calculated for each of the 7 days. Associations between ESM scores and end-of-day scores were tested using a linear mixed-effects model with end-of-day score as the dependent and ESM score as the independent variable, a random intercept, and correcting for repeated measures by using an autoregression (AR1) correlation structure. The level of agreement between these scores was assessed by calculating intraclass correlation coefficients (ICC), based on a single-rating, consistency, two-way model. Additionally, intercept-only linear mixed-effects models with the delta scores (i.e., difference between ESM and end-of-day diary) as the dependent variable were used to assess differences between assessment methods.

In order to compare ESM scores with end-of-week questionnaire scores, average ESM scores were calculated per subject. Paired samples *t*-test and Pearson correlations were calculated to assess the differences between measurement methods. A Pearson *r* above 0.7 reflects a strong correlation, a Pearson *r* of 0.50–0.70 reflects a good correlation, a Pearson *r* between 0.3 and 0.5 reflects a moderate correlation, and a Pearson *r* below 0.30 reflects a poor correlation.[37] For the PAGI-SYM questionnaire, the subscores for postprandial fullness, nausea/vomiting, bloating, upper abdominal pain, lower abdominal pain, and heartburn/regurgitation were used. Corresponding ESM scores that were used were fullness, nausea, bloating, upper abdominal pain, lower abdominal pain, and heartburn, respectively. For comparison between ESM scores and NDI or PAGI-SYM, end-of-week scores were rescaled to an 11-point scale by multiplying the NDI and PAGI-SYM scores by $\frac{11}{6}$.

Reliability was assessed with the internal consistency and test-retest reliability. For assessment of internal consistency with Cronbach's α coefficient, the ESM-PROM items were divided into five domains (i.e., upper gastrointestinal symptoms, lower gastrointestinal symptoms, physical non-GI symptoms, mental positive affect, and mental negative affect). Good internal consistency is reflected by Cronbach's α of 0.7–0.9.[38]

For assessment of test-retest reliability, we assumed that ESM scores during the first half of the study period (i.e., days 1, 2, and 3) would show moderate-to-good consistency with ESM scores during the second half of the study period (i.e., day 5, 6, and 7). For each symptom, mean scores were calculated per subject for these two time periods. A paired samples *t*-test was performed to test the differences between these study periods in order to exclude a time effect. Agreement was assessed by calculating an ICC between the time periods. For this, a two-way model based on average measures and absolute agreement was used. ICC values above 0.75 are considered good, whereas ICC values between 0.5 and 0.75 are considered moderate.[39]

3 | RESULTS

3.1 | Subjects

In total, 36 patients with FD met the inclusion criteria. One patient did not complete at least 1/3 of the total number of ESM assessments. Therefore, 35 patients (25 female [71.4%], age 44.7 [SD 15.7] years, GAD-7 4.8 [SD 4.5], HADS-Anxiety 5.3 [SD 4.5], HADS-Depression 4.9 [SD 4.3]) were included in the analyses. Ten patients fulfilled the criteria for EPS (28.6%), seven for PDS (20%), and 18 for OS (51.4%). Of the 35 included patients, 13 had comorbid IBS (37%, one in the EPS group, three in the PDS group, and nine in the overlap group). During the study, no adverse events were reported by the subjects.

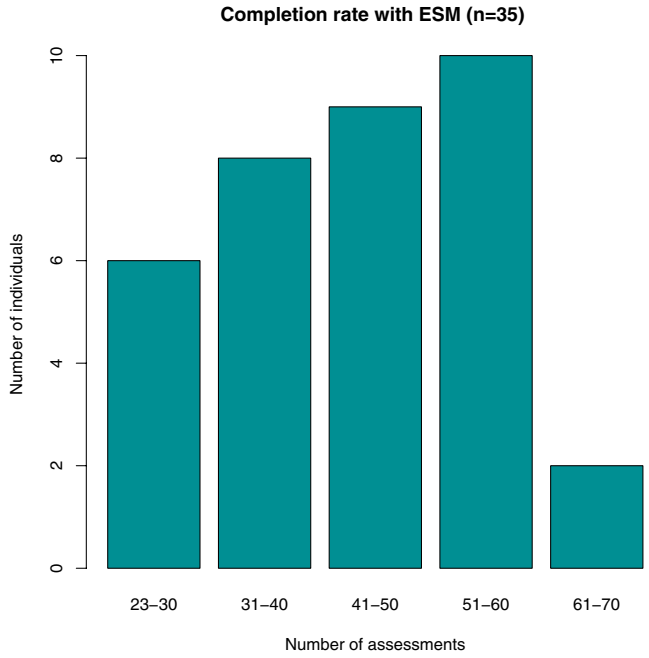

FIGURE 1 Number of individuals (*y*-axis) per category of completed number of assessments (*x*-axis)

TABLE 1 Associations and intraclass correlation coefficients between mean ESM scores and end-of-day diary scores

	Associations		Intraclass correlations	
Symptom	Estimate	SE	ICC	95%-CI
Fullness	0.952***	0.084	0.790	0.729–0.839
Heaviness	1.025***	0.060	0.803	0.745–0.849
Bloating	1.092***	0.054	0.865	0.824–0.898
Upper abdominal pain	0.953***	0.062	0.833	0.783–0.873
Upper abdominal burning	0.735***	0.069	0.885	0.849–0.913
Lower abdominal pain	1.103***	0.078	0.770	0.703–0.823
Nausea	1.147***	0.052	0.872	0.832–0.903
Belching	1.045***	0.061	0.861	0.818–0.894
Heartburn	1.107***	0.043	0.917	0.891–0.938
Regurgitation	1.154***	0.078	0.771	0.704–0.824

Note: Mixed linear models with end-of-day diary score as dependent variable and ESM mean scores as independent variable, corrected for repeated measures (AR1 covariate structure) were used to test significance. Strength and direction of the association is depicted by estimate.

