

MEASuRing symptoms in Irritable Bowel Syndrome

Citation for published version (APA):

Vork, L. (2021). *MEASuRing symptoms in Irritable Bowel Syndrome: towards an individualized and daily life approach*. Ridderprint. <https://doi.org/10.26481/dis.20210409lv>

Document status and date:

Published: 01/01/2021

DOI:

[10.26481/dis.20210409lv](https://doi.org/10.26481/dis.20210409lv)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

MEASuRing symptoms in Irritable Bowel Syndrome:
Towards an Individualized and Daily Life approach

© Lisa Vork, Maastricht, 2021.

All rights reserved. No parts of this thesis may be reproduced or transmitted in any form or by any means, without prior permission in writing by the author, or when appropriate, by the publishers of the publications.

ISBN 978-94-6423-175-5

Cover design: Jean Scheijen

Layout: Tiny Wouters

Printed by: Ridderprint, www.ridderprint.nl

The studies presented in this thesis were performed within the framework of NUTRIM School of Nutrition and Translational Research in Metabolism (Maastricht University). The research was partially funded by Grünenthal GmbH (Aachen, Germany) and COST program BM 1106.

Printing of this thesis was financially supported by Maastricht University and Nederlandse Vereniging voor Gastroenterologie (NVGE).

MEASuRing symptoms in Irritable Bowel Syndrome: Towards an Individualized and Daily Life approach

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Maastricht,
op gezag van de Rector Magnificus, Prof.dr. Rianne M. Letschert,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen op
vrijdag 9 april 2021 om 12:00 uur

door

Lisa Vork

Promotor

Prof. dr. A.A.M. Masclee

Copromotoren

Dr. D. Keszthelyi

Dr. J.W. Kruimel

Beoordelingscommissie

Prof. dr. B.P.F. Rutten (voorzitter)

Dr. S.O. Breukink

Prof. dr. J.W.L. Cals

Prof. dr. L. van Oudenhove (KU Leuven)

Prof. dr. A.J.P.M. Smout (Amsterdam UMC, locatie AMC)

Contents

Chapter 1	General introduction	9
PART I	DIAGNOSIS AND SYMPTOMS IN IBS	21
Chapter 2	Rome III versus Rome IV criteria for irritable bowel syndrome: A comparison of clinical characteristics in a large cohort study	23
Chapter 3	Reduction in IBS symptom severity is not paralleled by improvement in quality of life in patients with irritable bowel syndrome	43
Chapter 4	Does day-to-day variability in stool consistency link to the fecal microbiota composition?	65
PART II	THE EXPERIENCE SAMPLING METHOD – DEVELOPMENT AND VALIDATION	93
Chapter 5	The experience sampling method - A new digital tool for momentary symptom assessment in IBS: An exploratory study	95
Chapter 6	Development, content validity, and cross-cultural adaptation of a patient-reported outcome measure for real-time symptom assessment in irritable bowel syndrome	115
Chapter 7	Development of a real-time patient-reported outcome measure for symptom assessment in patients with functional dyspepsia using the experience sampling method	145
Chapter 8	The development of a patient-reported outcome measure for real-time symptom assessment in a population with functional urologic complaints - A focus group study	169
Chapter 9	Psychometric evaluation of an ESM-based patient-reported outcome measure for symptom assessment in irritable bowel syndrome	189

PART III	THE EXPERIENCE SAMPLING METHOD IN IBS – APPLICATIONS IN CLINICAL RESEARCH	209
Chapter 10	Patient specific stress-abdominal pain interaction in irritable bowel syndrome: An experience sampling method study	211
Chapter 11	The experience sampling method - Evaluation of treatment effect of escitalopram in IBS with comorbid panic disorder	229
Chapter 12	General discussion	249
Addendum	Summary	263
	Nederlandse samenvatting	271
	Impact paragraph	281
	Dankwoord	287
	List of publications	297
	Curriculum Vitae	305

Chapter 1

General introduction

General introduction

Irritable Bowel Syndrome

Irritable bowel syndrome (IBS) is one of the conditions defined as functional bowel disorders (FBD), which are a subcategory of the functional gastrointestinal disorders (FGID).^{1,2} According to the most recent definition, FGIDs are considered disorders of gut-brain interaction, in which symptoms most likely result from complex interactions between several biological, psychological, and social factors.²⁻⁴ FBDs present with chronic lower gastrointestinal (GI) symptoms in the absence of any structural abnormalities on clinical examinations, that could explain these symptoms. Since the separately described FBDs exist alongside a continuum, there can be considerable overlap between the conditions.⁵⁻⁷ In IBS, however, the main presenting GI symptom is recurrent abdominal pain, which is associated with altered bowel habits. Based on the predominant stool pattern, IBS patients are often subcategorized into four subgroups: diarrhea (IBS-D), constipation (IBS-C), a mix of diarrhea and constipation (IBS-M), or undefined predominant stool form (IBS-U).¹

Research during the past decades has led to the understanding that the mechanistic pathways underlying IBS are most likely multifactorial, with a central role for the gut-brain axis. Further etiological factors recognized, amongst others, are (epi)genetics, low-grade inflammation, altered intestinal permeability, increased visceral hypersensitivity, and alterations of the gut microbiome.^{4,8,9} However, the exact underlying pathophysiological mechanisms remain incompletely understood. Although IBS is not considered a severe illness, the high disease burden results in considerable impact on the quality of life of affected individuals as well as high (health care) costs for society.^{10,11} Considering an estimated world-wide prevalence of 5-11%¹²⁻¹⁴, depending on the diagnostic criteria used, this emphasizes the need for high quality IBS research in order to further unravel the pathophysiological mechanisms and seek for effective treatments that could reduce both the individual and societal burden. An important hurdle in the development of efficacious therapies is, however, the current lack of accurate non-invasive biomarkers for diagnosis, disease monitoring and treatment evaluation, as a consequence of the unelucidated etiology. Resulting, IBS diagnosis as well as follow-up of the disease course are currently symptom-based.

IBS diagnosis and symptomatology

Throughout the past decades, multiple sets of diagnostic criteria have been proposed to guide the diagnostic process for IBS patients in clinical practice as well as patient inclusion in clinical trials. The Manning criteria were the first to define IBS in 1978^{15,16} and since 1994, the Rome Foundation has been developing and revising the currently used Rome criteria.^{2,3} The most recent version are the Rome IV criteria, which were published in 2016.^{1,2} Rome IV criteria define IBS as recurrent abdominal pain, present on (on average) at least 1 day per week, and associated with two or more of three accompanying symptoms, *i.e.* 1) a relation to defecation, 2) an association with a change in stool frequency, and 3) an association with a change in stool consistency. These criteria should have been met during the past three months with symptom onset at least six months previously. Apart from these key-symptoms, and any concurrent FGID diagnoses, IBS patients often suffer from additional GI symptoms, such as abdominal bloating and distension, flatulence, and fecal urgency.^{17,18} Moreover, comorbid extra-intestinal and psychological disorders are often present within the IBS population.^{19,20} Although IBS typically presents as a chronic condition, the disorder is often of a relapsing-remitting nature, with symptoms fluctuating highly from time to time.^{21,22} Likewise, IBS subtypes have been demonstrated to change within subjects over time.²³ However, when it comes to diagnosis, subtyping, and determination of disease severity, we often rely on patient-reported symptoms at a single timepoint. With regard to patient health care, this approach might be sufficient as this allows for treatment strategies aiming at symptom improvement for the individual patient. In contrast, from a scientific point of view, such single-time evaluations might lack to capture a complete picture of the disease and taking into account the temporal instability could give better insight in the underlying pathophysiology.

Symptom assessment in IBS

Symptom assessment, and the method as used, is thus crucial in diagnosing IBS, monitoring disease severity, and evaluating therapy, since no valid biomarkers are currently available to rely on. The current “gold standard” for symptom assessment concerning disease severity and treatment evaluation in IBS studies are end-of-day symptom scores, as recommended by the drug regulatory authorities in both Europe and the USA.^{24,25} According to the key-features of IBS in the diagnostic criteria, both abdominal pain and stool consistency/frequency should be addressed in a composite primary endpoint. This results in the

reporting of a seven-day mean score of abdominal pain on an 11-point Numeric Rating Scale (NRS), assessed at the end of each day, as the common standard. However, this method of retrospective (*i.e.* end-of-day) self-reporting has some important shortcomings. Most importantly, these retrospective reports are based on autobiographical memory, in which the reported score is a reconstruction of a few specific moments during the day rather than a reliable reflection of symptoms over the past 24 hours, which is known as recall bias.²⁶⁻²⁸ Next, ecological bias arises when memory retrieval is influenced by the individual's environment and mental state at the time of recall.^{26,27} Another limitation is the inability of retrospective assessment methods to capture temporal symptom fluctuations during the day, since end-of-day scores are a summary of 24 hours as a whole. Lastly, low compliance with paper questionnaires or end-of-day diaries has been demonstrated, with a high risk for faked compliance, meaning that assessments are completed at a later time point than instructed (*i.e.* backfilling).²⁹ Altogether, this indicates that the currently used retrospective end-of-day symptom scores might not reliably reflect patients' experiences about their IBS symptoms during the day, as a result of considerable biases distorting symptom recall.

The Experience Sampling Method

Real-time assessment of symptoms, using the Experience Sampling Method (ESM), also referred to as Ecological Momentary Assessment (EMA), has been proposed to overcome these limitations.^{26,27,30} This method has been extensively used and studied in the fields of psychology and psychiatry^{31,32}, but has been minimally implemented in somatic fields.

The ESM is a questioning method assessing patients' symptoms randomly and repeatedly during regular daily life, focusing on the subject's current state and situation, herewith overcoming recall and ecological biases. By the concurrent assessment of momentary social and environmental factors, next to obtaining the actual symptom scores, ecological bias is further reduced. Traditionally, paper dairies were used, but nowadays electronic devices offer extra advantages to the concept of ESM.³⁰ Using auditory signals, a digital device prompts the individual each time an assessment is requested, followed by an electronic questionnaire being available during a preset time window. This methodology increases compliance rates, by prompting before each assessment, and eliminates faked compliance, by using electronic timestamps.

The ESM is generally completed for a number of consecutive days, resulting in multiple repeated assessments of identical items, which enables the possibility to look into symptom fluctuations extensively. Weinland *et al.* demonstrated that IBS patients reported higher abdominal pain scores in end-of-day diaries than by ESM³³, indicating the discrepancy between retrospective and real-time assessments in IBS. Furthermore, it is recognized that IBS symptoms are often influenced by the presence of affective symptoms, such as stress or anxiety.^{34,35} Further triggering and/or alleviating factors as food intake, exercise, and bowel movements have been described.³⁶ In this context, ESM could be particularly suitable to capture the interplay between GI and affective symptoms, as well as additional daily life factors being of influence on IBS, that might all vary during the day. Likewise, this method of symptom registration could be of benefit for other symptom-based gastrointestinal diagnoses such as functional dyspepsia, or functional disorders in other somatic fields like urology and gynecology. Ultimately, ESM could be helpful in the multidisciplinary approach of these functional disorders, by taking into account the overlap in symptoms in the different patient groups. Altogether, ESM has the potential to give more insight in individual symptom patterns and might lead to a more individualized approach in IBS management.

Outline and aims

IBS is a prevalent disorder with considerable impact on quality of life for affected individuals and important (health care) costs for society. Although IBS has been extensively studied in the past, major knowledge gaps still exist with regard to underlying pathophysiology and successful therapeutic strategies. As a result, determination of disease severity, follow-up of disease course, and evaluation of treatment are mainly based on patient-reported symptom scores. This thesis addresses the use of these scores for diagnosis and evaluation of temporal evolution of IBS symptoms, with a specific focus on the Experience Sampling Method (ESM) as a novel method for symptom registration.

The current thesis is divided into three major parts. As IBS diagnosis is symptom-based, disease prevalence largely depends on the diagnostic criteria as used. Resulting from this, disease symptomatology might also differ between groups of IBS patients that have been diagnosed using different criteria. Therefore, the first part of this thesis focuses on diagnosis and symptom patterns of IBS. In **chapter 2**, we compare IBS prevalence between Rome III and Rome IV criteria and make a comparison of clinical characteristics between Rome III-IBS patients and subjects that also fulfill the more stringent Rome IV criteria. In **chapter 3**, the evolution of symptoms over time is elucidated in a 5-year follow-up study of IBS patients. In addition, **chapter 4** focuses on a different perspective of symptom evolution over time, namely the short-term stability of gut microbiota from day to day in relation to fluctuations in stool consistency. Herewith, part I of the current thesis highlights some important aspects of assessment of symptom patterns and the interpretation of IBS diagnosis.

Due to the current lack of reliable biomarkers for diagnosis and follow-up of the disease course, symptom assessment is crucial both in IBS research and clinical patient care. In part II of this thesis the focus is specifically on symptom assessment in IBS. The current “gold standard” for symptom assessment, end-of-day symptom reporting, is considered suboptimal due to limitations as recall and ecological biases. In **chapter 5**, we explore possible advantages of the ESM over commonly used paper retrospective symptom assessment methods in an IBS population. In **chapter 6**, we describe the development of an IBS-specific patient-reported outcome measure (PROM) using ESM, and in **chapter 7** and **8** we elongate on this by describing the development of similar tools for functional dyspepsia (FD) and overactive bladder syndrome (OAB), both

functional disorders that often relate to or overlap with IBS. **Chapter 9** additionally elaborates on the further validation of the ESM-PROM that was developed for IBS, in a Dutch population. Part II of this thesis may form a basis for the use of ESM in future research as well as clinical practice in functional disorders such as IBS.

In the third part of this thesis we demonstrate applications of the ESM in IBS populations. In **chapter 10**, the role of ESM in identifying associations between psychological and/or environmental momentary factors and IBS symptoms is explored. The aim here is to capture the individual interplay between daily life triggers and the experienced symptoms in order to get better insight into symptom evolution. Understanding this symptom evolution and showing patients their individual bidirectional interaction between GI symptoms and other physical or psychological complaints, might be helpful for patients to understand and interfere on their vicious cycle of recurring symptoms. This may therefore play an important role in patient empowerment. In **chapter 11**, we additionally evaluate the potential of ESM to measure treatment effect. Using data of a randomized placebo-controlled trial of escitalopram in IBS patients, we compared the results of ESM with those of conventional paper retrospective methods. This last part of the thesis demonstrates that the real-time ESM has the potential to give insight into individual IBS symptom patterns, in relation to triggering daily life factors as well as subsequent behavioral patterns, which might lead to more personalized strategies in both clinical research and patient care.

Lastly, **chapter 12** serves as an integrative view on the key findings of the studies elaborated on in this thesis, by discussing those in the perspective of currently available literature, and expands on the future implications of this thesis' conclusions.

References

1. Mearin F, Lacy BE, Chang L, et al. Bowel Disorders. *Gastroenterology*. 2016(150):1393–1407.
2. Drossman DA. Functional Gastrointestinal Disorders: History, Pathophysiology, Clinical Features and Rome IV. *Gastroenterology*. 2016(150):1262–1279.
3. Drossman DA. The functional gastrointestinal disorders and the Rome III process. *Gastroenterology*. 2006;130(5):1377-1390.
4. Enck P, Aziz Q, Barbara G, et al. Irritable bowel syndrome. *Nat Rev Dis Primers*. 2016;2:16014.
5. Aziz I, Palsson OS, Tornblom H, Sperber AD, Whitehead WE, Simren M. The Prevalence and Impact of Overlapping Rome IV-Diagnosed Functional Gastrointestinal Disorders on Somatization, Quality of Life, and Healthcare Utilization: A Cross-Sectional General Population Study in Three Countries. *Am J Gastroenterol*. 2018;113(1):86-96.
6. Whitehead WE, Palsson O, Jones KR. Systematic review of the comorbidity of irritable bowel syndrome with other disorders: what are the causes and implications? *Gastroenterology*. 2002;122(4): 1140-1156.
7. Ford AC, Bercik P, Morgan DG, Bolino C, Pintos-Sanchez MI, Moayyedi P. Characteristics of functional bowel disorder patients: a cross-sectional survey using the Rome III criteria. *Aliment Pharmacol Ther*. 2014;39(3):312-321.
8. Camilleri M. Peripheral mechanisms in irritable bowel syndrome. *N Engl J Med*. 2012;367(17): 1626-1635.
9. Holtmann GJ, Ford AC, Talley NJ. Pathophysiology of irritable bowel syndrome. *Lancet Gastroenterol Hepatol*. 2016;1(2):133-146.
10. Canavan C, West J, Card T. Review article: the economic impact of the irritable bowel syndrome. *Aliment Pharmacol Ther*. 2014;40(9):1023-1034.
11. Spiegel BM. The burden of IBS: looking at metrics. *Curr Gastroenterol Rep*. 2009;11(4):265-269.
12. Lovell RM, Ford AC. Global prevalence of and risk factors for irritable bowel syndrome: a meta-analysis. *Clin Gastroenterol Hepatol*. 2012;10(7):712-721.e714.
13. Hungin AP, Whorwell PJ, Tack J, Mearin F. The prevalence, patterns and impact of irritable bowel syndrome: an international survey of 40,000 subjects. *Aliment Pharmacol Ther*. 2003;17(5):643-650.
14. Palsson OS, Whitehead W, Tornblom H, Sperber AD, Simren M. Prevalence of Rome IV Functional Bowel Disorders Among Adults in the United States, Canada, and the United Kingdom. *Gastroenterology*. 2020.
15. Manning AP, Thompson WG, Heaton KW, Morris AF. Towards positive diagnosis of the irritable bowel. *Br Med J*. 1978;2(6138):653-654.
16. Talley NJ, Phillips SF, Melton LJ, Mulvihill C, Wiltgen C, Zinsmeister AR. Diagnostic value of the Manning criteria in irritable bowel syndrome. *Gut*. 1990;31(1):77-81.
17. Spiegel BM, Bolus R, Agarwal N, et al. Measuring symptoms in the irritable bowel syndrome: development of a framework for clinical trials. *Aliment Pharmacol Ther*. 2010;32(10):1275-1291.
18. Ringel Y, Williams RE, Kalilani L, Cook SF. Prevalence, characteristics, and impact of bloating symptoms in patients with irritable bowel syndrome. *Clin Gastroenterol Hepatol*. 2009;7(1):68-72; quiz 63.
19. Clevers E, Vaes B, Henrard S, et al. Health problems associated with irritable bowel syndrome: analysis of a primary care registry. *Aliment Pharmacol Ther*. 2018;47(10):1349-1357.

20. Whitehead WE, Palsson OS, Levy RR, Feld AD, Turner M, Von Korff M. Comorbidity in irritable bowel syndrome. *Am J Gastroenterol.* 2007;102(12): 2767-2776.
21. Clevers E, Tack J, Tornblom H, et al. Development of Irritable Bowel Syndrome Features Over a 5-year Period. *Clin Gastroenterol Hepatol.* 2018;16(8): 1244-1251.e1.
22. Mearin F, Baro E, Roset M, Badia X, Zarate N, Perez I. Clinical patterns over time in irritable bowel syndrome: symptom instability and severity variability. *Am J Gastroenterol.* 2004;99(1):113-121.
23. Drossman DA, Morris CB, Hu Y, et al. A prospective assessment of bowel habit in irritable bowel syndrome in women: defining an alternator. *Gastroenterology.* 2005;128(3):580-589.
24. U.S. Department of Health and Human Services FaDA, Center for Drug Evaluation and Research (CDER). Guidance for Industry Irritable Bowel Syndrome - Clinical Evaluation of Drugs for Treatment. 2012.
25. (EMA) EMA. Guideline on the evaluation of medicinal products for the treatment of irritable bowel syndrome. 2014.
26. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol.* 2008;4:1-32.
27. 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.
28. Houtveen JH, Oei NY. Recall bias in reporting medically unexplained symptoms comes from semantic memory. *J Psychosomc Res.* 2007;62(3): 277-282.
29. 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.
30. Csikszentmihalyi M, Larson R. Validity and reliability of the Experience-Sampling Method. *J Nerv Ment Dis.* 1987;175(9): 526-536.
31. aan het Rot M, Hogenelst K, Schoevers RA. Mood disorders in everyday life: a systematic review of experience sampling and ecological momentary assessment studies. *Clin Psychol Rev.* 2012;32(6): 510-523.
32. Walz LC, Nauta MH, Aan Het Rot M. Experience sampling and ecological momentary assessment for studying the daily lives of patients with anxiety disorders: a systematic review. *JfAnxiety Dis.* 2014;28(8):925-937.
33. 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.
34. Mayer EA, Naliboff BD, Chang L, Coutinho SV. V. Stress and irritable bowel syndrome. *Am J Physiol Gastrointest Liver Physiology.* 2001;280(4):G519-524.
35. Levy RL, Olden KW, Naliboff BD, et al. Psychosocial aspects of the functional gastrointestinal disorders. *Gastroenterology.* 2006;130(5):1447-1458.
36. Ohman L, Simren M. New insights into the pathogenesis and pathophysiology of irritable bowel syndrome. *Dig Liver Dis.* 2007;39(3):201-215.



PART I
DIAGNOSIS AND SYMPTOMS IN IBS

Chapter 2

Rome III versus Rome IV criteria for irritable bowel syndrome: A comparison of clinical characteristics in a large cohort study

Lisa Vork, Zsa Zsa R.M. Weerts, Zlatan Mujagic, Joanna W. Kruimel,
Martine A.M. Hesselink, Jean W.M. Muris, Daniel Keszthelyi,
Daisy M.A.E. Jonkers, Ad A.M. Masclee

Neurogastroenterology & Motility. 2018 Feb;30(2).

Abstract

INTRODUCTION The Rome criteria for irritable bowel syndrome (IBS) have been revised and are expected to apply only to the subset of Rome III IBS subjects with abdominal pain as predominant symptom, occurring at least once a week. The aim of this study was to determine the percentage of Rome III IBS subjects that fulfill Rome IV criteria and to evaluate differences between Rome IV-positive and -negative subjects.

METHODS 404 Rome III IBS subjects completed a 14-day end-of-day symptom diary, the gastrointestinal symptom rating scale (GSRS), hospital anxiety and depression scale (HADS) and rand 36-item short-form health survey (SF-36). Diary-based surrogate Rome IV criteria were defined as occurrence of abdominal pain at least one day each week with a severity of ≥ 2 (*mild*; definition 1) or ≥ 3 (*considerable*; definition 2).

RESULTS Using surrogate Rome IV criteria, 353 (87.4%, definition 1) and 249 (61.6%, definition 2) subjects were defined as Rome IV-positive. These patients were more often female, younger and recruited from secondary/tertiary care compared to Rome IV-negative subjects. They also presented with higher abdominal pain scores and gastrointestinal (GI) symptom severity on both end-of-day diary and GSRS, higher psychological symptom scores and lower quality of life compared to Rome IV-negative subjects.

CONCLUSIONS The Rome IV IBS population likely reflects a subgroup of Rome III IBS patients with more severe GI symptomatology, psychological comorbidities and lower quality of life. This implies that results from Rome III IBS studies may not be directly comparable to those from Rome IV IBS populations.

Introduction

Irritable bowel syndrome (IBS) is a functional intestinal disorder characterized by abdominal pain associated with altered bowel habits. IBS has traditionally been subcategorized into four subtypes based on predominant stool pattern: diarrhea (IBS-D), constipation (IBS-C), a mix of diarrhea and constipation (IBS-M) or undefined predominant stool form (IBS-U). IBS is a prevalent disorder worldwide, with prevalence rates of 5-15% in the Western population.^{1,2} Symptoms most likely result from complex interactions between several biological, psychological and social factors.^{3,4} The exact underlying mechanisms of IBS pathophysiology are however not completely understood and as a consequence accurate non-invasive biomarkers for diagnosis, disease monitoring and treatment evaluation are not available.

At present, the diagnosis of IBS is symptom-based, using the Rome criteria. The Rome Foundation, a committee of international experts in the field of functional gastroenterology, has been working on the development and revision of diagnostic criteria for IBS, amongst other functional gastrointestinal (GI) disorders, since 1994. Recently, Rome III criteria (2006) have been updated to Rome IV criteria (2016). Major adjustments include removal of the term abdominal discomfort (Rome III), leaving only the occurrence of abdominal pain as the key requirement for Rome IV criteria. Furthermore, abdominal pain should be present on average at least one day per week in the Rome IV criteria (see Box 2.1).^{5,6} This new frequency threshold was based on a summary report on the distribution of symptom occurrence rates for all the Rome III symptoms.⁷ Most likely, fewer patients will fulfill the new Rome criteria compared to the previously set criteria. Indeed, using Rome IV criteria, a lower population prevalence of IBS has been reported by Whitehead and colleagues.^{8,9} Furthermore, IBS subtype identification has been revised, by only taking into account symptomatic stools (*i.e.* loose/watery stools and hard/lumpy stools), which might result in a shift in IBS subtypes.³

Studies on previous editions of the Rome criteria have demonstrated varying IBS prevalence rates depending on the diagnostic criteria employed. Also, differences in patient characteristics and symptomatology have been reported between several criteria.¹⁰⁻¹³ As a result of the requirement of weekly symptoms, Rome IV IBS patients are likely to be those with more severe symptomatology and possibly higher prevalence of psychiatric comorbidity and lower quality of life as compared to those fulfilling Rome III criteria. However,

data comparing clinical features between Rome III- and Rome IV-IBS populations are still lacking.

Rome criteria are widely used as a cornerstone for inclusion in IBS clinical trials and cohort studies. It is expected that in future studies only a subset of Rome III IBS patients, *i.e.* those with more severe abdominal pain, will be eligible for study participation by introducing Rome IV criteria. In order to generalize current data to Rome IV populations and to compare results from Rome III and Rome IV studies, it is important to evaluate which Rome III patients are likely to meet the Rome IV criteria. Therefore, the aim of this study was to determine the percentage of Rome III-positive IBS subjects that is also highly likely to fulfill Rome IV criteria, based on end-of-day symptom diaries, and to evaluate whether demographical, clinical and psychosocial differences exist between Rome IV-positive and Rome IV-negative subjects, in a well-defined Rome III IBS population.

Rome III IBS

- 1) Recurrent abdominal pain or discomfort* at least 3 days per month in the last 3 months and associated with 2 or more of the following:
 - a. Improvement with defecation;
 - b. Onset associated with a change in frequency of stool;
 - c. Onset associated with a change in form (appearance) of stool.
- 2) Criteria fulfilled for the last 3 months with symptom onset at least 6 months prior to diagnosis.

* Discomfort means an uncomfortable sensation not described as pain.

Rome IV IBS

- 1) Recurrent abdominal pain, on average, at least 1 day per week in the last 3 months and associated with 2 or more of the following:
 - a. Related to defecation;
 - b. Associated with a change in frequency of stool;
 - c. Associated with a change in form (appearance) of stool.
- 2) Criteria fulfilled for the last 3 months with symptom onset at least 6 months before diagnosis.

Box 2.1 Definition of IBS according to Rome III and Rome IV criteria.

Materials and methods

Study design

In the current analyses, data from a well-phenotyped Dutch cohort study, the Maastricht IBS (MIBS) Cohort, on the phenotypical and genotypical characterization of IBS patients¹⁴⁻¹⁶, were evaluated. The study protocol has been approved by the Maastricht University Medical Center + (Maastricht UMC+) Committee of Ethics in February 2009 and was executed according to the revised Declaration of Helsinki (64th WMA General Assembly, Fortaleza, Brazil, October 2013). Furthermore, the study has been registered in the US National Library of Medicine (<http://www.clinicaltrials.gov>, NCT00775060).

Study participants

Between July 2009 and May 2016, IBS patients aged 18-75 years were included in the Maastricht IBS cohort at the secondary/tertiary care outpatient department of Gastroenterology-Hepatology at the Maastricht UMC+ in Maastricht and via general practitioners practices in South-Limburg, the Netherlands. All subjects fulfilled the Rome III criteria (see Box 2.1) for IBS and were assigned the IBS subtype based on predominant bowel habit, *i.e.* diarrhea (IBS-D), constipation (IBS-C), a mix of diarrhea and constipation (IBS-M) or unspecified predominant bowel habit (IBS-U).^{3,17} Rome III criteria were evaluated in a face-to-face interview by a trained clinical researcher. Medical history was taken by a gastroenterologist and if indicated, GI endoscopy, abdominal imaging and/or blood, breath or fecal analyses were performed to exclude organic disease. A history of abdominal surgery, except for uncomplicated appendectomy, cholecystectomy or hysterectomy, was reason for exclusion. All subjects gave their written informed consent before participation.

Data collection

As subject inclusion was performed since 2009, Rome IV criteria were not collected in a face-to-face interview at the moment of inclusion. However, all participants completed an end-of-day diary on symptom severity and bowel habits, during 14 days, at the time of inclusion. Abdominal pain, amongst other symptoms, was scored using a 5-point Likert scale (1=not at all; 2=mild; 3=considerable; 4=severe; 5=extremely). Using this information, we

retrospectively determined which subjects were highly likely to fulfill the Rome IV criteria, based on the presence of abdominal pain on at least one day in both the first and the second seven days (*i.e.* abdominal pain at least once a week). Rome IV criteria do not take into account abdominal pain severity, however, as we will use surrogate Rome IV criteria to evaluate differences between Rome IV-positive and –negative IBS patients, we will report on two definitions for those criteria: ‘Definition 1: Abdominal pain score ≥ 2 once a week in each week’ and ‘Definition 2: Abdominal pain score ≥ 3 once a week in each week’. Only symptom diaries that were completed for at least 12 of the 14 days were considered eligible for analysis.

Information on demographics was collected using a predefined self-report questionnaire. Furthermore, subjects completed the gastrointestinal symptom rating scale (GSRS)¹⁸, hospital anxiety and depression scale (HADS)^{19,20} and rand 36-item short-form health survey (SF-36)^{21,22} for GI symptom severity, co-occurrence of depressive and/or anxiety symptoms and general quality of life, respectively.

In addition, in a subset of participants a rectal barostat procedure was performed. Measurement of rectal perception was performed using a standardized perception protocol, during which 17 pressure steps between 0 and 50 mm Hg, based on a semi-random staircase protocol, were applied. During each pressure step pain scores were reported on a 100 mm visual analogue scale (VAS). The cut-off value for visceral hypersensitivity was defined as a pain score ≥ 20 at pressure ≤ 26 mm Hg. A detailed description on the rectal barostat procedure was previously reported.²³

The end-of-day diary data were additionally used to perform an exploratory analysis on whether Rome III IBS patients, that do not fulfill (surrogate) Rome IV criteria for IBS, are likely to fulfill Rome IV criteria for other functional bowel disorders. Definitions that were used for retrospective evaluation of Rome IV diagnoses for functional constipation (FC), functional diarrhea (FD) and functional abdominal bloating/distension (FAB/D) are shown in **Table S2.1**. Since the symptom diary was designed for IBS and not specifically for assessing those other disorders, not all criteria could be definitively checked.

Data and statistical analyses

All analyses were performed using IBM SPSS Statistics, version 23 (IBM Statistics for Macintosh, Chicago, IL, USA).

The total study population is referred to as 'total'. Depending on whether subgroups are highly likely to fulfill Rome IV criteria based on the end-of-day diary or not they are referred to as 'Rome IV-positive' and 'Rome IV-negative', respectively. Results for 'Rome IV-positive' and 'Rome IV-negative' are presented separately for both definitions of fulfilling Rome IV criteria: 'Definition 1: Abdominal pain score ≥ 2 once a week in each week' and 'Definition 2: Abdominal pain score ≥ 3 once a week in each week'.

Categorical data are presented as proportions and differences between groups are tested using χ^2 or Fisher's exact test. Continuous data are presented as medians and interquartile ranges (IQR) and groups are compared using Mann-Whitney U test, taking into account asymmetric distribution of the data. A p-value of 0.05 was considered statistically significant.

Results

Study population

In total, 404 subjects that completed at least 12 days of the end-of-day symptom diary were included in the analyses: 293 (72.5%) were women and median age was 45 [IQR: 28-59] years. Seventy-two per cent (n=291) was recruited from secondary/tertiary care and IBS subtypes (based on Rome III criteria) were distributed as follows: 140 (34.7%) IBS-D, 81 (20%) IBS-C, 159 (39.4%) IBS-M and 24 (5.9%) IBS-U. Further characteristics of the total study population are shown in Table 2.1.

IBS according to Rome IV criteria – Definition 1: Abdominal pain score ≥ 2

Of the 404 IBS subjects diagnosed by Rome III criteria, 353 (87.4%) did meet the surrogate Rome IV criteria when assessed using the end-of-day symptom diary and the cut-off for abdominal pain severity of ≥ 2 .

Rome IV-positive subjects were more often female (74.5% vs. 58.8%, $P < 0.05$), younger (45 vs. 53 years, $P < 0.05$) and recruited from secondary/tertiary care (74.4% vs. 56.9%, $P < 0.05$) compared to Rome IV-negative subjects. Additionally, visceral hypersensitivity assessed by rectal barostat was present more often in Rome IV-positive subjects (47.5% vs. 11.8%, $P < 0.001$). Subtype distribution (*i.e.* based on Rome III criteria) was not different between both groups.

With regards to GI symptoms, Rome IV-positive versus -negative subjects reported higher scores in the end-of-day diary for all symptoms assessed (*i.e.*

abdominal pain, abdominal discomfort, abdominal bloating, flatulence, constipation and diarrhea) and higher symptom scores for all five GSRS-domains, however, this was not statistically significant for indigestion syndrome.

Furthermore, Rome IV-positive subjects showed a higher percentage of depressive (21.0% vs. 5.9%, $P < 0.001$) as well as anxiety (37.6% vs. 21.6%, $P < 0.05$) symptoms and lower physical composite scores (41.67 vs. 49.92, $P < 0.001$) with regards to quality of life. Mental composite scores of SF-36 did not show a significant difference between the groups. Results are shown in Table 2.1.

Table 2.1 Demographical and clinical characteristics of the total study population and Rome IV-positive and -negative subjects, separated for definition 1 and 2 of surrogate Rome IV criteria, i.e. assessed using end-of-day symptom diaries.

	Total n=404	Definition 1: Abdominal pain score ≥2		Definition 2: Abdominal pain score ≥3	
		Rome IV-positive n=353 (87.4%)	Rome IV-negative n=51	Rome IV-positive n=249 (61.6%)	Rome IV-negative n=55
Female sex , n (%)	293 (72.5)	263 (74.5)	30 (58.8) [#]	195 (78.3)	98 (63.2) [*]
Age , median [IQR]	45 [28 - 59]	45 [27 - 58]	53 [32 - 65] [#]	42 [26 - 57]	51 [36 - 63] [*]
BMI , median [IQR]	24.03 [21.41 - 27.63]	23.94 [21.38 - 27.69]	24.62 [21.58 - 27.52]	24.08 [21.40 - 27.47]	23.96 [21.45 - 27.98]
Subtype Rome III , n (%)					
IBS-D	140 (34.7)	118 (33.4)	22 (43.1)	85 (34.1)	55 (35.5)
IBS-C	81 (20.0)	73 (20.7)	8 (15.7)	51 (20.5)	30 (19.4)
IBS-M	159 (39.4)	142 (40.2)	17 (33.3)	101 (40.6)	58 (37.4)
IBS-U	24 (5.9)	20 (5.7)	4 (7.8)	12 (4.8)	12 (7.7)
Secondary/tertiary care , n (%)	291 (72.0)	262 (74.4)	29 (56.9) [#]	194 (78.2)	97 (62.6) [*]
Hypersensitive , n (%)	87 (44.4) ^a	85 (47.5) ^b	2 (11.8) [*]	72 (57.1) ^d	15 (21.4) ^e *
14-day mean score , median [IQR]					
Abdominal pain	2.21 [1.57 - 2.92]	2.36 [1.79 - 3.00]	1.07 [1.00 - 1.21] [*]	2.71 [2.29 - 3.29]	1.43 [1.17 - 1.77] [*]
Abdominal discomfort	2.43 [1.93 - 2.93]	2.50 [2.14 - 3.00]	1.50 [1.07 - 1.86] [*]	2.72 [2.36 - 3.21]	1.79 [1.43 - 2.21] [*]
Abdominal bloating	2.07 [1.50 - 2.79]	2.21 [1.64 - 2.86]	1.43 [1.07 - 2.00] [*]	2.43 [1.93 - 3.18]	1.71 [1.14 - 2.07] [*]
Flatulence	2.21 [1.71 - 2.86]	2.29 [1.79 - 2.93]	1.71 [1.29 - 2.33] [*]	2.43 [1.93 - 3.07]	1.93 [1.36 - 2.36] [*]
Constipation	1.29 [1.00 - 1.86]	1.36 [1.00 - 1.93]	1.00 [1.00 - 1.29] [*]	1.50 [1.07 - 2.07]	1.08 [1.00 - 1.43] [*]
Diarrhea	1.21 [1.00 - 1.64]	1.29 [1.07 - 1.71]	1.07 [1.00 - 1.29] [*]	1.31 [1.07 - 1.98]	1.08 [1.00 - 1.43] [*]
GSRs , median [IQR]					
Abdominal pain	3.33 [2.33 - 4.00]	3.33 [2.33 - 4.25]	2.67 [1.67 - 3.33] [*]	3.67 [2.67 - 4.33]	2.67 [2.00 - 3.67] [*]
Diarrhea syndrome	3.33 [2.00 - 4.33]	3.33 [2.33 - 4.67]	2.67 [1.67 - 4.00] [#]	3.67 [2.33 - 4.67]	3.00 [2.00 - 4.00] [*]
Constipation syndrome	3.00 [2.33 - 4.00]	3.00 [2.33 - 4.33]	2.00 [1.33 - 3.33] [*]	3.33 [2.33 - 4.33]	2.67 [1.67 - 3.67] [*]
Indigestion syndrome	4.00 [3.25 - 5.00]	4.00 [3.25 - 5.00]	3.75 [2.75 - 4.50]	4.25 [3.25 - 5.25]	3.75 [2.75 - 4.50] [*]
Regurgitation syndrome	1.50 [1.00 - 3.00]	1.50 [1.00 - 3.00]	1.00 [1.00 - 2.00] [#]	1.75 [1.00 - 3.50]	1.00 [1.00 - 2.50] [*]
HADS , n (%)					
Depression ≥8	76 (19.0)	73 (21.0)	3 (5.9) [*]	59 (24.1)	17 (11.0) [*]
Anxiety ≥8	142 (35.6)	131 (37.6)	11 (21.6) [#]	105 (42.9)	37 (24.0) [*]
SF-36 , median [IQR]					
Physical Composite Score (PCS)	42.83 [33.98 - 50.28]	41.67 [32.98 - 49.11]	49.92 [41.42 - 53.89] [*]	39.23 [29.84 - 46.95]	48.99 [39.41 - 52.62] [*]
Mental Composite Score (MCS)	49.75 [41.23 - 55.82]	49.41 [41.15 - 55.37]	51.03 [41.98 - 57.73]	48.31 [40.76 - 54.58]	51.34 [42.16 - 57.19] [#]

Differences tested using Mann-Whitney U test for continuous data and χ^2 or Fisher's exact test for categorical data. Significances are shown for both definitions of surrogate Rome IV criteria separately. [#] Significant difference between Rome IV-positive and Rome IV-negative subjects (p<0.05). ^{*} Significant difference between Rome IV-positive and Rome IV-negative subjects (<0.001). ^{a-e} Number of subjects that underwent rectal barostat. ^a196; ^b179; ^c17; ^d126; ^e70.



IBS according to Rome IV criteria – Definition 2: Abdominal pain score ≥ 3

When using the surrogate Rome IV criteria as defined by at least one day of abdominal pain each week using the cut-off for abdominal pain severity of ≥ 3 , of the 404 IBS subjects diagnosed by Rome III criteria, 249 (61.6%) were Rome IV-positive.

In line with the findings above, *i.e.* regarding the cut-off of abdominal pain severity of ≥ 2 , Rome IV-positive subjects were more often female, younger, recruited from secondary/tertiary care, more often hypersensitive on rectal barostat, showed higher symptom severity scores for all symptoms assessed in the end-of-day symptom diary and showed higher percentages of depressive as well as anxiety symptoms compared to Rome IV-negative subjects, when using the cut-off of ≥ 3 . Likewise, subtype distribution (*i.e.* based on Rome III criteria) did not differ between the groups.

Moreover, with regard to symptom severity on GSRS, Rome IV-positive versus – negative subjects scored significantly higher on all domains including indigestion syndrome when using the cut-off of ≥ 3 . Furthermore, Rome IV-positive subjects scored significantly lower on both physical composite score and mental composite score of SF-36 compared to Rome IV-negative subjects. Results are shown in Table 2.1.

Alternative Rome IV diagnoses in Rome IV-negative (IBS) subjects

Of the Rome IV-negative (IBS) subjects, according to the surrogate Rome IV criteria using the cut-off for abdominal pain severity of ≥ 3 , 34 (23.61%) fulfilled surrogate Rome IV-criteria for functional constipation, 50 (34.25%) for functional diarrhea and 37 (25.69%) for functional abdominal bloating/distension. Results are shown in Table S2.2.

Discussion

This study demonstrates that 61.6% to 87.4% of Rome III IBS patients is likely to also fulfill the new Rome IV criteria for IBS, depending on the cut-off for abdominal pain severity used, when applying surrogate Rome IV criteria based on end-of-day symptom diaries. Regardless of the cut-off chosen, Rome IV-positive subjects were more often female, younger and recruited from secondary/tertiary care than Rome IV-negative subjects. Not only did they present with higher abdominal pain scores, but overall symptom severity was

higher in Rome IV-positive subjects, including a higher percentage of visceral hypersensitivity as assessed by rectal barostat. Additionally, higher percentages of comorbid psychological symptoms and lower quality of life were found for Rome IV-positive subjects. Taken together, the current findings imply that the Rome IV IBS population will likely reflect a subgroup of Rome III IBS patients with more severe overall gastrointestinal symptomatology, psychological comorbidities and lower quality of life.

To date, a recent study by Bai *et al.* investigated the agreement between Rome III and Rome IV criteria for IBS in a GI outpatient population in China. They found a moderate consistency between Rome III and Rome IV criteria, with prevalences of 12.4% and 6.1% based on Rome III and Rome IV criteria, respectively.²⁴ Similarly, Whitehead and colleagues reported prevalences of 10.7% and 5.7%, respectively, in a large population-based study.^{8,9} Our current study confirms this decrease in prevalence by introducing Rome IV criteria, within a well-defined Rome III IBS population. Additionally, in line with these previously reported findings, our cut-off for abdominal pain severity of ≥ 3 may be most robust to define those subjects highly likely to fulfill Rome IV criteria, based on end-of-day diary.