Abbreviations: ICC, Intraclass Correlation Coefficient; SE, Standard Error.

***$p < 0.001$.

3.2 | Compliance

The completion rate of ESM assessments was 62.2%. Over the 7-day period, a mean number of 43.5 measurements was completed per individual (range: 23–68). The majority of subjects completed between 31 and 60 assessments during the study period (Figure 1).

3.3 | Concurrent validity

3.3.1 | ESM scores compared with end-of-day diary scores

Both ESM and the end-of-day diary scored the following gastrointestinal symptoms: upper abdominal fullness, upper abdominal heaviness, bloating, upper abdominal pain, upper abdominal burning sensation, lower abdominal pain, nausea, belching, heartburn, and regurgitation.

Mean scores on ESM and end-of-day scores were all significantly associated, which indicates that both assessment methods measure the same construct (Table 1). Furthermore, ICCs between

TABLE 2 Difference between ESM mean and maximum scores and end-of-day diary scores

	ESM mean versus end-of-day		ESM maximum versus end-of day	
Symptom	Difference	SE	Difference	SE
Fullness	1.031***	0.191	−1.001***	0.233
Heaviness	0.875***	0.127	−1.133***	0.164
Bloating	0.793***	0.133	−0.991***	0.101
Upper abdominal pain	0.563***	0.148	−0.712***	0.132
Upper abdominal burning	0.426**	0.144	−0.484***	0.124
Lower abdominal pain	0.562***	0.151	−0.459***	0.120
Nausea	0.524***	0.128	−0.597***	0.113
Belching	0.677***	0.150	−0.387***	0.115
Heartburn	0.459***	0.125	−0.398***	0.103
Regurgitation	0.329*	0.131	−0.359***	0.080

Note: A positive difference indicates a higher score, and a negative difference indicates a lower score in end-of-day diary compared with ESM. Mixed linear models with the delta score (i.e., difference between ESM mean or max score and end-of-day diary score) as the dependent variable corrected for repeated measures (AR1 covariate structure) were used to test significance. Strength and direction of the association is depicted by estimate.

Abbreviation: SE, Standard error.

*$p < 0.05$.; **$p < 0.01$.; ***$p < 0.001$.

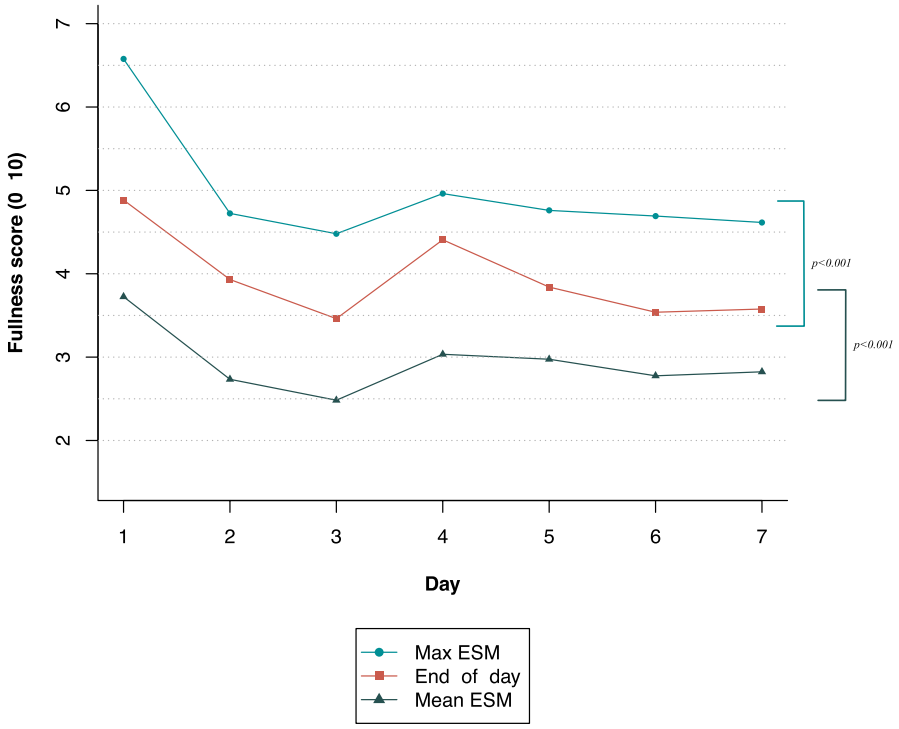

FIGURE 2 Daily fullness scores based on end-of-day diary scores and mean and maximum ESM scores (11-points NRS). Mixed linear models with the delta score (i.e., difference between ESM mean/max score and end-of-day diary score) as the dependent variable, corrected for repeated measures (AR1 covariate structure) were used to test significance

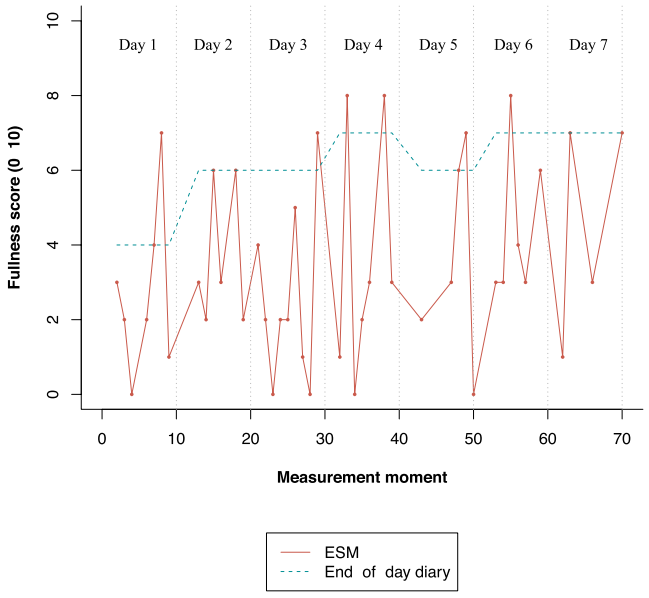

FIGURE 3 Fullness scores (on an 11-point NRS) for one FD patient on ESM and end-of-day diary over the seven-day study period. Each day, at maximum 10 assessments were available for ESM. One assessment was completed electronically at the end of each day

ESM scores and end-of-day diary scores were all above 0.75, indicating good agreement between these assessment methods (Table 1). However, symptom scores were all significantly higher in end-of-day diaries compared with mean ESM scores (Table 2). Furthermore, symptom scores were all significantly lower in end-of-day diaries compared with maximum ESM scores. Therefore, end-of-day diary scores were placed in between the mean and maximum ESM scores for all gastrointestinal symptoms. This concept is illustrated in Figure 2 for fullness scores.

These differences between end-of-day scores and both mean and maximum ESM scores can be made insightful for individual patients. Figure 3 depicts the fullness scores on the end-of-day diary and the ESM for one individual patient. A highly fluctuating pattern for fullness is observed during the 7-day study period with ESM: the patient reported multiple time-points without fullness (i.e., score 0) or with low feelings of fullness (i.e., scores below 5) and only a few time-points with higher symptom scores (i.e., above 5). Instead of real-time symptom assessment, end-of-day diary scores reflect scores of the entire day. For this individual, this resulted in end-of-day fullness scores higher than 5 for 6 out of 7 days. Figure 3 highlights the discrepancy between assessment methods.

3.3.2 | ESM scores compared with end-of-week scores

The comparison and correlation between ESM scores and end-of-week scores are depicted in Table 3. ESM scores for postprandial fullness, nausea, bloating, upper abdominal pain, lower abdominal pain, heartburn, upper abdominal burning, and belching were all lower compared with end-of-week scores on the NDI and PAGI-SYM. Scores for upper abdominal burning, heartburn, bloating, and belching on the NDI were strongly correlated with ESM scores. Scores for upper abdominal pain, fullness, and nausea on the NDI showed good correlation with ESM scores. Reflux scores on the NDI showed a moderate correlation with ESM scores. Bloating, upper abdominal pain, and heartburn/regurgitation subscores on the PAGI-SYM were

TABLE 3 ESM scores compared with end-of-week scores. Mean scores based on subject-level

NDI versus ESM	ESM score Mean ± SD	NDI score Mean ± SD	Pearson correlation
Upper abdominal pain	1.88 ± 2.05	3.72 ± 2.29***	0.617
Upper abdominal burning	1.37 ± 2.01	2.10 ± 2.56**	0.846
Heartburn	1.35 ± 1.96	2.51 ± 2.71***	0.737
Reflux	0.58 ± 1.05	1.94 ± 2.39***	0.465
Fullness	2.88 ± 1.98	4.66 ± 2.24***	0.574
Bloating	2.86 ± 1.95	4.03 ± 2.84***	0.838
Nausea	1.35 ± 1.91	3.25 ± 2.85***	0.556
Belching	1.23 ± 1.63	2.67 ± 2.79***	0.779
PAGI-SYM versus ESM		**PAGI-SYM subscale score Mean ± SD**	
Postprandial fullness	2.88 ± 1.98	3.89 ± 2.06**	0.636
Nausea/vomiting	1.35 ± 1.91	1.64 ± 2.05	0.556
Bloating	2.86 ± 1.95	4.64 ± 2.89***	0.728
Upper abdominal pain	1.81 ± 2.05	3.75 ± 2.44***	0.804
Lower abdominal pain	1.16 ± 1.56	2.20 ± 2.29***	0.680
Heartburn/regurgitation	1.35 ± 1.96	1.92 ± 1.87**	0.802
PHQ-9/GAD-7 versus ESM			
Down			0.828
Anxious			0.586
Worried			0.865
Irritated			0.724
Relaxed			0.836

Note: NDI versus ESM: NDI scores transformed from 6-point scale to 11-point scale by multiplying with $\frac{11}{6}$. PAGI-SYM sub scores versus ESM: PAGI-SYM scores transformed from 6-point scale to 11-point scale by multiplying with $\frac{11}{6}$. Corresponding ESM scores tested with PAGI-SYM scores: Early satiety and fullness, nausea/vomiting and nausea, bloating and bloating, upper abdominal pain and upper abdominal pain, lower abdominal pain and lower abdominal pain, heartburn/regurgitation and heartburn. PHQ-9/GAD-7 versus ESM: answering scales do not allow comparison between mean scores. Paired samples *t*-test was used to test for differences.