We observed that Rome IV-positive subjects were more often of female gender than Rome IV-negative subjects. These differences were not demonstrated by Bai and colleagues.²⁴ Differences in symptom severity between men and women, however, have been reported previously, indicating possible differences in pathophysiological mechanisms as well as in pain perception and coping strategies. A meta-analysis by Adeyemo and colleagues demonstrated that women are more likely to report abdominal pain than men.²⁵ Furthermore, higher overall IBS symptom severity in female IBS patients has been reported by Bjorkman *et al.*²⁶ These previous findings could explain the higher percentage of women in our Rome IV-positive IBS population compared to the Rome IV-negative subjects.

With regards to clinical differences between Rome III and Rome IV IBS subjects, Bai *et al.* reported higher abdominal pain scores for Rome IV, but no differences in abdominal discomfort, abdominal bloating or demographic characteristics. In contrast, our study shows significantly higher scores for abdominal discomfort, bloating, flatulence, diarrhea and constipation, apart from abdominal pain. A possible explanation for the discrepancy between these studies is the method

of data collection. Bai and colleagues retrospectively assessed the presence of above-mentioned symptoms at one time point, whereas in our study symptom severity scores were determined daily, using an end-of-day diary during 14 days.

As a study by Engsbro *et al.* demonstrated that IBS subtype classification differs depending on whether retrospective (*i.e.* Rome Diagnostic Questionnaire) or prospective methods (*i.e.* diary cards) are used²⁷, we did not use the symptom diary to assess the IBS subtypes. However, we feel that the end-of-day diary provides an objective overview of present symptoms during a 14-day period and therefore can be used to identify patients highly likely to fulfill the Rome IV criteria. Furthermore, Rome IV criteria require abdominal pain to be present at least once a week on average during the last three months. The end-of-day symptom diary as used only provides us with information about the past two weeks. However, all subjects fulfilled Rome III criteria assessed using the Rome III Diagnostic Questionnaire, which also requires the abdominal complaints to be present during the previous three months. Therefore, we think that the surrogate Rome IV criteria are a reliable reflection of the Rome IV Diagnostic Questionnaire.

This is the first study focusing on clinical differences between Rome III and Rome IV IBS subjects. In conclusion, the current study underlines a decrease of IBS prevalence when using (surrogate) Rome IV compared to Rome III criteria. In addition to these findings, the question arises how to deal with the Rome IV-negative patient population presenting with IBS-symptoms in both primary and secondary/tertiary care and with regard to future clinical and mechanistic studies.

First, it might be interesting to explore whether alternative Rome IV disorders can now be diagnosed in these subjects, for example functional constipation, functional diarrhea or functional abdominal bloating/distension. An exploratory analysis in the current study demonstrates that 24% is likely to fulfill Rome IV-criteria for FC, 34% for FD and 26% for FAB/D. These results should, however, be interpreted with caution, since these diagnoses are based on surrogate criteria, using an IBS-specific end-of-day symptom diary. Nevertheless, this suggests that an additional $\pm 16\%$ of Rome III IBS patients will not fulfill any of these Rome IV diagnoses. Possibly, a small subset might be defined as Rome IV unspecified functional bowel disorder. However, it is very likely that patients reporting abdominal pain or bloating/distension less than once weekly (without

predominant constipation or diarrhea), will not fulfill any Rome IV diagnosis due to the new frequency threshold.

Second, as the Rome criteria are not that universally used in routine clinical practice, in our opinion, patients with milder functional GI symptoms seeking health care, but not fulfilling one of the Rome IV diagnoses, should still be managed as functional GI disorders. With regard to future IBS research, patient inclusion should be based on the Rome IV criteria, in order to aim at agreement between studies.

Nevertheless, this study demonstrates that the Rome IV IBS population is represented by younger females with higher overall gastrointestinal symptom severity, including comorbid psychological symptoms and lower quality of life compared to the Rome III IBS population. Therefore, results from Rome III IBS studies may not be directly comparable to those from Rome IV IBS populations, which has implications for future IBS research in particular.

References

1. Lovell RM, Ford AC. Global prevalence of and risk factors for irritable bowel syndrome: a meta-analysis. *Clin Gastroenterol Hepatol*. 2012;10(7):712-721.e714.
2. Hungin AP, Whorwell PJ, Tack J, Mearin F. The prevalence, patterns and impact of irritable bowel syndrome: an international survey of 40,000 subjects. *Alimentary pharmacology & therapeutics*. 2003;17(5):643-650.
3. Drossman DA. The functional gastrointestinal disorders and the Rome III process. *Gastroenterology*. 2006;130(5):1377-1390.
4. Enck P, Aziz Q, Barbara G, et al. Irritable bowel syndrome. *Nat Rev Dis Primers*. 2016;2:16014.
5. Foundation R. Appendix A: Rome III Diagnostic Criteria for FGIDs.885-897.
6. Mearin F, Lacy BE, Chang L, et al. Bowel Disorders. *Gastroenterology*. 2016.
7. Whitehead W, Palsson O. Report on the Rome III Normative Gastrointestinal Symptom Survey. Chapel Hill, NC: University of North Carolina;2013.
8. Palsson O, van Tilburg MA, Simren M, Sperber A, Whitehead W. Population prevalence of Rome III and Rome IV Irritable Bowel Syndrome (IBS) in the United States (US), Canada and the United Kingdom (UK). *UEG Week*; 2016; Vienna.
9. Whitehead WE, Palsson OS, Simren M. Irritable bowel syndrome: what do the new Rome IV diagnostic guidelines mean for patient management? *Expert Rev Gastroenterol Hepatol*. 2017;11(4):281-283.
10. Boyce PM, Koloski NA, Talley NJ. Irritable bowel syndrome according to varying diagnostic criteria: are the new Rome II criteria unnecessarily restrictive for research and practice? *Am J Gastroenterol*. 2000;95(11):3176-3183.
11. Sperber AD, Shvartzman P, Friger M, Fich A. A comparative reappraisal of the Rome II and Rome III diagnostic criteria: are we getting closer to the 'true' prevalence of irritable bowel syndrome? *Eur J Gastroenterol Hepatol*. 2007;19(6):441-447.
12. Park DW, Lee OY, Shim SG, et al. The Differences in Prevalence and Sociodemographic Characteristics of Irritable Bowel Syndrome According to Rome II and Rome III. *J Neurogastroenterol Motil*. 2010;16(2):186-193.
13. Wang AJ, Liao XH, Hu PJ, Liu SC, Xiong LS, Chen MH. [A comparison between Rome III and Rome II criteria in diagnosing irritable bowel syndrome]. *Zhonghua nei ke za zhi*. 2007;46(8):644-647.
14. Ludidi S, Mujagic Z, Jonkers D, et al. Markers for visceral hypersensitivity in patients with irritable bowel syndrome. *Neurogastroenterol Motil*. 2014;26(8):1104-1111.
15. Mujagic Z, Ludidi S, Keszthelyi D, et al. Small intestinal permeability is increased in diarrhoea predominant IBS, while alterations in gastroduodenal permeability in all IBS subtypes are largely attributable to confounders. *Aliment Pharmacol Ther*. 2014;40(3):288-297.
16. Mujagic Z, Tigchelaar EF, Zhernakova A, et al. A novel biomarker panel for irritable bowel syndrome and the application in the general population. *Sci Rep*. 2016;6:26420.
17. Longstreth GF, Thompson WG, Chey WD, Houghton LA, Mearin F, Spiller RC. Functional bowel disorders. *Gastroenterology*. 2006;130(5):1480-1491.
18. Svedlund J, Sjodin I, Dotevall G. GSRS--a clinical rating scale for gastrointestinal symptoms in patients with irritable bowel syndrome and peptic ulcer disease. *Dig Dis Sci*. 1988;33(2):129-134.
19. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand*. 1983;67(6):361-370.

20. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *J Psychosom Res.* 2002;52(2):69-77.
21. Fariyar SS, Cunningham WE, Hays RD. Correlated physical and mental health summary scores for the SF-36 and SF-12 Health Survey, V.I. *Health Qual Life Outcomes.* 2007;5:54.
22. McHorney CA, Ware JE, Jr., Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care.* 1993;31(3):247-263.
23. Ludidi S, Conchillo JM, Keszthelyi D, et al. Rectal hypersensitivity as hallmark for irritable bowel syndrome: defining the optimal cutoff. *Neurogastroenterol Motil.* 2012;24(8):729-733, e345-726.
24. Bai T, Xia J, Jiang Y, et al. Comparison of the Rome IV and Rome III criteria for IBS diagnosis: A cross-sectional survey. *J Gastroenterol Hepatol.* 2017;32(5):1018-1025.
25. Adeyemo MA, Spiegel BM, Chang L. Meta-analysis: do irritable bowel syndrome symptoms vary between men and women? *Aliment Pharmacol Ther.* 2010;32(6):738-755.
26. Bjorkman I, Jakobsson Ung E, Ringstrom G, Tornblom H, Simren M. More similarities than differences between men and women with irritable bowel syndrome. *Neurogastroenterol Motil.* 2015;27(6): 796-804.
27. Engsbro AL, Simren M, Bytzer P. The Rome II and Rome III criteria identify the same subtype-populations in irritable bowel syndrome: agreement depends on the method used for symptom report. *Neurogastroenterol Motil.* 2012;24(7):604-611, e266.

Supplemental material

Table S2.1 Definition of Rome IV criteria⁶ and surrogate Rome IV criteria, based on 14-day end-of-day symptom diary, separately for functional constipation, functional diarrhea and functional abdominal bloating/distension.

Rome IV criteria	Surrogate Rome IV criteria
FUNCTIONAL CONSTIPATION	
1. Must fulfill two or more of the following: <ol style="list-style-type: none"> Straining with defecation (at least 25% of stools) Lumpy or hard stools (BSFS 1-2) (at least 25% of stools) Sensation of incomplete evacuation (at least 25% of stools) Sensation of anorectal obstruction/blockage (at least 25% of stools) Manual maneuvers to facilitate defecation (at least 25% of stools) Fewer than three spontaneous stools per week (BSFS 6-7) 	1. <ol style="list-style-type: none"> Not verifiable from diary data Checked in diary data: Percentage number of BM's with BSFS 1-2 from total number of BM's over 14 days; fulfilled if $\geq 25\%$ Not verifiable from diary data Not verifiable from diary data Not verifiable from diary data Checked in diary data: Number of BM's over 14 days; fulfilled if ≤ 6
2. Loose stools are rarely present (BSFS 6-7)	2. Checked in diary data: Percentage number of BM's with BSFS 6-7 from total number of BM's over 14 days; fulfilled if $\leq 20\%$
3. Insufficient criteria for IBS or OIC	3. Assumed true: Additional analysis only considers subjects that fulfill Rome III for IBS (no OIC) and do not fulfill Rome IV for IBS
4. Symptom onset at least 6 months prior to diagnosis	4. Assumed true: Subjects fulfilled Rome III for IBS

Table S2.1 (continued)

Rome IV criteria	Surrogate Rome IV criteria
FUNCTIONAL DIARRHEA	
<ol style="list-style-type: none"> 1. Loose or watery stools (BSFS 6-7) (at least 25% of stools) 2. Without predominant abdominal pain or bloating 3. Symptom onset at least 6 months prior to diagnosis 4. Insufficient criteria for IBS-D 	<ol style="list-style-type: none"> 1. Checked in diary data: Percentage number of BM's with BSFS 6-7 from total number of BM's over 14 days; fulfilled if $\geq 25\%$ 2. Not verifiable from diary data 3. Assumed true: Subjects fulfilled Rome III for IBS 4. Assumed true: Additional analysis only considers subjects that do not fulfill Rome IV for IBS
FUNCTIONAL ABDOMINAL BLOATING/DISTENSION	
<ol style="list-style-type: none"> 1. Recurrent bloating and/or distension (at least weekly) 2. Abdominal bloating/distension predominates over other symptoms 3. Insufficient criteria for IBS, FC, FD or PDS 4. Symptom onset at least 6 months prior to diagnosis 	<ol style="list-style-type: none"> 1. Checked in diary data: Symptom severity score for bloating ≥ 2 at least one day in both the first and the second seven days. 2. Not verifiable from diary data 3. Checked in diary data: Additional analysis only considers subjects that do not fulfill Rome IV for IBS. Subjects fulfilling surrogate criteria for FC or FD were excluded from FAB/D analysis. Diagnosis PDS was not verifiable from diary data. 4. Assumed true: Subjects fulfilled Rome III for IBS

BSFS: Bristol Stool Form Scale; FC: Functional Constipation; BM: Bowel movement; FD: Functional Diarrhea; OIC: Opioid induced constipation; FAB/D: Functional Abdominal Bloating/Distension; PDS: Post-prandial distress syndrome.

Table S2.2 Distribution of alternative Rome IV diagnoses, based on surrogate Rome IV criteria, within Rome IV-negative (IBS) population, *i.e.* assessed using end-of-day symptom diaries*.

		Rome IV-negative (n=155)
Functional constipation, n (%)	(n=373)	34 (23.61) ^a
Functional diarrhea, n (%)	(n=376)	50 (34.25) ^b
Functional bloating/distension, n (%)	(n=370)	37 (25.69) ^a
No alternative Rome IV diagnosis, n (%)	(n=370)	23 (15.97) ^a

^{a-d} Number of subjects included in analyses: ^a144; ^b146. * Rome IV IBS diagnosis according to 'Definition 2: Abdominal pain score ≥ 3 once a week in each week'.

Chapter 3

Reduction in IBS symptom severity is not paralleled by improvement in quality of life in patients with irritable bowel syndrome

Lisa Vork*, Zsa Zsa R.M. Weerts*, Zlatan Mujagic, Daniel Keszthelyi, Martine A.M. Hesselink, Joanna W. Kruimel, Carsten Leue, Jean W.M. Muris, Daisy M.A.E. Jonkers, Ad A.M. Masclee

* Both authors contributed equally to the manuscript

Neurogastroenterology & Motility. 2019;31(8):1-10.

Abstract

INTRODUCTION Irritable Bowel Syndrome (IBS) is a brain-gut-disorder, of which the natural course varies between patients and is difficult to predict. This study aimed to evaluate symptom evolution over a five-year follow-up period and to identify baseline predictors for symptom severity and quality of life (QoL) at follow-up.

METHODS Maastricht IBS Cohort participants completed questionnaires upon inclusion regarding demographics and lifestyle, gastrointestinal (GI) symptoms, anxiety and depression, and QoL. The same questionnaires, in addition to others, were completed after five years. Rome criteria were confirmed face-to-face at initial enrollment and through telephonic interviews at follow-up.

RESULTS At a mean follow-up of 4.7 years, 379 patients were approached of whom 203 (53.7%) responded. Of these, 161 were reached by telephone and analyzed; 49 (30.4%) did not fulfill the Rome III criteria at follow-up and had lower levels of GI symptoms and GI-specific anxiety compared to those remaining Rome III-positive ($P<0.001$). However, Rome III-negative patients had comparable levels of QoL and life satisfaction, comorbid anxiety and depression, work absenteeism and impaired productivity. No baseline predictors were found for being Rome III-positive or -negative. However, greater age and lower baseline physical QoL predicted lower physical QoL at follow-up ($P<0.005$ and $P<0.01$, respectively), while lower baseline mental QoL predicted lower mental QoL at follow-up ($P=0.005$). Additionally, higher anxiety and depression scores at follow-up were associated with lower QoL and life satisfaction at follow-up ($P<0.001$).

CONCLUSIONS Long-term QoL and general well-being might depend on concurrent psychological symptoms, rather than GI symptom improvement.

Introduction

Irritable Bowel Syndrome (IBS) is a brain-gut-disorder characterized by a chronic relapsing-remitting nature of symptoms, including abdominal pain and altered bowel habits. Global prevalence, based predominantly on the Rome III criteria, is estimated at 5-20%¹, with varying rates according to geographical area and diagnostic criteria used.² Recent studies using the more restrictive Rome IV criteria point to lower prevalence rates of 5-6%.^{3,4}

Although the exact pathophysiology of IBS remains incompletely understood, a multifactorial origin is generally recognized, in which dysregulation of the brain-gut-axis has a central role. Other factors include aberrant neuroimmune interactions, visceral hypersensitivity, genetic susceptibility, microbiome alterations, and psychosocial factors.^{5,6} As interference of IBS with patients' everyday lives is extensive⁷ and treatment results are often unsatisfactory, quality of life is low and comparable to chronic somatic diseases.⁸

Given the heterogeneous nature of the disorder, symptom patterns vary widely both between and within IBS patients and predicting individual disease courses remains challenging. IBS is known as a chronic, in many patients lifelong, condition with fluctuating gastrointestinal (GI) symptoms. Symptoms such as abdominal pain, constipation, and diarrhea are known to occur in episodes of several days followed by days without symptoms.⁹ In addition, transitions from one predominant bowel habit type to the other are common and occur in up to 75% of patients.¹⁰⁻¹² Quality of life, on the other hand, has been shown to be relatively stable over a three-month period.¹³ With regard to long-term symptom variability, several studies have investigated the disease course of IBS and have shown varying results with respect to symptom severity and quality of life. The majority of the prospective studies had a follow-up period of approximately one year, which is relatively short for a condition such as IBS. In addition, there is a lack of follow-up data for the Dutch population. As implications of IBS on quality of life have shown to vary considerably between different countries¹⁴, it is of added value to expand earlier findings and investigate the natural history of IBS in a Dutch, well-characterized population. Finally, as IBS is a symptom-based diagnosis, it is of interest to assess how symptoms evolve over time and how these symptoms and long-term quality of life relate to the Rome diagnostic criteria for IBS, in particular when they are confirmed in a telephonic interview rather than a purely survey-based assessment at follow-up measurements.

Hence, the current follow-up study, which is part of a large prospective cohort study on the pheno- and genotypical characterization of IBS in the Netherlands, aimed 1) to gain further insight into symptom evolution and long-term quality of life over a five-year follow-up period; 2) to identify baseline predictors for higher symptom severity and greater quality of life impairment at follow-up; and 3) to assess how point 1 and 2 differ between patients that remain Rome III-positive and those who are Rome-III negative at follow-up.

Materials and methods

Irritable bowel syndrome cohort study

Since 2009, patients with IBS between 18 and 75 years of age who visit the outpatient clinic of the Gastroenterology-Hepatology division of the Maastricht University Medical Center+ (MUMC+), The Netherlands, are requested to enroll in the Maastricht IBS (MIBS) Cohort study. The MUMC+ is a university hospital with a combined secondary and tertiary care service in the area of South Limburg, The Netherlands. In addition, patients with IBS are recruited via general practitioners in the area of South Limburg. The MIBS Cohort is an extensively phenotyped cohort of IBS patients with a follow-up measurement five years after initial inclusion.¹⁵⁻¹⁷ The research protocol had been approved by the MUMC+ Committee of Ethics and all study procedures were performed in compliance with Good Clinical Practice Guidelines and according to the revised Declaration of Helsinki.¹⁸ The study had been registered in the US National Library of Medicine (*NCT00775060*). All subjects gave a written informed consent prior to participation.

Subjects

Patients who had been included at least three years (five +/- two years) before the current study, *that is*, patients included in the MIBS cohort between September 2009 and September 2014, were eligible for participation in the current follow-up study. All patients fulfilled the Rome III criteria for IBS at the time of inclusion, which was confirmed by a trained clinical investigator in a face-to-face interview. Patients were assigned IBS subtypes based on their reported predominant stool-type, *i.e.* diarrhea (IBS-D), constipation (IBS-C), a mix of diarrhea and constipation (IBS-M), or unspecified predominant bowel habit (IBS-U). Their medical history was taken and, if any alarm symptoms were

present or if deemed necessary by the gastroenterologist, GI endoscopy, abdominal imaging, and/or blood, breath and/or fecal analyses were performed to exclude organic diseases. A history of abdominal surgery automatically led to exclusion, except for an uncomplicated appendectomy, hysterectomy, or cholecystectomy.

Data collection

At time of enrollment in the MIBS Cohort, patients completed several questionnaires, *i.e.* on demographics and lifestyle, GI symptoms, symptoms of comorbid anxiety and depression, and general quality of life. The data were obtained administering a predefined self-report questionnaire on demographics and lifestyle, an end-of-day 14 days symptom diary, the gastrointestinal symptom rating scale (GSRS, scale 1-7, generates symptom scores for the following subdomains: abdominal pain, reflux, diarrhea, constipation, indigestion)¹⁹, the hospital anxiety and depression scale (HADS, scale 0-3, a screening tool for anxiety and depression)^{20,21}, and the rand 36-item short-form health survey (SF-36, scale 1-6, generates a physical and a mental quality of life component summary).^{22,23}

Approximately five years (+/- two years) after inclusion, patients were invited to complete several follow-up questionnaires. These included the same questionnaires that were filled out at baseline (with the exception of the end-of-day diary), with the addition of other questionnaires to further standardize the phenotyping of our population and to allow valid comparisons to other cohorts (*i.e.* international harmonization). Additional questionnaires included: the GSRS specific for IBS (GSRS-IBS, scale 1-7, generates symptom scores for the following subdomains: abdominal pain, bloating, constipation, diarrhea, satiety)²⁴; the Visceral Sensitivity Index (VSI, scale 1-6, assesses GI-specific anxiety or, in other words, fear for GI symptoms)²⁵; and the Satisfaction With Life Scale (SWLS, scale 1-6, assesses individual overall satisfaction with life).²⁶⁻²⁸

As it is known that IBS can affect the ability to work²⁹, a self-report questionnaire regarding productivity (Productivity Cost Questionnaire (PCQ), validated for the Dutch situation) was included as well.^{30,31} Patients could opt for paper or digital web-based (invitation was sent via email) questionnaires to encourage participation. One reminder was sent when no response had been received within one month. A trained clinical investigator contacted all patients who did respond and confirmed the Rome III diagnostic criteria during a telephonic interview with them. Patients were reassigned IBS subtypes based

on their predominant bowel habit, during the follow-up measurement. Reliable information on treatment history was not available for the majority of patients as this was not registered systematically during the follow-up period. Furthermore, many patients underwent self-treatment or treatment via their primary care physician, of which data were not available and were therefore not included in the analysis of the current study.

Statistical analysis

All statistical analyses were carried out using IBM SPSS statistics 23.0 (Chicago IL, USA) and GraphPad Prism 6.0 (La Jolla, CA, USA) for Macintosh. Continuous data are presented as medians and interquartile ranges (IQR) and categorical data as proportions (%). To compare (non-parametric) continuous data between subgroups, Mann-Whitney U tests were used. Wilcoxon-signed-rank tests were used to evaluate differences within subjects over time. For categorical data, groups were compared using the Chi-square or Fisher's exact test and differences over time were evaluated using the McNemar's test. Correlations were assessed according to Spearman. To decrease the false discovery error rate induced by multiple testing, a post-hoc correction was applied using the Benjamini-Hochberg step-up procedure.³²

The patients participating in the follow-up were divided into two groups for comparison and further analyses, 1) Rome III-positive: patients who still fulfilled the Rome III criteria for IBS at follow-up; 2) Rome III-negative: patients who did not fulfill the Rome III criteria for IBS at follow-up. Multivariable logistic and linear regression models, respectively, were used to identify independent baseline predictors for 1) the presence of IBS according to the Rome III diagnostic criteria at follow-up; 2) quality of life at follow-up. We also used linear regression models to identify characteristics at follow-up that were associated with quality of life at follow-up. The β of the linear regression analyses signifies one-point change in the mental or physical quality of life component summary.

Results

Study population for follow-up of the Rome III irritable bowel syndrome cohort

At a mean follow-up time of 4.7 (SD 1.5) years, 379 patients with IBS, Rome III-positive at inclusion and participating at least three (five +/- two) years in the Maastricht IBS cohort, were approached for participation in the follow-up measurement. A total of 203 subjects (53.7%) responded by completing the electronic or paper questionnaires. Of these, 161 could be reached by telephone to confirm the Rome criteria, and these did not differ in demographics nor in symptom scores when compared to the 42 subjects that were not reached. A study flowchart is shown in Figure 3.1.

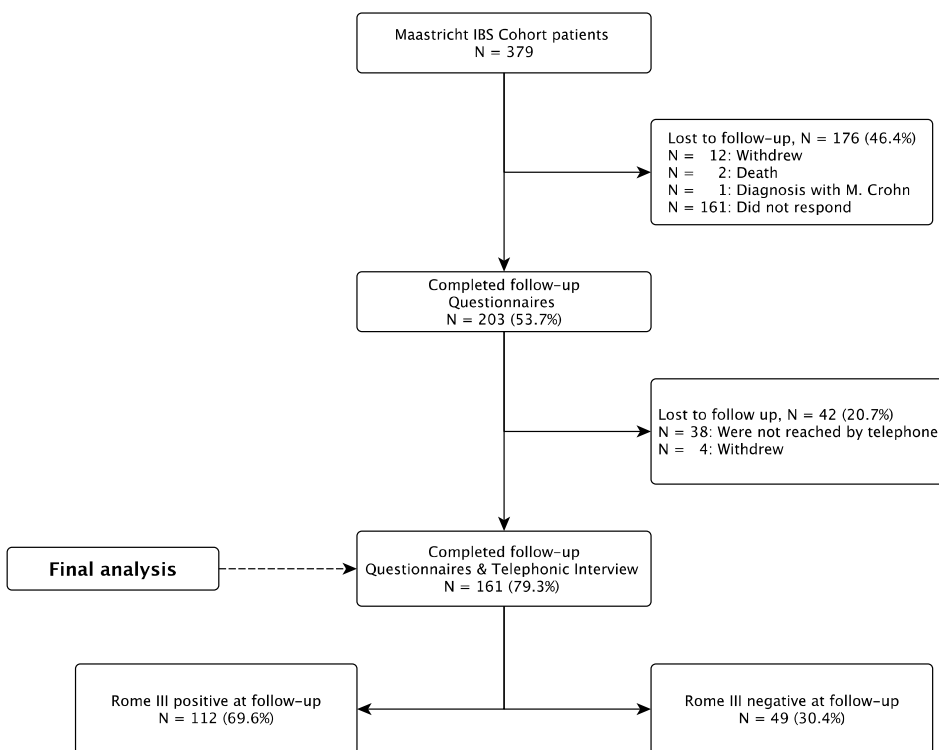


Figure 3.1 Flowchart of follow-up of n=379 IBS patients included in the Maastricht IBS Cohort study. N; number, IBS; Irritable Bowel Syndrome.

Follow-up – Demographics, clinical characteristics, and symptoms of the Rome III-positive and –negative patients

The 161 patients in whom IBS criteria could be checked by phone were included in the final analysis. The majority of patients was female (74.5%) and median age was 53 years old (IQR=36-65).

In total, 49 out of the 161 (30.4%) did no longer fulfill the Rome III criteria for IBS at follow-up (Figure 3.1). Demographical and clinical characteristics at follow-up of the Rome III-negative and Rome III-positive patients are shown in Table 3.1. The groups did not differ significantly in age or gender, although the Rome III-positive group was slightly younger. As expected, Rome III-negative patients had significantly lower GI symptom scores compared to Rome III-positive patients (assessed by both the GSRS and GSRS-IBS) at follow-up, which was shown on each of the following subdomains: abdominal pain, bloating, diarrhea, constipation, indigestion (all $P<0.001$), and satiety ($P<0.005$). Despite having less GI symptoms, they did not have significantly higher quality of life or more satisfaction with life. Although Rome III-negative patients had significantly less GI-specific anxiety at follow-up ($P<0.001$), they did not have less symptoms of general anxiety and depression. Moreover, levels of impaired work productivity and absence from work did not differ between both groups.

Baseline - Demographics, clinical characteristics, and symptoms of the Rome III-positive and –negative patients

Table 3.2 shows baseline demographics and clinical characteristics (at inclusion time into the cohort) of patients who were Rome III-positive and –negative at follow-up.

With regard to baseline levels of GI symptom severity (assessed by the GSRS, as baseline GSRS-IBS data were not available), it can be noted that Rome III-negative patients had slightly lower baseline scores for abdominal pain. Nevertheless, after correction for multiple comparisons, this did not reach statistical significance nor did other GI symptoms differ significantly between the groups at baseline. In terms of general anxiety, depression, and general quality of life scores, no baseline differences were observed between Rome III-positive and -negative patients (Table 3.2). Moreover, the proportion of patients that had been recruited for the cohort from the primary care setting did not differ between Rome III-positive and -negative patients (Table 3.2). In addition, the distribution in baseline IBS-subtypes did not differ between Rome III-positive and -negative patients (Table 3.2).

Table 3.1 Demographical and clinical characteristics, at a five-year follow-up period, of patients who still fulfilled the Rome III criteria for IBS at follow-up (FU Rome III-positive) and those who did not (FU Rome III-negative).

Characteristics at follow-up	FU Rome III-positive n=112	FU Rome III-negative n=49
Age , median years (IQR)	52 (34-65)	57 (45-67)
Female gender , n (%)	84 (75.00)	36 (73.47)
BMI , median kg m ⁻² (IQR)	25.15 (21.88-28.70)	25.35 (21.62-28.37)
GSRs , median (IQR)		
Abdominal pain	3.33 (2.67-4.33)	2.33 (1.17-3.00)***
Reflux syndrome	1.50 (1.00-3.00)	1.50 (1.00-3.00)
Diarrhea syndrome	3.33 (2.00-4.67)	1.67 (1.00-3.00)***
Constipation syndrome	3.00 (2.00-4.00)	2.00 (1.33-3.00)***
Indigestion syndrome	3.75 (2.00-4.67)	3.00 (2.25-4.50)***
GSRs-IBS , median (IQR)		
Abdominal pain	4.00 (3.5-5.00)	2.50 (1.50-3.50)***
Bloating	4.33 (3.33-5.00)	3.00 (2.00-4.33)***
Constipation	2.50 (1.00-4.00)	2.00 (1.00-2.50)
Diarrhoea	3.25 (2.25-4.50)	2.00 (1.50-3.00)**
Satiety	3.00 (2.00-4.50)	2.50 (1.25-3.50)**
Comorbid-FD , n (%)	61 (55.00)	20 (41.67)
Rome IV-positive IBS , n (%)	69 (61.61)	-
HADS, score > 8 , n (%)		
Depressive symptoms	26 (23.42)	8 (16.33)
General anxiety symptoms	33 (29.72)	8 (16.33)
GI-specific anxiety , median (IQR)	24.00 (13.00-37.00)	8.00 (4.00-23.00)***
Quality of life , median (IQR)		
PCS	45.10 (33.01-48.50)	48.96 (37.71-52.99)
MCS	51.27 (37.81-55.96)	54.75 (44.61-57.71)
Satisfaction with life , n (%)		
Very unsatisfied	3 (2.67)	0 (0.0)
Unsatisfied	7 (6.25)	2 (4.08)
Below average satisfaction	20 (17.86)	4 (8.16)
Averagely satisfied	25 (22.32)	14 (28.57)
Satisfied	37 (33.04)	20 (40.82)
Very satisfied	20 (17.86)	9 (18.37)
Absence from work , n (%)	18 (34.62)	6 (25.00)
Impaired productivity at work , n (%)	30 (61.22)	15 (62.50)

n, number of patients included in the analysis; IQR, Inter Quartile Range; BMI, Body Mass Index (kg m⁻²); GSRs, gastrointestinal symptom rating scale; FD, Functional Dyspepsia; HADS, hospital anxiety and depression scale; PCS, physical quality of life composite summary as assessed by SF-36; MCS, mental quality of life composite summary as assessed by SF-36; GI, gastrointestinal. Numbers may not add up to total due to missing. ***P<0.001 versus Rome III-positive; **P<0.005 versus Rome III-positive.

Table 3.2 Baseline demographical and clinical characteristics, of patients still fulfilling the Rome III criteria for IBS at follow-up (FU Rome III-positive) and those who did not (FU Rome IV negative).

Characteristics at inclusion	FU Rome III-positive n=112	FU Rome III-negative n=49
Age , median (IQR)	48 (29-61)	51 (39-61)
Gender , n female (%)	84 (75.0)	36 (73.47)
BMI , median kg m ⁻² (IQR)	23.70 (21.50-28.70)	24.40 (21.50-28.05)
Secondary/tertiary care , n (%)	72 (64.29)	35 (71.43)
IBS subtype , n (%)		
IBS-D	40 (35.7)	19 (38.8)
IBS-C	23 (20.5)	8 (16.3)
IBS-M	47 (42.0)	16 (32.7)
IBS-U	2 (1.8)	6 (12.2)
GSRS , median (IQR)		
Abdominal pain	3.67 (2.33-4.00)	2.83 (2.00-3.67)
Reflux syndrome	1.50 (1.00-2.63)	1.00 (1.00-2.88)
Diarrhea syndrome	3.00 (2.33-4.67)	2.83 (1.67-4.25)
Constipation syndrome	3.17 (2.33-4.00)	1.49 (2.00-3.67)
Indigestion syndrome	4.25 (3.38-5.00)	3.88 (2.75-5.00)
Diary , median score (IQR)		
Discomfort	2.29 (2.00-2.72)	2.21 (1.64-2.57)
Pain	2.23 (1.64-2.79)	1.93 (1.27-2.50)
Constipation	1.29 (1.00-1.79)	1.14 (1.00-1.59)
Diarrhea	1.21 (1.07-1.64)	1.07 (1.00-1.38)
Bloating	2.00 (1.52-2.64)	2.00 (1.25-2.43)
Flatulence	2.29 (1.89-2.89)	2.23 (1.64-2.80)
Belching	1.36 (1.00-2.00)	1.18 (1.00-1.92)
Nausea	1.21 (1.00-1.81)	1.14 (1.00-1.57)
HADS , score >8, n (%)		
Depressive symptoms	20 (19.05)	6 (5.71)
General anxiety symptoms	41 (38.68)	17 (38.64)
Quality of life , median (IQR)		
PCS	41.71 (33.90-48.98)	45.51 (36.32-51.74)
MCS	50.20 (41.45-55.19)	50.67 (41.38-57.44)

n, number of patients included in the analysis; IQR, Inter Quartile Range; BMI, Body Mass Index (kg m⁻²); secondary/tertiary care as compared to primary care; GSRS, gastrointestinal symptom rating scale; HADS, hospital anxiety and depression scale; PCS, physical quality of life composite summary as assessed by SF-36; MCS, mental quality of life composite summary as assessed by SF-36. Numbers may not add up to total due to missing. Please note that several variables shown in Table 3.1 derive from questionnaires not administered at baseline and could therefore not be considered in the comparison of baseline variables in Table 3.2.

Symptom evolution over time

Table 3.3 shows the evolution of clinical characteristics and symptoms over time by pairwise comparisons between baseline and follow-up measurements. Overall, improvements in GI symptom severity (assessed by the GSRS, as baseline GSRS-IBS data were not available), general quality of life, and general anxiety and depression over a five-year period were small and did not reach significance in patients that still met the Rome III-diagnostic criteria at follow-up. Symptom severity for abdominal pain, diarrhea, constipation, and indigestion did improve significantly over time in the Rome III-negative patients, although they still had 'low to moderate' levels of symptoms after five years, with a median abdominal pain GSRS score of 2.33 (IQR=1.33-3.00), and a median indigestion GSRS score of 2.88 (IQR=2.25-3.94). However, similar to the Rome III-positive patients, the Rome III-negative patients did not improve significantly in quality of life during the follow-up period. In addition, their comorbid depression scores were relatively unaltered.

With regard to IBS subtype based on predominant bowel habit, Supplemental Figure S3.1 illustrates the proportion of IBS subtypes at baseline and at follow-up. Sixty-two (55.4%) out of 112 Rome III-positive patients had changed IBS-subtype during the follow-up period. In Supplemental Figure S3.2, the within-patient subtype changes from baseline to follow-up are shown. IBS-D appeared the most stable subtype (72.5% of Rome III-positive unchanged), whereas IBS-M and IBS-C were least stable (76.6% and 60.9% of Rome III-positive changed respectively).

Multivariable regression model

No baseline predictors for fulfilling the Rome III diagnosis for IBS at follow-up could be identified with regression analyses. When looking for independent baseline predictors concerning quality of life at follow-up for both groups taken together (n=161), the combined regression model that included age, gender, baseline GI symptom severity, and baseline anxiety and depression, showed that only younger age (β -0.16, 95% CI -0.26; -0.06, $P=0.002$) and higher physical quality of life at baseline (β 0.59, 95% CI 0.41; 0.78, $P<0.001$) predicted higher physical quality of life at follow-up. Likewise, only higher mental quality of life at baseline (β 0.46, 95% CI 0.15; 0.77, $P=0.004$), but not baseline depression or anxiety scores, predicted higher mental quality of life at follow-up.

Table 3.3 Clinical characteristics, at baseline and at time of follow-up, of patients who still fulfilled the Rome III criteria for IBS at follow-up (Rome III-positive) and patients who did not (Rome III-negative).

	Baseline	Follow-up
FU Rome III-positive IBS patients, n=112		
BMI , median kg m ⁻² (IQR)	23.70 (21.45-28.70)	24.77 (21.92-28.60)**
GSRs , median (IQR)		
Abdominal pain	3.67 (2.33-4.00)	3.33 (2.50-4.33)
Reflux syndrome	1.50 (1.00-2.75)	1.50 (1.00-3.00)
Diarrhea syndrome	3.00 (2.33-4.67)	3.33 (1.83-4.67)
Constipation syndrome	3.00 (2.33-4.00)	3.00 (2.00-4.00)
Indigestion syndrome	4.00 (3.38-5.00)	3.75 (3.00-4.50)
HADS , score >8, n (%)		
Depressive symptoms	20 (19.05)	25 (23.81)
General anxiety symptoms	41 (39.05)	31 (29.52)
Quality of life , median (IQR)		
PCS	42.19 (34.36-48.90)	45.41 (33.44-48.78)
MCS	49.58 (37.87-54.96)	51.07 (37.57-55.93)
FU Rome III-negative IBS patients, n=49		
BMI , median kg m ⁻² (IQR)	24.50 (21.30-28.08)	25.35 (21.68-28.73)
GSRs , median (IQR)		
Abdominal pain	2.83 (2.00-3.67)	2.33 (1.33-3.00)*
Reflux syndrome	1.00 (1.00-2.88)	1.50 (1.00-3.00)
Diarrhea syndrome	2.83 (1.67-4.25)	1.67 (1.08-3.00)**
Constipation syndrome	2.50 (2.00-3.67)	2.00 (1.33-3.00)***
Indigestion syndrome	3.88 (2.75-5.00)	2.88 (2.25-3.94)***
HADS , score >8, n (%)		
Depressive symptoms	6 (13.64)	7 (15.91)
General anxiety symptoms	17 (38.64)	7 (15.91)∞
Quality of life , median (IQR)		
PCS	45.51 (36.32-51.74)	48.85 (33.14-52.99)
MCS	50.68 (41.38-57.44)	53.38 (43.89-57.30)

n, number of patients included in the analysis; IQR, Inter Quartile Range; BMI, Body Mass Index (kg m⁻²); GSRs, gastrointestinal symptom rating scale; HADS, hospital anxiety and depression scale; PCS, physical quality of life composite summary as assessed by SF-36; MCS, mental quality of life composite summary as assessed by SF-36. Numbers may not add up to total due to missing. ****P*<0.001 vs. baseline; ***P*<0.005 vs. baseline; * *P*<0.01 vs. baseline; ∞*P*< 0.05 vs. baseline.

When looking for characteristics at follow-up that were associated with quality of life at follow-up, general anxiety (β -1.49, 95% CI -1.90; -1.07, *P*<0.001), and depression levels (β -1.00, 95% CI -1.48; -0.53, *P*<0.001) at follow-up were the only two parameters independently associated with mental quality of life at follow-up after correcting the model for age, gender, IBS-subtype, GI symptom severity, GI-specific anxiety, and general anxiety and depression. No characteristics at follow-up were found to independently associate with physical quality of life at follow-up. Baseline IBS-subtypes were not included in the models that looked for baseline predictors, as univariate analyses showed no

significant differences in subtypes between groups, that is, the distribution in baseline subtypes did not differ significantly between Rome III-negative and -positive patients. Inclusion of these variables into the model would have decreased statistical validity and increased the probability of type I errors.

Discussion

In the current prospective study, we evaluated the natural symptom evolution of patients with IBS in the Maastricht IBS cohort over time.³³ We demonstrated that 30.4% of patients did no longer fulfill the Rome III criteria after a five-year follow-up period. The most salient finding is that quality of life did not improve significantly in patients who showed a decrease in gastrointestinal symptom severity (*i.e.* being Rome III-negative at follow-up), compared to patients who had unaltered symptom severity over time (Rome III-positive). In addition, general wellbeing in terms of comorbid general anxiety and depression, work absenteeism and productivity, and life satisfaction were also comparable in those who still fulfilled the Rome III criteria at follow-up, when compared to those who did not.

Several studies have investigated IBS symptom evolution in relation to diagnostic criteria in patient cohorts, with varying results. A study by Williams *et al.* using the Rome II criteria reported that 52% had no IBS two years after web-based diagnosis¹², Ford *et al.* found that 28% did not meet the Manning criteria after 10 years³⁴, Mearin *et al.* reported that abdominal pain frequency decreased in 26% after one year³⁵, and more recently, Card *et al.* reported that 27% of their large study population did not fulfill the Rome III criteria after one year.³⁶ Except for the study by Williams *et al.*, the proportion of patients who do not meet the diagnostic criteria after a follow-up period is thus in line with the findings of this follow-up study in our Dutch cohort.

Given that clinical treatment decisions often rely on prognostic predictions of the disease course, we sought to identify baseline predictors for a less favorable disease course. This study, however, did not show any baseline characteristics associated with meeting the Rome III criteria at follow-up. As IBS is a highly heterogeneous disorder, larger study populations might be necessary to investigate what factors influence and thereby predict the natural disease course in (subgroups of) patients with IBS.

Rather unexpectedly, we found that a decrease in GI symptom severity did not lead to an improved quality of life in our study. Clevers *et al.* recently evaluated

longitudinal symptom changes over time and, in contrast to our results, found that patients with lower GI symptom severity had significantly higher quality of life scores.³⁷ They also demonstrated that GI-specific anxiety is associated with an increase in GI symptom severity, which is in agreement with our results as GI-specific anxiety was significantly lower in the group that did not fulfill the Rome III criteria. The inconsistency between our data in terms of quality of life might be explained in part by the different questionnaires used. Clevers *et al.* used an IBS specific questionnaire, the IBS-QoL³⁸, which assesses more disease specific changes in quality of life in contrast to generic quality of life instruments, such as the SF-36 that we used. Our data can therefore also be compared with other diseases and with the general population. Both the Rome III-positive and -negative group showed lower mean quality of life than the mean of a Dutch population sample³⁹ and of a USA based population without a functional GI disorder.⁴⁰ Additionally, we used the validated SWLS to score overall life satisfaction. This has been used in healthy persons and in patients with Crohn's disease⁴¹, but, to our knowledge, not in IBS patients. We found scores comparable to the ones reported by Crohn's disease patients for both the Rome III-positive and -negative group.

In contrast to GI symptom severity, we found that quality of life did not improve over time in those that were Rome III-negative at follow-up. The data reported here suggest that concurrent, but not baseline, psychological comorbidities are more predictive of this impaired long-term quality of life than GI symptom severity. Several studies in IBS have found similar results, in particular for the mental health related quality of life; *i.e.* Naliboff *et al.* reported that psychological distress had a stronger effect on health related quality of life than GI symptoms⁴², Koloski *et al.* reported that depression was independently associated with mental quality of life⁴³, and Addante *et al.* reported that perceived stress, and anxiety and depression were significant predictors for mental health related quality of life.⁴⁴ This raises the question whether current treatment goals for IBS in daily clinical practice should be revised. Gaining insight into which symptoms specifically affect a patient's quality of life can aid in reprioritizing personal treatment goals and increase the chance of a successful individualized treatment trajectory.⁴⁵ Currently, treatment is merely targeted towards the patient's predominant GI symptoms and in many cases, this entails primary treatment of pain, constipation, or diarrhea. Targeting GI symptom improvement without characterizing the full extent of the disorder and possible psychosocial modifiers may therefore contribute to treatment failure concerning quality of life over time. A helpful and pragmatic framework

in this regard has been suggested by the Rome expert panel.⁴⁵ By following their five-step approach, clinicians can identify if and which psychological factors contribute to the disease burden in the individual patient and consider psychological treatments. In some patients, this might include the use of pharmacological neuromodulators⁴⁶, but other therapy options include cognitive behavioral therapy (CBT), gut-directed hypnotherapy⁴⁷, and dynamic psychotherapy.⁴⁸ Due to the relative lack in therapist availability, clinical implementation of these therapies is still limited. Recent attempts to reduce resource use without compromising effectiveness may assist in making psychological treatment more widely available; *e.g.* group hypnotherapy⁴⁹, internet-delivered exposure-based CBT⁵⁰, and home-based CBT.⁵¹ In that light, it will be interesting to explore whether these developments can affect long-term quality of life and well-being in patients with IBS.

Several limitations of the current study should be noted. Only two time-points within the five-year period were assessed; no further data were available about the period in between the baseline and follow-up measurement, including data on treatment received in this period. Therefore, we can neither comment about the frequency and duration of IBS flares within these five years, nor about the effect of treatment on symptoms, nor on the causal order of the impaired quality of life and psychological comorbidities. Moreover, selection bias cannot be excluded. Follow-up cohort studies depend on long-term dedication of their participants and the proportion of non-responders (loss to follow) in this study was 46.4%. To assess for potential selection bias, we tested for differences in baseline characteristics between responders and non-responders at follow-up. The groups did not differ in baseline demographic and lifestyle characteristics, nor in baseline symptom scores and subtypes (data not shown). Hence, there is no clear reason to assume that the non-responders would have had different outcomes at follow-up. Finally, a larger sample size would have been desirable. Due to a relatively small Rome III-negative group, small effects concerning quality of life or other parameters could have been missed, especially since we have corrected for multiple comparisons. Although we argue that this adjustment improves the reproducibility of our results, an unnecessarily high false negative rate cannot be ruled out. Because of the small sample size, it is important that our findings are corroborated in other cohorts of patients with IBS. Until then, findings should be considered exploratory and appropriate caution should be taken when interpreting the results of the current study.

Strengths of this prospective follow-up study include the heterogeneous and well-characterized Maastricht IBS cohort population, *i.e.* in-depth phenotyping enables comparison to other large cohorts and future studies should explore possibilities of pooling data from different centers to cluster phenotypes and validate the robustness of findings such as the ones reported here. Moreover, since we have recruited patients from both primary and secondary/tertiary care settings (Table 3.1), our population is representative for the general Dutch IBS population. Another strength is the evaluation of the Rome criteria in a telephonic follow-up interview with patients. Due to the current lack of validated biomarkers, IBS remains a symptom-based diagnosis for which the Rome criteria are used to aid diagnostic accuracy, both in research and clinical settings.⁵² Since the Rome criteria have not been developed nor validated as a self-administered questionnaire, asking patients to complete the Rome criteria without supervision by a trained researcher may lead to bias and over- or under-estimation of the actual diagnosis, as previously demonstrated by Lovell *et al.*² In contrast to earlier studies on follow-up of IBS, our study method implied a complete reevaluation of the diagnosis in a telephonic interview and, thus, allows for a more valid and less biased interpretation of Rome diagnosed IBS prevalence over time.

In conclusion, the current prospective study contributes to existing insight regarding symptom evolution over time in IBS patients and showed that 30% of patients did no longer fulfill the Rome III criteria after a five-year follow-up period. However, the decrease in GI symptom severity (*i.e.* being Rome III-negative at follow-up), did not impact quality of life nor life satisfaction. Our results indicate that long-term quality of life and general well-being might depend on comorbid psychological symptoms, *i.e.* affective states, rather than gastrointestinal symptom severity.

References

1. El-Salhy M. Irritable bowel syndrome: diagnosis and pathogenesis. *World J Gastroenterol.* 2012;18:5151-5163.
2. Lovell RM, Ford AC. Global prevalence of and risk factors for irritable bowel syndrome: a meta-analysis. *Clin Gastroenterol Hepatol.* 2012;10:712-721 e714.
3. Van den Houte K, Carbone F, Pannemans J, et al. Prevalence and impact of self-reported irritable bowel symptoms in the general population. *United European Gastroenterol J.* 2018;7:307-315.
4. Jossan N, Simren M, Sperber AD, et al. Health care utilization for Rome IV irritable bowel syndrome; a three-country survey in the general population. *Gastroenterology.* 2017;152:S68.
5. Camilleri M. Peripheral mechanisms in irritable bowel syndrome, *N Engl J Med.* 2012;367:1626-1635.
6. Holtmann GJ, Ford AC, Talley NJ. Pathophysiology of irritable bowel syndrome. *Lancet Gastroenterol Hepatol.* 2016; 1:133-146.
7. Ballou S, Keefer L. The impact of irritable bowel syndrome on daily functioning: Characterizing and understanding daily consequences of IBS. *Neurogastroenterol Motil.* 2017;29.
8. Gralnek IM, Hays RD, Kilbourne A, Naliboff B, Mayer EA. The impact of irritable bowel syndrome on health-related quality of life. *Gastroenterology.* 2000;119:654-660.
9. Palsson OS, Baggish J, Whitehead WE. Episodic nature of symptoms in irritable bowel syndrome. *Am J Gastroenterol.* 2014;109:1450-1460.
10. Drossman DA, Morris CB, Hu Y, et al. A prospective assessment of bowel habit in irritable bowel syndrome in women: defining an alternator. *Gastroenterology.* 2005;128:580-589.
11. Garrigues V, Mearin F, Badia X, et al. Change over time of bowel habit in irritable bowel syndrome: a prospective, observational, 1-year follow-up study (RITMO study). *Aliment Pharmacol Ther.* 2007;25:323-332.
12. Williams RE, Black CL, Kim HY, et al. Stability of irritable bowel syndrome using a Rome II-based classification. *Aliment Pharmacol Ther.* 2006;23:197-205.
13. Spiegel B, Harris L, Lucak S, et al. Developing valid and reliable health utilities in irritable bowel syndrome: results from the IBS PROOF Cohort. *Am J Gastroenterol.* 2009;104:1984-1991.
14. Corsetti M, Whorwell P. The global impact of IBS: time to think about IBS-specific models of care?, *Therap Adv Gastroenterol.* 2017;10:727-736.
15. Mujagic Z, Jonkers D, Ludidi S, et al. Biomarkers for visceral hypersensitivity in patients with irritable bowel syndrome. *Neurogastroenterol Motility.* 2017;29(12).
16. Ludidi S, Mujagic Z, Jonkers D, et al. Markers for visceral hypersensitivity in patients with irritable bowel syndrome, *Neurogastroenterol Motil.* 2014;26:1104-1111.
17. Thijssen AY, Mujagic Z, Jonkers DM, et al. Alterations in serotonin metabolism in the irritable bowel syndrome. *Aliment Pharmacol Ther.* 2016;43:272-282.
18. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA.* 2013;310(20):2191-2194.
19. Svedlund J, Sjodin I, Dotevall G. GRS--a clinical rating scale for gastrointestinal symptoms in patients with irritable bowel syndrome and peptic ulcer disease. *Dig Dis Sci.* 1988;33:129-134.

20. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand.* 1983;67:361-370.
21. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *J Psychosom Res.* 2002;52:69-77.
22. McHorney CA, Ware JE, Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care.* 1993;31:247-263.
23. Farivar SS, Cunningham WE, Hays RD. Correlated physical and mental health summary scores for the SF-36 and SF-12 Health Survey, V.I. *Health Qual Life Outcomes.* 2007;5:54.
24. Wiklund IK, Fullerton S, Hawkey CJ, et al. An irritable bowel syndrome-specific symptom questionnaire: development and validation, *Scand J Gastroenterol.* 2003;38:947-954.
25. Labus JS, Bolus R, Chang L, et al. The Visceral Sensitivity Index: development and validation of a gastrointestinal symptom-specific anxiety scale. *Aliment Pharmacol Ther.* 2004;20: 89-97.
26. Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. *J Pers Assess.* 1985;49:71-75.
27. Pavot W, Diener E, Colvin CR, Sandvik E. Further validation of the Satisfaction with Life Scale: evidence for the cross-method convergence of well-being measures. *J Pers Assess.* 1991;57:149-161.
28. Van Beuningen J. Satisfaction With Life Scale: examining construct validity. Heerlen (The Hague), 2012.
29. Frandemark A, Tornblom H, Jakobsson S, Simren M. Work productivity and activity impairment in irritable bowel syndrome (IBS): a multifaceted problem. *Am J Gastroenterol.* 2018;113(10):1540-1549.
30. Bouwmans C, Krol M, Severens H, Koopmanschap M, Brouwer W, Hakkaart-van Roijen L. The iMTA productivity cost questionnaire: a standardized instrument for measuring and valuing health-related productivity losses. *Value Health.* 2015;18:753-758.
31. Bouwmans C, Krol M, Brouwer W, Severens JL, Koopmanschap MA, Hakkaart L. IMTA productivity cost questionnaire (IPCQ), *Value Health.* 2014;17:A550.
32. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing, 1995.
33. Weerts Z, Vork L, Mujagic Z, Keszthelyi D, Hesselink M, Jonkers D, Masclee A. The natural history of IBS: A five-year follow-up of the Maastricht IBS Cohort study, Poster Sessions, Poster 276. *Neurogastroenterol Motil.* 2018;30:1-186.
34. Ford AC, Forman D, Bailey AG, Axon AT, Moayyedi P. Fluctuation of gastrointestinal symptoms in the community: a 10-year longitudinal follow-up study. *Aliment Pharmacol Ther.* 2008;28:1013-1020.
35. Mearin F, Badia X, Balboa A, et al. Predictive factors of irritable bowel syndrome improvement: 1-year prospective evaluation in 400 patients. *Aliment Pharmacol Ther.* 2006;23:815-826.
36. Card T, Enck P, Barbara G, et al. Post-infectious IBS: Defining its clinical features and prognosis using an internet-based survey. *United European Gastroenterol J.* 2018;6: 1245-1253.
37. Clevers E, Tack J, Tornblom H, et al. Development of irritable bowel syndrome features over a 5-year period. *Clin Gastroenterol Hepatol.* 2018;16(8).
38. Drossman DA, Patrick DL, Whitehead WE, et al. Further validation of the IBS-QOL: a disease-specific quality-of-life questionnaire. *Am J Gastroenterol.* 2000;95:999.

39. Ware JE Jr, Gandek B, Kosinski M, et al. The equivalence of SF-36 summary health scores estimated using standard and country-specific algorithms in 10 countries: results from the IQOLA Project. *International quality of life assessment. J Clin Epidemiol.* 1998;51:1167-1170.
40. Kanuri N, Cassell B, Bruce SE, et al. The impact of abuse and mood on bowel symptoms and health-related quality of life in irritable bowel syndrome (IBS). *Neurogastroenterol Motil.* 2016;28:1508-1517.
41. Sarid O, Slonim-Nevo V, Pereg A, et al. Coping strategies, satisfaction with life, and quality of life in Crohn's disease: A gender perspective using structural equation modeling analysis. *PLoS One.* 2017;12:e0172779.
42. Naliboff BD, Kim SE, Bolus R, Bernstein CN, Mayer EA, Chang L. Gastrointestinal and psychological mediators of health-related quality of life in IBS and IBD: a structural equation modeling analysis. *Am J Gastroenterol.* 2012;107: 451-459.
43. Koloski NA, Boyce PM, Jones MP, Talley NJ. What level of IBS symptoms drives impairment in health-related quality of life in community subjects with irritable bowel syndrome? *Qual Life Res.* 2012;21:829-836.
44. Addante R, Naliboff B, Shih W, et al. Predictors of Health-related Quality of Life in Irritable Bowel Syndrome Patients Compared With Healthy Individuals. *J Clin Gastroenterol.* 2018;53:e142-e149.
45. Max JS, It sup, et al. What Is New in Rome IV. *J Neurogastroenterol Motil.* 2017;23:151-163.
46. Tornblom H, Drossman DA. Psychotropics, antidepressants, and visceral analgesics in functional gastrointestinal disorders. *Curr Gastroenterol Rep.* 2018;20:58.
47. Miller V, Carruthers HR, Morris J, Hasan SS, Archbold S, Whorwell PJ. Hypnotherapy for irritable bowel syndrome: an audit of one thousand adult patients. *Aliment Pharmacol Ther.* 2015;41:844-855.
48. Guthrie E, Creed F, Dawson D, Tomenson B. A controlled trial of psychological treatment for the irritable bowel syndrome. *Gastroenterology.* 1991;100:450-457.
49. Flik CE, Laan W, Zuithoff NP, et al. Efficacy of individual and group hypnotherapy in irritable bowel syndrome (IMAGINE): a multicentre randomised controlled trial. *Lancet Gastroenterol Hepatol.* 2019;4:20-31.
50. Ljotsson B, Hedman E, Andersson E, et al. Internet-delivered exposure-based treatment vs. stress management for irritable bowel syndrome: a randomized trial. *Am J Gastroenterol.* 2011;106(8):1481-1491.
51. Lackner JM, Jaccard J, Keefer L, et al. Improvement in gastrointestinal symptoms after cognitive behavior therapy for refractory irritable bowel syndrome. *Gastroenterology.* 2018;155:47-57.
52. Mearin F, Lacy BE, Chang L, et al. Bowel disorders. *Gastroenterology.* 2016, DOI 10.1053/j.gastro.2016.02.031.

Supplemental material

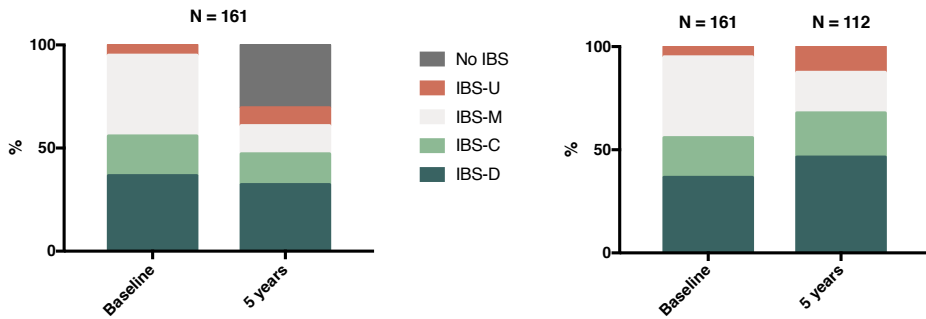


Figure S3.1 Proportion of patients with IBS-D, IBS-C, IBS-M, IBS-U, and No IBS (at follow-up) in the total population analyzed, n=161, and the proportion of patients with IBS-D, IBS-C, IBS-M, IBS-U in the population that was Rome III-positive at follow-up, n=112. The figure does not account for transitions of individual patients to different subtypes. N; number, IBS; Irritable Bowel Syndrome. IBS-D; Irritable Bowel Syndrome, diarrhea predominant subtype. IBS-C; Irritable Bowel Syndrome, constipation predominant subtype. IBS-M; Irritable Bowel Syndrome, mixed stool type subtype. IBS-U; Irritable Bowel Syndrome, undefined subtype.

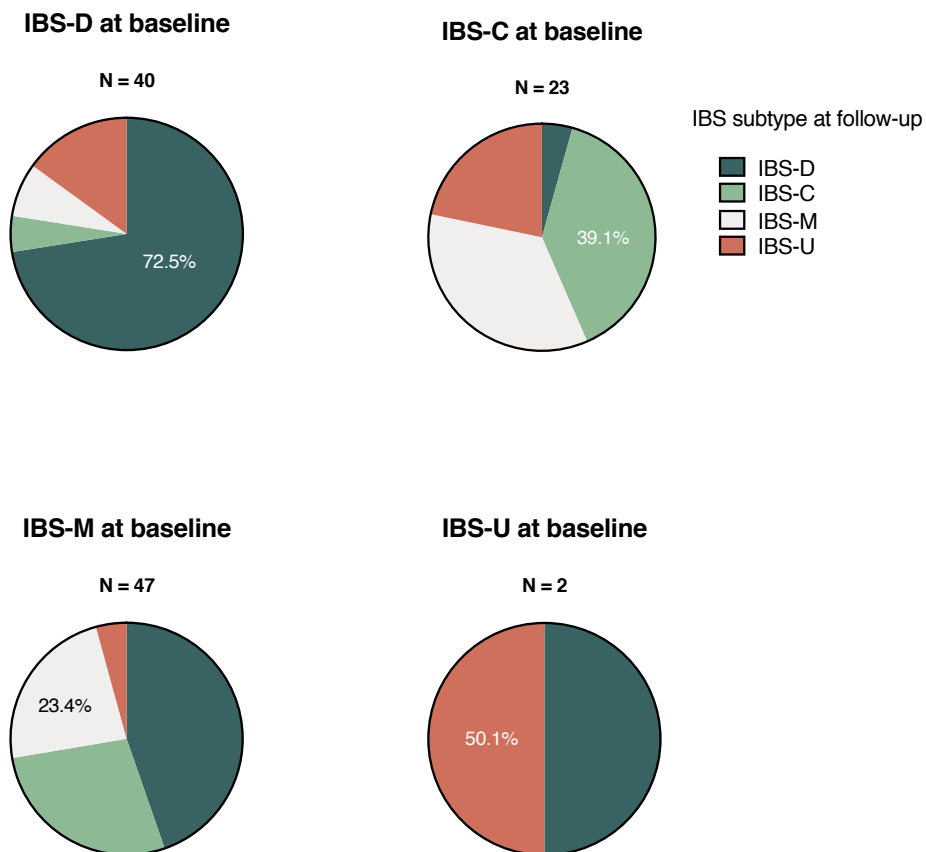


Figure S3.2 Pie slices represent the proportion of patients with a particular IBS subtype at follow-up (please refer to legend for information on which subtype). The percentage depicted within the pie represents the proportion of patients that were diagnosed as having the same IBS subtype at follow-up, as at baseline. IBS-D; Irritable Bowel Syndrome, diarrhea predominant subtype. IBS-C; Irritable Bowel Syndrome, constipation predominant subtype. IBS-M; Irritable Bowel Syndrome, mixed stool type subtype. IBS-U; Irritable Bowel Syndrome, undefined subtype.

Chapter 4

Does day-to-day variability in stool consistency link to the fecal microbiota composition?

Lisa Vork*, John Penders*, Jonna Jalanka, Ceca Bojic, Sander M.J. van Kuijk, Anne Salonen, Willem M. de Vos, Mirjana Rajilić-Stojanović, Ad A.M. Masclee, Marta Pozuelo, Chaysavanh Manichanh, Daisy M.A.E. Jonkers

On behalf of the COST Action GENIEUR / BM1106 Microbiome Working Group.

* Both authors contributed equally to the manuscript.

Submitted

Abstract

INTRODUCTION Stool consistency has been associated with fecal microbial composition. Temporal instability in stool consistency has been demonstrated in irritable bowel syndrome (IBS), raising the question whether variability in the microbial composition should be considered in microbiota studies. We evaluated within-subject day-to-day variability in stool consistency and the association with the fecal microbiota in IBS and healthy subjects, over seven days.

METHODS Twelve IBS patients and 12 healthy subjects collected fecal samples daily during seven consecutive days. Stool consistency was determined by the patient-reported Bristol Stool Scale (BSS) and fecal dry weight percentage. 16S rRNA V4 gene sequencing was performed and microbial richness (*i.e.* alpha diversity; Chao1 index, observed number of species, effective Shannon index) and microbial community structure (*i.e.* beta diversity; Bray-Curtis distance, generalized UniFrac and abundance of taxa on family level) were determined.

RESULTS Linear mixed-effects models showed significant associations between stool consistency and microbial richness, but no effect of time. This implies that between-subject but not within-subject variation in microbiota over time can partially be explained by variation in stool consistency. Redundancy analysis showed a significant association between stool consistency and microbial community structure, but additional linear mixed-effects models did not demonstrate an effect of time on this association.

CONCLUSIONS This study supports an association between stool consistency and fecal microbiota, but no effect of day-to-day fluctuations in stool consistency within seven days. This consolidates the importance of considering stool consistency in gut microbiota research, though confirms the validity of single fecal sampling to represent an individual's microbiota in short-term studies.

Introduction

Over the past decades, the gut microbiota has been studied extensively in the context of gastrointestinal (GI) function in both health and disease. Recently, stool consistency measured with Bristol Stool Scale (BSS) or fecal moisture content has been associated with gut microbiota diversity and composition^{1,2}, and has been demonstrated to be an important covariate of the microbiota composition in large population-based studies.^{3,4} These findings indicate the importance of taking into account stool consistency as a potential confounding factor in future intestinal microbiota analyses. However, stool consistency can fluctuate over time, possibly associated with temporal fluctuations in the microbiota composition. This raises the question whether within-subject day-to-day variability in microbial composition should be considered in microbiota studies, by collecting repeated fecal samples.

Irritable bowel syndrome (IBS), a functional GI disorder characterized by abdominal pain and altered bowel habits, can serve as a model for fluctuating stool consistency. According to the Rome criteria, IBS is typically divided into four subtypes: diarrhea predominant (IBS-D), constipation predominant (IBS-C), a combination of both (IBS-M; mixed type IBS), or an unspecified subtype in which both diarrhea and constipation are not predominantly present (IBS-U).⁵ Multiple studies have shown changes in gut microbiota composition as well as functionality in IBS patients compared to healthy subjects.⁶⁻⁸ Some results also point towards differences in microbiota between subgroups of IBS subjects⁹⁻¹¹, though findings on specific bacterial taxa are not always consistent. Additionally, changes in microbial composition have been correlated to IBS symptom scores in cross-sectional studies^{7,11,12}, which also generally vary over time.

So far, a few studies have focused on temporal (in)stability of the fecal microbiota in IBS populations and have found a more unstable microbial composition in IBS patients compared to healthy volunteers over a period of months.¹³⁻¹⁵ However, none have considered possible day-to-day variation and its association with stool consistency and IBS symptoms within subjects over time. Since both stool consistency and gastrointestinal symptoms can fluctuate from day to day in IBS subjects, a key question is whether a single fecal sample at one time point suffices or whether repeated fecal samples over a period of time should be collected in order to take into account day-to-day variability in microbial composition.

Therefore, we aimed to evaluate within-subject day-to-day variability in stool consistency and gastrointestinal symptoms, and the association with the fecal microbiota composition in IBS patients and healthy subjects, over a seven-day course.

Materials and methods

Study design and participants

The current project has been performed as part of a Dutch cohort study on the phenotypical and genotypical characterization of IBS patients, the Maastricht IBS (MIBS) Cohort. A subset of participants of the MIBS Cohort participated in the present subproject, which involves a prospective, longitudinal, observational study. The study protocol has been approved by the Maastricht University Medical Center+ (MUMC+) Committee of Ethics and was executed according to the revised Declaration of Helsinki (64th WMA General Assembly, Fortaleza, Brazil, October 2013). The study has been registered in the US National Library of Medicine (<http://www.clinicaltrials.gov>, NCT00775060).

Between January 2015 and March 2016, IBS patients aged 18-75 years were recruited at the outpatient department of Gastroenterology-Hepatology of MUMC+ in Maastricht, the Netherlands. All subjects fulfilled the Rome III criteria for IBS and were assigned to IBS subtypes based on predominant bowel habit, *i.e.* diarrhea (IBS-D), constipation (IBS-C), a mix of diarrhea and constipation (IBS-M), or unspecified predominant bowel habit (IBS-U).¹⁶⁻¹⁸ Medical history was taken by a gastroenterologist and if indicated, GI endoscopy, abdominal imaging, and/or blood, breath or fecal analysis were performed to exclude organic disease. A history of abdominal surgery, except for uncomplicated appendectomy, cholecystectomy, or hysterectomy, was reason for exclusion.

Age- and sex-matched healthy controls (HC) were recruited via public advertising in MUMC+ and Dutch study recruitment websites during the same period. A brief medical history was taken by a trained research physician to exclude the presence of any GI disorders or current GI symptoms. All study participants gave their written informed consent prior to inclusion.

Bio samples and symptom scores

A seven-day symptom diary was used to record daily symptom scores and bowel habits using the Bristol Stool Scale (BSS).^{19,20} Symptom scores (*i.e.* abdominal pain, bloating, fecal urgency, diarrhea, and constipation) were scored at the end of the day on an 11-point numeric rating scale (*i.e.* 0 (none) to 10 (severe))²⁰ and BSS for stool consistency was reported for every bowel movement. Additionally, medication use was reported every day and participants maintained their habitual dietary habits during the total study period.

The first fecal samples of each day were collected and one or two additional samples were collected when subjects reported diarrhea during that day, resulting in a total of 149 samples. Subjects were instructed to store the samples at -20°C in the home situation directly after collection. Following the seven-day study period, all samples were transported to the MUMC+ on dry ice and stored directly at -80°C directly upon arrival in the lab.

In addition to recording stool consistency using the BSS, the dry weight content of each fecal sample was determined, as a more objective measure of stool consistency. Of each sample, an aliquot of 0.5 g (*wet weight*) was dried at 60°C in a vacuum dryer for 5 hours. Directly after drying, the samples were weighed again (*dry weight*). The percentage of dry weight was calculated as: [dry weight(g) / wet weight(g)] * 100.

Fecal microbiota analysis

DNA extraction

Approximately 125 mg of fecal matter was used for DNA extraction using the validated Repeated Bead Beating (RBB) method^{21,22} with the following modifications for automated DNA purification: feces were suspended in 1.0 ml of sterile ice-cold PBS, and 175 µl of fecal suspension was combined with 235 µl of RBB lysis buffer (500 mM NaCl, 50 mM Tris-HCl (pH 8.0), 50 mM EDTA, 4% SDS) in a bead beating tube from the Ambion Magmax™ Total Nucleic Acid Isolation Kit (Life Technologies, Carlsbad, CA, USA). After repeated bead beating, 200 µl of the supernatant was used for DNA extraction with the KingFisher™ Flex automated purification system (ThermoFisher Scientific, Waltham, MA, USA) using MagMAX™ Pathogen High Vol Duo program. DNA was quantified using Quanti-iT™ Pico Green dsDNA Assay (Invitrogen, San Diego, CA, USA).

High-throughput DNA sequencing

The microbiota composition was profiled by amplifying the hypervariable V4-region of the bacterial and archaeal 16S rRNA gene by PCR. The detailed description of the high-throughput amplicon sequencing protocol can be found in a previous publication.²³ Briefly, forward V4F_515_19 (V4F_515_19: 5'-GTGCCAGCAMGCCGCGTAA -3') and 12-base Golay barcoded reverse primers V4R_806 (5'-GGACTACCAGGGTATCTAAT-3') were used to amplify the V4 region.

Amplicons were purified using the QIAquick PCR Purification Kit (Qiagen, Barcelona, Spain), quantified using a NanoDrop ND-1000 Spectrophotometer (Nucliber) and then pooled in an equimolar concentration. The pooled amplicon library was subsequently subjected to sequencing using Illumina MiSeq technology at the technical support unit of the Autonomous University of Barcelona (UAB, Spain), following standard Illumina platform protocols.

Sequence data analysis

Raw sequences were loaded in a unique file into the QIIME software. Upon removal of low-quality reads with a Phred score smaller than 20, reads were demultiplexed. This resulted in 10.145.380 high quality reads (from 14.642.422 input sequences) for the 149 samples ranging from 36.107 to 93.479 reads per sample. Next, sequences were clustered into Operational Taxonomic Units (OTUs) with 97% of similarity using the USEARCH algorithm.²⁴ Chimeric sequences were removed with UCHIME²⁵ and OTUs corresponding to singletons were eliminated.

To reduce the data to work with, a representative sequence from each OTU was picked and these were aligned against a GreenGenes template alignment using PyNAST as recommended by Navas-Molina *et al.*²⁶. Using the Basic Local Alignment Search Tool (BLAST) and a database that combines GreenGenes (gg_13_8 release) and PATRIC (Pathosystems Resource Integration Center) databases, taxonomy was assigned to each OTU. The OTU table, containing the taxonomical assignment per OTU and their abundances in each sample, was built and at this point, a total of 8.054.504 reads for the 149 samples ranging from 29.926 to 73.203 reads per sample, remained for downstream analyses. Data were normalized using a rarefaction technique, by randomly selecting 29.926 sequences (minimum number of reads in a sample) in each sample, in order to obtain the same number of reads per sample.

Using the FastTree program, the phylogenetic tree was created, followed by computing Chao1 and Shannon indexes as measures of alpha diversity. Furthermore, (un)weighted and generalized UniFrac metrics, and Bray Curtis index were computed to evaluate between-sample diversity (beta diversity).

Statistical analysis

All statistical analyses were performed using QIIME version 1.9.1²⁷ or R version 3.4.2 (R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria). Categorical patient characteristics are presented as proportions and differences between groups were tested using χ^2 or Fisher's exact test. Continuous characteristics are presented as mean and standard deviation (SD) or as median with first and third quartiles in case of skewness. Differences between groups were tested using the independent t-test or the Mann-Whitney U test, depending on the normality of the distribution.

Alpha diversity data are expressed by means of Chao1 index, observed species and effective Shannon index ($\exp[\text{Shannon index}]$), all being normally distributed. In order to evaluate the within-subject variability in alpha diversity over time, inter-item (Pearson) correlations between the different, consecutively collected fecal samples, as well as intra-class correlations (ICC) were calculated for IBS and healthy subjects separately. Data from subjects that collected at least five consecutive samples were included in these analyses. The Bray-Curtis distance and generalized UniFrac (*i.e.* beta diversity) were used to quantify the (dis)similarity in microbial community structure between samples. As the taxonomical abundance data were not normally distributed, non-parametric tests were used for analyses on these data. Principal coordinates analyses (PCoA) were performed on beta diversity indices to evaluate possible clustering of the microbial community structure. A Mann Whitney-U test was used to compare the average within-subject diversity distance between groups (IBS patients vs. healthy subjects).

In order to evaluate the correlation between stool consistency and the microbiota, constrained redundancy analysis (RDA) was carried out, using routines from R package "vegan"²⁸. For this purpose, the zeros from the count data (summarized on family level) were imputed using R package "zCompositions"²⁹ and data were clr (centered log ratio) transformed. In addition, two-level mixed-effects linear regression models (level 1: consecutive stool samples; level 2: subjects) were used to examine the association between

stool consistency and microbiota using all longitudinal measurements. Separate models were used for different measures of microbial diversity and composition, with alpha-diversity (*i.e.* Chao1-index, observed species and effective Shannon index) and clr transformed taxonomical abundance data (family level), respectively, as the dependent variables. The two-way interaction term “stool consistency*time” and stool consistency were used as independent variables and a random intercept was chosen to correct for clustering of multiple measurements within each participant. A p-value of ≤ 0.05 was considered statistically significant.

In all analyses, stool consistency was primarily based on percentage of fecal dry weight, since this is a more objective measure of stool consistency. Additional analyses using BSS scores were performed to assess whether results were different between the two measures of stool consistency.

Results

Study population

Twelve IBS patients and 12 healthy subjects were included in this study; demographic characteristics of the study population are shown in Table 4.1. In both groups, 10 subjects were female and mean age was comparable between the groups. One IBS-C patient collected only two samples during the study period, in line with her IBS-subtype, for one healthy subject four consecutive samples were available, while all other subjects (*i.e.* IBS and healthy) collected at least five samples over the seven-day period (see Table 4.1). Three subjects provided eight samples, meaning that they collected a second sample on a day on which they reported diarrhea. Samples were numbered consecutively, regardless of the day on which they were collected. In total, 71 and 78 samples were available for the IBS and healthy group, respectively.

The day-to-day variability in stool consistency, measured by fecal dry weight percentage, was found to be high in the IBS group (ICC of 0.223) and moderate (ICC of 0.622) in the healthy population. Bristol Stool Scale scores showed high variability from day to day in both groups (ICC for IBS of 0.397; ICC for healthy of 0.276). This variability in stool consistency is illustrated in Supplementary Figure S4.1, by depicting fecal dry weight percentage and BSS scores for each sample and per individual.

The predominant phylum in both groups was *Firmicutes* with 81.88% in IBS and 79.46% HC. The composition of the microbiota on the phylum level, separately for IBS and healthy subjects, is represented in Supplementary Figure S4.2.

Table 4.1 Demographic characteristics of study population.

	Healthy (n=12)	IBS (n=12)
Female sex , n (%)	10 (83.3)	10 (83.3)
Age in years , median [Q1-Q3]	36.78 [30.09 – 47.19]	45.0 [35.37 – 48.56]
BMI in kg/m² , median [Q1-Q3]	21.99 [21.80 – 23.25]	23.65 [22.65 – 24.57]
Current smoking , n (%)	0 (0)	4 (33.3)
Current alcohol use (<15 units/week) , n (%)	6 (50)	7 (58.3)
Number of samples , n (%)		
1	-	-
2	-	1 (8.3)
3	-	-
4	1 (8.3)	-
5	-	4 (33.3)
6	4 (33.3)	2 (16.7)
7	6 (50.0)	3 (25.0)
8	1 (8.3)	2 (16.7)
GSRS , median [Q1-Q3]		
Abdominal Pain	1.67 [1.08-2.25]	3.33 [2.00-4.67] [§]
Regurgitation Syndrome	1.00 [1.00-1.00]	2.50 [1.00-3.50] [#]
Diarrhea Syndrome	1.00 [1.00-1.33]	3.33 [1.00-1.33] [§]
Indigestion Syndrome	1.75 [1.50-2.50]	4.75 [3.75-5.13] [§]
Constipation Syndrome	1.67 [1.00-1.92]	3.33 [2.33-4.67] [§]
Medication use , n (%)		
PPI	-	3 (25)
NSAID	-	-
Prokinetic	-	-
Spasmolytic	-	2 (16.7)
Laxative	-	1 (8.3)
Antidiarrheal	-	-
Antibiotic	-	-
Probiotic	-	-
Prebiotic	-	-
Other	(58.3)	7 (58.3)

Differences tested using Mann-Whitney U test for continuous data and χ^2 - or Fisher's exact test for categorical data. Significances are shown for IBS versus healthy. [#] $P < 0.01$; [§] $P < 0.001$.

Day-to-day variability of the microbiota

For both IBS and healthy subjects, microbial richness was found to show high correlations between all subsequent samples, demonstrating low within-subject variability from day to day. Inter-item correlations between different samples were all higher than 0.800. Following, a high degree of agreement in microbial

richness was found between the different samples of one subject, presented by single measure ICCs of at least 0.893. Inter-item matrices and ICCs for Chao1 index, for IBS patients and healthy subjects, are shown in Table 4.2. Similar results were found for observed species, and correlations were moderate for effective Shannon index (Supplementary Table S4.1A and S4.1B).

A PCoA demonstrated clustering of the microbial community structure per individual (Figure 4.1A), suggesting that the dissimilarity in microbiota between consecutive samples is larger between than within subjects. No clear separation between IBS and healthy subjects was demonstrated (Figure 4.1A) based upon the microbial community structure. Moreover, no significant difference was found in the average within-subject beta diversity between IBS and healthy subjects, indicating that the day-to-day variation in overall microbial community structure is similar between IBS and healthy subjects (Figure 4.1B). Comparable results were found for the generalized UniFrac as a measure of beta diversity (Supplementary Figure S4.3A and S4.3B).

Table 4.2 Inter-item correlations (Pearson correlations) between Chao1 index of consecutive samples, for healthy subjects and IBS patients separately.

Chao1 index	Healthy subjects				IBS patients			
	Intraclass Correlations Coefficient = 0.940				Intraclass Correlations Coefficient = 0.893			
	Sample 1	Sample 2	Sample 3	Sample 4	Sample 1	Sample 2	Sample 3	Sample 4
Sample 2	0.967				0.923			
Sample 3	0.866	0.922			0.870	0.882		
Sample 4	0.972	0.954	0.907		0.853	0.910	0.908	
Sample 5	0.946	0.989	0.936	0.925	0.916	0.895	0.916	0.915

Association between microbiota and stool consistency

A linear mixed-effects model with the Chao1 index as the dependent variable demonstrated no significant effect of the two-way interaction “fecal dry weight*time” on microbial richness (B: 0.030, SE: 0.072, $P=0.676$), indicating that the association between stool consistency and microbial richness is not different between subsequent time points. After removal of this term from the model, stool consistency was found to be a significant predictor of microbial richness (B: 1.231, SE: 0.200, $P<0.001$) (Table 4.3), showing an overall significant association between stool consistency and the microbiota, but independent of time. Subsequent linear mixed-effects models with observed species and the effective Shannon index as dependent variables, respectively, showed similar results (Supplementary Table S4.2A and S4.2B). In addition, models using BSS as

a measure of stool consistency (*i.e.* instead of fecal dry weight percentage) showed similar results (Table 4.3, Table 4.2A and Table 4.2B).

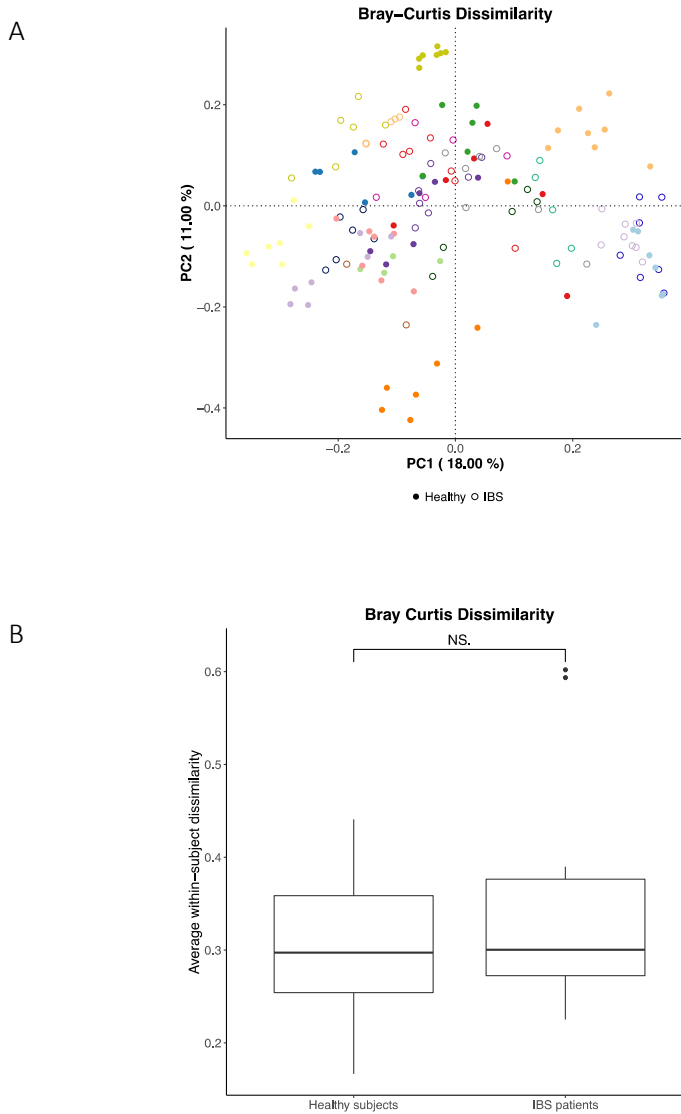


Figure 4.1 **A)** Bray Curtis PCoA plot per individual (different colors) and for healthy subjects vs. IBS patients (see figure annotations). **B)** Average within-subject beta diversity (Bray-Curtis Dissimilarity) for healthy subjects vs. IBS patients, significance tested using Mann Whitney U test, $P=0.71$.

Table 4.3 Results of linear mixed-effects models (with random intercept, fixed slopes, and scaled identity covariance structure). Regression coefficient indicates the direction and strength of the association between the predictor and dependent variable.

Dependent variable	Predictor	Regression coefficient [95%-CI]	SE	P-value
Chao1 index	Fecal dry weight*time	0.030 [-0.113; 0.173]	0.072	0.676
	Fecal dry weight ¹	1.231 [0.835;1.628]	0.200	<0.001
	BSS*time	0.135 [-0.710; 0.980]	0.426	0.752
	BSS ¹	-2.518 [-4.728; -0.308]	1.116	0.026
	Abdominal pain*time	-0.034 [-0.436; 0.368]	0.203	0.869
	Abdominal pain ¹	0.473 [-1.188; 2.135]	0.839	0.574
	Abdominal bloating*time	0.471 [-0.019; 0.962]	0.248	0.059
	Abdominal bloating ¹	0.182 [-1.572; 1.937]	0.886	0.837

¹ Insignificant interaction terms, respectively, “fecal dry weight*time”, “BSS*time” “abdominal pain*time”, and “abdominal bloating*time” were removed from the models. BSS: Bristol Stool Scale. SE: standard error.

In addition, a redundancy analysis showed a significant association between stool consistency and microbial composition on the family level mainly driven by *Christensenellaceae*, *Enterobacteriaceae*, *Verrucomicrobiaceae*, *Methanobacteriaceae*, and *Veillonellaceae* (Figure 4.2). In order to take into account within-subject day-to-day variability in these associations between specific bacterial groups and stool consistency, linear mixed-effects models with the respective bacterial groups as the dependent variable were performed. For *Christensenellaceae*, *Enterobacteriaceae*, *Verrucomicrobiaceae*, and *Methanobacteriaceae*, stool consistency was indeed a significant predictor (*i.e.* higher abundance in firmer stools), but not for *Veillonellaceae*.