**$p < 0.01$.; ***$p < 0.001$.

strongly correlated with ESM scores. All other PAGI-SYM subscores showed good correlation with ESM scores.

In addition to gastrointestinal symptoms, psychological factors were assessed with ESM. Several psychological factors corresponded with the PHQ-9 or GAD-7, namely 'feeling down', 'feeling anxious', 'feeling worried', 'feeling irritated', and 'feeling relaxed'. Answering scales were substantially different and, therefore, did not allow for harmonization of the scores. Hence, no mean scores could be compared between assessment methods. However, correlations between the two assessment methods could be calculated. Strong correlations between ESM and end-of week scores were found for 'feeling down', 'feeling worried', 'feeling irritated', and 'feeling relaxed'. Good correlations were found for 'feeling anxious'.

3.4 | Internal consistency

Internal consistency of the ESM-PROM for FD was determined by categorizing the items in five constructs, namely upper gastrointestinal symptoms, lower gastrointestinal symptoms, physical non-gastrointestinal symptoms, positive affect, and negative affect. Table 4 lists Cronbach's α coefficients for these constructs. Very good internal consistency was found for upper gastrointestinal symptoms and negative affect. Good internal consistency was found for physical non-gastrointestinal symptoms. An acceptable internal consistency was found for positive affect. Consistency was relatively low for lower gastrointestinal symptoms, indicating that these two symptoms might not perfectly reflect the same construct for lower gastrointestinal symptoms.

3.5 | Test-retest reliability

Mean ESM scores for the first (i.e., day 1, 2, and 3) and second (i.e., day 5, 6, and 7) half-week of the study period are depicted in Table 5. Only the scores for 'feeling down', 'feeling anxious', 'feeling irritated', 'feeling stressed', and 'feeling worried' differed, with all scores lower in the second half-week. All other scores did not differ between the

TABLE 4 Internal consistency of five ESM-PROM symptom domains, reflected by Cronbach's α coefficient

Symptoms	Cronbach's α
Upper gastrointestinal symptoms	
Fullness	0.842
Upper abdominal heaviness	
Upper abdominal pain	
Upper abdominal burning	
Nausea	
Vomiting	
Belching	
Heartburn	
Regurgitation	
Lower gastrointestinal symptoms	
Bloating	0.483
Lower abdominal pain	
Physical—non-gastrointestinal	
Palpitations	0.843
Sweating	
Dyspnoea	
Dizziness	
Pressure on chest	
Mental—Positive affect	
Good	0.584
Relaxed	
Mental—Negative affect	
Down	0.917
Anxious	
Irritated	
Stressed	
Worried	

first half-week and the second half-week. All symptoms showed good consistency between the measurements, as reflected by the ICCs. These findings suggest sufficient test-retest reliability.

4 | DISCUSSION

The present study evaluated the validity and reliability of an ESM-based PROM for symptom assessment in patients with FD. The development of this tool has been previously described.[18] This study demonstrated significant associations between ESM scores for gastrointestinal symptoms and end-of-day scores, and moderate-to-strong correlations between ESM scores and end-of-week scores, confirming concurrent validity. Besides validity, reliability was considered adequate based on moderate-to-good internal consistency and excellent test-retest reliability.

Prior to the implementation of a novel PROM in patient care or clinical trials, adequate validity and reliability have to be demonstrated. Throughout the years, a plethora of statistical methods to test psychometric properties of novel PROMs have been described.[6,38,40] In this study, measures of validity and reliability that are most applicable to the novel ESM-based PROM are used. Regarding concurrent validity, significant associations for all mean ESM and end-of-day scores for gastrointestinal symptoms were found. Given the good agreement between these assessment methods, it can be stated that the ESM-PROM and end-of-day diary measure similar constructs regarding gastrointestinal symptoms. Interestingly, in this study, no peak symptom score in end-of-day reporting was shown, contrary to studies performed in IBS patients.[15,16] End-of-day diary symptom scores tended to be in between the mean and maximum ESM scores. This points toward over-reporting of gastrointestinal complaints when subjects need to provide one score over the entire day. The difference between mean ESM scores and end-of-week scores was even more pronounced. This indicates that subjects tend to remember the moments that they were aware of complaints and neglect the moments without complaints when providing scores over a longer period of time, emphasizing the usefulness of ESM in generating accurate individual symptom patterns.

This study demonstrated very good internal consistency for upper gastrointestinal symptoms, whereas a poor consistency was found for lower gastrointestinal symptoms. This suggests that the two items chosen to represent lower gastrointestinal symptoms (i.e., bloating and lower abdominal pain) should possibly be considered to reflect different constructs. It should be noted that the items were selected based on focus group interviews with FD patients. A previous study developed an ESM-based PROM for IBS patients that included more items reflecting lower gastrointestinal symptoms.[33] It could have been possible to combine both PROMs. However, this would have increased the number of total symptom items substantially and, therefore, the patient's burden. Instead, we decided to focus on the symptoms deemed essential by patients with FD, as reflected by the item selection in the focus groups. However, internal consistency might have been improved by expanding this domain with lower gastrointestinal symptoms that are deemed more appropriate.