Association between microbiota and GI symptoms

Within the IBS population, linear mixed-effects models showed that the association between the interaction term “abdominal pain*time” and Chao1 index was not significant (B: -0.034, SE: 0.203, *P*=0.869). Furthermore, no significant association was found between abdominal pain and Chao1 index (B: 0.473, SE: 0.839, *P*=0.574), after removal of the two-way interaction term from the model (Table 4.3). In addition, in models assessing abdominal bloating as an independent variable (including both the interaction “abdominal bloating*time” and abdominal bloating as fixed effects), also no significant association between abdominal bloating and fecal microbial richness was found (Table 4.3). Similar results were found for observed species and the effective Shannon index as dependent variables (Supplementary Table 4.2A and 4.2B).

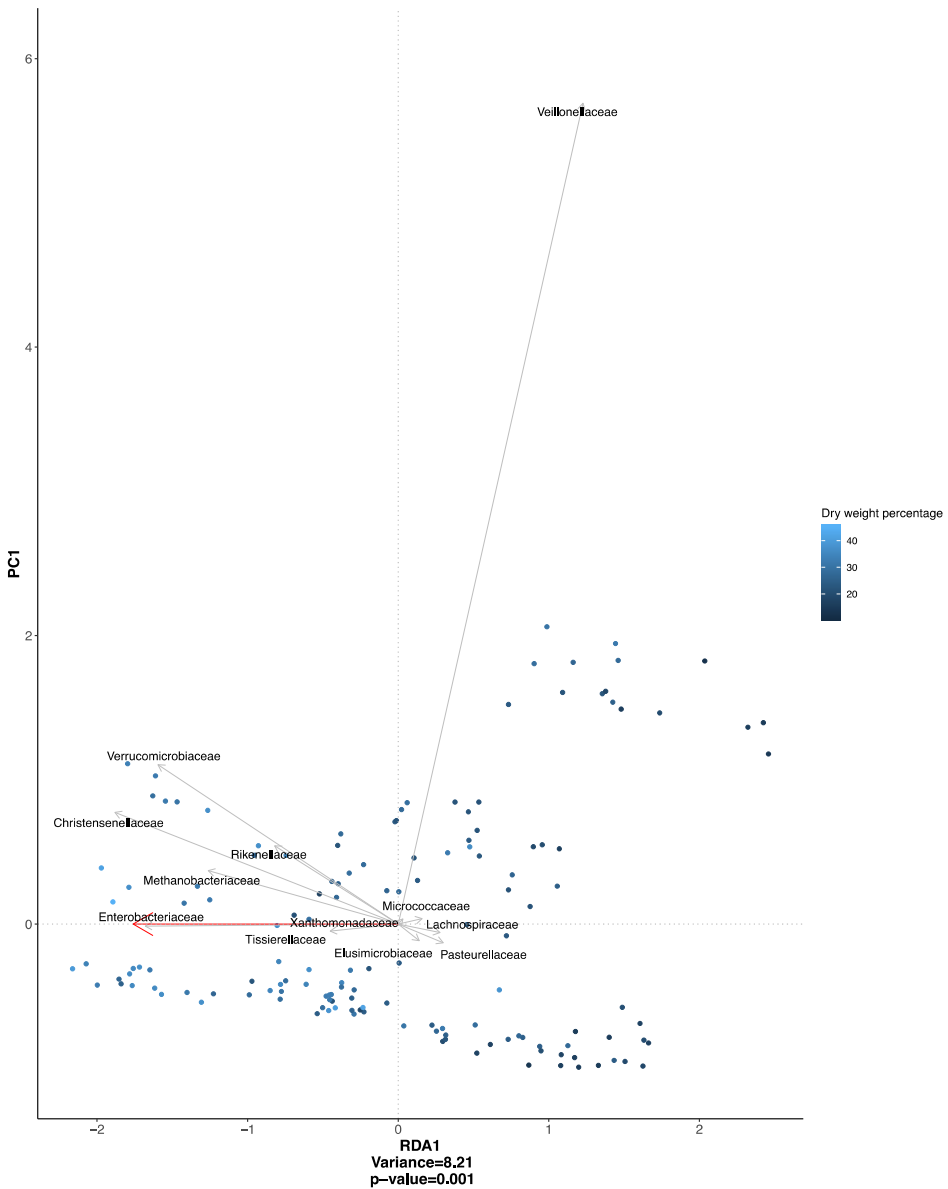


Figure 4.2 Redundancy analysis (RDA) plot based on clr transformed abundancies, and constrained on stool consistency (dry weight percentage), with individual variation partialled-out. Significant association between stool consistency and microbial composition ($P=0.001$), mainly driven by bacterial families depicted in the figure.

Moreover, redundancy analysis again showed clear clustering per individual, with no significant association between abdominal pain and microbiota ($P=0.545$), nor between abdominal bloating and microbiota ($P=0.855$) (Figure 4.3). Since these redundancy analyses showed no significant association between abdominal symptoms and microbial composition, no subsequent analyses (*i.e.* linear mixed-effects models) on specific associations between different taxa and GI symptoms were performed.

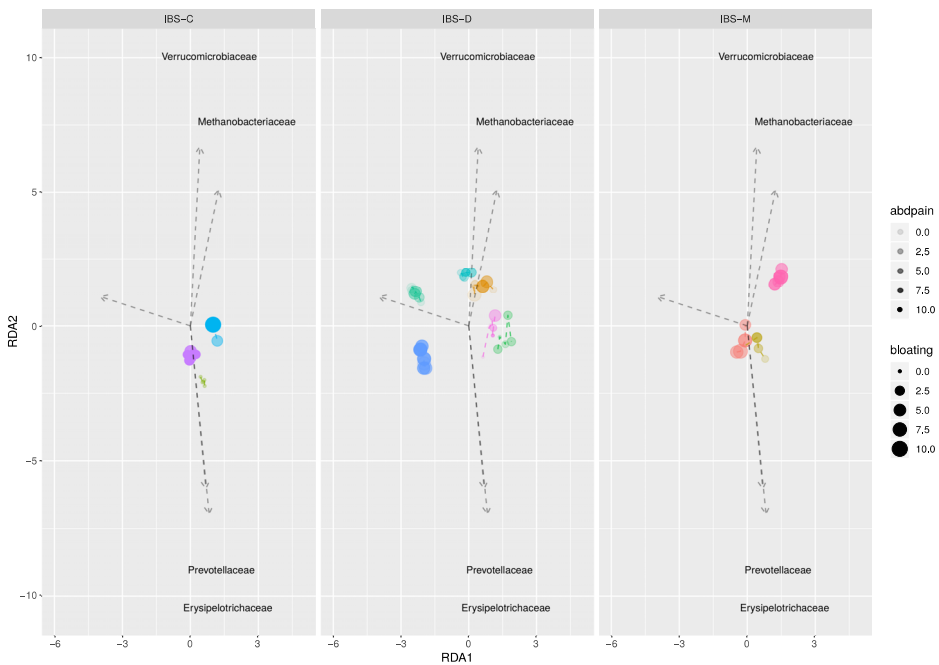


Figure 4.3 Redundancy analysis (RDA) plot based on clr transformed abundancies, and constrained on individual, abdominal pain, and abdominal bloating. Each dot represents an individual sample; IBS patients are depicted by different colors. No significant association between both GI symptoms ($P=0.368$ for abdominal pain; $P=0.521$ for bloating) and microbial composition.

Discussion

Previous findings of strong correlations between stool consistency and the fecal microbiota pointed towards the importance of considering measures of stool consistency and/or gut transit time as confounding factors in microbiota analyses.^{1,3,4} This was confirmed by the current study. However, we

demonstrated that the within-subject variation in the fecal microbiota composition between consecutive samples over seven days was limited, in both IBS and healthy subjects, even in case of fluctuating stool consistency or GI symptoms.

Stool consistency is known to vary within individuals over time¹³, especially in subjects with GI disorders, suggesting that the fecal microbiota might also change over short periods of time, related to the changes in stool form. Therefore, this study focused on the microbial stability from day to day and the association between this (in)stability and stool consistency. Both IBS patients and healthy subjects were evaluated, since IBS is known to show high within-subject variability in stool consistency from day to day. In summary, we here demonstrate low within-subject variability in fecal microbiota from day to day, over a period of seven consecutive days. Overall variability in the microbial community structure was found to be mainly driven by between-subject variability, with rather small within-subject changes between consecutive samples. Although significant associations between stool consistency and microbial richness and composition were found, no important significant effect of time on this association could be demonstrated. Altogether, this indicates that, regardless of fluctuations in stool consistency, the variation in fecal microbiota composition within subjects over a one-week period, in both IBS patients and healthy subjects, is rather limited when compared to variations between subjects.

Results from earlier studies on temporal stability of the microbial composition are contradictory. Instability in both IBS and healthy subjects has been demonstrated, with possibly a more instable microbiota in IBS patients, assessed over three to six months.^{13,30} Others found low variability between microbial profiles of fecal samples in healthy subjects with intervals of 14 days to several months, though with intra-individual temporal variations for some specific microbial genera.³¹⁻³⁵ Furthermore, inter-individual differences have been demonstrated particularly important when it comes to microbial variability, as compared to intra-individual differences.³⁴⁻³⁶ According to our results, the within-subject variation in microbial profiles is limited in both IBS and healthy subjects, over a period of seven days.

In the current study, we used fecal dry weight percentage as well as patient-reported Bristol Stool Scale scores as measures of fecal stool consistency. From day to day, moderate-to-high levels of variability in both measures of stool consistency in the IBS and healthy populations were found. This confirms that if any day-to-day changes in the microbial composition would be related to

variability in stool consistency, our data should be suitable to detect this. Previous findings on a correlation between fecal microbial richness and stool consistency demonstrated that looser stools, using the Bristol Stool Scale, were associated with lower species richness.¹ Our findings, both using fecal dry weight percentage and the Bristol Stool Scale, confirm this association, but in addition indicate that there is no major within-subject shifting in microbial richness and composition from day to day, associated with short-term changes in stool consistency in our study population. Also, the previously demonstrated correlation between specific genera abundances and stool consistency was not dependent of time in our study. An increase in abundance of members of the *Methanobacteriaceae* and *Verrucomicrobiaceae*, and the *Enterobacteriaceae* and *Christensenellaceae* families in association with firmer stools has been described previously.^{1,2,4} However, linear mixed-effects model analysis could not confirm an association between *Veillonellaceae* and stool consistency, suggesting that this association is mainly driven by between-subject differences and becomes less important when correcting for within-subject variation. In addition to these previous findings, we could not demonstrate important within-subject shifting in these bacterial groups in relation to fluctuations in stool consistency.

We also evaluated the impact of the large day-to-day variability in symptom scores as often found in IBS. No significant interaction was found from day to day between abdominal symptoms (*i.e.* abdominal pain and abdominal bloating) and microbial richness as well as overall microbial community structure, in the IBS population. Previous studies have suggested a link between the microbiota and clinical characteristics of IBS, such as abdominal symptoms, but results are inconsistent^{7,11,12}. Our results indicate that there are no daily shifts in the gut microbiota related to daily individual symptom patterns, when focusing on abdominal pain and abdominal bloating as the key symptoms of IBS. Previous studies on the stability of the gut microbiota have shown rapid changes of microbial composition over the course of several days, for example after considerable dietary changes.^{37,38} This demonstrates that the gut microbial community can change rapidly and underlines the importance of considering the time factor when evaluating potential confounders. In the current study, participants maintained their habitual diet and were not allowed to introduce major changes in dietary habits during the study period. Therefore, we do not expect this to be a factor of importance in our results. The fact that we did not find a(n) (within-subject) effect of time here suggests that the previously demonstrated association between stool consistency and microbiota is to a

large extent reflected by between-subject differences. Presumably, changes in microbiota are related to changes in stool consistency on the longer term, but the microbial community does not change rapidly over short periods of time in relation to variability in stool consistency. An explanation for this could be that short-term perturbations, reflected by intra-individual changes in stool consistency, do not immediately result in altered microbiota due to the resilience of the human gut microbial ecosystem.³⁹ On the longer term, persistent perturbations might ultimately result in changes of the gut microbiota as well.

This is the first study to examine a short-term within-subject association between stool consistency and the gut microbiota, using repeated fecal sampling over one week. Both IBS and healthy subjects were evaluated in order to capture highly fluctuating as well as more stable day-to-day patterns of stool consistency. Fecal dry weight percentage was used as an objective measure of stool consistency, which might have an advantage over the commonly used Bristol Stool Scale, since the latter was developed as a surrogate marker of whole-gut transit time and is subject to inter-individual differences in interpretation.¹⁹ A more detailed recommendation on the use of fecal dry weight percentage to correct for stool consistency in future (microbiota) studies was recently published elsewhere.⁴⁰ However, results were similar when using BSS scores instead of fecal dry weight percentage, confirming that our results are suitable to make comparisons with previous studies.

The results on the day-to-day stability of microbial richness were slightly less pronounced for the effective Shannon index when compared to Chao1 index and observed species. The Shannon index is a measure of biodiversity, taking into account the evenness next to the richness, and is therefore more affected by shifts in abundance of microbial taxa. Nevertheless, we could not demonstrate a significant effect of time on the association between stool consistency and microbial richness for any of the three measures.

Slight baseline differences in smoking habits and the current use of medication (in particular proton pump inhibitors) were found between IBS patients and healthy subjects, although those were not statistically significant. Both characteristics could potentially affect the microbiota.^{41,42} However, since we found no important dissimilarities in microbial composition between both groups, we are confident that these differences did not influence the results as presented. Furthermore, the current study was not designed and powered to draw any conclusions on differences in the microbial composition between IBS and healthy subjects.

In conclusion, this study supports an association between stool consistency and the fecal microbiota, but the overall microbial composition was not significantly related to day-to-day fluctuations in stool consistency. This consolidates the importance of taking into account stool consistency in future gut microbiota research, though confirms the validity of single fecal sampling in the context of short-term studies. As between-subject variation is substantially larger than within-subject variation from day to day over seven days, a single fecal sample appears representative for between-subject comparisons over a short period of time. However, this does not imply that there is no need for longitudinal studies with repeated fecal sampling over longer time-periods to gain further insight into the dynamics of the microbiota in relation to disease pathophysiology and disease course. The current results merely suggest that a single sample is sufficient to gain a reliable picture of the microbiota within a short time-period, even within subjects with large fluctuations in stool consistency.

References

1. Vandeputte D, Falony G, Vieira-Silva S, Tito RY, Joossens M, Raes J. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut*. 2016;65(1):57-62.
2. Tigchelaar EF, Bonder MJ, Jankipersadsing SA, Fu J, Wijmenga C, Zhernakova A. Gut microbiota composition associated with stool consistency. *Gut*. 2016;65(3):540-542.
3. Falony G, Joossens M, Vieira-Silva S, et al. Population-level analysis of gut microbiome variation. *Science*. 2016;352(6285):560-564.
4. Vandeputte D, Kathagen G, D'Hoe K, et al. Quantitative microbiome profiling links gut community variation to microbial load. *Nature*. 2017;551(7681):507-511.
5. Longstreth GF, Thompson WG, Chey WD, Houghton LA, Mearin F, Spiller RC. Functional bowel disorders. *Gastroenterology*. 2006;130(5):1480-1491.
6. Salonen A, de Vos WM, Palva A. Gastrointestinal microbiota in irritable bowel syndrome: present state and perspectives. *Microbiology (Reading, England)*. 2010;156(Pt 11): 3205-3215.
7. Rajilic-Stojanovic M, Biagi E, Heilig HG, et al. Global and deep molecular analysis of microbiota signatures in fecal samples from patients with irritable bowel syndrome. *Gastroenterology*. 2011;141(5):1792-1801.
8. Jeffery IB, Quigley EM, Ohman L, Simren M, O'Toole PW. The microbiota link to irritable bowel syndrome: an emerging story. *Gut Microbes*. 2012;3(6):572-576.
9. Malinen E, Rinttila T, Kajander K, et al. Analysis of the fecal microbiota of irritable bowel syndrome patients and healthy controls with real-time PCR. *Am J Gastroenterol*. 2005;100(2):373-382.
10. Kassinen A, Krogius-Kurikka L, Makivuokko H, et al. The fecal microbiota of irritable bowel syndrome patients differs significantly from that of healthy subjects. *Gastroenterology*. 2007;133(1):24-33.
11. Jeffery IB, O'Toole PW, Ohman L, et al. An irritable bowel syndrome subtype defined by species-specific alterations in faecal microbiota. *Gut*. 2012;61(7):997-1006.
12. Malinen E, Krogius-Kurikka L, Lyra A, et al. Association of symptoms with gastrointestinal microbiota in irritable bowel syndrome. *World J Gastroenterol*. 2010;16(36):4532-4540.
13. Matto J, Maunukela L, Kajander K, et al. Composition and temporal stability of gastrointestinal microbiota in irritable bowel syndrome--a longitudinal study in IBS and control subjects. *FEMS Immunol Med Microbiol*. 2005;43(2):213-222.
14. Lyra A, Rinttila T, Nikkila J, et al. Diarrhoea-predominant irritable bowel syndrome distinguishable by 16S rRNA gene phylotype quantification. *World J Gastroenterol*. 2009;15(47):5936-5945.
15. Durban A, Abellan JJ, Jimenez-Hernandez N, et al. Instability of the faecal microbiota in diarrhoea-predominant irritable bowel syndrome. *FEMS Microbiol Ecol*. 2013;86(3): 581-589.
16. Foundation R. Appendix A: Rome III Diagnostic Criteria for FGIDs.885-897.
17. Drossman DA. The functional gastrointestinal disorders and the Rome III process. *Gastroenterology*. 2006;130(5):1377-1390.
18. Ford AC, Bercik P, Morgan DG, Bolino C, Pintos-Sanchez MI, Moayyedi P. Validation of the Rome III criteria for the diagnosis of irritable bowel syndrome in secondary care. *Gastroenterology*. 2013;145(6):1262-1270.e1261.
19. Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit time. *Scand J Gastroenterol*. 1997;32(9):920-924.

20. U.S. Department of Health and Human Services FaDA, Center for Drug Evaluation and Research (CDER). Guidance for Industry Irritable Bowel Syndrome - Clinical Evaluation of Drugs for Treatment. 2012.
21. Salonen A, Nikkila J, Jalanka-Tuovinen J, et al. Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis. *J Microbiol Methods*. 2010;81(2):127-134.
22. J. D, Ehrlich SD, Levenez F, et al. IHMS_SOP 06 V1: Standard operating procedure for fecal samples DNA extraction, Protocol Q. . International Human Microbiome Standards 2015.
23. Pascal V, Pozuelo M, Borruel N, et al. A microbial signature for Crohn's disease. *Gut*. 2017;66(5):813-822.
24. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics (Oxford, England)*. 2010;26(19):2460-2461.
25. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics (Oxford, England)*. 2011;27(16):2194-2200.
26. Navas-Molina JA, Peralta-Sanchez JM, Gonzalez A, et al. Advancing our understanding of the human microbiome using QIIME. *MethodsEnzymol*. 2013;531:371-444.
27. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nature methods*. 2010;7(5):335-336.
28. Oksanen J, Blanchet FG, Friendly M, et al. Vegan: Community Ecology Package. R package version 2.5-1. 2018.
29. Palarea-Albaladejo J, Martin-Fernandez JA. zCompositions — R package for multivariate imputation of left-censored data under a compositional approach. *Chemom Intell Lab Syst*. 2015;143:85-96.
30. Maukonen J, Satokari R, Matto J, Soderlund H, Mattila-Sandholm T, Saarela M. Prevalence and temporal stability of selected clostridial groups in irritable bowel syndrome in relation to predominant faecal bacteria. *J Med Microbiol*. 2006;55(Pt 5):625-633.
31. Vanhoutte T, Huys G, Brandt E, Swings J. Temporal stability analysis of the microbiota in human feces by denaturing gradient gel electrophoresis using universal and group-specific 16S rRNA gene primers. *FEMS Microbiol Ecol*. 2004;48(3):437-446.
32. Zoetendal EG, Akkermans AD, De Vos WM. Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals stable and host-specific communities of active bacteria. *Appl Environ Microbiol*. 1998;64(10):3854-3859.
33. Zoetendal EG, Akkermans AD, Akkermans - van Vliet WM, de Visser AGM, de Vos WM. The host genotype affects the bacterial community in the human gastrointestinal tract. *Microbial Ecology in Health and Disease*. 2001;13:129-134.
34. Jalanka-Tuovinen J, Salonen A, Nikkila J, et al. Intestinal microbiota in healthy adults: temporal analysis reveals individual and common core and relation to intestinal symptoms. *PLoS One*. 2011;6(7):e23035.
35. Rajilic-Stojanovic M, Heilig HG, Tims S, Zoetendal EG, de Vos WM. Long-term monitoring of the human intestinal microbiota composition. *Environ Microbiol*. 2012.
36. David LA, Materna AC, Friedman J, et al. Host lifestyle affects human microbiota on daily timescales. *Genome Biol*. 2014;15(7):R89.
37. David LA, Maurice CF, Carmody RN, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature*. 2014;505(7484):559-563.
38. Wu GD, Chen J, Hoffmann C, et al. Linking long-term dietary patterns with gut microbial enterotypes. *Science*. 2011;334(6052):105-108.
39. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. *Nature*. 2012;489(7415):220-230.

40. Vork L, Wilms E, Penders J, Jonkers D. Stool Consistency: Looking Beyond the Bristol Stool Form Scale. *J Neurogastroenterol Motil.* 2019;25(4):625.
41. Imhann F, Bonder MJ, Vich Vila A, et al. Proton pump inhibitors affect the gut microbiome. *Gut.* 2016;65(5):740-748.
42. Benjamin JL, Hedin CR, Koutsoumpas A, et al. Smokers with active Crohn's disease have a clinically relevant dysbiosis of the gastrointestinal microbiota. *Inflamm Bowel Dis.* 2012;18(6):1092-1100.

Supplemental material

Table S4.1A Inter-item correlations (Pearson correlations) between observed species of consecutive samples, for healthy subjects and IBS patients separately.

Observed species	Healthy subjects				IBS patients			
	Sample 1	Sample 2	Sample 3	Sample 4	Sample 1	Sample 2	Sample 3	Sample 4
	Intraclass Correlations Coefficient = 0.935				Intraclass Correlations Coefficient = 0.879			
Sample 2	0.965				0.886			
Sample 3	0.868	0.927			0.849	0.891		
Sample 4	0.971	0.943	0.904		0.818	0.908	0.946	
Sample 5	0.946	0.990	0.929	0.912	0.907	0.877	0.932	0.917

Table S4.1B Inter-item correlations (Pearson correlations) between effective Shannon index of consecutive samples, for healthy subjects and IBS patients separately.

Effective Shannon index	Healthy subjects				IBS patients			
	Sample 1	Sample 2	Sample 3	Sample 4	Sample 1	Sample 2	Sample 3	Sample 4
	Intraclass Correlations Coefficient = 0.612				Intraclass Correlations Coefficient = 0.519			
Sample 2	0.633				0.020			
Sample 3	0.598	0.913			0.312	0.872		
Sample 4	0.780	0.516	0.712		0.056	0.746	0.611	
Sample 5	0.432	0.831	0.613	0.173	0.537	0.616	0.866	0.386

Table S4.2A Results from linear mixed-effects models (with random intercept, fixed slopes, and scaled identity covariance structure). Regression coefficient indicates the direction and strength of the association between the predictor and dependent variable.

Dependent variable	Predictor	Regression coefficient [95%-CI]	SE	P-value
Observed species	Fecal dry weight*time	0.028 [-0.095; 0.151]	0.062	0.651
	Fecal dry weight ¹	1.066 [0.725; 1.407]	0.172	<0.001
	BSS*time	-0.050 [-0.786; 0.687]	0.372	0.894
	BSS ¹	-2.190 [-4.116; -0.264]	0.973	0.026
	Abdominal pain*time	-0.074 [-0.420; 0.273]	0.175	0.674
	Abdominal pain ¹	0.626 [-0.806; 2.058]	0.723	0.388
	Abdominal bloating*time	0.418 [-0.005; 0.840]	0.213	0.053
	Abdominal bloating ¹	0.507 [-1.004; 2.019]	0.763	0.508

¹ Insignificant interaction terms, respectively, "fecal dry weight*time", "BSS*time" "abdominal pain*time", and "abdominal bloating*time" were removed from the models. BSS: Bristol Stool Scale. SE: standard error.

Table S4.2B Results from linear mixed-effects models (with random intercept, fixed slopes, and scaled identity covariance structure). Regression coefficient indicates the direction and strength of the association between the predictor and dependent variable.

Dependent variable	Predictor	Regression coefficient [95%-CI]	SE	p-value
Effective Shannon index	Fecal dry weight *time	-0.289 [-0.839; 0.261]	0.278	0.300
	Fecal dry weight ¹	3.159 [1.720; 4.598]	0.727	<0.001
	BSS*time	-0.2085 [-5.324; 1.153]	1.635	0.205
	BSS ¹	-2.921 [-11.143; 5.302]	4.152	0.483
	Abdominal pain*time	-0.537 [-1.961; 0.886]	0.719	0.456
	Abdominal pain ¹	-4.283 [-9.694; 1.128]	2.733	0.120
	Abdominal bloating*time	0.713 [-1.062; 2.488]	0.896	0.428
	Abdominal bloating ¹	-2.459 [-8.280; 3.364]	2.940	0.405

¹ Insignificant interaction terms, respectively, “fecal dry weight*time”, “BSS*time” “abdominal pain*time”, and “abdominal bloating*time” were removed from the models. BSS: Bristol Stool Scale. SE: standard error.

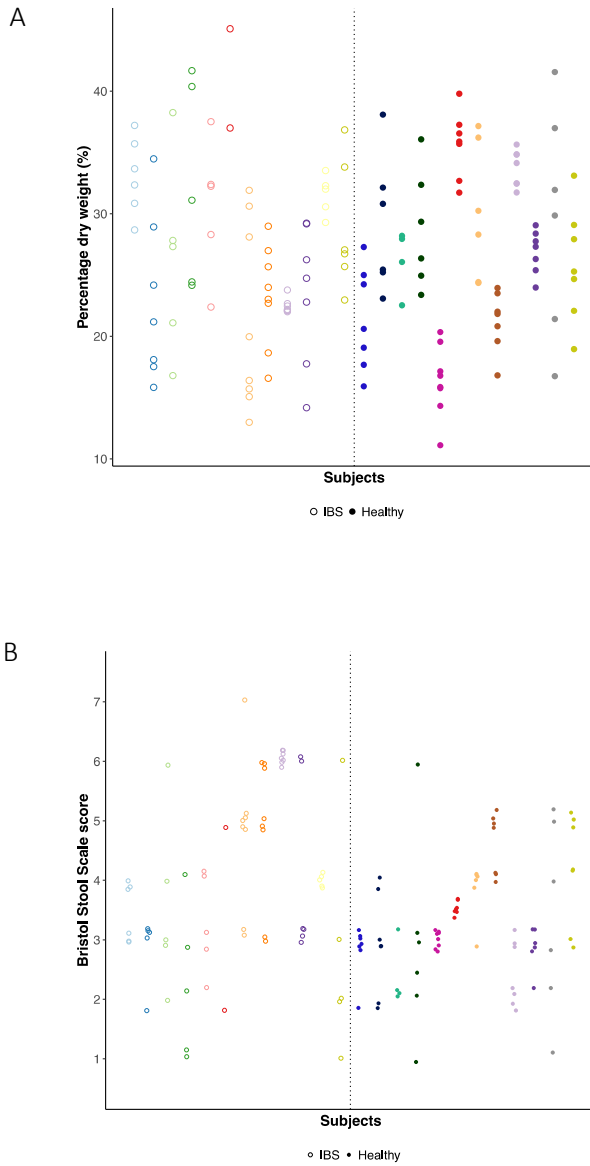


Figure S4.1 Stool consistency for each sample and separately for all subjects (i.e. each color represents a subject), measured by fecal dry weight percentage (A) and Bristol Stool Scale (B).

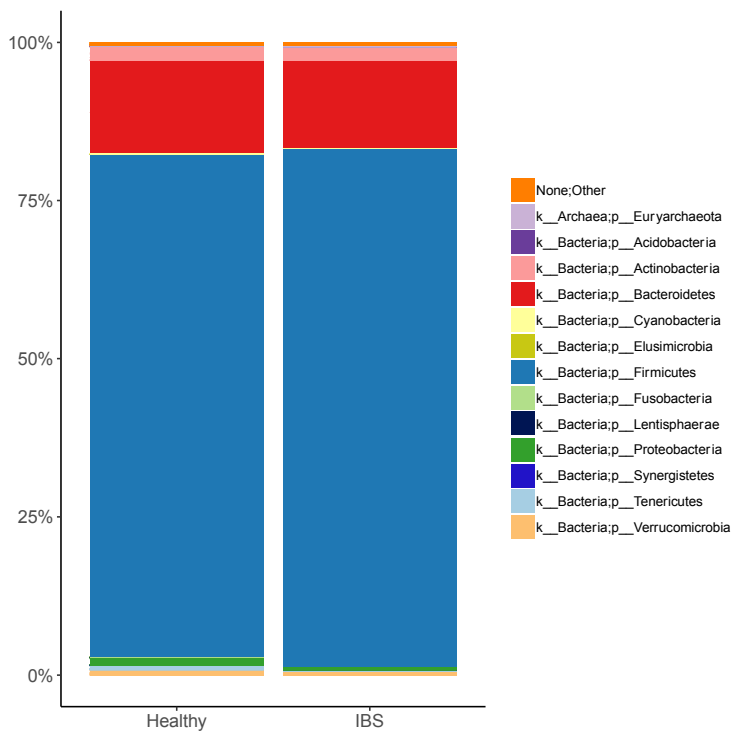


Figure S4.2 Overview of the microbial composition, separately for IBS and healthy subjects, based on relative abundances on the phylum level.

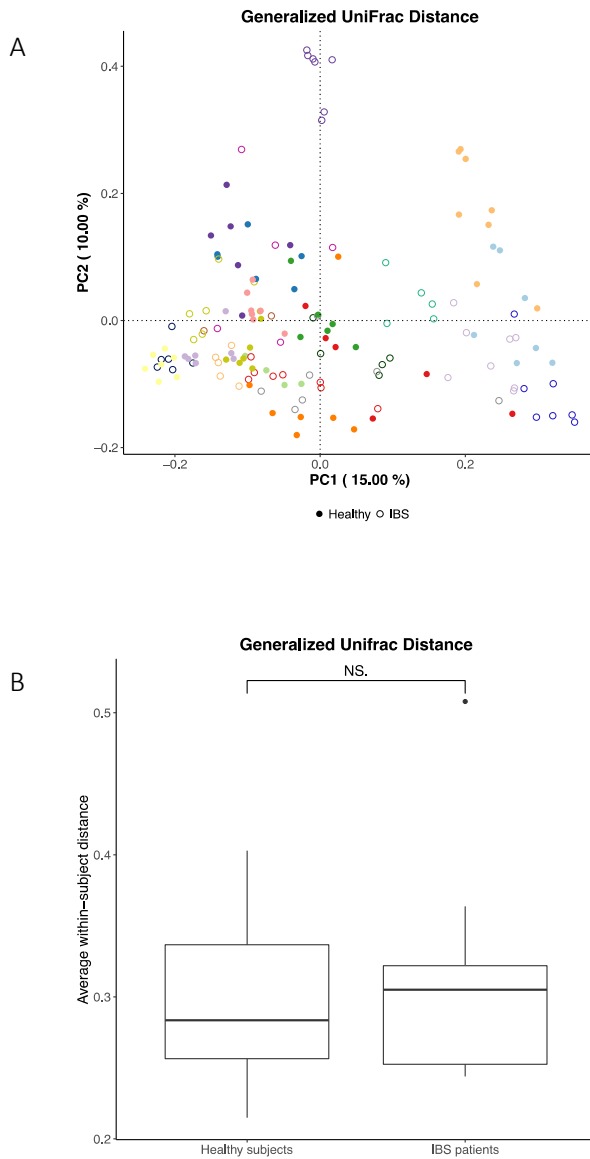


Figure S4.3 **A)** Generalized UniFrac PCoA plot per individual (different colors) and for healthy subjects vs. IBS patients (see figure annotations). **B)** Average within-subject beta diversity (generalized UniFrac) for healthy subjects vs. IBS patients, significance tested using Mann Whitney U test, $P=0.80$.



PART II
THE EXPERIENCE SAMPLING METHOD
DEVELOPMENT AND VALIDATION

Chapter 5

The Experience Sampling Method - A new digital tool
for momentary symptom assessment in IBS:
An exploratory study

Zlatan Mujagic, Lisa Vork*, Carsten Leue*, Richel Lousberg, Daisy MAE Jonkers,
Daniel Keszthelyi, Martine AM Hesselink, Thirza JC van Schagen, Jim van Os,
Ad AM Masclee, Joanna W Kruimel

* Both authors contributed equally to the manuscript

Neurogastroenterology & Motility. 2015;27(9):1295-302.

Abstract

INTRODUCTION Retrospective questionnaires are frequently used for symptom assessment in Irritable Bowel Syndrome (IBS) patients, but are influenced by recall bias and circumstantial and psychological factors. These limitations may be overcome by random, repeated, momentary assessment during the day, using electronic Experience Sampling Methodology (ESM). Therefore, we compared symptom assessment by ESM to retrospective paper questionnaires in IBS patients.

METHODS Twenty-six IBS patients (Rome III) were included, of which 16 were diagnosed with panic disorder (DSM-IV-TR). Patients scored symptoms using end-of-day diaries during 14 days and the gastrointestinal symptom rating scale (GSRS) once. ESM was used on 7 consecutive days during the same time period.

RESULTS End-of-day diary abdominal pain scores were 0.4 (SE 0.1, $P<0.001$) point higher (on a 1-to-5-point scale) compared to corresponding ESM mean-scores in IBS patients. The difference was even more pronounced for upper abdominal pain scores assessed by the GSRS (4.77 ± 1.50) compared to ESM mean-scores (2.44 ± 1.30 , $P<0.001$), both on 1-to-7-point scale. For flatulence, comparable results were found. Nausea and belching scores showed small, but significant differences between end-of-day diary and ESM. All tested symptoms were scored higher on GSRS compared to ESM mean-scores ($P<0.01$). Affective comorbidity did not influence differences in pain reporting between methods.

CONCLUSIONS IBS patients report higher scores for abdominal pain in retrospective questionnaires compared to ESM, with a tendency to report peak rather than average pain scores. ESM can provide more insight in symptom course and potential triggers, and may lead to a better understanding of IBS symptomatology.

Introduction

A major hurdle for the development of new therapies in Irritable Bowel Syndrome (IBS) is the current lack of reliable instruments to measure symptoms over time and quantify therapy response. At present, retrospective self-reported paper questionnaires based on daily (*i.e.* by end-of-day diaries) or weekly symptom monitoring (*i.e.* by end-of-week questionnaires) are widely used as the standard method to assess symptoms in IBS patients.¹⁻⁴ The Food and Drugs Administration (FDA) recommends a 7-day mean of an 'end-of-day 11-point Numeric Rating Scale' as a primary endpoint for clinical trials assessing abdominal pain in IBS.⁵ In a recently published study, Lackner *et al.* demonstrated that (electronic) end-of-day diaries may more accurately quantify gastrointestinal (GI) symptoms in IBS patients compared to weekly symptom monitoring, in particular for average abdominal pain over a predefined period of time.⁴ However, it should be taken into account that all retrospective questionnaires (*e.g.* end-of-day or end-of-week symptom assessment) are based on autobiographical memory, thereby leading to recall bias, representing a reconstruction of specific moments, rather than a reliable reflection of symptoms over a time period.⁶ Memory of pain can be influenced by contextual, as well as by emotional and cognitive factors. This can affect perception and interpretation of symptoms, especially pain, and may result in state-congruent and ecological recall bias in retrospective questionnaires.⁷⁻⁹ Affective comorbidity such as anxiety disorders, are prevalent in IBS patients and the presence of psychological comorbidity may contribute to a distorted perception and memory of GI symptoms as well.^{10,11} Furthermore, the inability to record temporal dynamics of symptoms renders retrospective questionnaires less suitable for measurement of symptoms fluctuating over time. Finally, it has been shown that paper diaries are subject to low compliance rates, as low as approximately 10%, due to backfilling of questionnaires.¹² The Experience Sampling Method (ESM),¹³ also referred to as Ecological Momentary Assessment (EMA),¹⁴ may overcome the significant limitations of retrospective questionnaires.¹⁵ This assessment method is characterized by the following criteria: data are collected repeatedly and randomly during the day, for several days, using a digital device, in the subject's natural environment, with a focus on the subject's in-the-moment (*i.e.* momentary) physical and mental state and behaviour. Digital momentary assessment reduces recall and ecological bias, may increase compliance rates and can take into account emotional, cognitive and contextual factors impacting IBS symptom reporting in

general and abdominal pain in particular. One previous study by Weiland and colleagues used electronic momentary assessment in a group of 58 IBS patients to describe differences in symptom episodes between IBS subtypes. As secondary outcome they reported that abdominal pain scores were significantly higher in the traditional paper end-of-day diaries when compared to the electronic momentary assessment.¹⁶ More research is needed to confirm their observation for abdominal pain, but also to compare ESM to current standards (*i.e.* daily and weekly retrospective symptom reporting) with regard to other GI symptoms, as well as to investigate possible patient characteristics, such as anxiety, that may affect pain memory and therefore the differences between momentary and retrospective symptom assessment.

The aims of the present study were to explore possible advantages of electronic ESM over paper retrospective questionnaires for GI symptom measurement in IBS patients, by i) comparing abdominal pain and other GI symptom scores assessed by a paper based end-of-day diary to corresponding scores assessed by ESM, by ii) comparing the GSRS, which includes a recall period of one week, to ESM, and by iii) investigating patient characteristics, that may affect memory of pain and may therefore contribute to a possible difference in scores between the momentary and retrospective symptom assessment.

We hypothesized that a) IBS patients will report higher levels of abdominal pain, and possibly also other GI symptoms, in end-of-day diaries and the GSRS compared to ESM average scores, and b) that the presence of a comorbid panic disorder, more severe mental symptoms, and negative cognitions will influence this difference in symptom reporting between the methods.

Materials and methods

The present study is part of a larger cohort study on the phenotypical and genotypical characterization of IBS (Maastricht IBS cohort) and of a 'Randomized Controlled Trial of Escitalopram versus Placebo for Patients with Irritable Bowel Syndrome and Panic Disorder', of which the baseline data (*i.e.* before start of any intervention) were used for the present analysis. Study protocols were approved by the Maastricht University Medical Center+ (MUMC+) Ethics Committee, were in compliance with the revised Declaration of Helsinki (64th WMA General Assembly, Fortaleza, Brazil, 2013) and were

registered in the US National Library of Medicine (<http://www.clinicaltrials.gov>, NCT00775060 and NCT01551225, respectively).

Study participants

Subjects between 18 and 75 years of age with IBS, diagnosed by a gastroenterologist using the Rome III criteria,¹⁷ were recruited via the MUMC+ Gastroenterology-Hepatology outpatient clinic, which is a secondary and tertiary referral centre for patients with functional GI disorders. For the diagnosis of IBS, GI endoscopy with biopsies, abdominal imaging by ultrasonography or CT scan, and/or blood, breath, and faecal analysis were performed to exclude organic disease, when indicated by the physician. Patients with a history of abdominal surgery were excluded, with the exception of appendectomy, laparoscopic cholecystectomy or hysterectomy. The presence of co-morbid panic disorder was diagnosed by an experienced psychiatrist based on the DSM-IV-TR criteria, by using the Mini-International-Neuropsychiatric-Interview (M.I.N.I.).¹⁸ All subjects gave written informed consent prior to participation.

5

Paper questionnaires and ESM

Participants completed questionnaires on demographics, lifestyle factors, and use of medication. Furthermore, they completed scales indexing psychopathology and negative cognitions. The Hospital Anxiety and Depression Scale (HADS), consisting of 14 items clustered into two subscales, *i.e.* anxiety and depression scale, with a score of 0-21 per subscale, and a recall period of one week. Higher scores indicate higher symptom burden.^{19,20} The Cognitive Scale for Functional Bowel Disorders (CS-FBD), consisting of 31 items, scored on a 7-point Likert scale, and a total score of 31-217, with a recall period of one month, was used to assess dysfunctional cognitions related to symptoms in FBD. A higher score indicates more dysfunctional cognitions.²¹

GI symptoms were assessed over 14 consecutive days using an end-of-day diary. Complaints of abdominal discomfort, abdominal pain, nausea, bloating, belching, flatulence, diarrhoea, constipation, and overall symptom burden were scored on 5-point Likert scales.^{2,3,22}

Electronic ESM¹³ was used to assess real-time GI symptoms (*e.g.* abdominal pain, heartburn, nausea, belching, bloating, flatulence) in daily life, as well as momentary physical and mental experiences, activities, and contextual aspects, during 7 consecutive days of the same time period. ESM was started at a

random day during the first week of the 14-day end-of-day diary period. Subjects carried a digital device, the PsyMate (comparable to a mobile phone).²³ The device gives an auditory signal (a “beep”) at ten random time points per day. Participants were instructed to complete 60 short questions on the device following every signal, using a 7-point Likert scale. The completion of the question set took 4-5 minutes on each occasion. The same set of questions was repeated in the same order at every beep. As subjects are usually not able to respond after every beep (for example because of driving a car), completion of ≥ 6 of the 10 beep-questionnaires per day was considered as being compliant.²⁴ In addition, participants were instructed to complete the Gastrointestinal Symptom Rating Scale (GSRS), consisting of 15 items, scored on a 7-point Likert scale, with a recall period of one week, during the same time period.²⁵ Individual questions of both retrospective methods were compared to individual questions of ESM. Only GI symptoms assessed by ESM that corresponded to questions in the end-of-day diary and/or GSRS were included in the current analyses. The phrasing of the questions was similar, apart from the time frame explanation, and for the end-of-day diary with regard to the used scale, *i.e.* a 1-to-5 point scale for the end-of-day diary compared to a 1-to-7 point scale for the ESM and GSRS.

Data and statistical analysis

4D software was used to program PsyMate devices and extract the data. Statistical analyses were performed using IBM SPSS Statistics version 22.0 (IBM Statistics for Windows, Armonk, New York).

In order to check if completing the ESM questionnaires at multiple time points each day influenced the scores of the end-of-day diary, we compared the end-of-day diary scores (*i.e.* 7-day mean of each symptom) of the 7 days during which participants also carried the PsyMate device to the 7 days during which subjects only completed the end-of-day diary, using a paired samples t-test.

ESM data are based on a multi-level structure, including the variation between subjects and repeated measures within subjects. Associations between GI symptom scores of the end-of-day diary and corresponding ESM scores were tested, using linear mixed models, based on a random intercept and an autoregressive (AR1) covariate structure. Currently the end-of-day diary is the standard method and was therefore used as dependent variable in the analysis. For each participant, ESM day-mean scores of individual symptoms were used

as predictor, to test whether average symptom scores of the day reflect the end-of-day-diary scores.

Scoring scales of the end-of-day diaries and ESM were harmonized by rescaling the ESM data from a 7-point to a 5-point scale, *i.e.* a score of 1 was 1, 2 or 3 was rescaled to 2, 4 to 3, 5 or 6 to 4 and 7 to 5. In order to assess whether the rescaling affected the reliability of GI symptom related questions in ESM, a Cronbach's α was calculated for the original and rescaled data.

To quantify the difference of scores between the two methods, ESM data were subtracted from corresponding end-of-day diary scores, resulting in a delta score, which was statistically tested using the same multi-level analysis, adjusting for repeated measures. In like manner, the differences between end-of-day diary scores and ESM maximum scores of the day were tested.

In order to investigate if certain patient characteristics lead to a different retrospective perception of abdominal pain, possible predictors of a larger delta-score (*i.e.* the differences in abdominal pain reporting between ESM day-mean scores and the end-of-day diary scores) were tested: *i.e.* sex, age, presence of co-morbid panic disorder, psychological symptoms (HADS), and negative cognitions (CS-FBD). To further test the hypothesis that the presence of panic disorder in patients may influence GI symptom reporting, linear mixed models were used to investigate whether there was a difference in symptom score levels between IBS patients with and without a diagnosis of panic disorder, within ESM or end-of-day diary data.

Differences between GI symptom scores assessed by the GSRS and ESM could only be analysed on subject level, by paired sample t-test, due to lack of intra-subject variation in the GSRS scores (only one measurement per subject). Subject-means (*i.e.* means over 7 days) of non-rescaled ESM data were calculated and compared to the corresponding GSRS scores (both ESM and GSRS use a 7-point scale).

Results

Twenty-six IBS patients were included in the analysis. Female gender was predominant (69.2%), mean age was 41.8 ± 15.5 years and mean BMI was 25.5 ± 5.2 kg/m². IBS subtypes were defined based on bowel habits according to the ROME III criteria: 14 subjects were classified as diarrhoea predominant (IBS-

D), 4 as constipation predominant (IBS-C), 7 as mixed (IBS-M), and 1 as undefined subtype (IBS-U). Sixteen patients were diagnosed with co-morbid panic disorder. Mean scores for CS-FBD, measuring negative cognitions related to GI symptoms, and the HADS subscales for depression and anxiety were 112±44, 4.2±3.7, and 7.4±4.4, respectively.

The scores of the 14 day paper end-of-day diary did not differ significantly between the 7 days on which also ESM was used compared to 7 days without ESM (Supplementary Table S5.1), indicating that carrying the PsyMate device did not influence the end-of-day diary scores.

All subjects reported some level of abdominal pain during the test period, and the rate of symptomatic days, on which abdominal pain was present, *i.e.* end-of-day diary score or at least once daily ESM score of ≥2, was 84% and 86%, respectively. The compliance rate was 76.8% for ESM, *i.e.* out of the total number of days on which ESM was used by all study participants, on 76.8% of the days at least 6 beep-questionnaires had been completed. Not completed beep-questionnaires were considered as missing values in the analyses.

To compare end-of-day diary to ESM scores, the scales of the two methods were harmonized (*i.e.* ESM from a 7-point to a 5-point scale). A Cronbach's α was calculated for the original and rescaled ESM data; $\alpha=0.647$ versus 0.648, respectively, demonstrating that the rescaling did not influence the data strongly.

End-of-day diary compared to ESM scores

There was a significant association between end-of-day diary scores and corresponding ESM day-mean scores for all GI symptoms analysed (Table 5.1), demonstrating that both methods measure the same construct, with estimates (indicating the direction and strength of the association) ranging between 0.38 (SE 0.10) and 1.08 (SE 0.06), $P<0.001$.

Table 5.1 Association between end-of-day diary scores and corresponding ESM day-mean scores, using linear mixed models (random intercept and AR1 covariate structure) and corrected for repeated measures: the dependent variable is the end-of-day diary score and the predictor is the corresponding ESM item. Estimate indicates the direction and strength of the association.

GI symptom	Estimate	Std. Error	P-value	95% CI
Abdominal pain	0.38	0.10	<0.001	0.19; 0.59
Nausea	0.93	0.07	<0.001	0.80; 1.07
Belching	1.08	0.06	<0.001	0.95; 1.20
Bloating	0.75	0.08	<0.001	0.59; 0.90
Flatulence	0.49	0.08	<0.001	0.33; 0.66

In order to calculate possible differences in scores between both methods, the delta score was tested using the same multi-level analysis (Table 5.2). The estimate of the intercept represents the difference between the scores of the two methods. On a 1-to-5-point scale, abdominal pain and flatulence scores were 0.42 (SE 0.11) and 0.41 (SE 0.06) point higher, respectively ($P<0.001$), for the end-of-day diary compared to the corresponding ESM scores. Small, though statistically significant differences were also observed for nausea (0.07 (SE 0.03), $P=0.032$), and belching (0.13 (SE 0.03), $P<0.001$), but the difference was not significant for bloating. In contrast, ESM day-maximum scores did not differ when compared to paper end-of-day diary scores for abdominal pain (Figure 5.1A), but were significantly higher for bloating (Figure 5.1B), nausea, belching, and flatulence (Supplementary Figure S5.1).

Table 5.2 Difference between end-of-day diary scores and corresponding ESM scores, *i.e.* the delta score (as dependent), tested using linear mixed models (random intercept and AR1 covariate structure) and corrected for repeated measures. Estimate represents the difference between the scores of the two methods on a 1-to-5-point scale. A positive estimate indicates a higher score in the end-of-day diary when compared to ESM.

GI symptom delta score	Estimate intercept	Std. Error	P-value	95% CI
Abdominal pain	0.42	0.11	<0.001	0.19; 0.65
Nausea	0.07	0.03	0.032	0.01; 0.13
Belching	0.13	0.03	<0.001	0.06; 0.20
Bloating	0.09	0.06	0.113	-0.02; 0.20
Flatulence	0.41	0.06	<0.001	0.30; 0.54

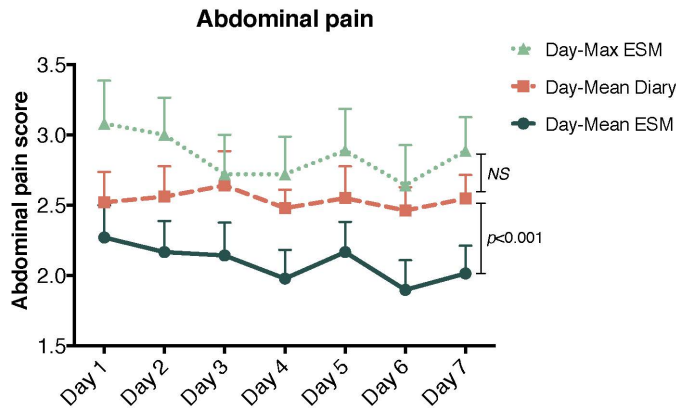
Predictors of differences in abdominal pain reporting between ESM and end-of-day diary

In order to identify possible predictors for the difference in abdominal pain reporting between the two methods, delta scores were used as dependent variable in the multilevel analysis. Only the interaction between the presence of panic disorder and male sex resulted in a statistically significant larger difference between end-of-day diary and the corresponding ESM pain score (estimate 1.88 (SE 0.87), $P=0.041$). When taking into account multiple testing, this result was no longer statistically significant. No significant effect could be demonstrated for age, presence of panic disorder in general, higher anxiety and depression scores (HADS), and negative cognitions (CS-FBD).

However, within each method (*i.e.* ESM and end-of-day diary), IBS patients with panic disorder reported higher abdominal pain scores compared to those without; mean difference between groups was 1.1 ± 0.2 point ($P<0.01$) on a

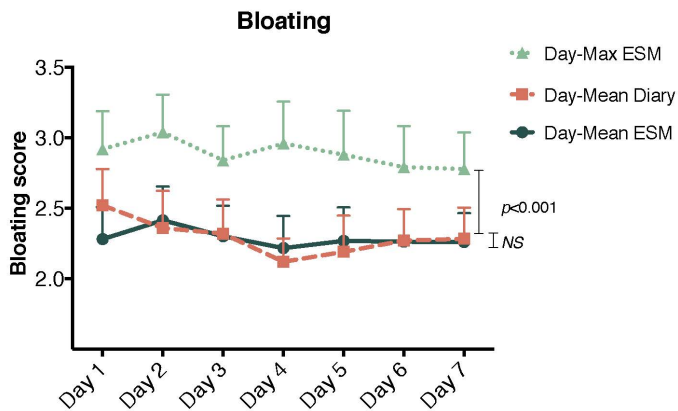
1-to-7-point scale in ESM, and 0.8 ± 0.4 point ($P < 0.01$) on a 1-to-5-point scale in the end-of-day diary. This was also found for bloating (ESM: 1.6 ± 0.2 , $P < 0.01$, and end-of-day diary: 1.2 ± 0.2 , $P < 0.01$), while there were no significant differences for nausea, belching, and flatulence between scores of patients with and without the diagnosis of panic disorder.

A



Figures 5.1 (A) End-of-day diary scores compared to ESM day-mean and day-maximum scores, for abdominal pain. Differences tested for both ESM day-mean and ESM day-max scores vs. end-of-diary scores, using linear mixed models.

B



Figures 5.1 (B) End-of-day diary scores compared to ESM day-mean and day-maximum scores, for bloating. Differences tested for both ESM day-mean and ESM day-max scores vs. end-of-diary scores, using linear mixed models.

GSRs compared to corresponding ESM scores

Symptom scores of the GSRs, reflecting a one-week period, were also compared to corresponding ESM 7-day-mean scores per subject. For all symptoms (*i.e.* upper abdominal pain, nausea, heartburn, bloating, and flatulence), GSRs scores were significantly higher compared to ESM scores ($P<0.01$), with the most pronounced difference noted for pain and flatulence (Table 5.3).

Table 5.3 GSRs scores compared to corresponding ESM symptom scores.

Scale GI symptom (mean \pm SD)	GSRs IBS patients (n=26)	ESM IBS patients (n=26)
7 point scale:		
Upper abdominal pain	4.77 \pm 1.50	2.44 \pm 1.30 ^b
Heartburn	2.54 \pm 1.61	1.11 \pm 0.28 ^b
Nausea	2.69 \pm 1.69	1.45 \pm 0.88 ^a
Bloating	4.35 \pm 1.81	2.94 \pm 1.67 ^a
Flatulence	4.69 \pm 1.38	2.13 \pm 1.38 ^b

Means calculated on subject level, SD indicates variation between subjects. Scores were not rescaled (original 1-to-7-point scale in GSRs and ESM).^a $P<0.01$, ^b $P<0.001$.

Discussion

The current study demonstrates that retrospective symptom measurement by end-of-day diaries as well as an end-of-week questionnaire (GSRs), leads to higher scores when compared to electronic momentary symptom assessment (*i.e.* by ESM), in patients with IBS. This difference is most pronounced for abdominal pain and flatulence. Abdominal pain scores in the end-of-day diaries were similar to maximum pain reporting in ESM, indicating that subjects tend to report peak pain rather than average pain scores of the day in an end-of-day diary. This was not found for the other tested GI symptoms. With regard to GSRs versus ESM, in addition to abdominal pain and flatulence, also nausea, heartburn, and bloating scores were higher in the retrospective method compared to momentary assessment. Furthermore, affective complaints in general did not lead to a larger difference in abdominal pain reporting between the end-of-day diary and ESM. However, within each method, IBS patients with panic disorder reported higher abdominal pain and bloating scores compared to those without.

The reporting of *peak* rather than *average* abdominal pain has been described previously by Weinland *et al.*¹⁶ for IBS patients, and by Redelmeier *et al.*²⁶ studying abdominal pain of other origin, *i.e.* colonoscopy and lithotripsy associated pain. The observed patient preferences of peak pain reporting provide more understanding with regard to pain memory and perceived relevance in IBS patients. Interestingly, peak reporting was not found for nausea, bloating, belching, and flatulence. The current findings may imply that pain is more prone to recall bias than other GI symptoms in IBS and may also be related to a considerable affective component in pain perception. However, it should be noted that some symptoms, especially nausea and belching, were less prevalent, and consequently the present study may lack power to identify potential differences between the methods for these symptoms. Accordingly it might be interesting to assess a symptom as nausea in a population of functional dyspepsia patients in order to investigate whether a larger difference between ESM and end-of-day diaries can be observed.

When comparing ESM scores to retrospective end-of-week GSRS scores, even larger differences were found in comparison to ESM versus end-of-day diary scores, indicating that questionnaires with a longer recall period are associated with higher symptom scores in IBS patients. This is in line with previous research describing differences between end-of-day and end-of-week symptoms assessment in IBS.⁴ When analysing ESM data in more detail, GI symptoms in IBS patients were found to fluctuate considerably during the day and during the week. The 7-day mean of ESM data is an average of days with higher and lower symptom burden. Subjects are prone to base their scores on symptomatic days when filling in the GSRS, equivalent to the peak-pain tendency in the end-of-day diaries. This tendency may underlie the recall bias in retrospective methods. The differences between ESM and GSRS were tested parametrically on subject-level. However, non-parametric testing did not lead to different outcomes. The standard deviations of GSRS and ESM 7-day mean scores, indicating variation between subjects, in this analysis were broad, but differences between methods were highly significant.

Prevalent psychological factors in IBS, in particular anxiety, have been reported to affect pain processing, memory, and interpretation.¹¹ To investigate how anxiety may influence the difference between the retrospective (based on memory) and momentary assessment of abdominal pain, we included IBS patients with and without a diagnosis of comorbid panic disorder in the current

study. Although, the overall scores for abdominal pain were higher in patients with compared to the group without panic disorder, only the presence of panic disorder in male subjects was found to be a significant predictor of a higher pain score difference between ESM and end-of-day diary. However, this predictor was no longer significant when corrected for multiple testing and should therefore be interpreted carefully. Furthermore, given the higher abdominal pain scores in IBS patients with panic disorder compared to those without, anxiety should be seen as a stable trait factor, which may less depend from momentarily fluctuating contextual influences in IBS. It has to be acknowledged that in the current study the IBS patients with panic disorder were overrepresented relative to the general IBS population. Therefore, our data need to be replicated by future studies in large non-selected IBS cohorts.

In addition to higher overall scores and peak-symptom reporting in paper end-of-day diaries, this method is also compromised by high fake compliance, as demonstrates previously by Stone *et al.*¹² Therefore, serious concern has risen with respect to the validity of retrospective paper end-of-day diaries. Fake compliance is excluded in ESM, since subjects can only respond within 10 minutes after the beep, and moreover, actual compliance can be monitored. We considered a subject to be compliant on a specific day if at least 6 out of 10 beep-questionnaires had been completed and found a good compliance rate for ESM, *i.e.* 76.8% for the total study population.

Certain limitations of the present study should be mentioned. In the current study we have not assessed differences in reporting of consistency and frequency of defecation between the symptom assessment methods, while disordered defecation is a key symptom of IBS. Furthermore, the use of ESM may lead to higher patient burden compared to end-of-day paper diaries. This may lead to selection bias, because patients may decide not to participate beforehand. The number of questions asked was high and should be reduced in future investigations. The time needed to complete a beep-questionnaire should be less than 2 minutes. Questions used in ESM were selected based on available questionnaires, such as the end-of-day diary^{2,3,22} and the GSRS,^{25,27} and were previously not psychometrically tested for momentary symptom assessment. Therefore, future studies evaluating the potential of ESM for symptom assessment in IBS should study psychometric properties, regarding validity and reliability of the instrument, which were not subject to analysis in this preliminary study. It has to be acknowledged that for the comparison of

ESM and end-of-day diaries, it was necessary to rescale the ESM data from a 7- to a 5-point scale, which could cause a distortion of the results. However, the Cronbach's α for ESM GI symptom scores did not change after rescaling, and moreover, comparable results were found by comparing ESM (non-rescaled data) with GSRS scores. Finally, the exploratory aim of the current study has resulted in a small sample size. However, this has not hampered the multilevel analyses, which are based on repeated measures.

Our results demonstrate certain advantages of ESM as a new digital symptom assessment tool in IBS patients, regardless of the presence of psychological complaints. Nevertheless, further research is needed to optimize the content and number of questions for the assessment of IBS symptoms and to validate the tool as well as the selected questions within the target population, before ESM can be implemented in medical science and clinical practice.

Because of the repeated measures, ESM offers the advantage to study symptom fluctuation during the day and to investigate possible triggers of pain or other symptoms in future investigations, including for example ecological and psychosocial factors as well as linkage to bowel habits. This insight may lead to improvement of therapies. Previously has been found that IBS symptom episodes on average last several hours,¹⁶ which indicates that repeated measures over the day by ESM are suitable to capture the symptom variability in these patients. All patients in the current study had a debriefing interview with the researcher, and acknowledged that the number of repeated measures was sufficient to adequately measure the GI symptom variation over the day. Furthermore, ESM could be used for more detailed evaluation of treatment response or for intervention purposes in behavioural therapies, in the context of personalized medicine, by giving IBS patients specific feedback with regard to triggers of symptoms identified by ESM. This has recently been successfully applied in a group of individuals with depression.²⁸

Conclusion

IBS patients report higher abdominal pain and flatulence scores in end-of-day diaries when compared to day-average scores of momentary symptom assessment by ESM, with a preference of peak-pain reporting at the end of the day. In retrospective symptom questionnaires that cover a one-week period,

the difference of symptom reporting when compared to electronic ESM is even higher for abdominal pain and the other GI symptoms. Affective comorbidity in IBS leads to higher abdominal pain and bloating scores compared to IBS patients without, but did not strongly influence differences in abdominal pain reporting between tested symptom assessment methods.

References

1. Spiegel B, Bolus R, Harris LA, Lucak S, Naliboff B, Esrailian E, Chey WD, Lembo A, Karsan H, Tillisch K, Talley J, Mayer E, Chang L. Measuring irritable bowel syndrome patient-reported outcomes with an abdominal pain numeric rating scale. *Aliment Pharmacol Ther.* 2009;30:1159-1170.
2. Thijssen AY, Jonkers DM, Leue C, van der Veek PP, Vidakovic-Vukic M, van Rood YR, Clemens CH, Masclee AA. Dysfunctional cognitions, anxiety and depression in irritable bowel syndrome. *J Clin Gastroenterol.* 2010;44:e236-241.
3. Ludidi S, Mujagic Z, Jonkers D, Keszthelyi D, Hesselink M, Kruijmel J, Conchillo J, Masclee A. Markers for visceral hypersensitivity in patients with irritable bowel syndrome. *Neurogastroenterol Motil.* 2014; 26:1104-1111.
4. Lackner JM, Jaccard J, Keefer L, Firth R, Carosella AM, Sitrin M, Brenner D; Representing the IBSOS Research Group. The accuracy of patient-reported measures for GI symptoms: a comparison of real time and retrospective reports. *Neurogastroenterol Motil.* 2014;26:1802-1811.
5. U.S. Department of Health and Human Services FaDA, Center for Drug Evaluation and Research (CDER). Guidance for Industry Irritable Bowel Syndrome - Clinical Evaluation of Drugs for Treatment 2012. <http://www.fda.gov/downloads/Drugs/Guidances/UCM205269.pdf>.
6. Bradburn NM, Rips LJ, Shevell SK. Answering autobiographical questions: the impact of memory and inference on surveys. *Science.* 1987;236:157-161.
7. Bower GH. Mood and memory. *Am Psychol.* 1981;36:129-148.
8. Babel P, Pieniazek L, Zarotynski D. The effect of the type of pain on the accuracy of memory of pain and affect. *Eur J Pain.* 2015;19:358-368.
9. Bolger N, Davis A, Rafaeli E. Diary methods: capturing life as it is lived. *Annu Rev Psychol.* 2003;54: 579-616.
10. Grzesiak M, Beszlej JA, Mulak A, Szechiński M, Szewczuk-Bogusławska M, Waszczuk E, Kantorska M, Frydecka D. The lifetime prevalence of anxiety disorders among patients with irritable bowel syndrome. *Adv Clin Exp Med.* 2014;23:987-992.
11. Gorczyca R, Filip R, Walczak E. Psychological aspects of pain. *Ann Agric Environ Med.* 2013;Spec no. 1:23-27.
12. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. *Control Clin Trials.* 2003;24:182-199.
13. Csikszentmihalyi M, Larson R. Validity and reliability of the Experience-Sampling Method. *J Nerv Ment Dis.* 1987;175:526-536.
14. Stone A, Shiffman S. Ecological Momentary Assessment (EMA) in behavioral medicine. *Ann Behav Med.* 1994;16(199-202).
15. 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;39:1533-1547.
16. 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: 1813-1820.
17. Longstreth GF, Thompson WG, Chey WD, Houghton LA, Mearin F, Spiller RC. Functional bowel disorders. *Gastroenterology.* 2006;130:1480-1491.
18. 4th. ed., text rev.; DSM-IV-TR; American Psychiatric Association. 2000.
19. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand.* 1983;67: 361-370.

20. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *J Psychosom Res.* 2002;52:69-77.
21. Toner BB, Stuckless N, Ali A, Downie F, Emmott S, Akman D. The development of a cognitive scale for functional bowel disorders. *Psychosom Med.* 1998;60:492-497.
22. Mujagic Z, Ludidi S, Keszthelyi D, Hesselink MA, Kruijmel JW, Lenaerts K, Hanssen NM, Conchillo JM, Jonkers DM, Masclee AA. Small intestinal permeability is increased in diarrhoea predominant IBS, while alterations in gastroduodenal permeability in all IBS subtypes are largely attributable to confounders. *Aliment Pharmacol Ther.* 2014;40:288-297.
23. Myin-Germeys I, Birchwood M, Kwapil T. From environment to therapy in psychosis: a real-world momentary assessment approach. *Schizophr Bull.* 2011;37:244-247.
24. Courvoisier DS, Eid M, Lischetzke T. Compliance to a cell phone-based ecological momentary assessment study: the effect of time and personality characteristics. *Psychol Assess.* 2012;24:713-720.
25. Revicki DA, Wood M, Wiklund I, Crawley J. Reliability and validity of the Gastrointestinal Symptom Rating Scale in patients with gastroesophageal reflux disease. *Qual Life Res.* 1998;7:75-83.
26. Redelmeier DA, Kahneman D. Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. *Pain.* 1996;66:3-8.
27. Svedlund J, Sjodin I, Dotevall G. GSRS--a clinical rating scale for gastrointestinal symptoms in patients with irritable bowel syndrome and peptic ulcer disease. *Dig Dis Sci.* 1988;33:129-134.
28. Kramer I, Simons CJ, Hartmann JA, Menne-Lothmann C, Viechtbauer W, Peeters F, Schruers K, van Bommel AL, Myin-Germeys I, Delespaul P, van Os J, Wichers M. A therapeutic application of the experience sampling method in the treatment of depression: a randomized controlled trial. *World Psychiatry.* 2014;13:68-77.

Supplemental material

Table S5.1 End-of-day diary scores on 7 days with and 7 days without ESM, differences tested using paired samples t-test.

End-of-day GI symptom diary			
GI symptoms (mean ± SD)	Days with ESM	Days without ESM	p-value
Abdominal pain	2.53 ± 1.01	2.43 ± 1.03	NS
Discomfort	2.54 ± 1.01	2.47 ± 0.97	NS
Nausea	1.31 ± 0.76	1.41 ± 0.82	NS
Bloating	2.35 ± 1.20	2.42 ± 1.18	NS
Belching	1.41 ± 0.89	1.44 ± 0.88	NS
Flatulence	2.16 ± 0.99	2.40 ± 1.16	NS
Constipation	1.58 ± 1.02	1.48 ± 0.83	NS
Diarrhoea	1.27 ± 0.70	1.36 ± 0.82	NS

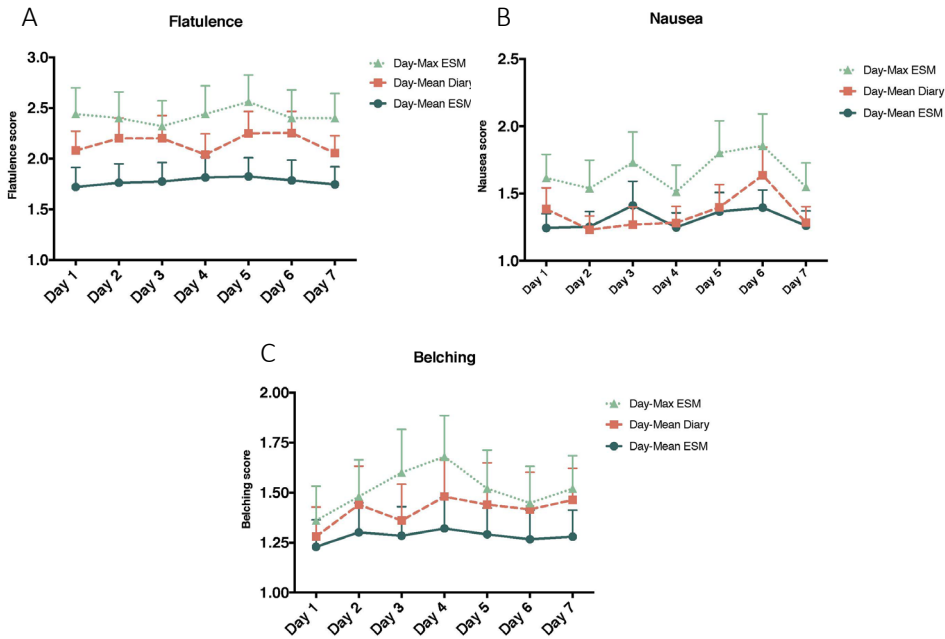


Figure S5.1 End-of-day diary scores compared to ESM day-mean and day-maximum scores, for flatulence (A), nausea (B), and belching (C). Differences tested for both ESM day-mean and ESM day-max scores vs. end-of-day scores, using linear mixed models.

Chapter 6

Development, content validity, and cross-cultural adaptation of a patient-reported outcome measure for real-time symptom assessment in irritable bowel syndrome

Lisa Vork, Daniel Keszthelyi, Zlatan Mujagic, Joanna W. Kruiemel, Carsten Leue, Irina Pontén, Hans Törnblom, Magnus Simrén, Ahmed Albu-Soda, Qasim Aziz, Maura Corsetti, Lieselot Holvoet, Jan Tack, Satish S. Rao, Jim van Os, Emilio G. Quetglas, Douglas A. Drossman, Ad A.M. Masclee

Neurogastroenterology & Motility. 2018;30(3).

Abstract

INTRODUCTION End-of-day questionnaires, which are considered the gold standard for assessing abdominal pain and other gastrointestinal (GI) symptoms in irritable bowel syndrome (IBS), are influenced by recall and ecological bias. The Experience Sampling Method (ESM) is characterized by random and repeated assessments in the natural state and environment of a subject, and herewith overcomes these limitations. This report describes the development of a patient-reported outcome measure (PROM) based on the ESM-principle, taking into account content validity and cross-cultural adaptation.

METHODS Focus group interviews with IBS patients and expert meetings with international experts in the fields of neurogastroenterology & motility and pain were performed in order to select the items for the PROM. Forward-and-back translation and cognitive interviews were performed to adapt the instrument for the use in different countries and to assure on patients' understanding with the final items.

RESULTS Focus group interviews revealed 42 items, categorized into five domains: physical status, defecation, mood and psychological factors, context and environment, and nutrition and drug use. Experts reduced the number of items to 32 and cognitive interviewing after translation resulted in a few slight adjustments regarding linguistic issues, but not regarding content of the items.

CONCLUSIONS An ESM-based PROM, suitable for momentary assessment of IBS symptom patterns was developed, taking into account content validity as well as cross-cultural adaptation. This PROM will be implemented in a specifically designed smartphone application and further validation in a multicenter setting will follow.

Introduction

Irritable bowel syndrome (IBS) is a functional gastrointestinal (GI) disorder, characterized by recurrent abdominal pain and altered bowel habits. It affects up to 20% of the Western population, with a female predominance.¹⁻⁴ IBS leads to a reduction of quality of life, increased utilization of health care and higher (in)direct costs for society.^{5,6} Since specific biological markers are currently unavailable, IBS is a symptom-based diagnosis⁷⁻¹⁰ and reliable symptom assessment methods are essential to investigate natural disease course and evaluate response to therapy.

Both the United States Food and Drug Administration (FDA) and the European Medicines Agency (EMA) recommend the use of a composite primary endpoint for clinical trials, in which abdominal pain as well as stool consistency/frequency is addressed. With regard to abdominal pain, a seven-day mean score on an 11-point Numeric Rating Scale (NRS) assessed at the end of each day is recommended. However, both guidelines address the challenge of capturing all clinically important symptoms associated with IBS in a reliable measure of treatment benefit, since IBS is a heterogeneous disorder with diverse and fluctuating symptom patterns. Therefore, the development of a multi-item patient-reported outcome (PRO) instrument, that captures all clinically important symptoms of the IBS population, is recommended.^{11,12}

Furthermore, currently recommended retrospective (*i.e.* end-of-day), self-reported outcomes have important limitations. First, there is a high risk for recall bias, since retrospective information consists of a reconstruction of a few specific moments rather than a reliable reflection of symptoms over a predefined period of time.¹³⁻¹⁵ Second, it is well described that memory retrieval is influenced by the individual's environment and mental state at the time of recall, known as ecological bias.^{13,14} In conclusion, retrospectively reported patient experiences about symptoms in the past can be significantly distorted.

The Experience Sampling Method (ESM), also referred to as Ecological Momentary Assessment (EMA), may overcome these limitations. ESM is an electronic questioning method characterized by random, repeated assessments in the subject's current state and environment for several consecutive days.^{13,14,16} Hereby, ESM offers the opportunity to reduce the risk for recall and ecological bias and to capture symptom variability over time, while taking into account contextual, social and psychological factors, which might have an impact on IBS symptom reporting. ESM has been applied in a range of disorders, *i.e.* chronic pain, chronic fatigue and psychiatric disorders,¹⁷⁻¹⁹ but has

not been validated as a symptom assessment method for IBS. Only one study has been published reporting the use of ESM in an IBS population. Weinland *et al.* demonstrated that higher abdominal pain scores were reported in end-of-day diaries compared to ESM in IBS patients.²⁰ Additionally, our previously performed pilot trial on the correlation of Experience Sampling Method with currently used retrospective patient-reported outcomes supported our hypothesis that ESM provides more accurate information about IBS symptoms.²¹

The current report describes the development of a patient-reported outcome measure (PROM) for real-time measurement of abdominal and extra-abdominal symptoms, and their potential triggers, in IBS, using the ESM principle. In agreement with FDA guidelines on PROM development patient as well as expert input obtained by structured focus group interviews and expert meetings were used in the item selection process, herewith addressing content validity. Furthermore, cross-cultural adaptation was taken into account by performing cognitive interviews after forward-and-back translation of the developed instrument.²²

Materials and methods

Study design

The development of a novel PROM for symptom assessment in IBS, using the ESM principle, was performed in four phases: I) initial item selection, II) focus group interviews to obtain patients' input, III) expert meetings to obtain the input of international experts in the fields of neurogastroenterology and motility and pain, and IV) translations and cognitive interviews to assure on patients' understanding with the developed tool (cross-cultural adaptation). Throughout this report, the developed instrument will be referred to as 'ESM-PROM'.

Phase I: Item selection

In our previous pilot study, real-time ESM was compared to retrospective paper symptom assessments in IBS patients.²¹ The IBS-specific ESM items used in this study were based on a developed framework of theoretical constructs for symptom assessment in an IBS population. These IBS constructs were obtained from currently used retrospective symptom surveys and qualitative studies,

selected after an extended literature search.²³⁻²⁹ The ESM-specific constructs, *i.e.* items regarding psychological status, social factors and context and environment at the moment of assessment, were derived from ESM questionnaires as used in previous studies at the department of Psychiatry of Maastricht University Medical Center + (Maastricht UMC+).^{30,31} Additionally, subjects, participating in the pilot study, gave their feedback about the feasibility and burden of ESM as an assessment tool and about the relevance of the items on the initial draft instrument (developed after first item selection). This feedback was taken into account for further selection.

Phase II: Focus group interviews

Recruitment of focus group participants

Subjects between the age of 18 and 75 years and diagnosed with IBS, according to the Rome III criteria⁷⁻⁹, were recruited via the outpatient department of the Gastroenterology and Hepatology division of Maastricht UMC+ as well as general practitioner practices in South-Limburg, the Netherlands. All subjects were previously included in the Maastricht IBS Cohort.^{32,33} Subjects with organic diseases possibly explaining the gastrointestinal symptoms, as diagnosed by a physician, were not eligible for inclusion. Furthermore, subjects with prior abdominal surgery (except appendectomy, laparoscopic cholecystectomy and hysterectomy) were not included. Participants needed to be able to understand written Dutch and speak the Dutch language, since focus groups were conducted in Dutch. All subjects gave written informed consent prior to participation.

Conducting and moderating focus group interviews

Focus group meetings were arranged in order to select a relevant set of items for real-time symptom assessment in IBS using ESM, according to patient opinions. Subjects eligible for participation were invited to Maastricht UMC+ to participate in a focus group interview. Meetings took 90 minutes and were planned one by one, each preferably including six to eight participants, continuing until saturation of input was reached. Saturation was determined as meetings no longer augmenting any new items or information in addition to previous meetings.^{22,34} Focus groups were guided by a moderator (LV) and at least one assistant-moderator (ZM or DK) and were performed according to a predefined, two-stage framework, based on guidelines described in literature.³⁴⁻

³⁶ The first stage consisted of an open discussion, in which participants were instructed to bring forward every item they considered to be essential in an IBS symptom assessment tool aiming to get insight in their symptom pattern during daily life. This concept, rather than introducing specific items, was chosen to prevent bias to arise. During the second stage, participants were requested to criticize the ESM items, that were initially selected during phase I. We specifically asked them to argue which questions were relevant in a real-time symptom assessment method for IBS and which could be omitted to their opinion. Also, they were asked to define any incomprehensible items and to give their opinion about the most appropriate response scales.

Analysis of focus group data

Focus group discussions were recorded by video camera and additionally, attending investigators made notes. Video recordings were transcribed verbatim after the meetings and subsequently summarized in a spreadsheet database. In this, also the notes of the investigators and feedback of subjects in our previous pilot trial were added. Subsequently, this spreadsheet was used to evaluate whether to preserve, modify, delete or add items from the questionnaire to develop a draft instrument. Results from all focus group meetings were combined in the spreadsheet and are presented together as the results of phase II in this report.

Phase III: Expert meeting

Investigators with extensive experience in the fields of neurogastroenterology and motility and pain were gathered to obtain expert opinions regarding the further development of the ESM-PROM. The international experts were invited for a face-to-face meeting and were asked to discuss all items that were selected during initial item selection (phase I) and the focus group meetings (phase II). They specifically criticized the items on relevance in clinical practice (*i.e.* does an item add something to the insight in a patient's symptom pattern and disease severity?) and suitability for a real-time symptom assessment method (*i.e.* will repeated, momentary assessment of this item result in measuring variability during the day?). On the basis of this discussion, a final set of questions was selected. The authors LV, DK, JK, CL, QA, MC, EQ and AM were present at the meeting and MS and DD gave their input via telephone conference.

Phase IV: Translations and cognitive interviews

In order to perform a further validation study in a multicenter, international context the ESM-PROM, originally developed in Dutch, was translated into English and Swedish. Translation was performed using forward-and-back procedures, as described by the World Health Organization (WHO).³⁷ Additional individual cognitive interviews with native-speaking IBS patients were arranged with regard to each language (*i.e.* Maastricht UMC+ regarding Dutch, Queen Mary University of London regarding English and University of Gothenburg regarding Swedish) to verify patient understanding with the developed items in each language. Participants were instructed to read out loud all items and response scales and to communicate items or answers they did not understand. Furthermore, they were asked to identify any revisions that could make the instrument more appropriate. All cognitive interviews were transcribed independently and decisions about revisions were made separately for each language.

Interviews were planned one by one continuing until saturation of input was reached. Inclusion- and exclusion criteria for participants in this phase of the project were similar to those described under 'Phase II: Focus group interviews'.

Results

Study population

Patient characteristics of all subjects that participated in the different phases of the project are summarized in Table 6.1. With regard to the focus group interviews a total of 67 patients diagnosed with IBS according to the Rome III criteria⁷⁻⁹ were invited to one of the focus group sessions. Thirty-two of them agreed to participate, however, only seventeen subjects were in fact present during four different meetings. Reasons for cancelation were illness (n=6) and family circumstances (n=1). Two subjects did not inform us about their reason for cancelation and another six did not cancel at all.

To take into account possible selection bias, demographical characteristics and symptom severity scores of the 17 focus group participants were compared to those of subjects, included in the Maastricht IBS cohort, that did not participate in the focus group meetings. Focus group participants were slightly older (median age 58.0 vs. 45.0 years, $P<0.05$), but no further clinical differences were found. Results are shown in Table S6.1.

Table 6.1 Patient characteristics of subjects that participated in the different phases of the project.

	Focus group interviews, Netherlands n=17	Cognitive interviews, Netherlands n=5	Cognitive interviews, Sweden n=6	Cognitive interviews, UK n=6
Age , median [IQR]	58.0 [39.5 – 63.0]	49 [24.5 – 57.5]	37.5 [31.0 – 64.25]	34 [22.5 – 44.0]
Female , n (%)	13 (76.5)	4 (80)	3 (50)	3 (50)
IBS-subtype , n (%)				
IBS-D	8 (47.0)	2 (40)	0	2 (33.3) [#]
IBS-C	2 (11.8)	2 (40)	2 (33.3)	1 (16.7) [#]
IBS-M	7 (41.2)	1 (20)	4 (66.7)	1 (16.7) [#]
IBS-U	0	0	0	0 [#]
Level of education				
Primary	1 (5.9) [#]			
Secondary	12 (70.6) [#]	4 (80.0) [#]		1 (16.7)
University	2 (11.8) [#]		6 (100)	5 (83.3)
Work status				
School/student				[#]
Employed	8 (47.1) [#]	2 (40.0) [#]	5 (83.3) [#]	[#]
Unemployed	4 (23.5) [#]	2 (40.0) [#]		[#]
Retired	3 (17.6) [#]			[#]

[#] Data of other participants are missing.

Phase I: Item selection

The initially developed ESM instrument contained 60 items, based on five domains: physical status, defecation, mood, environmental & social context, and drug use.

The twenty-six participants that completed ESM assessments during our previous pilot study were asked to evaluate practical issues of ESM as well as the relevance of all specific items on the instrument. With regard to practical aspects, a subset of the participants stated that the number (*i.e.* ten) of assessments during the day (n=14; 53.8%) and the time (*i.e.* ±5 minutes) spent completing each assessment (n=10; 38.5%) were quite burdensome. Furthermore, a few questions (mostly questions about psychological status, such as 'I feel enthusiastic' and 'I feel strong') were considered difficult to define, making these questions inappropriate for a PROM. In general, participants stated that questions about nutrition and relation between food and GI complaints should be included. The set of items constructed during phase I is shown in Table S6.2.

Phase II: Focus group interviews

Focus group interviews resulted in a reduction of the number of items from 60 on the initial draft instrument to 42, with some slight modifications to the five

domains: physical status, defecation, mood and psychological factors, context and environment, nutrition and drug use (Figure 6.1). The suggested modifications regarding each domain are summarized below. The resulting set of items is shown in Table S6.3.

Physical status

Participants emphasized that abdominal pain is one of the most important symptoms to assess and they agreed on two essential elements in a real-time symptom assessment method; location and character of the pain. For pain location, all subjects supported the concept of a schematic picture of the abdomen in which they can indicate a particular abdominal region. One participant noted that it might be confusing which side of the picture reflects the left versus the right side of the abdomen, so it is important to clearly point this out within the picture (Figure 6.2). Furthermore, it was stated that it should be possible to indicate more than one abdominal region at a time. For assessment of the character of the pain participants suggested to insert several categories of which patients can choose one or more suitable categories for their pain at that moment. Suggested categories were cramp-like, sharp, stabbing, dull and bruise-like. Moreover, subjects agreed that it is important to assess whether the abdominal pain is continuous and prolonged or acute and short lasting.

In addition to abdominal pain, subjects agreed on the importance of scoring and reporting other GI symptoms associated with IBS: bloating, visible abdominal distension, abdominal rumbling, flatulence, belching, nausea and heartburn. Further, non-abdominal complaints as stated in the draft instrument (*i.e.* dizziness, palpitations, shortness of breath and sweating) were considered relevant for the majority of participants.

Lastly, fatigue, general illness (*i.e.* chill, the flu, fever) and comorbid symptoms were mentioned as possible factors provoking IBS symptoms. Regarding questioning methods, participants suggested solely asking after these conditions without suggesting any association with IBS symptoms, *i.e.* 'Do you feel tired?' instead of 'Do you think fatigue is influencing your IBS symptoms at this moment?' Since comorbid symptoms can be very widespread, a momentary question could be 'Do you have complaints of comorbid symptoms right now?', with answers on a Likert scale.

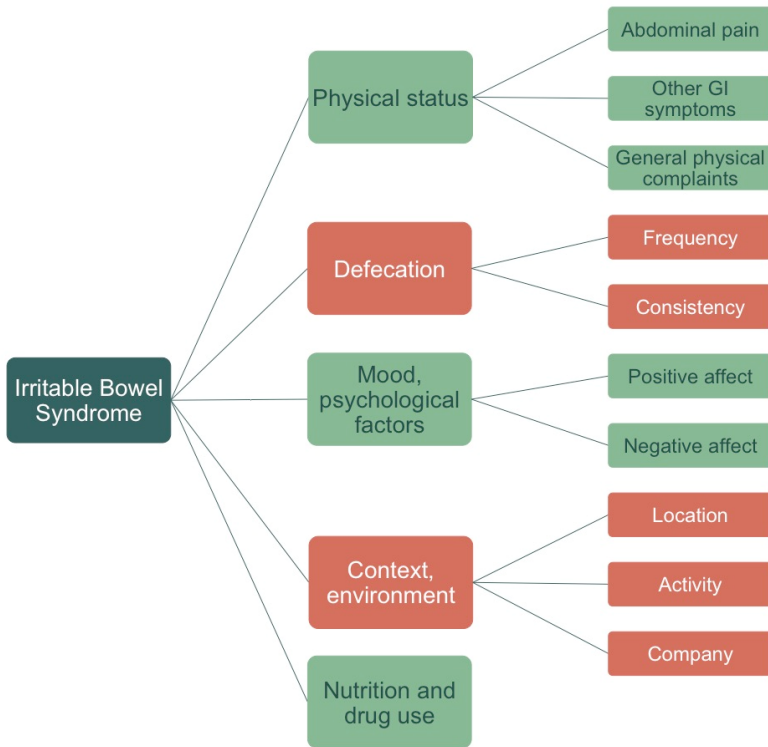


Figure 6.1 Framework of the developed PROM, consisting of 25 items (in the figure summarized in 10 constructs) based on five domains.