In the present study, continuous repeated measurements were performed for 1 week. The strength and purpose of ESM is real-time assessment. Test-retest reliability of ESM was investigated by comparing and correlating mean scores of the first half-week with those of the second half-week. We hypothesized that test-retest reliability of the ESM would not be perfect due to the fluctuating nature of FD symptoms and the influence of subjects' daily life on (gastrointestinal) symptoms. The probability that within-subject differences are smaller than between-subject differences is plausible. Therefore, it is possible that a subject's own symptom pattern can be identified by using ESM. For all upper and lower gastrointestinal symptoms, and non-gastrointestinal symptoms and positive affect, no significant differences in mean scores were found, when the first half of the week was compared with the second half. Symptoms reflecting negative affect tended to be significantly lower during the second half-week. For all scores, high correlations between the first half-week and second

TABLE 5 Mean scores for ESM-reported gastrointestinal symptoms, non-gastrointestinal physical symptoms, and mental status in the first half-week and second half-week of the study period

	First half-week Mean score ± SD	Second half-week Mean score ± SD	ICC [95%-CI]
Gastrointestinal symptoms			
Fullness	2.94 ± 1.93	2.79 ± 2.08	0.913 [0.828–0.956]
Upper abdominal heaviness	2.58 ± 1.72	2.48 ± 1.91	0.935 [0.871–0.967]
Bloating	2.99 ± 1.86	2.74 ± 2.13	0.934 [0.869–0.966]
Upper abdominal pain	1.74 ± 1.89	1.86 ± 2.22	0.956 [0.913–0.978]
Upper abdominal burning	1.40 ± 1.90	1.36 ± 2.11	0.957 [0.915–0.978]
Lower abdominal pain	1.24 ± 1.57	1.13 ± 1.65	0.947 [0.896–0.973]
Nausea	1.43 ± 2.00	1.28 ± 1.82	0.970 [0.940–0.985]
Belching	1.27 ± 1.67	1.24 ± 1.77	0.942 [0.884–0.971]
Heartburn	1.30 ± 1.82	1.41 ± 2.16	0.915 [0.832–0.957]
Regurgitation	0.53 ± 1.17	0.58 ± 0.96	0.931 [0.864–0.965]
Non-gastrointestinal physical symptoms			
Palpitations	0.68 ± 1.33	0.67 ± 1.52	0.953 [0.907–0.976]
Sweating	1.50 ± 2.19	1.37 ± 2.27	0.958 [0.918–0.979]
Dyspnoea	1.20 ± 1.96	1.35 ± 2.12	0.975 [0.950–0.987]
Dizziness	0.65 ± 1.34	0.67 ± 1.47	0.955 [0.910–0.977]
Pressure on chest	0.89 ± 1.62	1.06 ± 1.95	0.955[0.910–0.977]
Tired	4.77 ± 2.36	5.03 ± 2.35	0.966 [0.930–0.983]
Mental status			
Good	6.11 ± 1.62	6.32 ± 1.50	0.931 [0.862–0.965]
Down	1.41 ± 1.85	0.98 ± 1.67**	0.929 [0.822–0.968]
Anxious	0.85 ± 1.58	0.48 ± 1.44**	0.914 [0.804–0.959]
Irritated	1.32 ± 1.90	1.02 ± 1.80*	0.963 [0.914–0.982]
Stressed	2.26 ± 2.27	1.79 ± 2.33*	0.938 [0.858–0.971]
Relaxed	6.15 ± 1.80	6.22 ± 1.83	0.929 [0.859–0.964]
Worried	2.07 ± 2.61	1.66 ± 2.54*	0.941 [0.878–0.971]

Note: Agreement between first and second half-week is reflected by the ICC. First half-week reflects day 1, 2, and 3. Second half-week reflects day 5, 6, and 7. Paired samples *t*-test was used to test for significance.

Abbreviations: ICC, intraclass correlation coefficient; SD, standard deviation.

*$p < 0.05$.; **$p < 0.01$.

half-week were demonstrated. Most subjects started the test period in the beginning of the week. Consequently, many of the second half-week assessments were scheduled on one or more weekend days. It is possible that feelings of negative affect were lower in the second half-week, due to less daily life hassles (i.e., no work pressure).

The regulatory authorities recommend the use of end-of-day reporting in FD.[41] ESM represents a higher burden for patients than end-of-day reporting, as multiple assessments during the day are likely to be more time-consuming than most conventional assessment methods. The accuracy of the assessment method would be at stake if this had resulted in low adherence. A completion rate of 33% for ESM-based PROMS is conventionally accepted.[35,36] In the present study, 62.2% of the total assessments was completed. Reported completion rates are reliable and accurate, since ESM assessments were only available for a set period of time and electronic date and timestamps were registered for each assessment.