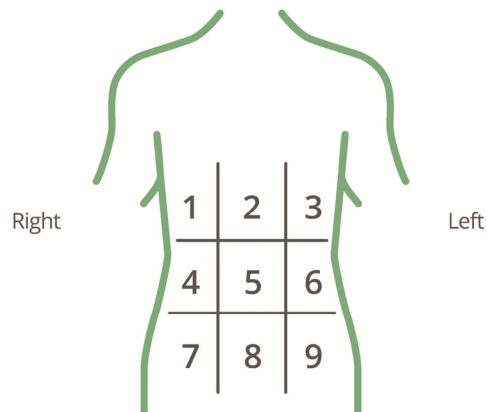


Figure 6.2 Picture on which subjects can indicate the location of abdominal pain according to nine abdominal regions.

Defecation

Assessment of several aspects of bowel habits was considered essential since irregular bowel patterns can be one of the main symptoms itself as well as be associated with and influence abdominal symptoms. Regarding defecation, the participants came up with at least three essential elements: frequency, stool form and urgency. Frequency can be assessed asking 'How many times did you defecate since the previous assessment?' and answering using categories, *i.e.* 0 times, 1 time, etcetera. Participants concluded that stool form is properly assessed using the Bristol Stool Form Scale (BSFS) for each time of defecation, which they are familiar with due to existing symptom surveys they have completed before. Other items that were raised as potentially of relevance, but not essential, are painful and difficult defecation, whether or not being able to postpone defecation in case of urgency and bloody or mucinous stool.

The items about defecation on the initial draft instrument were considered relevant, but were lacking the possibility to report the frequency of defecation since the previous assessment. Therefore, subjects suggested asking after defecation frequency and subsequently consistency according to the BSFS (for each reported defecation). Furthermore, nearly all subjects agreed on the relevance of adding an item regarding the degree of fecal urgency and about whether the abdominal complaints did diminish after defecation. A few aspects of defecation were proposed by small subsets of participants, *i.e.* bloody stool (n=1), mucus in stool (n=2), painful defecation (n=3) and whether patients were able to postpone a bowel movement (n=3).

The last comment with regard to defecation was that in some cases having a bowel movement relieves the abdominal symptoms. Therefore, participants suggested adding a question asking 'Did having a bowel movement relieve your abdominal complaints?'

Mood and psychological factors

All participants agreed on the importance of assessing psychological factors given the often-experienced association between for instance negative emotions, fear for developing abdominal complaints or stress in general and IBS symptomatology. Furthermore, given the possibility that ESM offers to simultaneously assess for social, environmental and psychological features at the moment of symptom reporting (*i.e.* to overcome ecological bias as well as to evaluate for factors of influence on IBS symptoms), it is essential to include psychological items in the instrument. In total, ten items regarding a person's

current mood or mental status were selected from the initial draft instrument. Additionally, an eleventh item, 'I feel good right now', was considered relevant.

Context and environment

Participants stated that the presence or experienced severity of IBS symptoms at a certain moment can depend on the concerning situation and therefore, as psychological factors, also contextual and environmental factors at the moment of assessment were considered relevant. Items 'Where am I?', 'Who is with me?' and 'What am I doing?' were therefore kept in the instrument (including the answer options). Additionally, all four focus group meetings resulted in adding one extra question in this section: 'My symptoms are limiting my current activities.' In total, based on input from the participants, this section is left with ten relevant items.

Nutrition and drug use

Participants stressed that diet should be taken into account when assessing IBS symptoms and associated factors. Two components of nutrition were found to be essential, namely specific food items that may provoke symptoms and regularity in dietary pattern. Subjects suggested to insert a question about what someone ate, *i.e.* 'What did you eat since the last beep?', with categorical answers, in which more than one answer is possible at a time. Categories could be: carbohydrates, dairy products, spicy foods, fatty foods, onions/leek, cabbage, fruits, sweets, carbonated beverages and nuts. The participants were not able to suggest the best way to assess the regularity of the dietary pattern. Some suggested to insert a question asking 'How much did you eat since the last beep signal?' with answers on a Likert scale, in which the mean score is a normal amount of food for that particular person. Therewith, patients can score whether they ate less or more than what is usual for them.

With regard to drug use several important factors were determined: coffee, tea, alcohol, nicotine and medication use. Obviously, regular medication use should not be included in the momentary assessments, but subjects stated that it is important to ask whether patients used any occasional drugs. Answers could be categorical, *i.e.* painkillers, antacids, spasmolytics/antispasmodics, laxatives and antidiarrheals.

Practical issues

Lastly, we asked participants to discuss some practical issues concerning ESM. Most frequently discussed were the number of assessments per day and the acceptable time spent completing the assessments. Some subjects stated that 10 times a day, with 3 to 5 minutes per assessment would be acceptable. However, frequently, participants stated that this would not be possible for them due to their occupational situation. Some stated that 6 times a day during 5 minutes might be possible, but that there should be a longer period of time in which a person can respond after the beep signal. For example, it would be advisable to have the questions accessible for one hour after the signal, rather than only for 10 minutes.

Furthermore, the possibility of standard personalized answers was discussed. Participants raised the option of keeping former answers to the next assessment moment, so that questions have to be completed only in case of changes since the previous assessment. This might decrease the time needed for each participant.

Phase III: Expert meeting

All 42 items resulting from the previous phases of the project were discussed with regard to relevance in a momentary assessment method. It is important to emphasize that the developed ESM-PROM concerns not only real-time assessments of IBS symptomatology, but additionally takes into account momentary aspects, as contextual and psychological factors, regarding the moment of symptom reporting. Regarding abdominal pain, experts concluded that character of abdominal pain is less relevant to be assessed ten times a day. Therefore, it was concluded that severity and location of abdominal pain should be included as the only abdominal pain-relating items. Some nuances were added with regard to the GI symptoms, such as besides asking solely for the severity of bloating, it was suggested also to assess whether this is associated with a feeling of discomfort and whether this is accompanied by visible distension of the abdomen. The latter was also mentioned during focus group interviews.

Regarding comorbid symptoms as “feeling tired” and “general illness”, as were mentioned during the focus group interviews, experts agreed that those are not suitable for momentary assessments per se. Comorbid symptoms should rather be assessed cross-sectionally, as fatigue and general illness are not likely to

change very much during the day. Therefore, those items were not included in the ESM-PROM.

With regard to bowel movements, frequency, stool form and fecal urgency were included as the most important items. Furthermore, straining and (in)complete evacuation were added in case a bowel movement is reported. Other issues regarding defecation were not included. Focus group participants mentioned that it is important to ask whether the severity of abdominal complaints decreased after a bowel movement. However, it was decided that this is not a momentary question and that ESM itself offers the opportunity to evaluate a possible association (relief as well as aggravation of abdominal pain before or following a bowel movement) between bowel movements and abdominal symptoms by repeated measurements over the day.

The ten items regarding psychological status, resulting from focus group interviews, were thoroughly discussed, since experts stated that ten items would be too much of a burden with respect to the time needed to complete one assessment. Therefore, this was reduced to seven items, including both positive and negative affect. Contextual items were discussed as well, but were not changed.

Experts all agreed that nutritional factors are of important interest with regard to IBS symptoms. However, it was discussed that the explicit evaluation of associations between GI symptoms and specific food products is beyond the scope of the current project. Including items on specific nutritional factors would result in too many items. Therefore, it was decided to only include an item on whether or not someone ate between the current and the previous assessment in order to evaluate a possible role for food intake in individual IBS symptomatology.

With respect to the practical issues that were raised during focus group meetings, several options to change the number of assessments during the day and the time during which an assessment is available after each beep were considered. Since in our pilot study²¹ median response time after the signal was 21 seconds (IQR: 11-40 seconds) and in only 0.7% of total assessments response time was longer than 10 minutes, it was decided that extending this time period would probably not result in a much higher response rate. Furthermore, the total compliance rate with ten assessments each day of 76.8%, *i.e.* on 76.8% of the total days at least six assessments were completed, was considered high. Therefore, it was decided not to change both of these practical features.

In conclusion, experts' opinions resulted in a further reduction of the number of items to 32, categorized into the same domains as discussed before. The final instrument, as developed in phase III, is shown in Table S6.4.

Phase IV: Translations and cognitive interviews

Demographic characteristics of the participants in each center are shown in Table 6.1. Cognitive interviews did not reveal any large-scale problems regarding the patients' understanding of the instrument. With respect to the Dutch instrument only a few modifications were suggested in order to verify the patients' understanding. No adjustments were made to the content or the order of the items. The English as well as Swedish instrument needed some modifications with respect to wording after the translation. Also, a few items were adapted to make them more suitable for momentary assessments. Besides these minor changes no alterations of content or order of items were done.

Discussion

The current study was conducted as part of a multicenter project aiming to develop and validate a new patient-reported assessment tool for real-time evaluation of IBS symptoms. Given the lack of biological markers for IBS, evaluation of potential interventions aimed at symptom improvement is dependent on symptom assessments by using patient-reported outcome measures.^{25,38} The current lack of a reliable symptom assessment method, that considers all clinically important signs and symptoms of IBS, requires development of a new multi-item tool.^{11,12,29}

The PROM development in this study was executed according to FDA guidelines and the NIH Patient Reported Outcome Measurement Information System (PROMIS) consortium, recommending PRO instruments to be developed according to a conceptual framework with obtaining patient input to ensure content validity and reliability.^{22,23,25} The development was therefore performed in four consecutive phases: initial item selection from literature, obtaining patients' input by structured focus group interviews, considering experts' opinions by an (international) expert meeting and cognitive interviewing with IBS patients after forward-and-back translation.

Focus group interviews revealed a broad spectrum of items that, according to the participants, are relevant for a momentary IBS-specific PROM. After

transcription of focus group discussions, a total of 42 items and another 21 optional items were defined. Optional items comprised questions suggested or defined as relevant only by a subset of the participants. All defined items were understandable for the total focus group population, positively contributing to content validity. Since in the experience sampling methodology momentary assessments are collected ten times a day, the number of instrument items should be limited in order to reduce patient burden and thereby maintain patient compliance. Although compliance with ESM was reasonably good (76.8%) in our pilot study using an ESM instrument containing as much as 60 items, ideally, each assessment should take no longer than 2 minutes.³⁰ Therefore, selected items were reviewed by an international expert team in the field of neurogastroenterology and motility, leading to further selection. Finally, a momentary questionnaire containing 32 items was developed.

Items could be clustered into five categories: physical status, defecation, mood and psychological factors, context and environment, and nutrition and drug use. The categories “physical status” (including abdominal symptoms) and “defecation” are the key-elements in actual IBS symptom assessment, whereas the other three categories are aiming to evaluate influencing factors of these symptoms. Specifically, the items covered by “mood and psychological factors” will give insight into a possible link between mental disturbances and IBS symptoms, taking into account the individual moment-to-moment variability of both aspects. Likewise, “context and environment” can be linked to symptom severity, herewith taking into account possible effects of ecological factors on symptom response. “Nutrition and drug use” were raised among the most important daily life elements influencing GI symptoms. To assess the effect of these possible influencing factors on GI symptoms, multivariate and multilevel analytical techniques will be required.^{30,39}

Several PROM's for IBS symptom assessment have been developed recently. Spiegel and colleagues delineated a conceptual framework for IBS-PROM's, consisting of 35 items, subdivided over four GI-related domains and an extra-intestinal domain.²³ These five categories are also represented in our ESM-PROM, however, not all individual items have been implemented. All items were discussed during focus group meetings, but given the repeated character of ESM, in particular items regarding symptom-bothersomeness, -interference, -impact and -predictability were thought less suitable for the ESM-PROM.

The PROMIS consortium has developed the PROMIS GI symptom scale, which consists of short outcome measures for eight GI symptom complexes.²⁵ All symptom complexes, but ‘disrupted swallowing’ and ‘bowel incontinence/

soilage', are taken into account in the ESM-PROM. Focus groups revealed that problems with swallowing are not very frequently encountered in the context of IBS. Bowel incontinence might occur, but was not reported by the majority of IBS patients and was therefore thought not generalizable to the target population of the ESM-PROM.

Recently, the 'Diary for Irritable Bowel Syndrome Symptoms' was developed by the Patient-Reported Outcome Consortium IBS Working Group (PRO Consortium IBS-WG).⁴⁰ The conceptual framework for this 24-hour diary is based on two core symptoms: abdominal symptoms and bowel movement-related symptoms. Of the elicited specific items, abdominal discomfort and abdominal cramping are not included in our ESM-PROM, since these were found difficult to differentiate from abdominal pain – in line with recent Rome IV definitions. Therefore, these items will not add considerable extra information and were omitted in order to reduce the number of items.

All in all, the differences between the ESM-PROM and previously developed PROM's for IBS mainly relate to differences in methodology and the fact that ESM concerns repeated assessments over the day.

Next to the items, also possible scoring scales for all items were discussed. Participants supported the use of a Numeric Rating Scale, but could not agree on seven (as formerly used in ESM³⁰) or eleven endpoints. In comparison, the PROMIS scales make use of a 5-point Likert scale⁴¹, however, an end-of-day 11-point NRS is currently recommended by the FDA and EMA (European Medicines Agency) as a primary endpoint in IBS clinical trials.^{11,12} In order to prevent confusion by using different response scales, this scale was chosen to be incorporated into the instrument for all items.

A number of study limitations ought to be noted. First, the risk of self-selection (sampling) bias, *i.e.* the participants' decision whether or not to participate may influence the representativeness of the study population. As an example, subjects with a stronger opinion than others or with more severe symptoms might be more willing to participate in the focus group meetings. Furthermore, only 53% (17/32) of the initially attending participants of the focus groups did eventually participate. However, since the Maastricht IBS cohort concludes a large heterogeneous IBS population, we expect to cover this heterogeneity even in this small subset of the population. Additionally, possible differences between subjects who did participate in the focus group meetings and subjects who did not, were explored. As shown in Table S1 no statistically significant differences in demographical characteristics or symptom severity scores, possibly indicating self-selection bias, were demonstrated. It appears that focus

group participants less often scored positive for depressive symptoms on the HADS, using a cut-off of ≥ 8 , and showed a slightly higher Mental Composite Score on the SF-36, than the subjects who did not participate in focus group meetings. Though these differences were not statistically significant, this may indicate that the focus group population did not perfectly reflect the heterogeneity of the general IBS population with regard to mental health. However, since the HADS is a screening tool using a cut-off value, a score < 8 does not mean that no depressive symptoms were present in our population at all. Furthermore, participants agreed on the importance of assessing psychological factors additionally to GI symptoms, and therefore, we are confident that the content of the developed ESM-PROM includes the relevant questions for symptom assessment in IBS patients. Another limitation might be the distribution of IBS subtypes amongst the participants, since only 12% (2/17) of subjects were defined as constipation-predominant IBS (IBS-C). Nevertheless, we believe that constipation-associated symptoms will as well be experienced by mixed-type (IBS-M) patients and so, we expect to have captured both diarrhea- and constipation-associated symptoms. Also, during the cognitive interviews IBS-C patients were attending and did not mention a lack of any constipation-specific items.

The number of subjects attending focus group meetings should be mentioned. The relatively small number of participants per focus group (*i.e.* 6-8) was chosen, since it is considered to invite all subjects to actively participate in the discussion on the one hand and to gain a variety of perspectives on the other hand.⁴² Furthermore, as discussed previously in this report, focus group meetings were arranged one by one until reaching saturation of input after four meetings and therefore, we feel that the total sample of 17 participants suffices with regard to this project. To date, successful focus group studies with similar sample sizes have been reported previously.^{35,43} Likewise, rather small numbers of participants were included for the cognitive interviews in each center, however, also in this phase of the project saturation of input was ensured. Since focus groups and the expert meeting already determined the content of the tool, cognitive interviews were particularly intended to evaluate patients' understanding with the individual items of the ESM-PROM with regard to linguistic issues (*i.e.* after forward-and-back translation). Therefore, the resulting number of cognitive interviews per center meets our expectations. With regard to level of education and employment status, the Dutch participants reflect the general Dutch population quite well. The Swedish and English participants were, however, relatively highly educated. This should be

taken into account when recruiting subjects for future studies regarding the developed questionnaire, since extra instructions on how to interpret and answer the questions might be needed in lower educated individuals.

In conclusion, we here report on the development of a disease-targeted PROM suitable for real-time assessment of abdominal pain and other gastrointestinal symptoms as well as daily life factors that may influence IBS symptoms. A smartphone application has been specifically developed to implement this PROM and future research will comprise further, multicenter validation of this instrument, with regard to the different languages and cross-cultural aspects. ESM appears particularly suitable for the evaluation of individual symptom patterns over the course of a predefined period of time and the identification of specific triggers of gastrointestinal symptoms in daily life. It can hereby function in supporting diagnostic as well as therapeutic trajectories, in clinical practice. Moreover, once ESM is demonstrated to be a valid and reliable PRO instrument in this particular context, it can be implemented in clinical research to assess the effect of therapeutic interventions in IBS and to stratify IBS patients based on their symptoms.

References

1. Choung RS, Locke GR, 3rd. Epidemiology of IBS. *Gastroenterol Clin North Am.* 2011;40(1):1-10.
2. Grundmann O, Yoon SL. Irritable bowel syndrome: epidemiology, diagnosis and treatment: an update for health-care practitioners. *J Gastroenterol Hepatol.* 2010;25(4):691-699.
3. Camilleri M, Lasch K, Zhou W. Irritable bowel syndrome: methods, mechanisms, and pathophysiology. The confluence of increased permeability, inflammation, and pain in irritable bowel syndrome. *Am J Physiol Gastrointest Liver Physiol.* 2012;303(7):G775-785.
4. Lovell RM, Ford AC. Global prevalence of and risk factors for irritable bowel syndrome: a meta-analysis. *Clin Gastroenterol Hepatol.* 2012;10(7):712-721.e714.
5. Canavan C, West J, Card T. Review article: the economic impact of the irritable bowel syndrome. *Aliment Pharmacol Ther.* 2014;40(9):1023-1034.
6. Simren M, Brazier J, Coremans G, et al. Quality of life and illness costs in irritable bowel syndrome. *Digestion.* 2004;69(4):254-261.
7. Drossman DA. The functional gastrointestinal disorders and the Rome III process. *Gastroenterology.* 2006;130(5):1377-1390.
8. Longstreth GF, Thompson WG, Chey WD, Houghton LA, Mearin F, Spiller RC. Functional bowel disorders. *Gastroenterology.* 2006;130(5):1480-1491.
9. Ford AC, Bertic P, Morgan DG, Bolino C, Pintos-Sanchez MI, Moayyedi P. Validation of the Rome III criteria for the diagnosis of irritable bowel syndrome in secondary care. *Gastroenterology.* 2013;145(6):1262-1270.e1261.
10. Mearin F, Lacy BE, Chang L, et al. *Bowel Disorders.* Gastroenterology. 2016.
11. U.S. Department of Health and Human Services FaDA, Center for Drug Evaluation and Research (CDER). *Guidance for Industry Irritable Bowel Syndrome - Clinical Evaluation of Drugs for Treatment.* 2012.
12. (EMA) EMA. *Guideline on the evaluation of medicinal products for the treatment of irritable bowel syndrome.* 2014.
13. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol.* 2008;4: 1-32.
14. 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.
15. Houtveen JH, Oei NY. Recall bias in reporting medically unexplained symptoms comes from semantic memory. *J Psychosom Res.* 2007;62(3):277-282.
16. Moskowitz DS, Young SN. Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology. *J Psychiatry Neurosci.* 2006;31(1):13-20.
17. Jamison RN, Raymond SA, Levine JG, Slawsby EA, Nedeljkovic SS, Katz NP. Electronic diaries for monitoring chronic pain: 1-year validation study. *Pain.* 2001;91(3):277-285.
18. Stone AA, Broderick JE, Schwartz JE, Shiffman S, Litcher-Kelly L, Calvanese P. Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. *Pain.* 2003;104(1-2):343-351.
19. Stone AA, Broderick JE, Shiffman SS, Schwartz JE. Understanding recall of weekly pain from a momentary assessment perspective: absolute agreement, between- and within-person consistency, and judged change in weekly pain. *Pain.* 2004;107(1-2):61-69.

20. 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.
21. 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.
22. 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.
23. Spiegel BM, Bolus R, Agarwal N, et al. Measuring symptoms in the irritable bowel syndrome: development of a framework for clinical trials. *Aliment Pharmacol Ther.* 2010;32(10):1275-1291.
24. Spiegel BM, Bolus R, Harris LA, et al. Characterizing abdominal pain in IBS: guidance for study inclusion criteria, outcome measurement and clinical practice. *Aliment Pharmacol Ther.* 2010;32(9):1192-1202.
25. Spiegel BM, Hays RD, Bolus R, et al. Development of the NIH Patient-Reported Outcomes Measurement Information System (PROMIS) gastrointestinal symptom scales. *Am J Gastroenterol.* 2014;109(11):1804-1814.
26. Marquis P, Lasch KE, Delgado-Herrera L, et al. Qualitative development of a patient-reported outcome symptom measure in diarrhea-predominant irritable bowel syndrome. *Clin Transl Gastroenterol.* 2014;5:e59.
27. Wiklund IK, Fullerton S, Hawkey CJ, et al. An irritable bowel syndrome-specific symptom questionnaire: development and validation. *Scand J Gastroenterol.* 2003;38(9):947-954.
28. Labus JS, Bolus R, Chang L, et al. The Visceral Sensitivity Index: development and validation of a gastrointestinal symptom-specific anxiety scale. *Aliment Pharmacol Ther.* 2004;20(1):89-97.
29. Mujagic Z, Keszthelyi D, Aziz Q, et al. Systematic review: instruments to assess abdominal pain in irritable bowel syndrome. *Aliment Pharmacol Ther.* 2015;42(9):1064-1081.
30. Verhagen SJ, Hasmi L, Drukker M, van Os J, Delespaul PA. Use of the experience sampling method in the context of clinical trials. *Evid Based Mental Health.* 2016;19(3):86-89.
31. Delespaul P. Assessing schizophrenia in daily life the experience sampling method.: UPM, Universitaire Pers Maastricht, Maastricht University; 1995.
32. Ludidi S, Mujagic Z, Jonkers D, et al. Markers for visceral hypersensitivity in patients with irritable bowel syndrome. *Neurogastroenterol Motil.* 2014;26(8):1104-1111.
33. Mujagic Z, Ludidi S, Keszthelyi D, et al. Small intestinal permeability is increased in diarrhoea predominant IBS, while alterations in gastroduodenal permeability in all IBS subtypes are largely attributable to confounders. *Aliment Pharmacol Ther.* 2014;40(3):288-297.
34. Morgan DL. *The Focus Group Guidebook. Vol 1.* California: Sage Publications; 1998.
35. Drossman DA, Chang L, Schneck S, Blackman C, Norton WF, Norton NJ. A focus group assessment of patient perspectives on irritable bowel syndrome and illness severity. *Dig Dis Sci.* 2009;54(7): 1532-1541.
36. A. KR. *Moderating Focus Groups. Vol 4.* California: Sage Publications; 1998.
37. (WHO) WHO. Process of translation and adaptation of instruments. 2015; http://www.who.int/substance_abuse/research_tools/translation/en/.
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. 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 Psychiatr Scand*. 2003;107(2):124-131.
40. Fehnel SE, Ervin CM, Carson RT, Rigoni G, Lackner JM, Coons SJ. Development of the Diary for Irritable Bowel Syndrome Symptoms to Assess Treatment Benefit in Clinical Trials: Foundational Qualitative Research. *Value Health*. 2017;20(4):618-626.
41. DeWalt DA, Rothrock N, Yount S, Stone AA. Evaluation of item candidates: the PROMIS qualitative item review. *Med Care*. 2007;45(5 Suppl 1):S12-21.
42. Rabiee F. Focus-group interview and data analysis. *Proc Nutr Soc*. 2004;63(4):655-660.
43. Carbone F, Holvoet L, Vandenberghe A, Tack J. Functional dyspepsia: outcome of focus groups for the development of a questionnaire for symptom assessment in patients suffering from postprandial distress syndrome (PDS). *Neurogastroenterol Motil*. 2014;26(9):1266-1274.

Supplemental material

Table S6.1 Demographical characteristics and symptom severity scores of participants of the Maastricht IBS-cohort, separate for 'focus group participation' versus 'no focus group participation'.

Maastricht IBS Cohort	Focus group participation (n=17)	No focus group participation (n=412)
Female sex , n (%)	13 (76.5)	295 (71.6)
Age , median [IQR]	58.0 [39.5-63.0]	45.0 [28.0-59.0]*
BMI , median [IQR]	26.5 [22.2-30.9]	24.1 [21.5-27.4]
IBS subtype , n (%)		
IBS-D	8 (47.1)	135 (32.8)
IBS-C	2 (11.8)	82 (19.9)
IBS-M	7 (41.2)	169 (41.0)
IBS-U	0	26 (6.3)
14-day mean symptom score , median [IQR]		
Abdominal pain	2.14 [1.64-2.64]	2.21 [1.57-2.86]
Abdominal discomfort	2.29 [2.04-2.69]	2.36 [1.92-2.93]
Abdominal bloating	2.07 [1.72-2.46]	2.09 [1.50-2.86]
Constipation	1.07 [1.00-1.59]	1.29 [1.00-1.89]
Diarrhea	1.14 [1.00-1.48]	1.21 [1.07-1.67]
HADS , n (%)		
Depression ≥8	1 (5.9)	86 (20.9)
Anxiety ≥8	5 (29.4)	154 (37.4)
SF-36 , median [IQR]		
Physical Composite Score (PCS)	41.21 [36.05-53.18]	42.63 [33.99-49.60]
Mental Composite Score (MCS)	54.34 [47.30-56.97]	48.67 [39.85-55.41]

Differences tested using Mann-Whitney U test for continuous data and χ^2 - or Fisher's exact test for categorical data. * Significant difference between 'focus group participation' and 'no focus group participation' ($p < 0.05$).

Table S6.2 Set of items for the ESM-PROM developed during phase I: Initial item selection.

Phase I: Initial item selection		Answer scale
	Physical status	
1	I feel hungry	1 (none) - 7 (very much)
2	I feel tired	1 (none) - 7 (very much)
3	I do not feel well	1 (none) - 7 (very much)
4	I am having palpitations	1 (none) - 7 (very much)
5	I am sweating	1 (none) - 7 (very much)
6	I feel numbness/tingling	1 (none) - 7 (very much)
7	I am short of breath	1 (none) - 7 (very much)
8	I have a choking sensation	1 (none) - 7 (very much)
9	I feel chest pain	1 (none) - 7 (very much)
10	I feel dizzy	1 (none) - 7 (very much)
11	I am having hot flashes	1 (none) - 7 (very much)
12	I am having chills	1 (none) - 7 (very much)
13	I feel upper abdominal pain	1 (none) - 7 (very much)
14	I feel lower abdominal pain	1 (none) - 7 (very much)
15	I am feeling heartburn	1 (none) - 7 (very much)
16	I am feeling sick	1 (none) - 7 (very much)
17	I feel bloated	1 (none) - 7 (very much)
18	I am having intestinal gas	1 (none) - 7 (very much)
19	I am suffering from burping	1 (none) - 7 (very much)
20	I have been physically active since the previous beep signal	1 (none) - 7 (very much)

Table S6.2 (continued)

Defecation		Answer scale
21	Have you defecated since the previous beep signal?	Yes / No
22	If yes, what did the stool look like?	Bristol Stool Score 1 - 7
Mood		
23	I feel enthusiastic	1 (none) - 7 (very much)
24	I feel sad	1 (none) - 7 (very much)
25	I feel irritated	1 (none) - 7 (very much)
26	I feel content	1 (none) - 7 (very much)
27	I feel stressed	1 (none) - 7 (very much)
28	I feel restless	1 (none) - 7 (very much)
29	I feel happy	1 (none) - 7 (very much)
30	I feel lonely	1 (none) - 7 (very much)
31	I feel strong	1 (none) - 7 (very much)
32	I feel anxious	1 (none) - 7 (very much)
33	I feel relaxed	1 (none) - 7 (very much)
34	I feel worried	1 (none) - 7 (very much)
35	I feel inspired	1 (none) - 7 (very much)
36	I feel disappointed	1 (none) - 7 (very much)
37	I feel insecure	1 (none) - 7 (very much)
38	I feel cheerful	1 (none) - 7 (very much)
39	I feel guilty	1 (none) - 7 (very much)
40	I am afraid to lose self-control	1 (none) - 7 (very much)
41	I can not drop my thoughts	1 (none) - 7 (very much)
42	I am worrying	1 (none) - 7 (very much)
43	Overall, I feel good	1 (none) - 7 (very much)
Environmental and social context		
44	Where am I?	At home, at someone else's house, work/school, public place, on the go, somewhere else
45	Who is with me?	Partner, friends, residents, colleagues, family (not residents), acquaintances, unknown/others, nobody
46	I find their company pleasant	1 (none) - 7 (very much)
47	I feel accepted	1 (none) - 7 (very much)
48	I feel uncomfortable in this situation	1 (none) - 7 (very much)
49	I would rather be alone	1 (none) - 7 (very much)
50	We are doing something together	1 (none) - 7 (very much) Resting, work/school, housekeeping/grocery, hygiene/self-care, eating/drinking, relaxing, something else
51	What am I doing (right before the beep)?	1 (none) - 7 (very much)
52	I would rather like to do something else	1 (none) - 7 (very much)
53	I am active right now	1 (none) - 7 (very much)
54	I like to do this activity	1 (none) - 7 (very much)
55	It takes effort to do this	1 (none) - 7 (very much)
56	I am good at this	1 (none) - 7 (very much)
57	This is a challenge for me	1 (none) - 7 (very much)

Table S6.2 (continued)

Think of the most important event since the previous beep signal:	
58	This event was: -3 (very unpleasant) – 3 (very pleasant)
59	This event was: -3 (very unimportant) – 3 (very important)
Drug use	
60	I used ... since the previous beep signal <i>Caffeine, nicotine, medication, alcohol, cannabis, other drugs, none of those</i>

Red color indicates items that were removed from the instrument during phase II: Focus group interviews.

Table S6.3 Set of items for the ESM-PROM, adjusted according to patients' input during phase II: Focus group interviews.

Phase II: Focus group interviews		
	Physical status	Answer scale
1	I feel hungry	0 (none) - 10 (very much)
2	I feel tired	0 (none) - 10 (very much)
3	I am having palpitations	0 (none) - 10 (very much)
4	I am sweating	0 (none) - 10 (very much)
5	I am short of breath	0 (none) - 10 (very much)
6	I feel dizzy	0 (none) - 10 (very much)
7	I am having abdominal pain	0 (none) - 10 (very much)
7a	This pain is located in the following part(s) of my abdomen	Figure abdominal regions
8	I am feeling heartburn	0 (none) - 10 (very much)
9	I am feeling sick	0 (none) - 10 (very much)
10	My abdomen feels bloated	0 (none) - 10 (very much)
11	I am having intestinal gas	0 (none) - 10 (very much)
12	I am suffering from burping	0 (none) - 10 (very much)
13	I am having rumbling sounds coming from my abdomen	0 (none) - 10 (very much)
14	I have been physically active since the previous beep signal	0 (none) - 10 (very much)
Defecation		
15	How many times did you open your bowels since the previous beep signal?	0, 1, 2, 3, 4, more than 4 times
16	What did the stool look like?	Bristol Stool Score 1 - 7
17	I have felt an urge to open my bowel since the previous beep signal	0 (none) - 10 (very much)
18	The bowel movement(s) was/were associated with a relief of my abdominal symptoms	0 (none) - 10 (very much)
19	I have been able to postpone a bowel movement since the previous beep signal	0 (none) - 10 (very much)
Mood and psychological factors		
20	I feel sad	0 (none) - 10 (very much)
21	I feel irritated	0 (none) - 10 (very much)
22	I feel content	0 (none) - 10 (very much)
23	I feel stressed	0 (none) - 10 (very much)
24	I feel lonely	0 (none) - 10 (very much)
25	I feel anxious	0 (none) - 10 (very much)
26	I feel relaxed	0 (none) - 10 (very much)
27	I feel insecure	0 (none) - 10 (very much)
28	I feel cheerful	0 (none) - 10 (very much)
29	I am worrying	0 (none) - 10 (very much)

Table S6.3 (continued)

Context and environment	
30	Where am I? <i>At home, at someone else's house, work/school, public place, on the go, somewhere else</i>
31	Who is with me? <i>Partner, friends, residents, colleagues, family (not residents), acquaintances, unknown/others, nobody</i>
32	I find this company (un)pleasant <i>-5 (very unpleasant) – 5 (very pleasant)</i>
33	I find this situation (un)pleasant <i>-5 (very unpleasant) – 5 (very pleasant)</i>
34	I feel (un)comfortable in this situation <i>-5 (very uncomfortable) – 5 (very comfortable)</i>
35	What am I doing (right before the beep)? <i>Resting, work/school, housekeeping/grocery, hygiene/self-care, eating/drinking, relaxing, sports, something else</i>
36	I like to do this activity <i>0 (none) - 10 (very much)</i>
37	My complaints restrict me in doing my normal activities right now <i>0 (none) - 10 (very much)</i>
	Think of the most important event since the previous beep signal:
38	This event was: <i>-5 (very unpleasant) – 5 (very pleasant)</i>
39	This event was: <i>-5 (very unimportant) – 5 (very important)</i>
Nutrition and drug use	
40	I used ... since the previous beep signal <i>Caffeine (coffee) / theine (tea), nicotine (smoking), medication, alcohol, drugs, none of those</i>
40a	This was following medication: <i>Medication for abdominal pain, other pain relief, medication for stomach acidity, medication for nausea, medication to stop diarrhea, medication for constipation, something else</i>
41	I ate ... since the previous beep signal <i>Wheat, spicy food, fatty food, cabbage/leek/onion, dairy products, none of these</i>
42	I ate ... ago <i>Less than 15 min; 15-30 min; 30 min-1 hr; more than 1 hr</i>

Green color indicates items that were newly added to the instrument during phase II: Focus group interviews.

Red color indicates items that were removed from the instrument during phase III: Expert meeting.

Table S6.4 Set of items for the ESM-PROM, adjusted according to experts' input during phase III: Expert meeting.

Phase III: Expert meeting		
	Physical status	Answer scale
1	I am having abdominal pain	0 (none) - 10 (very much)
1a	This pain is located in the following part(s) of my abdomen:	Figure abdominal regions
2	I am having intestinal gas	0 (none) - 10 (very much)
2a	The intestinal gas is causing discomfort	0 (none) - 10 (very much)
3	I am having rumbling sounds coming from my abdomen	0 (none) - 10 (very much)
4	My abdomen feels bloated	0 (none) - 10 (very much)
4a	The bloating makes me feel uncomfortable	0 (none) - 10 (very much)
4b	The bloating is accompanied by a swollen abdomen	0 (none) - 10 (very much)
5	I have the urge to open my bowels	0 (none) - 10 (very much)
6	I am feeling sick	0 (none) - 10 (very much)
7	I am suffering from burping	0 (none) - 10 (very much)
8	I am feeling heartburn	0 (none) - 10 (very much)
9	I am feeling full	0 (none) - 10 (very much)
10	I am having palpitations	0 (none) - 10 (very much)
11	I am sweating	0 (none) - 10 (very much)
12	I am short of breath	0 (none) - 10 (very much)
13	I feel dizzy	0 (none) - 10 (very much)
14	My muscles are hurting	0 (none) - 10 (very much)
15	My joints are hurting	0 (none) - 10 (very much)
16	I feel an urge to pass urine	0 (none) - 10 (very much)
Defecation		
17	Since the last beep, I have had the feeling that I had to open my bowels ... time(s)	0, 1, 2, 3, 4, more than 4 times
17a	Since the last beep, I have actually opened my bowels ... time(s)	0, 1, 2, 3, 4, more than 4 times
17aI	It looked like this: ...	Bristol Stool Score 1 - 7
17aII	I had to strain	0 (none) - 10 (very much)
17aIII	It feels like my bowels are not completely empty	0 (none) - 10 (very much)
Mood and psychological factors		
18	I am feeling good	0 (none) - 10 (very much)
19	I am feeling low	0 (none) - 10 (very much)
20	I am feeling anxious	0 (none) - 10 (very much)
21	I am feeling irritated	0 (none) - 10 (very much)
22	I am feeling stressed	0 (none) - 10 (very much)
23	I am feeling relaxed	0 (none) - 10 (very much)
24	I am worried	0 (none) - 10 (very much)
	Context and environment	
25	Where am I?	At home, at someone else's home, work/school, public place, on my way, somewhere else

Table S6.4 (continued)

Context and environment (continued)	
	<i>Resting, work/school work, household work/shopping, hygiene/self-care, eating/drinking, relaxing/recreation, sports, travelling, something else</i>
26	What am I doing (just before the beep)? <i>-5 (extremely uncomfortable) –</i>
27	I feel (un)comfortable doing this <i>5 (extremely comfortable)</i>
28	My symptoms are limiting my current activities <i>0 (none) - 10 (very much)</i>
	<i>Partner/children, friends, housemates, colleagues, family (other than those who live in your house), acquaintances, strangers/others, no one</i>
29	Who is with me? <i>-5 (extremely unpleasant) - 5 (extremely pleasant)</i>
29a	I find this company (un)pleasant <i>pleasant</i>
Nutrition and drug use	
	<i>Breakfast, lunch, dinner, a snack, none of these</i>
30	I have eaten ... since the last beep <i>More than 15 min., 15–30 min.,</i>
31	I ate ... ago <i>30 min.-1 hour, more than 1 hour</i>
	<i>Caffeine (coffee) / theine (tea), nicotine (smoking), alcohol, drugs, medication, none of these</i>
32	Since the last beep I have used ... <i>Medication for abdominal pain, other pain relief, medication for stomach acidity, medication for nausea, medication to stop diarrhea, medication for constipation,</i>
32a	This was following medication: <i>something else</i>

Green color indicates items that were newly added to the instrument during phase III: Expert meeting.

Chapter 7

Development of a real-time patient-reported outcome measure for symptom assessment in patients with functional dyspepsia using the experience sampling method

Fabiënne G.M. Smeets, Daniel Keszthelyi, Lisa Vork, Jan Tack, Nick J. Talley, Magnus Simrén, Qasim Aziz, Alex C. Ford, José M. Conchillo, Joanna W. Kruimel, Jim Van Os, Ad A.M. Masclee

Neurogastroenterology & Motility. 2019; 31(2).

Abstract

INTRODUCTION Patient-reported outcome measures (PROMs) are used to assess symptoms in patients with functional dyspepsia. Current end-of-day questionnaires have several limitations including sensitivity to recall and ecological bias. The Experience Sampling Method (ESM) is characterized by random and repeated assessments across momentary states in daily life and therefore less sensitive to these limitations. This study describes the development of a novel PROM based on ESM technology.

METHODS An initial draft of the PROM was developed based on literature. Focus group interviews with FD patients according to Rome IV criteria, and an expert meeting with international opinion leaders in the field of functional gastrointestinal disorders were conducted in order to select items for the PROM. Cognitive interviews were performed to evaluate patients' understanding of the selected items and to create the definitive PROM.

RESULTS A systematic literature search revealed 59 items across four domains (*i.e.* physical status, mood and psychological factors, context and environment, and nutrition, medication and substance use). After patient focus group interviews and an international expert meeting, the number of items was reduced to 33. Cognitive interviews resulted in some minor linguistic changes in order to improve patients' understanding.

CONCLUSIONS A novel digital ESM-based PROM for real-time symptom assessment in patients with functional dyspepsia was developed. This novel PROM has the potential to identify individual symptom patterns and specific triggers for dyspeptic symptoms, and optimize treatment strategies.

Introduction

Functional dyspepsia (FD) is a common functional gastrointestinal disorder with an estimated prevalence of 8-12% in the general population.¹⁻³ Symptom presentation of dyspeptic patients is heterogeneous although four core symptoms have been defined according to the Rome IV criteria: postprandial fullness, early satiation, epigastric burning and epigastric pain.^{4,5} These symptoms lead to impaired quality of life, reduced work productivity and increased health care costs, which underlines the need for (development of) effective treatment options.⁵⁻⁷

Functional dyspepsia is a symptom-based diagnosis and patient-reported outcome measures (PROMs) are used to assess treatment efficacy.⁸ In a recently published systematic review, twenty available retrospective outcome measures were described for the evaluation of dyspeptic symptoms.⁹ However, these outcome measures do not fulfill all criteria for adequate psychometric validation, as defined by regulatory authorities.

Moreover, several limitations of retrospective end-of-day questionnaires are apparent. First, retrospective outcome measures are prone to recall bias as retrieval of information is based on autobiographical memory.¹⁰ Second, dyspeptic symptoms vary over time due to the influence of certain circumstances and triggers (*e.g.* food intake and psychosocial factors).^{11,12} Lack of ecological validity may occur when questionnaires are completed in another environment or situation, compared with situations in which symptoms were triggered.¹³ Third, non-compliance is a major limitation of retrospective paper questionnaires, and this could potentially be eliminated by use of an electronic sampling method.¹⁴ These limitations of available retrospective questionnaires together with the lack of a single universally accepted PROM, underline the need for development of a novel PROM.

The Experience Sampling Method (ESM) is an attractive electronic method for real-time assessment of symptoms, which may overcome some of the limitations of retrospective questionnaires.¹³ Random, repeated assessments are used for several consecutive days to capture symptom variability over time, and this takes into account contextual, social and psychological factors, which might influence dyspeptic symptoms.¹³ Although ESM has been applied in several disorders (*e.g.* mental disorders), use in patients with functional gastrointestinal diseases is currently limited to two studies in patients with irritable bowel syndrome (IBS).¹⁵⁻¹⁷ Both studies reported higher symptom

scores in end-of-day diaries when compared to day-average scores of momentary assessments with ESM.^{16,17}

However, the ESM has not been previously applied in patients with functional dyspepsia. In this report we describe the development of a novel PROM based on ESM technology for real-time assessment of symptoms in patients with functional dyspepsia.

Methods

Development of a novel ESM-based PROM for symptom assessment in patients with functional dyspepsia was undertaken in several stages according to the FDA guidelines for the development of PROMs.⁸

Phase I: Item selection

A conceptual framework of theoretical constructs was developed for assessment of symptoms in patients with functional dyspepsia. For development of this framework, an extensive systematic literature search was performed to obtain available retrospective questionnaires for selection of FD-specific items.⁹ Moreover, ESM-specific constructs (*i.e.* psychological status, social, and contextual factors) were derived from previously used ESM questionnaires at the departments of Gastroenterology and Hepatology, and Psychiatry and Psychology of Maastricht University Medical Center (MUMC+).^{17,18}

Phase II: Focus group interviews

Selection of participants for focus groups

Consecutive ambulatory patients between the age of 18 and 75 years with a diagnosis of functional dyspepsia were recruited at the outpatient department of Gastroenterology and Hepatology at Maastricht University Medical Center (MUMC+). Patients were diagnosed with functional dyspepsia according to Rome IV criteria and an upper gastrointestinal endoscopy was performed up to 12 months prior to the focus group interview in order to exclude organic abnormalities.⁵ Patients with FD and a comorbid diagnosis of irritable bowel syndrome (IBS) were not excluded due to the considerable overlap of both disorders, and to acquire a representative sample of the general population.

Participants spoke Dutch as their mother tongue as focus groups were conducted in Dutch. All participants gave written informed consent and the study protocol was approved by the medical ethical committee of Maastricht University Medical Center (METC 17-4-056).

Conducting and moderating focus group interviews

Patient focus group sessions were organized in order to identify a relevant set of items for symptom assessment in functional dyspepsia and discuss practical issues of the ESM-procedure.

Eligible participants were invited to participate in a focus group meeting which took approximately 90 minutes. Focus group interviews were planned and continued until saturation of input was achieved.¹⁹ Preferably, five to ten participants were included per focus group, in order to obtain the full spectrum of perspectives.^{19,20}

After an introduction about the aim of the study, one moderator (F.S.) and one assistant-moderator (L.V.) guided the focus groups with use of a power point presentation to ensure adequate consistency in interview content. In order to minimize the probability of bias induced by the moderator, the first part of the focus group consisted of an open discussion in which participants were questioned about which items they considered to be essential in an outcome measure for dyspeptic patients. This section of the focus group aimed to (1) acquire information about experiences of gastrointestinal symptoms, (2) obtain information about the influence of symptoms on daily life, and (3) identify which factors or triggers influence dyspeptic symptoms.

During the second part of the focus group, 59 items selected from previously used outcome measures and ESM questionnaires were presented on the power point presentation, and participants were asked to criticize these items. They were asked to consider which items were relevant in a real-time symptom assessment method for functional dyspepsia, and which items could be excluded. Moreover, patients were asked to identify incomprehensible items and provide alternative terms or descriptions. Care was taken to obtain feedback from all participants on each item, and the number of participants that considered each item relevant was assessed. At the end of each domain open-ended questions were asked to evaluate whether additional symptoms required mentioning.

Fatigue and sleeping problems are frequently mentioned by patients with functional gastrointestinal complaints. Assessment of sleep and fatigue is however not suitable for repeated assessment over the day. In line with the

recently developed ESM-based IBS PROM, we intended to include a ‘morning’ questionnaire for assessment of sleep once a day.

Besides evaluation of items to be included in an outcome measure for assessment of dyspeptic symptoms, several practical components of the ESM procedure were discussed. The proposed 11-point (0-10) Numeric Rating Scale (NRS), which is currently recommended for use as endpoint in clinical trials of patients with IBS, was discussed.²¹ Second, the number of assessments during the day and the time spent for completing each assessment (*i.e.* 3-5 minutes) were discussed to assess the feasibility and burden of the ESM-based outcome measure. Random and repeated symptom assessment is the key feature of ESM. Ten times a day a beep signal is produced by the application on the smartphone between 07.30 and 22.30 hour. Auditory cues are submitted in 90-minute blocks, and the minimum interval between two beeps is 15 minutes. Beeps occur completely random without relation to meal intake in order to prevent anticipation of symptom assessment.

Phase III: Expert meeting

After conducting the focus groups, a teleconference meeting was organized with international experts in the fields to discuss the relevance of the individual items that were selected from the systematic literature search and patient focus groups. Experts criticized the items regarding their relevance and suitability for a real-time symptom assessment method. Based on this discussion, a definitive draft with questions for the ESM-based outcome measure was developed.

Phase IV: Cognitive interviews

Interviews focused on cognitive understanding were performed to evaluate patients’ grasp of the items included in the definitive draft. Again, native Dutch-speaking patients with functional dyspepsia according to the Rome IV criteria were included, but who had not previously taken part in the focus group interviews. The definitive ESM-draft was presented to the participants on paper. Participants were asked to read the items, speak them out loud, and consider whether the meaning of the items was clear (*i.e. verbal probing*).²² Moreover, participants were asked for recommendations to improve the outcome measure. Cognitive interviews were performed on a one-to-one basis, and were continued until additional cognitive interviews did not lead to substantial suggestions or recommendations from the participants.

Data analysis

Analysis of focus group data

Focus groups were recorded by a voice-recorder and notes were taken by both the moderator and assistant-moderator. Immediately after the focus group sessions, a debriefing was performed between the moderator and assistant-moderator to review and summarize the acquired information and observations.^{19,20} In addition, voice recordings were transcribed verbatim and summarized in a database together with the moderators' notes. This database was used to identify saturation of input, which is achieved when the core symptoms discussed in the sessions are stable, and subsequent focus groups only produce repetitive information.^{19,23} Finally, the database was used to decide which items should be included in the definitive draft instrument.

Analysis of cognitive grasp interviews

For the analysis of interviews focusing on cognitive grasp, a database was constructed to transcribe and summarize participants' statements, in order to facilitate comparison of interpretations for all individual items. This information was used to decide whether any modification of items was necessary.²²

Results

Phase I: Item selection

Based on a systematic literature search and collection of previously used ESM questionnaires, 59 items were included in the initial ESM instrument. These items were divided into four domains: (1) physical status, (2) mood, (3) context and environment, and (4) nutrition, medication and substance use (Figure 7.1). The "physical status" domain could be divided into three categories: (1) upper gastrointestinal symptoms, including symptoms of dyspepsia, gastroesophageal reflux disease, and other disorders (*e.g.* nausea, belching, vomiting), (2) lower gastrointestinal symptoms, and (3) general physical complaints.

Phase II: Focus group interviews

Study population

Forty-four patients with functional dyspepsia according to Rome IV criteria were invited for the focus group meetings. Eighteen patients confirmed their participation, and fourteen patients were present during three different focus group meetings (group 1: n=6, group 2: n=3, group 3: n=5). Family circumstances (n=2) and lack of time (n=2) were reasons for cancelation. Patient characteristics are described in Table 7.1.

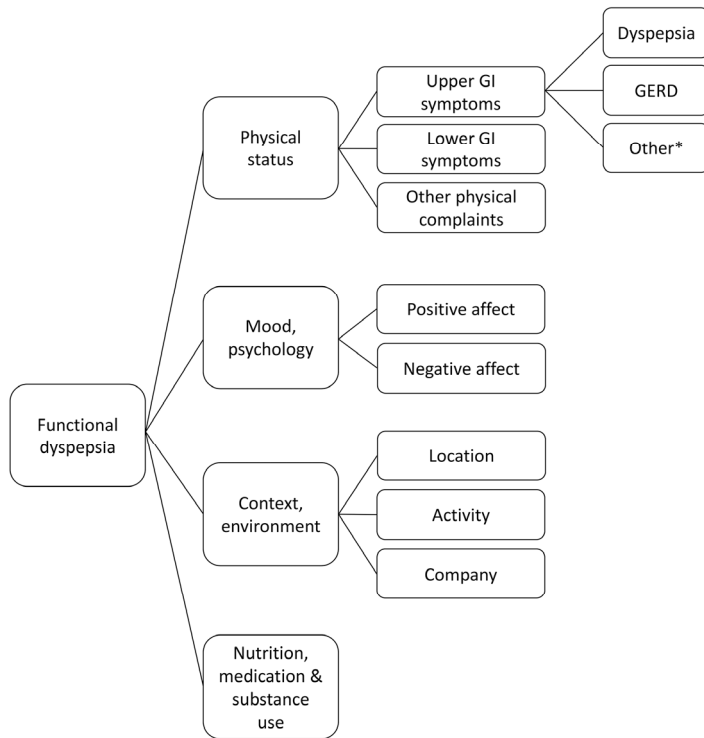


Figure 7.1 Conceptual framework for development of a novel patient-reported outcome measure using the Experience Sampling Technology.

* Nausea, vomiting, belching. GI = gastrointestinal, GERD = gastroesophageal reflux disease.

Table 7.1 Baseline characteristics of patients which participated in focus group interviews and cognitive interviews.

	Participants focus group interviews (n=14)	Participants cognitive interviews (n=6)
Age in years , median (range)	64 (39-73)	47 (27-58)
Female , n (%)	10 (71%)	5 (83%)
FD-subtype based on Rome IV criteria , n (%)		
FD-PDS*	1 (7%)	3 (50%)
FD-EPS‡	6 (43%)	1 (17%)
Mixed subtype	7 (50%)	2 (33%)
Comorbid diagnosis of IBS , n (%)	3 (21%)	1 (17%)

* FD-PDS: functional dyspepsia of the postprandial distress subtype; ‡ FD-EPS: functional dyspepsia of the epigastric pain subtype.

Focus group interviews

Physical status

- Upper abdominal symptoms

Postprandial fullness, early satiation, epigastric pain, and epigastric burning were considered important symptoms, and were described by the majority of patients (86%, 64%, 64% and 86% respectively). Participants considered two aspects essential for real-time symptom assessment of pain: (1) pain severity and (2) location of pain. Assessment of pain severity using an 11-point NRS was considered adequate. To localize pain, participants suggested to add a schematic picture of the abdomen in which study subjects could select separate abdominal regions. In addition, several participants suggested description of the character of the pain (*e.g.* cramp-like, dull) using multiple-choice options.

In addition to the four core FD symptoms, participants agreed on the importance of scoring additional frequently occurring upper gastrointestinal symptoms including bloating (86%), nausea (100%), and belching (93%). The open discussion revealed that bloating was sometimes considered equivalent to postprandial fullness. After clarification of the meaning of bloating (*i.e.* an unpleasant sensation of gaseous distension), patients considered this an essential item for real-time symptom assessment separate from postprandial fullness. Not only bloating, but also nausea required some further explanation as several participants correlated nausea with general malaise, instead of a desire to vomit. Heartburn was described by 12 patients (86%) and was considered distinct from epigastric pain.

- Lower abdominal symptoms
In addition to upper gastrointestinal symptoms, four participants reported lower abdominal pain and three patients had a comorbid diagnosis of IBS according to the Rome IV criteria. As a consequence, assessment of lower abdominal pain, defecation frequency, and consistency were considered important by four and three patients respectively.
- General complaints
Several non-abdominal complaints were considered relevant by the majority of participants including palpitations, sweating, shortness of breath, and dizziness. Moreover, fatigue was reported by all participants and was considered an essential question. However, due to the often long lasting nature of fatigue, this item was considered not eligible for repeated assessment. However, assessment of sleep with a 'morning' questionnaire once a day was considered important by participants. Complaints reported by individual subjects were headache, weight loss, and fainting.

Mood and psychological factors

Assessment of psychological factors was considered important by all subjects, as they often experienced an association between psychological factors and gastrointestinal symptoms. 'I feel anxious' (79%), 'I feel irritated' (79%), and 'I feel stressed' (100%) were considered important negative emotions; whereas 'I feel relaxed' (100%) was considered a relevant positive affect. Participants discussed the item 'I feel excited' as several patients found this item too positive and preferred 'I feel good right now'.

Context and environment

Contextual and environmental factors were considered relevant as these might influence the presence and/or severity of gastrointestinal symptoms. Questions with regard to location ('Where am I?'), activity ('What am I doing?'), and company ('Who is with me?'), with multiple choice answers, were considered important by all participants. In addition, the majority of patients (86%) found it important to ask whether participants were restricted in their daily activities by current complaints.

Nutrition, medication and substance use

Participants stressed that dietary factors and medication use should be taken into account when assessing dyspeptic symptoms. Two components of nutrition were considered essential, namely (1) type of food products, and (2) amount of food intake. All subjects agreed on the question 'Since the last beep I used ...' with categorical answers (*i.e.* breakfast, lunch, dinner, or a snack). As several types of food and substance use can trigger symptoms, the question 'What did you use since the last beep?' with categorical answer options (*i.e.* coffee, tea, nicotine, drugs, medication, carbonated beverages) was considered appropriate to all participants. In patients who took medication, a categorical answer option was preferable (*i.e.* painkillers for abdominal pain, painkillers for another cause, acid suppressive medication, anti-emetics, or something else). Although early satiety is a core symptom of FD, assessment of the amount of food intake is difficult. We suggested 'I was able to finish a normal sized meal', which was appropriate for participants.

Practical issues

Several practical issues relating to the ESM procedure were discussed. An 11-point NRS was considered adequate and appropriate to grade the presence and severity of individual symptoms. Divergent opinions were present with regard to the proposed number of 10 measurements per day. A subgroup of participants indicated that 10 assessments per day would be quite burdensome and suggested that five to six measurements per day would be more acceptable. However, other participants considered it an investment in order to get additional information about their disorder and potentially optimize treatment.

Moreover, participants indicated that a 3 to 5-minute period was acceptable for completing each questionnaire. In order to reduce the time needed to complete the questionnaires, several participants suggested the use of an algorithm, in order to skip questions if possible. For instance, if participants answer '0 times' on the question 'How many times did you defecate since the previous assessment?' then the question with regard to stool consistency could be omitted.

Phase III: International expert meeting

All items resulting from the previous phases of the project were discussed by international experts in the field to assess the relevance for inclusion of individual items in an ESM-based FD-PROM.

With regard to upper gastrointestinal symptoms, several items were excluded as they were considered not relevant for inclusion in an FD-specific questionnaire (*e.g.* dysphagia) or were not eligible for repeated assessments (*e.g.* loss of appetite). In addition, some nuances were added for several items. For instance, patients in focus group interviews could not agree whether 'I am having a full feeling' and 'I am having a heavy feeling' were considered as identical or separate symptoms. Therefore, experts suggested to include both items and evaluate them in a future validation study. Moreover, experts decided to merge 'nausea' and 'the feeling that I have to vomit', and recommended that a schematic pictogram of the abdomen should be given to patients before the start of the ESM-questionnaire, in order to define the 'upper' and 'lower' abdomen.

Inclusion of items related to the lower gastrointestinal tract was extensively debated, as opinions of the international experts were heterogeneous. Although the goal was to develop an FD-specific PROM, considerable overlap is present between FD and IBS, which was also demonstrated in the patient focus groups, as 21% had a comorbid diagnosis of IBS. As it was supposed that it would be beneficial to evaluate the (potential) relation between upper and lower gastrointestinal symptoms with this ESM-principle, experts agreed with inclusion of the most important lower gastrointestinal items (*i.e.* lower abdominal pain, number and consistency of bowel movements, and incomplete evacuation).

The international experts agreed on inclusion of general physical complaints, as previous studies found associations between dyspeptic symptoms and somatic complaints.

Selection and inclusion of psychological items was extensively discussed as these ESM-psychological items have not been validated before in patients with gastrointestinal disorders. Item selection was therefore performed based on a recent publication evaluating psychometric validation properties of the ESM in patients with mental disorders, the previously developed ESM-IBS questionnaire, together with patients' opinion in focus group interviews.^{18,24}

Contextual items were discussed as well, but no changes were made. Moreover, experts agreed that nutritional items were of special interest with regard to dyspeptic symptoms. Although patients in the focus groups mentioned specific

kinds of food as triggers for dyspeptic symptoms (*e.g.* onions, spicy food), evaluation of associations between dyspeptic symptoms and specific kinds of food was considered beyond the scope of this project.

In conclusion, focus group interviews and an international expert meeting resulted in a reduction of the number of included items from 59 to 33, categorized in the before-mentioned four domains.

Phase IV: Cognitive grasp interviews

Demographic characteristics of the participants are shown in Table 7.1. One participant had a comorbid diagnosis of IBS according to Rome IV criteria. Only a few minor modifications were suggested to improve patients' understanding with the items. For instance, the wording of items regarding dyspnea and chest pain were modified.

The final ESM-based instrument for repeated symptom assessment is shown in Table 7.2. In addition, the morning questionnaire for assessment once a day is shown in Table 7.3. Content validity was assessed according to the following items: (1) development of a conceptual framework, (2) extensive item generation by a systematic literature search and organisation of patient focus group interviews in the target population of FD patients according to Rome IV criteria with use of open-ended questions, continuation until saturation of input was achieved and verbatim transcription, (3) discussion of practical issues of the novel ESM-based PROM during focus group interviews (*e.g.* digital mode of data collection, number of assessments per day, duration of data collection per day, recall period and response scale), (4) teleconference with international experts in the field of functional dyspepsia and ESM, and (5) cognitive interviews in order to assess the participants' understanding of the PROM and evaluate the comprehensiveness of content.

Table 7.2 Definite set of items for the ESM-PROM.

Physical status – Upper abdomen		Answer scale
1	I am having a full feeling in my upper abdomen	0 (none) - 10 (very much)
2	I am having a heavy feeling in my upper abdomen	0 (none) - 10 (very much)
3	My abdomen feels bloated	0 (none) - 10 (very much)
4	I am having pain in my upper abdomen	0 (none) - 10 (very much)
5	I am having a burning feeling in my upper abdomen	0 (none) - 10 (very much)
6	I am having pain in my lower abdomen	0 (none) - 10 (very much)
7	I feel nauseous	0 (none) - 10 (very much)
7a	Since the last beep, I have actually vomited ... time(s)	0, 1, 2, 3, 4, 5, more than 5 times
8	I am suffering from bothersome burping	0 (none) - 10 (very much)
9	I am having a burning feeling behind the breastbone	0 (none) - 10 (very much)
10	I am bringing up stomach contents into the mouth and/or nose	0 (none) - 10 (very much)
Nutrition, medication, and substance use		
11	I have eaten ... since the last beep	<i>Breakfast, lunch, dinner, a snack, none of these</i>
11a	I ate ... ago	<i>Less than 15 min., 15–30 min., 30 min.-1 hour, more than 1 hour</i>
11b	I am able to finish a normal sized meal	0 (none) - 10 (very much)
12	Since the last beep I have used ...	<i>Caffeine (coffee) / theine (tea), nicotine (smoking), alcohol, drugs, medication, carbonated beverages, milk products, fruit juices, none of these</i>
12a	This was medication for:	<i>abdominal pain, other pain relief, stomach acidity, nausea, diarrhea, constipation, something else</i>
Physical status – General complaints		
13	I am having palpitations	0 (none) - 10 (very much)
14	I am sweating	0 (none) - 10 (very much)
15	I am short of breath	0 (none) - 10 (very much)
16	I feel dizzy	0 (none) - 10 (very much)
17	I have pain on my chest	0 (none) - 10 (very much)
18	I feel tired	0 (none) - 10 (very much)
19	Since the last beep, I have had the feeling that I had to open my bowels ... time(s)	0, 1, 2, 3, 4, 5, more than 5 times
20	Since the last beep, I have actually opened my bowels ... time(s)	0, 1, 2, 3, 4, 5, more than 5 times
20a	It looked like this:	<i>Bristol Stool Form Scale</i>
20b	I had to strain	0 (none) - 10 (very much)
20c	It feels like my bowels are not completely empty	0 (none) - 10 (very much)
Psychological aspects		
21	I am feeling good	0 (none) - 10 (very much)
22	I am feeling relaxed	0 (none) - 10 (very much)
23	I am feeling low	0 (none) - 10 (very much)
24	I am feeling anxious	0 (none) - 10 (very much)
25	I am feeling irritated	0 (none) - 10 (very much)
26	I am feeling stressed	0 (none) - 10 (very much)
27	I am worried	0 (none) - 10 (very much)

Table 7.2 (continued)

Context and environment		
28	Where am I?	<i>At home, at someone else's home, work/school, public place, on my way, somewhere else</i>
29	What am I doing (just before the beep)?	<i>Resting, work/school, household work/shopping, hygiene/self-care, eating/drinking, relaxing/recreation, sports, travelling, something else</i>
30	I feel (un)comfortable doing this	<i>-5 (extremely uncomfortable) – +5 (extremely comfortable)</i>
31	My symptoms are limiting my current activities	<i>0 (none) - 10 (very much)</i>
32	Who is with me?	<i>Partner/children, friends, housemates, colleagues, family (other than those who live in your house), acquaintances, strangers/others, no one</i>
33	I find this company (un)pleasant	<i>-5 (extremely unpleasant) – +5 (extremely pleasant)</i>

Table 7.3 Definitive set of items for the morning questionnaire for assessment once a day.

Morning questionnaire		Answer scale
1	How long did it take you to fall asleep last night?	<i>0-5 min.; 5-15 min.; 15-30 min.; 30-45 min.; 45 min.-1 hour; 1-2 hours; 2-4 hours; more than 4 hours</i>
2	Last night, how many times did you wake up during the night?	<i>0, 1, 2, 3, 4, 5, more than 5 times</i>
3	I slept well	<i>0 (none) - 10 (very much)</i>
4	I feel rested	<i>0 (none) - 10 (very much)</i>
5	I had abdominal problems during the night	<i>0 (none) - 10 (very much)</i>
5a	During the night I suffered from:	<i>a full feeling, a heavy feeling, bloating, pain in my upper abdomen, a burning feeling in my upper abdomen, pain in my lower abdomen, nausea, vomiting, a burning feeling behind the breastbone, nocturnal defecation</i>

Discussion

The current study was conducted in order to develop a novel PROM for assessment of symptoms in functional dyspepsia based on the Experience Sampling Method.

Development of the novel ESM-based PROM was executed according to FDA guidelines over four executive phases. A systematic literature search, patient focus group interviews, and international expert meeting led to a novel ESM-

based PROM, containing 33 items. Interviews were performed to confirm patients' understanding, although adequate briefing before use of the novel PROM remains necessary as variable interpretations of the meaning of the terms nausea and bloating were reported during focus group interviews. In the future, addition of pictograms in the digital ESM-based PROM may be useful to improve comprehension of verbal symptom descriptors, as previously demonstrated by Tack et al.²⁵ Moreover, forward-and-backward translation with additional cognitive interviews are necessary to verify patient understanding with the developed items in each language.

In line with the Rome IV criteria, the four core FD symptoms were described by the majority of patients. Additional frequently reported symptoms were bloating, nausea, and belching. These symptoms were also reported by a substantial subgroup of FD patients during previously performed focus group interviews, cognitive interviews, and validation studies (e.g. bloating 86-93%, nausea 40-73%, belching 27-69%).^{26,27} Moreover, bloating was considered the most important symptom for improvement with effective therapy during development of the '*Functional Dyspepsia Symptom Diary*'.²⁷ As the FDA states that the effect of treatment should be measured at the level of each symptom in order to ensure that treatment does not negatively affect symptoms, we suggest that a PROM should, as a minimum, evaluate the four FD core symptoms as well as the three frequently occurring additional symptoms.⁸ Food intake, medication use, and psychosocial factors were considered important triggers for generation and/or severity of gastrointestinal symptoms, which is in line with previous studies.^{26,28-30} Repeated assessment of dyspeptic symptoms, together with symptom triggers, offers the potential to optimize and individualize treatment strategies, and this is a potential advantage of the novel ESM-based PROM, compared with the recently developed '*Leuven Postprandial Distress Scale*' (LPDS) and '*Functional Dyspepsia Symptom Diary*'.^{27,31}

Besides dyspeptic symptoms, lower gastrointestinal symptoms were studied, as evidence has been found for an increased prevalence of IBS in patients with FD.^{3,32} In our patient focus groups, 21% of patients had a comorbid diagnosis of IBS according to the Rome IV criteria, and this subgroup recommended inclusion of questions with regard to lower abdominal pain and defecation pattern. Although we intended to develop a disease-specific PROM for functional dyspepsia, the substantial overlap with IBS and potential implications for individualized therapy lead to incorporation of IBS core symptoms (i.e. lower abdominal pain and defecation pattern) in the newly developed PROM.

Psychological factors such as anxiety and depression are associated with functional gastrointestinal disorders, and may precede or exacerbate symptoms.³³⁻³⁵ One of the main advantages of ESM compared with retrospective questionnaires is the fact that ESM offers the opportunity to improve ecological validity by taking into account these psychological factors.²⁴ However, one point of discussion during the international expert meeting was the fact that no validated psychological ESM-items are currently available. In clinical trials, the Hospital Anxiety and Depression Scale (HADS) is often used to screen for the presence of anxiety or depressive disorders. However, HADS items are not eligible for momentary symptom assessment due to the retrospective nature. Verhagen et al. recently described psychometric validation properties of ESM in patients with mental disorders. Thirteen mood items were included in their ESM-questionnaire which could be divided into positive affect (4 items) and negative affect (9 items) based on factorial analysis. The authors demonstrated excellent reliability, and significant correlations were found between mean scores for positive and negative affect, and the HADS scores, demonstrating concurrent validity.²⁴ However, inclusion of thirteen psychological items in an outcome measure for FD was considered too much of a burden, and we therefore reduced the psychological ESM-items from 13 to seven, in line with the previously developed ESM-IBS PROM.¹⁸ Additional research is necessary to validate these psychological ESM-items in patients with functional gastrointestinal disorders.

Somatization is also considered a factor contributing to symptom generation in functional gastrointestinal disorders.³⁶⁻³⁸ The Patient Health Questionnaire 12 (PHQ-12) is frequently used in clinical trials to assess somatic complaints.³⁹ In the final ESM-FD PROM, five of the 12 items of the PHQ-12 were included because participants mentioned these complaints in focus group interviews (*i.e.* chest pain, dizziness, palpitations, dyspnea, and fatigue). The remaining seven items were not included, because they were not mentioned by focus group participants (back pain, pain in extremities or joints), they were mentioned by only one participant (headache, fainting), or the items were not eligible for assessment in a momentary questionnaire (sleeping problems).

Currently available outcome measures use several types of response options including Visual Analogue Scales (VAS), Likert scales, pictorial scales, and binary endpoints (adequate relief of symptoms: yes/no).²⁵ Recently, our research group performed focus groups in IBS patients for the development of an ESM-based PROM, and discussed several endpoints including a 7-point NRS (formerly used in ESM) and 11-point NRS.¹⁸ Finally, the 11-point NRS scale was

incorporated, as the use of an end-of-day 11-point NRS is recommended by the FDA as primary endpoint in clinical trials of IBS patients.²¹ In contrast to IBS, a recommended response scale for FD is currently lacking. The 11-point NRS, which has been incorporated into the ESM-based IBS-PROM, was discussed during the focus group meetings and was considered appropriate by FD patients. Therefore, the 11-point NRS was chosen, in order to have a uniform response option for ESM-based PROMs in functional gastrointestinal disorders. Strengths of this study include the fact that the ESM-based PROM was developed according to steps recommended by the FDA, with development of a conceptual framework after an extensive literature search, organization of patient focus group interviews and an international expert meeting, and arrangement of cognitive interviews.⁸ Due to the extensive literature search, a broad spectrum of items was identified and only a few additional symptoms (*e.g.* headache, fainting) were mentioned by individual patients during the focus group interviews.

Several limitations should be noted. First, we intended to invite five to 10 participants per focus group in order to ensure active participation of patients and gain a variety of perspectives. Although the final sample of 14 participants could be considered relatively small, focus group interviews were continued until saturation of input was achieved. Moreover, comparable sample sizes have been described in previous studies.^{25,26} Second, there may be a risk of limited representativeness, as only 32% of the initially invited patients eventually participated, and all patients were secondary and tertiary care patients. One could hypothesize that more confident patients and/or patients with more severe symptoms may be more willing to participate in focus group meetings. However, we assume that the risk of altered representativeness is limited, due to the inclusion of patients with a variety of symptoms matching different FD subtypes, the extensive item selection, and the limited additional input of symptoms by patients during focus group interviews. Another limitation may reside in the distribution of FD subtypes among participants since only one patient with the FD-PDS subtype participated in patient focus group sessions, whereas during cognitive grasp interviews only one patient with the EPS subtype was included. Nevertheless, in total four patients with PDS (20%), seven patients with EPS (35%), and nine patients with PDS-EPS (45%) were included. Therefore, we assume that the entire spectrum of dyspeptic symptoms was captured.⁴⁰⁻⁴² Participants did have to have sufficient command of the Dutch language, which means minority groups were underrepresented. Due to the inclusion of dyspeptic patients according to the Rome IV criteria, it is

not possible to extrapolate this PROM to patients with a selected number of dyspeptic symptoms not fulfilling Rome criteria. Another potential limitation was the fact that during the cognitive interviews the draft of the PROM was presented on paper, as the digital application was not yet available.

Next step in the development of this novel ESM-based PROM is a validation study in patients with FD according to Rome IV criteria. In this study, psychometric validation properties (*e.g.* reliability, validity and responsiveness) will be assessed, and the final goal is to define a minimum clinically important difference for use in clinical trials. Another point of interest during this validation study is the compliance rate, as a subgroup of patients during the focus group interviews thought that ten assessments per day would be quite burdensome. However, an adequate compliance rate of 76.8% was found in a small pilot study in IBS patients using a 60-item ESM questionnaire.¹⁷

In conclusion, we report on the development of a digital disease-specific PROM for real-time assessment of symptoms in patients with FD using the ESM. Although future studies with this novel PROM are necessary to assess construct validity, reliability, and responsiveness, we suggest that the newly developed ESM-based PROM might be able to assess FD, identify specific triggers for symptoms, optimize (individualized) treatment strategies, and could potentially be used as an instrument to quantify therapeutic efficacy.

References

1. Tack J, Talley NJ, Camilleri M, Holtmann G, Hu P, Malagelada JR, et al. Functional gastroduodenal disorders. *Gastroenterology*. 2006;130(5):1466-1479.
2. Mahadeva S, Goh KL. Epidemiology of functional dyspepsia: a global perspective. *World J Gastroenterol*. 2006;12(17):2661-2666.
3. 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.
4. Taylor F, Reasner DS, Carson RT, Deal LS, Foley C, Iovin R, 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.
5. Stanghellini V, Chan FK, Hasler WL, Malagelada JR, Suzuki H, Tack J, et al. Gastroduodenal Disorders. *Gastroenterology*. 2016;150(6):1380-1392.
6. Sander GB, Mazzoleni LE, Francesconi CF, Balbinotto G, Mazzoleni F, Wortmann AC, et al. Influence of organic and functional dyspepsia on work productivity: the HEROES-DIP study. *Value Health*. 2011;14(5 Suppl 1):S126-129.
7. 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.
8. Administration FaD. Guidance for Industry; Patient-Reported Outcome Measures: use in Medical Product Development to support Labeling Claims 2009.
9. Smeets FGM, Masclee AAM, Conchillo JM, Keszthelyi D. Systematic review: Disease-specific instruments to assess gastrointestinal symptoms in functional dyspepsia. *Neurogastroenterol Motil*. 2018.
10. van Os J, Verhagen S, Marsman A, Peeters F, Bak M, Marcelis M, 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.
11. Feinle-Bisset C, Horowitz M. Dietary factors in functional dyspepsia. *Neurogastroenterol Motil*. 2006;18(8):608-618.
12. Mak AD, Wu JC, Chan Y, Chan FK, Sung JJ, Lee S. Dyspepsia is strongly associated with major depression and generalised anxiety disorder - a community study. *Aliment Pharmacol Ther*. 2012;36(8):800-810.
13. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol*. 2008;4:1-32.
14. 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.
15. 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.
16. 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.
17. Mujagic Z, Leue C, Vork L, Lousberg R, Jonkers DM, 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.

18. Vork L, Keszthelyi D, Mujagic Z, Kruiemel JW, Leue C, Ponten 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).
19. Rabiee F. Focus-group interview and data analysis. *Proc Nutr Soc.* 2004;63(4):655-660.
20. Krueger RA. *Designing and conducting focus group interviews.* 2002.
21. (FDA) FaDA. *Guidance for industry, Irritable bowel syndrome - Clinical evaluation of drugs for treatment 2012.*
22. Knafelz K, Deatrick J, Gallo A, Holcombe G, Bakitas M, Dixon J, et al. The analysis and interpretation of cognitive interviews for instrument development. *Res Nurs Health.* 2007;30(2):224-234.
23. Krueger RA. *Developing questions for focus groups.* SAGA Publications. 1998.
24. Verhagen SJW, Berben JA, Leue C, Marsman A, Delespaul P, 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.
25. Tack J, Carbone F, Holvoet L, Vanheel H, Vanuytsel T, Vandenbergh A. The use of pictograms improves symptom evaluation by patients with functional dyspepsia. *Aliment Pharmacol Ther.* 2014;40(5):523-530.
26. Carbone F, Holvoet L, Vandenbergh A, Tack J. Functional dyspepsia: outcome of focus groups for the development of a questionnaire for symptom assessment in patients suffering from postprandial distress syndrome (PDS). *Neurogastroenterol Motil.* 2014;26(9):1266-1274.
27. Taylor F, Higgins S, Carson RT, Eremenco S, Foley C, Lacy BE, et al. Development of a Symptom-Focused Patient-Reported Outcome Measure for Functional Dyspepsia: The Functional Dyspepsia Symptom Diary (FDSD). *Am J Gastroenterol.* 2018;113(1):39-48.
28. Talley NJ. Functional dyspepsia: new insights into pathogenesis and therapy. *Korean J Intern Med.* 2016;31(3):444-456.
29. Fang YJ, Liou JM, Chen CC, Lee JY, Hsu YC, Chen MJ, et al. Distinct aetiopathogenesis in subgroups of functional dyspepsia according to the Rome III criteria. *Gut.* 2015;64(10):1517-1528.
30. Bisschops R, Karamanolis G, Arts J, Caenepeel P, Verbeke K, Janssens J, et al. Relationship between symptoms and ingestion of a meal in functional dyspepsia. *Gut.* 2008;57(11):1495-1503.
31. Carbone F, Vandenbergh A, Holvoet L, Vanuytsel T, Van Oudenhove L, Jones M, et al. Validation of the Leuven Postprandial Distress Scale, a questionnaire for symptom assessment in the functional dyspepsia/postprandial distress syndrome. *Aliment Pharmacol Ther.* 2016;44(9):989-1001.
32. Ford AC, Marwaha A, Lim A, Moayyedi P. Systematic review and meta-analysis of the prevalence of irritable bowel syndrome in individuals with dyspepsia. *Clin Gastroenterol Hepatol.* 2010;8(5):401-409.
33. Wouters MM, Boeckstaens GE. Is there a causal link between psychological disorders and functional gastrointestinal disorders? *Expert Rev Gastroenterol Hepatol.* 2016;10(1):5-8.
34. 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.
35. Pinto-Sanchez MI, Ford AC, Avila CA, Verdu EF, Collins SM, Morgan D, et al. Anxiety and Depression Increase in a Stepwise Manner in Parallel With Multiple FGIDs and Symptom Severity and Frequency. *Am J Gastroenterol.* 2015;110(7):1038-1048.

36. Van Oudenhove L, Vandenberghe J, Geeraerts B, Vos R, Persoons P, Fischler B, et al. Determinants of symptoms in functional dyspepsia: gastric sensorimotor function, psychosocial factors or somatisation? *Gut*. 2008;57(12):1666-1673.
37. Simren M, Tornblom H, Palsson OS, van Tilburg MAL, Van Oudenhove L, Tack J, et al. Visceral hypersensitivity is associated with GI symptom severity in functional GI disorders: consistent findings from five different patient cohorts. *Gut*. 2018;67(2):255-262.
38. Gracie DJ, Bercik P, Morgan DG, Bolino C, Pintos-Sanchez MI, Moayyedi P, et al. No increase in prevalence of somatization in functional vs organic dyspepsia: a cross-sectional survey. *Neurogastroenterol Motil*. 2015;27(7):1024-1031.
39. Spiller RC, Humes DJ, Campbell E, Hastings M, Neal KR, Dukes GE, et al. The Patient Health Questionnaire 12 Somatic Symptom scale as a predictor of symptom severity and consulting behaviour in patients with irritable bowel syndrome and symptomatic diverticular disease. *Aliment Pharmacol Ther*. 2010;32(6):811-820.
40. Vakil N, Halling K, Ohlsson L, Wernersson B. Symptom overlap between postprandial distress and epigastric pain syndromes of the Rome III dyspepsia classification. *Am J Gastroenterol*. 2013;108(5):767-774.
41. Carbone F, Holvoet L, Tack J. Rome III functional dyspepsia subdivision in PDS and EPS: recognizing postprandial symptoms reduces overlap. *Neurogastroenterol Motil*. 2015;27(8):1069-1074.
42. Carbone F, Holvoet L, Vanuytsel T, Tack J. Rome III functional dyspepsia symptoms classification: Severity vs frequency. *Neurogastroenterol Motil*. 2017;29(6).

Chapter 8

The development of a patient-reported outcome measure for real-time symptom assessment in a population with functional urologic complaints - A focus group study

Alexandra Herrewegh, Lisa Vork, Eline Eurelings, Carsten Leue,
Joanna W. Kruiemel, Gommert A. van Koeveringe, Desiree M.J. Vrijens
Neurourology & Urodynamics. 2018;37(8):2893-2903.

Abstract

INTRODUCTION In the current diagnostic process for overactive bladder syndrome (OAB), biased retrospective questionnaires are often used. There is a need for a new assessment tool that embraces the heterogeneity of the OAB complex. A momentary assessment tool, the Experience Sampling Method (ESM), is promising, capturing random repetitive measurements during the day in the context of daily life and is capable to measure potential contextual triggers and psychological aspects. A focus group study was set up to evaluate which items should be implemented in a urological ESM.

METHODS Focus group interviews were arranged to assess the suitability and comprehensibility of a newly developed urological patient-reported outcome measurement (PROM), 'Uromate'. 'Uromate' was created based on ESM literature. A multidisciplinary expert meeting was conducted to gain consensus on item relevance.

RESULTS The initial ESM questionnaire contained 58 items, but was eventually reduced to 39 items after focus group sessions and expert meeting. Thirty-seven items are repeated questions, including 3 gender-dependent items. Two items are one-time questions about the use of incontinence material. Additionally, a morning questionnaire was included. Depending on the symptom pattern, a minimum of 26 items and a maximum of 36 items will be repeatedly assessed with 'Uromate'.

CONCLUSIONS There is a need for a modern assessment tool for OAB which overcomes the limitations of today's retrospective questionnaires. Therefore, a urological ESM tool, the 'Uromate', is being developed as a PROM, following the FDA PROM development guidelines, to measure real-time symptoms in the context of daily life.

Introduction

Overactive bladder syndrome (OAB) is “urinary urgency, with or without urgency urinary incontinence, usually accompanied by frequency and nocturia.”¹ OAB is associated with comorbidities and even increased mortality.² Currently, there is increasing evidence that OAB is multifactorial in aetiology and has a complex pathogenesis in which OAB is associated with various somatic illnesses and is affected by genetic and environmental factors.³ OAB and gastrointestinal complaints often coexist. Chronic constipation and faecal incontinence have been shown to occur more frequently in OAB patients,⁴ and approximately 30% of OAB patients are concurrently diagnosed with IBS.⁵ Additionally, OAB is associated with fibromyalgia, migraine, chronic fatigue, eating disorders³ and affective disorders.⁶ Recent studies suggest that a spectrum or continuum of functional urological and intestinal disorders across organ systems is present, with a shared underlying central hypersensitivity or alarm falsification pathway.^{7,8} This theory is supported by activation or reduced deactivation of the anterior cingulate cortex that is seen in OAB patients as well as in irritable bowel syndrome (IBS) patients or affective conditions.⁷

Furthermore, patients with OAB have different strategies to cope with complaints, such as timed voiding, locating toilets and avoidance of drinking. Changing of voiding pattern may exacerbate their symptoms. In addition, OAB symptoms may be increased by caffeine-containing products, smoking and excessive fluid intake.⁹ Urgency and incontinence are influenced by environmental cues and mood seems to play a role in the exacerbation of urgency.¹⁰

Thus, OAB is a heterogeneous symptom complex and patients show fluctuating symptom patterns in which urological complaints might be influenced by contextual factors in daily life as well as psychological and other somatic comorbidities.

Due to the lack of a biological marker, the diagnosis of OAB and the evaluation of therapeutic responses rely on symptom assessment using retrospective questionnaires and bladder diaries. Although there is a strong correlation between existing patient-reported outcome measures (PROMs) and diaries,¹¹ there is no assessment tool currently available that assesses all contextual factors that could have influence on OAB complaints, including symptom triggers and overlapping symptoms with other comorbidities. In addition, currently used questionnaires ask about the presence of symptoms over weeks (for example: OABSS over the last week, OABq or ICIQ-FLUTS/MLUTS over the

last 4 weeks). A shorter recall period allows to evaluate the variability and the effect of time-varying factors between multiple measurements on the outcome. This is recommended by Stull et al. to capture the variability of urge or frequency of urination associated with OAB.¹²

Hence, there is a need for the development of a multi-item patient-reported outcome tool to capture all clinically important aspects of OAB in the context of daily life, preferably in agreement with the FDA guidelines on PROM development.¹³

The Experience Sampling Method (ESM) is a momentary assessment tool, which randomly measures multiple times (8-10) during the day in a real life environment.¹⁴ ESM ensures the assessment of symptom variability over time, taking social, psychological and contextual factors into account, which is currently missing in the assessment methods for OAB. This real-time assessment reduces information bias, especially recall bias. Furthermore, ESM can overcome the risk of ecological bias, which occurs when a questionnaire is used to obtain information with regard to patient's daily life, thereby only recording one moment of life in a specific context but generalizing this given answer to more than this specific moment in life.¹⁵ Hence, ESM is expected to be an attractive and valid method to assess urological complaints in the context of patient's daily life. ESM has already proven its clinical value to assess affective states and the influence of contextual factors at the level of individual subjects in psychiatric patients with affective disorders.¹⁴ Moreover, ESM has been successfully piloted in IBS patients, where ESM has been proven to provide more accurate information about IBS symptoms than current assessment with retrospective questionnaires.^{14,16,17} The purpose of the urological ESM is to investigate the multifactorial character of OAB, as well as to assess symptom triggers to better select adequate treatment options.

The current article describes the development process of 'Uromate', a PROM suitable for real-time assessment of urological and non-urological symptoms in OAB, and potential triggers, using the ESM principle. Focus group interviews and an expert meeting were conducted to broadly explore the knowledge and experience from patients and experts for item selection, in agreement with FDA guidelines on PROM development.

Materials and methods

The development of the questions to be used in 'Uromate', started with an initial draft based on a theoretical framework from the literature.^{13,18} Items concerning social factors, contextual and environmental factors and psychological status following the ESM specific construct, were derived from questions used in previous ESM validation studies.^{14,16,17,19} A specific list of questions, that incorporated urological complaints like urgency and frequency, was compiled by our research team. Several questions were derived from validated questionnaires (OAB-q²⁰, ICIQ-FLUTS/MLUTS^{21,22}, ICIQ-OAB^{21,22}, KHQoL²³, BFLUTS²², SF-36²⁴) and made suitable for momentary assessment. Fifty-eight initial ESM questions were developed regarding different domains: somatic items (urological and otherwise), nutrition and drug use, psychological items and situation and companionship (Table 8.1). The possible end-points for all items were ranged with the use of an 11-points Numeric Rating Scale (NRS), if applicable. The 11-points scale was used according to FDA recommendations.¹³ Furthermore, it creates homogeneity of end-points, making it easier and faster for patients to fill out 'Uromate' repeatedly. Focus group interviews were the next step to assess the suitability of the proposed items. An expert meeting followed to finalize the item selection process.