A particular strength of the present study is the validation of this novel tool according to the recommendations of de FDA.[6] An important addition is that conventionally static measures of health status at specific time-points are used for comparison. In contrast to end-of-day or end-of-week health questionnaires, ESM is able to detect short-term fluctuations in symptoms by its dynamic assessment of symptomatology. Therefore, ESM has the ability to provide a more detailed and individual assessment of symptom patterns. Moreover, this study demonstrated the suitability of ESM to provide an overview of patients' symptoms over the 7-day period and a detailed insight into within-day fluctuations of symptoms. Furthermore, by assessing other symptoms at the exact same moments, ESM offers the opportunity to investigate associations between concurrent symptoms, environmental factors, and psychological symptoms.[13,42] This means that ESM is capable of investigating symptom formation, in other words, how symptoms impact

on symptoms. More insight in symptom dynamics may reveal a better understanding of the underlying pathophysiology in FD and a more customized treatment trajectory. A recent study in psychiatry even demonstrated the capability of ESM-based self-monitoring combined with positive emotion enhancement to enhance treatment effects in patients treated for depression.[43] This emphasizes the capability of ESM to aid in disease insight, self-management, and improved shared decision making.

Moreover, FD is characterized as a heterogeneous disorder often accompanied by general somatic complaints and/or psychological disturbances. This is reflected by a considerable number of studies describing lower levels of physical and mental quality of life.[44–48] Therefore, items reflecting frequently reported physical complaints and mental state were included in addition to gastrointestinal symptoms. Previous studies have described the assessment of patients' mental state by using ESM-based PROMs.[32,49,50] However, the main focus of the current study was the gastrointestinal complaints as reflected in this PROM.

A potential shortcoming of the present study is the relatively small sample size. For the evaluation of validity and reliability of outcome measures, it has been recommended to include at least 50 subjects.[51] However, the large number of repeated measures per subject provides a significant increase in power, which is a substantial strength of ESM concerning required sample sizes in clinical trials. Moreover, in this small sample size, adherence was reasonable. However, adherence should be evaluated in a larger pragmatic trial in order to adequately evaluate compliance to the present smartphone application. Furthermore, potential user bias should be considered, as this study required subjects to own a smartphone and to be able to adequately operate the smartphone application requiring sufficient digital skills. Additionally, intensive recording using this smartphone application was mandatory. Another important aspect in quality testing of PROMs is assessment of responsiveness (i.e., the sensitivity to detect change over time).[38,40,52] In the present study, responsiveness was not evaluated, and this has to be performed before the use of this novel ESM-based PROM as a tool to evaluate treatment efficacy in patients with FD. In addition, cross-cultural validation remains to be performed.

In conclusion, adequate concurrent validity, moderate-to-good internal consistency, and very good test-retest reliability were demonstrated for the novel ESM-based PROM for patients with FD. Moreover, the ESM-based PROM has the advantage of evaluating individual symptom patterns, providing the opportunity to evaluate interactions between symptoms and environmental factors. It must be noted that this was not evaluated in the present study as the goal of the present study was merely the validation of the used smartphone application. Future studies should evaluate the potential of this tool to evaluate these interactions, as this may lead to increased insight into their illness and tools for self-management, and improved shared decision making. Thus, this novel ESM-based tool has the ability to aid in the transition toward more personalized health care for patients with FD. Future research for assessment of responsiveness of this novel ESM-based PROM is warranted, in order to determine its place in the evaluation of treatment efficacy.

ACKNOWLEDGMENTS

We thank MEMIC (centre for data and information management at Maastricht UMC+) for the development of the smartphone application and their support regarding data management. We also thank all patients with FD who participated in this study.

DISCLOSURES

T.K.: personal fees from Will Pharma, outside the submitted work. F.G.M.S.: none to declare. L.V.: none to declare. J.T.: personal fees from Adare, personal fees from Arena, personal fees from Christian Hansen, personal fees from Devintec, personal fees from Ironwood, personal fees from Shire, personal fees from Truvion, personal fees from Abbott, personal fees from Menarini, grants from Shire, grants from Tsumura, grants from Sofar, grants from Mylan, outside the submitted work. N.J.T.: Consultancy: Allergan, IM Health Sciences, Takeda, Theravance, Danone, Sanofi, outside the submitted work. M.S.: grants and personal fees from Danone Nutricia Research, grants and personal fees from Glycom A/S, personal fees from Nestlé, personal fees from Ironwood, personal fees from Menarini, personal fees from Biocodex, grants from Genetic Analysis AS, from null, personal fees from Arena, personal fees from Adnovate, personal fees from Shire, personal fees from Tillotts, personal fees from Kyowa Kirin, personal fees from Takeda, personal fees from Alimentary Health, personal fees from AlfaSigma, personal fees from Falk Foundation, outside the submitted work. Q.A.: none to declare. A.C.F.: acted as a consultant for, and has received research funding from, Almirall outside of the submitted work. J.W.K.: none to declare. J.M.C.: none to declare. C.L.: none to declare. A.A.M.: unrestricted grant from Grunenthal for development of ESM on IBS; grant from ZonMw and Will Pharma for RCT on peppermint oil in IBS, outside the submitted work. D.K.: grants from Will Pharma, Allergan, Grunenthal, ZonMw, Maag-Lever-Darmstichting, United Europe Gastroenterology, EU Horizon 2020, outside the submitted work.

AUTHOR CONTRIBUTIONS

Conceptualization: all authors; Methodology: Tim Klaassen, Lisa Vork, Fabienne Smeets, Adrian Masclee, and Daniel Keszthelyi; Formal analysis and investigation: Tim Klaassen; Writing—original draft preparation: Tim Klaassen; Writing—review and editing: all authors; Supervision: Adrian Masclee and Daniel Keszthelyi.

ORCID

Tim Klaassen https://orcid.org/0000-0002-1519-4026
Fabienne G. M. Smeets https://orcid.org/0000-0002-1761-8312
Lisa Vork https://orcid.org/0000-0001-7316-3191
Jan Tack https://orcid.org/0000-0002-3206-6704
Nicholas J. Talley https://orcid.org/0000-0003-2537-3092
Magnus Simrén https://orcid.org/0000-0002-1155-1313

Qasim Aziz https://orcid.org/0000-0002-2718-2065
Alexander C. Ford https://orcid.org/0000-0001-6371-4359
Joanna W. Kruimel https://orcid.org/0000-0001-7102-1720
José M. Conchillo https://orcid.org/0000-0003-0659-0358
Carsten Leue https://orcid.org/0000-0003-1553-5705
Adrian A. M. Masclee https://orcid.org/0000-0001-5819-0116
Daniel Keszthelyi https://orcid.org/0000-0001-9856-9980

REFERENCES

1. Aziz I, Palsson OS, Tornblom H, Sperber AD, Whitehead WE, Simren M. Epidemiology, clinical characteristics, and associations for symptom-based Rome IV functional dyspepsia in adults in the USA, Canada, and the UK: a cross-sectional population-based study. *Lancet Gastroenterol Hepatol*. 2018;3(4):252–262.
2. Stanghellini V, Chan FK, Hasler WL, et al. Gastroduodenal disorders. *Gastroenterology*. 2016;150(6):1380–1392.
3. Taylor F, Reasner DS, Carson RT, et al. Development of a symptom-based patient-reported outcome instrument for functional dyspepsia: a preliminary conceptual model and an evaluation of the adequacy of existing instruments. *Patient*. 2016;9(5):409–418.
4. Sander GB, Mazzoleni LE, Francesconi CF, et al. Influence of organic and functional dyspepsia on work productivity: the HEROES-DIP study. *Value Health*. 2011;14(5 Suppl 1):S126–S129.
5. Ford AC, Forman D, Bailey AG, Cook MB, Axon AT, Moayyedi P. Who consults with dyspepsia? Results from a longitudinal 10-yr follow-up study. *Am J Gastroenterol*. 2007;102(5):957–965.
6. U.S. Department of Health and Human Services FaDA, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH). Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims. 2009.
7. Smeets FGM, Masclee AAM, Conchillo JM, Keszthelyi D. Systematic review: disease-specific instruments to assess gastrointestinal symptoms in functional dyspepsia. *Neurogastroenterol Motil*. 2018;30:e13327.
8. Bradburn NM, Rips LJ, Shevell SK. Answering autobiographical questions: the impact of memory and inference on surveys. *Science*. 1987;236(4798):157–161.
9. Feinle-Bisset C, Azpiroz F. Dietary and lifestyle factors in functional dyspepsia. *Nat Rev Gastroenterol Hepatol*. 2013;10(3):150–157.
10. Feinle-Bisset C, Horowitz M. Dietary factors in functional dyspepsia. *Neurogastroenterol Motil*. 2006;18(8):608–618.
11. Aro P, Talley NJ, Johansson SE, Agreus L, Ronkainen J. Anxiety is linked to new-onset dyspepsia in the Swedish population: a 10-year follow-up study. *Gastroenterology*. 2015;148(5):928–937.
12. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. *Control Clin Trials*. 2003;24(2):182–199.
13. Myin-Germeys I, Oorschot M, Collip D, Lataster J, Delespaul P, van Os J. Experience sampling research in psychopathology: opening the black box of daily life. *Psychol Med*. 2009;39(9):1533–1547.
14. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol*. 2008;4:1–32.
15. Weinland SR, Morris CB, Hu Y, Leserman J, Bangdiwala SI, Drossman DA. Characterization of episodes of irritable bowel syndrome using ecological momentary assessment. *Am J Gastroenterol*. 2011;106(10):1813–1820.
16. Mujagic Z, Leue C, Vork L, 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.
17. Chan Y, So SH, Mak ADP, Siah KTH, Chan W, Wu JCY. The temporal relationship of daily life stress, emotions, and bowel symptoms in irritable bowel syndrome-Diarrhea subtype: a smartphone-based experience sampling study. *Neurogastroenterol Motil*. 2019;31(3):e13514.
18. Smeets FGM, Keszthelyi D, Vork L, et al. Development of a real-time patient-reported outcome measure for symptom assessment in patients with functional dyspepsia using the experience sampling method. *Neurogastroenterol Motil*. 2019;31(2):e13496.
19. Wang AJ, Liao XH, Xiong LS, et al. The clinical overlap between functional dyspepsia and irritable bowel syndrome based on Rome III criteria. *BMC Gastroenterol*. 2008;8:43.
20. Agreus L, Svardsudd K, Nyren O, Tibblin G. Irritable-Bowel-syndrome and dyspepsia in the general-population - overlap and lack of stability over time. *Gastroenterology*. 1995;109(3):671–680.
21. Park JM, Choi MG, Cho YK, et al. Functional gastrointestinal disorders diagnosed by Rome III questionnaire in Korea. *J Neurogastroenterol Motil*. 2011;17(3):279–286.
22. Outlaw WM, Koch KL. Dyspepsia and its overlap with irritable bowel syndrome. *Curr Gastroenterol Rep*. 2006;8(4):266–272.
23. Choi YJ, Kim N, Yoon H, et al. Overlap between irritable bowel syndrome and functional dyspepsia including subtype analyses. *J Gastroen Hepatol*. 2017;32(9):1553–1561.
24. Talley NJ, Haque M, Wyeth JW, et al. Development of a new dyspepsia impact scale: the Nepean Dyspepsia Index. *Aliment Pharmacol Ther*. 1999;13(2):225–235.
25. Jones MP, Sato YA, Talley NJ. The Nepean Dyspepsia Index is a valid instrument for measuring quality-of-life in functional dyspepsia. *Eur J Gastroenterol Hepatol*. 2019;31(3):329–333.
26. Talley NJ, Verlinden M, Jones M. Validity of a new quality of life scale for functional dyspepsia: a United States multicenter trial of the Nepean Dyspepsia Index. *Am J Gastroenterol*. 1999;94(9):2390–2397.
27. Rentz AM, Kahrilas P, Stanghellini V, et al. Development and psychometric evaluation of the patient assessment of upper gastrointestinal symptom severity index (PAGI-SYM) in patients with upper gastrointestinal disorders. *Qual Life Res*. 2004;13(10):1737–1749.
28. Spitzer RL, Kroenke K, Williams JB, Lowe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med*. 2006;166(10):1092–1097.
29. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand*. 1983;67(6):361–370.
30. 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.
31. Kimhy D, Delespaul P, Corcoran C, Ahn H, Yale S, Malaspina D. Computerized experience sampling method (ESMc): assessing feasibility and validity among individuals with schizophrenia. *J Psychiatr Res*. 2006;40(3):221–230.
32. Myin-Germeys I, Peeters F, Havermans R, et al. Emotional reactivity to daily life stress in psychosis and affective disorder: an experience sampling study. *Acta Psychiat Scand*. 2003;107(2):124–131.
33. Vork L, Keszthelyi D, Mujagic Z, 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.
34. Vork L, Keszthelyi D, van Kuijk SMJ, 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.
35. Delespaul P. *Assessing Schizophrenia in Daily Life the Experience Sampling Method*. Maastricht: UPM, Universitaire Pers Maastricht, Maastricht University; 1995.
36. Palmier-Claus JE, Myin-Germeys I, Barkus E, et al. Experience sampling research in individuals with mental illness: reflections and guidance. *Acta Psychiatr Scand*. 2011;123(1):12–20.
37. Hazra A, Gogtay N. Biostatistics series module 6: correlation and linear regression. *Indian J Dermatol*. 2016;61(6):593–601.

38. Alrubaiy L, Hutchings HA, Williams JG. Assessing patient reported outcome measures: a practical guide for gastroenterologists. *United European Gastroenterol J*. 2014;2(6):463–470.
39. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155–163.
40. Verhagen SJW, Berben JA, Leue C, 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.
41. Clinical Outcome Assessments (COA) Qualification Program DDTCOA: Functional Dyspepsia Symptom Diary (FDSD). 2017. https://www.fda.gov/media/125024/download. Accessed October 27, 2017.
42. 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 Heal*. 2016;19(3):86–89.
43. Kramer I, Simons CJ, Hartmann JA, et al. A therapeutic application of the experience sampling method in the treatment of depression: a randomized controlled trial. *World Psychiatry*. 2014;13(1):68–77.
44. Aro P, Talley NJ, Agreus L, et al. Functional dyspepsia impairs quality of life in the adult population. *Aliment Pharmacol Ther*. 2011;33(11):1215–1224.
45. Van Oudenhove L, Vandenberghe J, Vos R, Holvoet L, Demyttenaere K, Tack J. Risk factors for impaired health-related quality of life in functional dyspepsia. *Aliment Pharmacol Ther*. 2011;33(2):261–274.
46. Lee HJ, Lee SY, Kim JH, et al. Depressive mood and quality of life in functional gastrointestinal disorders: differences between functional dyspepsia, irritable bowel syndrome and overlap syndrome. *Gen Hosp Psychiatry*. 2010;32(5):499–502.
47. Mahadeva S, Goh KL. Anxiety, depression and quality of life differences between functional and organic dyspepsia. *J Gastroen Hepatol*. 2011;26:49–52.
48. Talley NJ, Locke GR, Lahr BD, et al. Functional dyspepsia, delayed gastric emptying, and impaired quality of life. *Gut*. 2006;55(7):933–939.
49. van Os J, Verhagen S, Marsman A, et al. 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.
50. van Winkel M, Nicolson NA, Wichers M, Viechtbauer W, Myin-Germeys I, Peeters F. Daily life stress reactivity in remitted versus non-remitted depressed individuals. *European Psychiatry*. 2015;30(4):441–447.
51. Arrindell WA, van der Ende J. An empirical test of the utility of the observations-to-variables ratio in factor and components analysis. *Appl Psychol Meas*. 1985;9:165–178.
52. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34–42.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Klaassen T, Smeets FG, Vork L, et al. Psychometric evaluation of an experience sampling method–based patient-reported outcome measure in functional dyspepsia. *Neurogastroenterology & Motility*. 2021;33:e14136. https://doi.org/10.1111/nmo.14136