Table 8.1 Set of items for the urological ESM-PROM developed by the research team.

Urological ESM items developed by the research team		
	Somatic items urological	Answer scale
1	Since last beep I had to void ... times.	0, 1 time, 2 times, 3 times, 4 times, 5 times, > 5 times.
2	I am sensing that my bladder is filled.	0 (none) - 10 (very much)
3	I am experiencing urgency.	0 (none) - 10 (very much)
3a	The urgency is untenable.	Yes / No
3b	I can postpone voiding now.	0 (no urgency), 1 (30 min), 2 (10 min), 3 (now/untenable)
4	Since the last beep I have lost urine involuntarily.	No loss, drops, dash, wet pants, wet outerwear, walking along legs, walking to the floor.
4a	This happened while experiencing urgency.	0 (none) - 10 (very much)
4b	This happened during physical activity, coughing, sneezing.	0 (none) - 10 (very much)
5	<u>Since the last beep I had to change incontinence material.</u>	<u>No incontinence material, panty liner changed, incontinence bandage changed, changed pants, changed clothes.</u>
6	I am having abdominal pain.	0 (none) - 10 (very much)
6a	The pain is located in...	Figure abdomen

Table 8.1 (continued)

Nutrition & drug use		
7	Since the last beep I have had...glasses/consumptions.	0, 1, 2, 3, 4, >4
7a	These were ... glasses of caffeine or beverages like coke / coffee / tea / energy drink or alcohol.	0, 1, 2, 3, 4, >4
8	<u>Since the last beep I have used....</u>	<u>Caffeine (coffee)/ theine (tea), nicotine (smoking), alcohol, drugs, none.</u>
Somatic items		
1	I have no other physical complaints.	0 (none) - 10 (very much)
2	I am having palpitations.	0 (none) - 10 (very much)
3	I am sweating.	0 (none) - 10 (very much)
4	I am short of breath.	0 (none) - 10 (very much)
5	I feel dizzy.	0 (none) - 10 (very much)
6	<u>I have muscle pain.</u>	<u>0 (none) - 10 (very much)</u>
7	<u>I have painful joints.</u>	<u>0 (none) - 10 (very much)</u>
8	I have an urge to defecate.	0 (none) - 10 (very much)
Psychological items		
1	I feel cheerful.	0 (none) - 10 (very much)
2	I feel sad.	0 (none) - 10 (very much)
3	I feel anxious.	0 (none) - 10 (very much)
4	I feel energetic.	0 (none) - 10 (very much)
5	I feel irritated.	0 (none) - 10 (very much)
6	I feel stressed.	0 (none) - 10 (very much)
7	I feel relaxed.	0 (none) - 10 (very much)
8	I am worrying.	0 (none) - 10 (very much)
9	I am satisfied.	0 (none) - 10 (very much)
10	I feel enthusiastic.	0 (none) - 10 (very much)
11	I feel rushed.	0 (none) - 10 (very much)
12	I am restless.	0 (none) - 10 (very much)
13	I am happy.	0 (none) - 10 (very much)
14	I am lonely.	0 (none) - 10 (very much)
15	I feel strong.	0 (none) - 10 (very much)
16	I feel worried.	0 (none) - 10 (very much)
17	I am inspired.	0 (none) - 10 (very much)
18	I am disappointed.	0 (none) - 10 (very much)
19	I feel insecure.	0 (none) - 10 (very much)
20	I feel ashamed.	0 (none) - 10 (very much)
21	I feel guilty.	0 (none) - 10 (very much)
22	I am anxious.	0 (none) - 10 (very much)
Situation/ Companionship		
1	Where am I?	<i>At home, someone else's home, work/school, public place, in transit, somewhere else.</i>
2	What am I doing (just before the beep)?	<i>Resting, work/school, housekeeping/ groceries, hygiene/self-care, eating, recreation, sports, in transit, something else.</i>
3	I feel comfortable in this situation.	<i>-5 (very uncomfortable) – 5 (very comfortable)</i>
4	My symptoms are limiting me in my current activities.	<i>0 (none) - 10 (very much)</i>
5	With whom am I?	<i>Partner/kids, friends, roommates, colleagues, family, acquaintances, unknowns/others, nobody.</i>

Table 8.1 (continued)

Morning questionnaire		
1	How long did it take before you were asleep?	<i>0-15 min; 15-30 min; 30-45 min; 45 min-1 hrs; 1-2 hrs; 2-4 hrs; >4 hrs.</i>
2	How often were you awake during the night?	<i>0, 1 time, 2 times, 3 times, 4 times, 5 times, >5 times.</i>
3	How long were you awake before you went out of bed?	<i>0-15 min; 15-30 min; 30-45 min; 45 min-1 hr; 1-2 hrs; 2-4 hrs; > 4hrs.</i>
4	I have slept well.	<i>0 (none) - 10 (very much)</i>
5	I feel rested.	<i>0 (none) - 10 (very much)</i>
6	Last night I went ... times out of bed to void.	<i>0, 1 time, 2 times, 3 times, 4 times, 5 times, >5 times.</i>
7	Last night I have lost urine involuntarily.	<i>No loss, drops, dash, wet pants, wet underwear, walking along legs, walking to the floor.</i>
Evening questionnaire		
7	This evening I have had ... glasses to drink.	<i>0, 1, 2, 3, 4, >4</i>
7a	These were ... glasses caffeine, like coke/coffee/ energy drink.	<i>0, 1, 2, 3, 4, >4</i>

Italic, bold, and underlined indicate items that were removed from the instrument during the focus group interviews. The items are developed in Dutch and used in a Dutch population. The questions are translated to English for this article. A validation for English language is still needed.

The Experience Sampling Method

According to this method, patients answer questions in real-time, usually by use of a mobile app. Patients get access to the ESM via a smartphone or another electronic device to the 'Uromate' app. 'Uromate' will repetitively evaluate during the day, *i.e.* at 10 randomly chosen moments between 07.30 am and 22.30 pm. The device sends an auditory signal ('beep') every time the questionnaire can be completed. Each questionnaire should be filled out during the 10 following minutes after the beep to guarantee momentary assessment.¹⁶ Preferably, completion of ≥ 6 momentary repeated questionnaires a day should be achieved in order to consider the patient as compliant.^{16,25} The completion of each question set will approximately take 4-5 minutes.¹⁶

Focus group recruitment

Patients older than 18 years and diagnosed with OAB with or without incontinence by a urologist were included. Patients with a clear anatomical or other urological cause for their complaints (*e.g.* bladder cancer, benign prostate hyperplasia (BPH), prostate cancer, urolithiasis, chronic or acute urinary tract infections (UTI) or neurological disease) were excluded. Pregnancy was also a reason for exclusion. Patients needed to be able to speak and understand

(written) Dutch. Eligible patients were informed about the focus group meetings by their treating physician at the urological outpatient clinic.

Focus group method

Focus groups were organized according to international guidelines on the development of PROMs¹³ and the literature on performing focus group interviews.¹⁸ The meetings took about 90 minutes each and were scheduled one by one until saturation of input was reached: the moment that meetings no longer augmented any new items or information in addition to previous meetings.^{13,18} Ideally, 6-10 patients were attending each meeting.^{13,18} The focus groups were performed in two phases based on Focus Group guidelines, with the guidance of a moderator (AH) and at least one assistant-moderator (LV, EE).^{13,18} At first, an open discussion took place. Patients were instructed to define what living with OAB meant to them and to discuss which items should be essential in a urological ESM assessment tool in order to get good insight into the daily symptom pattern.

During the second part of the meetings, the predefined questionnaire was shown, so that the patients could give their opinion on this list of questions and discuss and criticize different items. It was stressed that the items should be relevant and easy to understand. In addition, questions had to be suitable for momentary assessment. The focus group discussions and the expert meeting were voice recorded and transcribed verbatim. The attending investigators made notes as well.

Expert meeting

An expert meeting was organized to finalize the list of questions. The experts in the meeting were specialists in the field of functional urology or other functional disorders including two urologists, a gynaecologist, a gastroenterologist and a hospital psychiatrist. The primary goal of the meeting was to critically discuss and convert the information and findings from the focus groups to applicable questions for clinical practice. A second goal was to shorten the list of ESM items, in order to minimize response fatigue and non-compliance of the patients consequently.²⁶ The final items are presented in this report (Table 8.2).

Table 8.2 Set of items for the urological ESM-PROM after focus groups and expert meeting.

Urological ESM items developed by the research team	
Somatic items urological	Answer scale
<i>One-time questionnaire items</i>	
1 I use incontinence material during the day.	<i>No incontinence material, panty liner, thin incontinence bandage, thick incontinence bandage, incontinence pants.</i>
2 I use incontinence material during the night.	<i>No incontinence material, panty liner, thin incontinence bandage, thick incontinence bandage, incontinence pants.</i>
<i>Momentary</i>	
1 I am sensing that my bladder is filled.	<i>0 (none) - 10 (very much)</i>
2 I am experiencing urgency.	<i>0 (no urgency), 1 (30 min delay), 2 (10 min delay), 3 (now, no delay possible)</i>
2a The urgency is untenable.	<i>0 (no urgency) -10 (very untenable)</i>
3 I am experiencing the urge to defecate.	<i>0 (none) - 10 (very much)</i>
4 I am having abdominal pain.	<i>0 (none) - 10 (very much)</i>
4a The abdominal pain is located ... in my abdomen.	<i>Figure abdomen</i>
<i>Retrospective</i>	
1 Since the last beep I have had to void ... times.	<i>0 times, 1 time, 2 times, 3 times, 4 times, >4 times.</i>
1a I have had the feeling that I could completely empty my bladder by voiding.	<i>0 (empty) - 10 (not empty)</i>
1b The urinary stream was strong...	<i>Figure VPSS</i>
2 I have involuntarily lost urine.	<i>Not, drops, shoot, wet underwear, wet outer clothes.</i>
2a This was accompanied by urgency.	<i>0 (no urgency) -10 (very untenable)</i>
2b This was during...	<i>(Unexpected) posture change, physical activity, coughing, sneezing, squeezing, sitting.</i>
2c Since the last beep I have had to change the incontinence material ... times.	<i>0 times, 1 time, 2 times, 3 times, 4 times, > 4 times.</i>
Nutrition & drug use (retrospective)	
1 Since the last beep I have had...glasses/consumptions.	<i>0, 1, 2, 3, 4, >4</i>
1a These were ... glasses of caffeine containing beverages like coke / coffee / tea / energy drink or alcohol.	<i>0, 1, 2, 3, 4, >4</i>
Somatic items non-urological	
1 I am feeling physically well.	<i>0 (none) - 10 (very much)</i>
2 I am having palpitations.	<i>0 (none) - 10 (very much)</i>
3 I am sweating.	<i>0 (none) - 10 (very much)</i>
4 I am having shortness of breath.	<i>0 (none) - 10 (very much)</i>
5 I am feeling dizzy.	<i>0 (none) - 10 (very much)</i>
6 I am having muscle or joint pain.	<i>0 (none) - 10 (very much)</i>
7 I am feeling that something is hanging out of the vagina.	<i>0 (none) - 10 (very much), N/A</i>
8 I am having pain in the scrotum.	<i>0 (none) - 10 (very much), N/A</i>

Table 8.2 (continued)

Psychological items		
1	I am feeling cheerful.	0 (none) - 10 (very much)
2	I am feeling sad.	0 (none) - 10 (very much)
3	I am feeling anxious.	0 (none) - 10 (very much)
4	I am feeling irritated.	0 (none) - 10 (very much)
5	I am feeling stressed/ rushed.	0 (none) - 10 (very much)
6	I am feeling relaxed.	0 (none) - 10 (very much)
7	I am worrying.	0 (none) - 10 (very much)
8	I am feeling ashamed.	0 (none) - 10 (very much)
Situation/ Companionship		
1	Where am I?	At home, someone else's home, work/school, public place, in transit, somewhere else.
2	What am I doing (just before the beep)?	Resting, work/school, housekeeping/groceries, hygiene/ self-care, eating, recreation, sports, in transit, something else.
3	I feel (un)comfortable in this situation.	-5 (very uncomfortable) – 5 (very comfortable)
4	My complaints are limiting me in my current activities.	0 (none) -10 (very much)
5	With whom am I?	Partner/kids, friends, roommates, colleagues, family, acquaintances, unknown people / others, nobody.
5a	I find this company (un)pleasant / I feel at ease with this company.	-5 (very unpleasant) – 5 (very pleasant)
Morning questionnaire (retrospective)		
1	How long did it take before you were asleep?	0-15 min; 15-30 min;30-45 min; 45 min - 1 hr; 1-2 hrs; 2-4 hrs; >4 hrs.
2	How often did you wake up during the night?	0 times, 1 time, 2 times, 3 times, 4 times, >4 times.
2a	I woke up because of...	Urgency, urinary incontinence, abdominal pain, no reason.
2b	Last night I had to get up times to go to the toilet.	0 times, 1 time, 2 times, 3 times, 4 times, > 4 times.
3	Last night I have lost urine involuntarily.	None, drops, shoot, wet underwear, wet mattress.
Sexuality in morning questionnaire (retrospective)		
1	I have had intercourse.	Yes, no.
1a	During intercourse, I experienced urgency.	0 (no urgency), 1 (30 min delay), 2 (10 min delay), 3 (now, no delay possible).
1b	After intercourse, I experienced urgency.	0 (no urgency), 1 (30 min delay), 2 (10 min delay), 3 (now, no delay possible).
1c	During intercourse, I have lost urine involuntarily.	No incontinence, incontinence during sexual relations, incontinence after sexual relations.
1d	I experienced pain during intercourse.	0 (no pain) - 10 (much pain)

The items are developed in Dutch and used in a Dutch population. The questions are translated to English for this article. A validation for English language is still needed.

Results

Study population

Patient characteristics are summarized in Table 8.3. A total of 132 patients with either OAB or mixed urinary incontinence (MUI) were approached to participate. Seventeen patients confirmed to participate and 13 female subjects were eventually present during the focus group meetings, of which 9 suffered from OAB and 4 from MUI. The main reason of declining to participate in the focus groups was embarrassment with regard to discussing their bladder problems in a group. Saturation was reached after 2 meetings, with a total population of 13 subjects.

Table 8.3 Patient characteristics of the focus groups participants.

	Focus group 1 (n=7)	Focus group 2 (n=6)
Mean age±SD, years	53.9±6.4	54.5±7.6
Female, %	100	100
OAB wet, n	2	1
OAB dry, n	2	4
Mixed urinary incontinence, n	3	1

Focus group ESM questionnaire

During the focus group meetings ESM repeated items were reduced from 49 to 36 items. The list of morning questions went from 9 to 6 items. To limit the risk that some key OAB symptoms would be missed with momentary assessment, it was decided that some retrospective questions should be added with a short recall period of at least 15 minutes and maximum 90 minutes (Table 8.2). Additionally, patients suggested that clustering of questions made them easier to understand. Furthermore, 2 one-time questions were selected to record the use of incontinence material, because the number of daily-used products did not significantly change over time.

Somatic items: urgency

Most patients experienced urgency as an intensified sensation of the normal urge. One patient experienced no specific urgency but only pain in the lower abdomen prior to urinary leakage. Therefore, the figure of the abdomen was retained from the draft instrument (Figure 8.1). Another patient had sudden

urinary leakage without prodromal signs or sensations. Urgency to defecate was also incorporated in the list of momentary questions, after some patients stated that both urinary and faecal urgency could arise together.

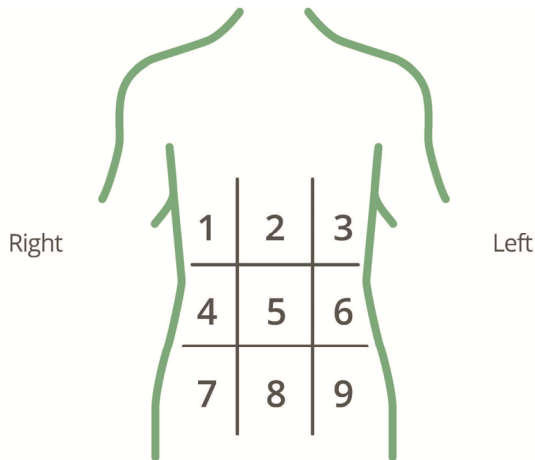


Figure 8.1 Figure abdomen for momentary assessment.

Nutrition and substance use

Patients noted that liquid intake was a very important item and most patients stated that they adjusted the amount of fluid intake when going outside. Some patients mentioned that carbonated drinks increased the urgency and alcohol increased the voiding frequency. Coffee or tea was not mentioned as voiding triggers. Patients stated that in order to trigger OAB symptoms it was more important how much they drank, rather than what they drank.

Somatic complaints other than OAB

Patients appreciated that questions concerning other somatic complaints were included in the proposed list of questions for 'Uromate'. In general, their treating physician had not explored these symptoms. It is of particular interest that a few patients experience the whole range of proposed somatic complaints. However, focus group patients could not mention one specific non-urological somatic complaint that they associate with OAB. Furthermore, 'painful muscles' was changed to 'muscle cramps', and 'painful joints' was not

found to be relevant to patients and was thus removed. Additionally, patients advised to add vaginal pain as a somatic symptom to the list of questions.

Mood and psychological factors

Patients were asked to point out the least relevant items, leading to a substantially shortened list. 'Energetic', 'enthusiastic', 'happy', 'strong', 'worried', 'inspired', 'disappointed', 'insecure' and 'guilty' were removed because patients found those items not to be associated with OAB. Of the 22 initial proposed items concerning mood and mental status, 8 questions in total were selected.

Environment

Situation and company were very important factors, influencing the psyche and severity of complaints. One patient stated: "I feel uncomfortable to be in a situation where people do not show understanding of my urological complaints". Another patient told: "When my colleagues make jokes about my bladder problems, it feels like they don't take me seriously."

Morning questionnaire

The focus group participants found the proposed 'Uromate' questions more suitable for sleeping disorders rather than urological complaints. Patients considered the frequency of awakening and whether awakening was due to urological symptoms (such as urgency, abdominal pain or urinary incontinence) most important. They agreed with implementing questions about the frequency of awakening to void (nocturia) and nightly urinary loss.

Sexuality

Since OAB patients experience diminished sexual activity and enjoyment of sex,²⁷ questions about sexuality were proposed to be incorporated into the initial list of questions. Eventually, sexuality questions were not incorporated, because repeated assessment was not considered to be useful. Nevertheless, patients missed questions about sexuality in the list of 'Uromate' items. To them, sexuality was an important item, because their sexual functioning was impaired due to OAB complaints. Integration in the morning questionnaire was proposed.

Expert meeting ESM questions

The list of questions was shortened by making sub questions in the case of positive answers. For example: if a patient did not void, subsequently, the question about the quality of the urinary stream would not pop up. A validated icon of the Visual Prostate Symptom Score (VPSS) depicting the urinary stream was added (Figure 8.2) for retrospective repeated assessment of the urinary stream force in men, however this was not assessed in the current study.²⁸

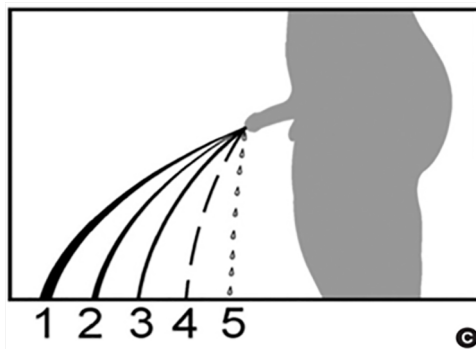


Figure 8.2 The Visual Prostate Symptom Score (VPSS).

The experts decided to use a validated 4-points urgency scale.²⁹ Regarding urinary loss, some of the original answering options for quantification were removed, because they had no additional value. Additionally, it was decided to evaluate the degree of untenability as well, using an 11-points NRS. Moreover, there was agreement between the experts and the focus group participants on merging muscle- and joint complaints together, with the aim to identify a possible link between OAB and fibromyalgia. Experts decided to add scrotal pain and prolapse sensation as gender-specific questions to the PROM. Most of the initial positive affect psychological items were removed during the focus group interviews. However, experts agreed that a couple of positive options must be present in the ESM-questionnaire, otherwise a negative response bias might be introduced.²⁶ Therefore, the items 'cheerful' and 'relaxed' were kept as positive affect items to maintain an overall neutral question tone. The items 'anxious', 'lonely' and 'nervous' were removed to prevent response fatigue.²⁶ Carbonated drinks were not incorporated in the questions by the experts, because there is little evidence on the association with urinary complaints. Morning questions

about sexuality were added, such as pain, urinary urgency and incontinence during intercourse.

Discussion

Until now, the clear etiological nosology or specific structural organic cause in OAB is unknown. The definition of OAB consists of a symptom complex, and does not involve biomarkers or urodynamic clinical findings.¹ Hence, OAB is difficult to diagnose and relies on patient reported outcomes. However, there is a large variety of measurement tools, all retrospective in nature, and, often measuring only a few symptoms of the OAB complex. Thus, a new assessment method is needed that fully captures the multifactorial complexity of the OAB syndrome as a whole.⁸ 'Uromate' may prove to be the assessment method with the potential to overcome these limitations and meet these requirements.

The repeated measurements of the ESM, which are collected in different contexts (social situations and environment), allow us to understand the influence of psychological factors and mental state variability on complaint pattern and voiding behaviour. Multivariate and multilevel analysis will give us more insight into the effect of possible disease factors. It enables patients to create awareness of factors contributing to the emerging and maintaining of complaints. Using the 'Uromate' may make self-monitoring easier, which can be useful in the evaluation of treatment outcome.

Following the development of an ESM tool specific for psychiatric conditions^{14,19} and functional gastrointestinal disorders as IBS¹⁶, 'Uromate' is being developed for urological disorders.

To limit selection bias and attrition bias, which increases due to lack of motivation and compliance by patients, the number of questions were reduced by use of the focus groups and an expert meeting. This yielded a concise list of items that were all deemed relevant in the assessment of OAB. In this respect, an advantage of the ESM is that non-compliance can be directly monitored, when questions remain unanswered in the app. Additionally, a good compliance has been reported concerning the use of ESM in previous studies, even in a population of high vulnerability such as schizophrenia (87%) and anxiety disorders (98%).³⁰

Voiding dairies are a reliable measure of outcome in OAB, are sensitive to change³¹ and therefore widely used. Currently, voiding parameters can be measured real-time with the use of an electronic voiding diary. Although real-

time electronic monitoring seem to be more accurately reflecting patient symptoms³², it does not take account of social, psychological and contextual factors, as is done with 'Uromate'.

Since the initial questionnaire was compiled with questions derived from validated questionnaires, the 'Uromate' is comparable to other urological validated PROMs, but without the risk of ecological or recall bias.

The use of patient focus groups in the process of developing a urological PROM strengthens the validity of the questionnaire.¹³ Our focus group procedure revealed that many factors, especially contextual factors, have an important influence on OAB complaints. Hence, focus group research is essential in PROM development.

A limitation of the study is that only women were included in the focus groups, while the prevalence of OAB in men and women is almost equal.³³ However, no men were included, because they felt ashamed to talk about their OAB complaints in a group.

Additionally, only 13 patients participated, although 132 patients were approached. However, the main aim of focus groups is "to understand and explain the meanings, beliefs and cultures that influence the feelings, attitudes and behaviours of individuals."³⁴ In PROM development, focus groups are conducted to generate items, evaluate if all items are covered and to ensure clarity and readability of the items.¹³ In focus group research, the group dynamic is essential to obtain deep and rich data.³⁴ To ensure that active participation is possible, but perspective variety is maintained, 6-8 participants per focus group is recommended.¹⁸ Hence, 13 participants were sufficient to obtain the aim of our focus groups for PROM development. Most importantly, saturation of input was achieved.

'Uromate' is in an early stage of development and obviously further studies are needed to validate and prove its feasibility and usability. Further modifications may be needed.

Since OAB is a clinical diagnosis, the best assessment method is the use of PROMs to capture the full scope and complexity of OAB complaints. Therefore, a specific urological PROM ('Uromate') according to the ESM principle, is being developed for OAB patients, using a focus group procedure followed by an expert meeting. 'Uromate' assesses OAB complaints and triggers in daily life and captures psychological and contextual factors, influencing the severity and variability of these symptoms. Further studies on 'Uromate' are needed and are planned in the near future to demonstrate the feasibility and to validate this tool in the diagnosis and treatment of OAB.

References

1. Abrams P, Cardozo L, Fall M, et al. The standardisation of terminology of lower urinary tract function: report from the Standardisation Sub-committee of the International Continence Society. *Neurourol Urodyn.* 2002;21(2):167-78.
2. Foley AL, Loharuka S, Barrett JA, et al. Association between the Geriatric Giants of urinary incontinence and falls in older people using data from the Leicestershire MRC Incontinence Study. *Age and ageing.* 2012;41(1):35-40.
3. Altman D, Iliadou AN, Lundholm C, Milsom I, Pedersen NL. Somatic comorbidity in women with overactive bladder syndrome. *J Urol.* 2016;196(2):473-7.
4. Coyne KS, Cash B, Kopp Z, et al. The prevalence of chronic constipation and faecal incontinence among men and women with symptoms of overactive bladder. *BJU Int.* 2011;107(2):254-61.
5. Matsumoto S, Hashizume K, Wada N, et al. Relationship between overactive bladder and irritable bowel syndrome: a large-scale internet survey in Japan using the overactive bladder symptom score and Rome III criteria. *BJU Int.* 2013;111(4):647-52.
6. Vrijens D, Drossaerts J, van Koeveeringe G, Van Kerrebroeck P, van Os J, Leue C. Affective symptoms and the overactive bladder - a systematic review. *Journal of psychosomatic research.* 2015;78(2):95-108.
7. Reynolds WS, Dmochowski R, Wein A, Bruehl S. Does central sensitization help explain idiopathic overactive bladder? *Nat Rev Urol.* 2016;13(8):481-91.
8. Leue C, Kruimel J, Vrijens D, Masclee A, van Os J, van Koeveeringe G. Functional urological disorders: a sensitized defence response in the bladder-gut-brain axis. *Nat Rev Urol.* 2016.
9. Wyman J, Burgio K, Newman D. Practical aspects of lifestyle modifications and behavioural interventions in the treatment of overactive bladder and urgency urinary incontinence. *Int J Clin Pract.* 2009;63(8):1177-91.
10. Victor E, O'connell KA, Blaivas JG. Environmental cues to urgency and leakage episodes in patients with overactive bladder syndrome: a pilot study. *J Wound Ostomy Continence Nurs.* 2012;39(2):181-6.
11. Chapple CR, Kelleher CJ, Evans CJ, et al. A narrative review of patient-reported outcomes in overactive bladder: what is the way of the future? *Eur Urol.* 2016;70(5):799-805.
12. Stull DE, Leidy NK, Parasuraman B, Chassany O. Optimal recall periods for patient-reported outcomes: challenges and potential solutions. *Curr Med Res Opin.* 2009;25(4):929-42.
13. 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. In:2009.
14. 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.
15. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol.* 2008;4: 1-32.
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):1295-302.
17. 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).

18. Morgan DL. *The Focus Group Guidebook*. Sage Publications, Inc. 1998.
19. Verhagen SJ, Hasmi L, Drukker M, van Os J, Delespaul PA. Use of the experience sampling method in the context of clinical trials. *Evid Based Ment Health*. 2016;19(3):86-9.
20. Coyne K, Revicki D, Hunt T, et al. Psychometric validation of an overactive bladder symptom and health-related quality of life questionnaire: the OAB-q. *Qual Life Res*. 2002;11(6):563-74.
21. Donovan JL, Abrams P, Peters TJ, et al. The ICS-'BPH' Study: the psychometric validity and reliability of the ICSmale questionnaire. *Br J Urol*. 1996;77(4):554-62.
22. Jackson S, Donovan J, Brookes S, Eckford S, Swithinbank L, Abrams P. The Bristol Female Lower Urinary Tract Symptoms questionnaire: development and psychometric testing. *Br J Urol*. 1996;77(6):805-12.
23. Okamura K, Nojiri Y, Osuga Y. Reliability and validity of the King's Health Questionnaire for lower urinary tract symptoms in both genders. *BJU Int*. 2009;103(12):1673-8.
24. McHorney CA, Ware JE, Jr., Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care*. 1993;31(3):247-63.
25. Courvoisier DS, Eid M, Lischetzke T. Compliance to a cell phone-based ecological momentary assessment study: the effect of time and personality characteristics. *Psychol Assess*. 2012;24(3):713-20.
26. Choi BC, Pak AW. A catalog of biases in questionnaires. *Prev Chronic Dis*. 2005;2(1):A13.
27. Heidler S, Mert C, Wehrberger C, et al. Impact of overactive bladder symptoms on sexuality in both sexes. *Urol Int*. 2010;85(4):443-6.
28. van der Walt CL, Heyns CF, Groeneveld AE, Edlin RS, van Vuuren SP. Prospective comparison of a new visual prostate symptom score versus the international prostate symptom score in men with lower urinary tract symptoms. *Urology*. 2011;78(1):17-20.
29. Bowden A CS, Sabounjian L, Sandage B, Schwiderski U, Zayed H. Psychometric validation of an urgency severity scale (IUSS) for patients with overactive bladder. *Neurourol Urodyn*. 2003;22(5):119.
30. Johnson EI, Grondin O, Barrault M, et al. Computerized ambulatory monitoring in psychiatry: a multi-site collaborative study of acceptability, compliance, and reactivity. *Int J Methods Psychiatr Res*. 2009;18(1):48-57.
31. Burkhard F.C. BJLHR, Cruz F, Lemack G.E., Nambiar A.K., Thiruchelvam N., Tubaro A. EAU Guidelines on Urinary Incontinence in Adults. European Association of Urology Guidelines. 2017;2017 edition:1-82.
32. Abrams P, Paty J, Martina R, et al. Electronic bladder diaries of differing duration versus a paper diary for data collection in overactive bladder. *Neurourol Urodyn*. 2016;35(6):743-9.
33. Milsom I, Abrams P, Cardozo L, Roberts RG, Thuroff J, Wein AJ. How widespread are the symptoms of an overactive bladder and how are they managed? A population-based prevalence study. *BJU Int*. 2001;87(9):760-6.
34. Rabiee F. Focus-group interview and data analysis. *Proc CNutr Soc*. 2004;63(4):655-60.

Chapter 9

Psychometric evaluation of an ESM-based patient-reported outcome measure for symptom assessment in irritable bowel syndrome

Lisa Vork, Daniel Keszthelyi, Sander M.J. van Kuijk, Zlatan Mujagic,
Emilio G. Quetglas, Carsten Leue, Joanna W. Kruimel, Ad A.M. Masclee

Abstract

INTRODUCTION Currently used end-of-week or end-of-day reports of gastrointestinal (GI) symptoms in irritable bowel syndrome (IBS) are considered suboptimal due to important biases. The experience sampling method (ESM) has been proposed as a more accurate measurement method. The aim of the current study is to evaluate the validity and reliability of a previously developed patient-reported outcome measure (PROM), based on the ESM, for symptom assessment in an IBS population.

METHODS Thirty-seven IBS patients (26 female; mean age 36.7) completed the electronic ESM-PROM at a maximum of ten random moments during the day and an end-of-day symptom diary, both during seven consecutive days. End-of-week questionnaires included the Gastrointestinal Symptom Rating Scale for IBS (GSRS-IBS) and IBS Severity Scoring System (IBS-SSS).

RESULTS ESM scores for GI symptoms were significantly associated to corresponding end-of-day scores, with moderate-to-good agreement between the methods. However, end-of-day scores were significantly higher than *mean* ESM scores, and for abdominal pain, end-of-day scores did not differ significantly from the *maximum* ESM level. The difference between ESM and end-of-week scores was even more pronounced and correlations were weaker. Cronbach's α coefficients were good for lower GI symptoms and moderate-to-good for four other domains within the ESM-PROM. Good consistency was shown between scores of the first half-week and scores of the second-half week.

CONCLUSIONS This study supports the validity and reliability of the previously developed electronic ESM-PROM for measuring GI symptoms in an IBS population. This PROM has the advantage of a more detailed view on individual symptom patterns, with the option to analyze symptom-symptom and symptom-environment interaction, in order to effectively monitor symptom formation, disease course and treatment response in both clinical practice and the research setting.

Introduction

Diagnosis, monitoring of disease course, and evaluation of treatment efficacy in irritable bowel syndrome (IBS) are currently primarily based on symptom severity, since reliable biomarkers for the disorder are lacking.^{1,2} As patient-reported symptomatology is the only clinical read-out available, accurate symptom monitoring is essential in this patient population. In the context of clinical trials, the use of a composite primary endpoint of both abdominal pain and stool consistency/frequency is currently recommended,^{3,4} thus addressing the key symptoms of IBS as defined by the Rome IV criteria.² The common standard for the evaluation of abdominal pain is an end-of-day 11-point numeric rating scale (NRS), averaged over seven consecutive days.^{3,4}

This retrospective information is, however, based on patients' autobiographical memory and therefore reflects certain specific moments during the day rather than a representative summary of symptoms over the entire day, leading to recall bias.⁵⁻⁷ This way of recalling also hampers the monitoring of daily fluctuations in symptoms. In addition, memory retrieval can be influenced by contextual and psychological factors at the time of recall, leading to ecological bias.^{5,6} A further disturbing factor in retrospective assessments is a lack of compliance and, especially when using paper-and-pencil methods, the occurrence of fake compliance (*i.e.* backfilling reports at a later timepoint than instructed).⁸ Altogether, documentation of patients' past symptoms can be evidently distorted when performed retrospectively, resulting in inaccurate representations of symptom patterns.

The use of real-time symptom assessment can overcome the shortcomings of these retrospective measurements. The Experience Sampling Method (ESM), also referred to as Ecological Momentary Assessment (EMA), collects in-the-moment assessments concerning the subject's current state and the real-life environment, randomly and repeatedly during the day. Compliance can be further supported by the use of an electronic device, which emits an auditory signal each time an ESM assessment is required.^{5,8}

An extensive review by Myin-Germeys *et al.* on this method concluded that the ESM allows the whole of the patient's daily life reality to be reconstructed rather than just a snapshot at the end of the day,⁶ and the method has been applied for symptom assessment in a range of disorders, *e.g.* chronic somatic pain, chronic fatigue, and psychiatric disorders.^{6,9-12} Previous studies in IBS populations have demonstrated that patients report higher abdominal pain scores in end-of-day diaries than by using ESM, suggesting a tendency to report

peak rather than average abdominal pain scores when retrospectively assessed at the end of the day.^{13,14} Altogether, electronic, repeated in-the-moment symptom assessments during regular daily life should assure ecological validity and eliminate recall bias, lead to higher compliance rates, and capture symptom variability during the day.

Recently, a patient-reported outcome measure (PROM) specific for the use of the ESM in IBS patients was developed. Content validity was assured, and the tool was cross-culturally adapted for the use in the Netherlands, Belgium (Flemish), Sweden, the UK, and the USA.¹⁵ However, further validation should be undertaken in order to assess the psychometric properties of this ESM-PROM for symptom assessment in IBS populations.

Therefore, the aim of the current study is to evaluate the validity and reliability of the previously developed ESM-PROM for symptom assessment in an IBS population, with the potential of being used as an instrument to effectively monitor disease course and treatment response.

Materials and methods

This prospective observational study was executed as part of a larger international, multicenter project. The current report presents data of one center, the Maastricht University Medical Center (Maastricht UMC), Maastricht, the Netherlands. The study protocol has been approved by the Maastricht UMC Committee of Ethics in November 2016 and was executed according to the revised Declaration of Helsinki (64th WMA General Assembly, Fortaleza, Brazil; October 2013). The study has been registered in the US National Library of Medicine (<http://www.clinicaltrials.gov>, NCT02880722).

Study participants

Recruitment of IBS patients, aged between 18 and 70 years old, took place at the outpatient clinic of Gastroenterology and Hepatology of Maastricht UMC+. In addition, subjects that previously participated in the Maastricht IBS Cohort¹⁶⁻¹⁸ were contacted to participate in the current study. IBS, including subtype assignment, was diagnosed according to the Rome IV criteria,^{2,19} which were evaluated by a trained clinical researcher in a face-to-face interview. Reason for exclusion was abdominal surgery in the past (except for uncomplicated appendectomy, cholecystectomy, and/or hysterectomy). All study subjects could only participate if they could understand the Dutch

language and were able to use the smartphone application. Subjects that did not have a smartphone able to run the application, were lent a suitable device for the duration of the study period. Participants did not change their medication use within one month before the start of study participation until the end of the study period. All subjects gave written informed consent prior to participation.

Data collection

All study participants completed the ESM and a paper end-of-day symptom diary during seven consecutive days. At the end of this week, validated symptom questionnaires were additionally completed using an electronic eCRF system (CastorEDC).

ESM

Participants completed the ESM for seven consecutive days during their regular daily life. A digital application (MEASuRE), that was specifically developed for the use of ESM in IBS patients,¹⁵ was downloaded on the participants' smartphones and was activated for the course of the study period. Subjects were instructed to carry their smartphone with them during the week and to complete the real-time questionnaires as often as possible. The MEASuRE application was set to send out an auditory and written signal ten times a day at randomly chosen moments between 07:30 a.m. and 10:30 p.m., with at least 15 minutes between consecutive signals. The ESM-questionnaire was available for 10 minutes following a signal. Questionnaires that were not completed within 10 minutes following a signal were considered as missing data. Therefore, subjects were instructed to complete as many questionnaires as possible each day, as soon as possible following each signal, but to skip questionnaires when considering completion impossible at the moment of signaling (for example when driving a car).

The development of this ESM-questionnaire, momentarily assessing GI symptoms, the affective state, and environmental factors, was previously described.¹⁵ This questionnaire was repeated in the same order at all measurement moments and questions were scored on an 11-point Numeric Rating Scale (0=*not at all* to 10=*very severely*).

End-of-day diary

A 7-day end-of-day symptom diary was used to evaluate GI symptom severity on a daily basis over seven consecutive days. GI symptoms (*i.e.* abdominal pain, abdominal discomfort, abdominal bloating, belching, flatulence, constipation, diarrhoea, as well as urge to defecate) were scored using an 11-point Numeric Rating Scale (0=*not at all* to 10=*very severely*) at the end of each day. Stool consistency for each bowel habit was recorded during the day using the Bristol Stool Chart.^{3,20} For means of comparison between conventional end-of-day diaries and the newly developed electronic, real-time ESM, the symptom diary was completed on paper.

End-of-week questionnaires

At the end of the study period, validated questionnaires were completed, assessing GI symptoms and mental health status. Regarding GI symptom severity, the Gastrointestinal Symptom Rating Scale for Irritable Bowel Syndrome (GSRS-IBS; 1-7 scale; composes sub-scores for abdominal pain, abdominal bloating, constipation, diarrhea, satiety; recall period of one week)²¹ and the Irritable Bowel Syndrome Severity Scoring System (IBS-SSS; 100mm VAS; assesses abdominal pain, bloating, and bowel habits; recall period of ten days)²² were completed. The Patient Health Questionnaire-9 (PHQ-9; 0-3 scale; calculates a total composite score for severity of depressive symptoms; recall period of two weeks)²³ and Generalized Anxiety Disorder-7 (GAD-7; 0-3 scale; calculates a total composite score for severity of anxious symptoms; recall period of two weeks)²⁴ were collected regarding depression and anxiety.

Statistical analysis

All analyses were performed using R version 3.5.1 (R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.) Continuous outcomes are presented as mean \pm standard deviation (SD), and tested using the independent or paired samples t-test. Proportions for categorical variables were tested using the χ^2 -test. A *P*-value of ≤ 0.05 was considered statistically significant. Subjects that completed at least 1/3 of the total number of assessments (*i.e.* 23 out of 70) were included in the analyses.^{25,26}

Concurrent validity was assessed by comparing ESM scores to end-of-day diary scores (*i.e.* on day-level; no repeated measures within the day) and to end-of-week questionnaire scores (*i.e.* on week-level; no repeated measures within the

week). To compare ESM scores with end-of-day diary scores, the maximum and mean scores for ESM were calculated for each of the seven days, *i.e.* combining all repeated measurements of all subjects for the concerning day. Associations between end-of-day scores and mean 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, and correcting for repeated measures by using an autoregression (AR1) correlation structure. The level of agreement between end-of-day and mean ESM scores was evaluated by calculating intra-class correlation coefficients (ICC), based on a single-rating, consistency, two-way model. In addition, differences between the assessment methods were tested using intercept-only linear mixed-effects models with the delta scores (*i.e.* difference between end-of-day diary and ESM) as the dependent variable.

To compare ESM scores with end-of-week questionnaire scores, all ESM measures were averaged to one score per subject. Differences between the measurement methods were tested using the paired samples t-test and Pearson correlations were calculated. In order to harmonize the scores of ESM and GSRS, the ESM scores were rescaled from an 11-point to a 7-point NRS. For the comparison between ESM and IBS-SSS, the scores of the latter were divided by 10.

The internal consistency of the ESM-PROM was evaluated by dividing the items into five domains and calculating the Cronbach's alpha coefficient per domain. Cronbach's alpha values of 0.7-0.9 are generally considered to reflect good internal consistency.²⁷

In order to assess test-retest reliability of the ESM-PROM we assumed that ESM scores during the first half-week of the study period (*i.e.* days 1, 2, and 3) would show moderate-to-good consistency with the scores during the second half-week (*i.e.* days 5, 6, and 7). For each subject and for each symptom a mean score for these two time-periods was calculated. The difference between these was tested using a paired samples t-test to exclude a time effect, and an ICC between the time-periods was calculated to assess agreement, using a two-way model based on average-measures and absolute agreement. ICC values less than 0.75 are considered moderate and values above 0.75 good.²⁸

Results

Study population

Thirty-seven IBS patients were enrolled and completed the study. Female sex was predominant (70%) and mean age was 36.7 ± 13.6 years. IBS subtypes were represented as follows: 14 diarrhea predominant (IBS-D), 10 constipation predominant (IBS-C), 8 mixed stool pattern (IBS-M), and 5 undefined predominant stool form (IBS-U).

Compliance

Completion rate of the ESM assessments was 68.6%, which corresponds to a mean number of completed measurements of 48 per individual over the seven-day period. The lowest number of completed measurements per subjects was 23; the highest number was 63. Figure 9.1 depicts the completion rate with ESM, showing that the majority of subjects completed between 41 and 60 beeps during the study period.

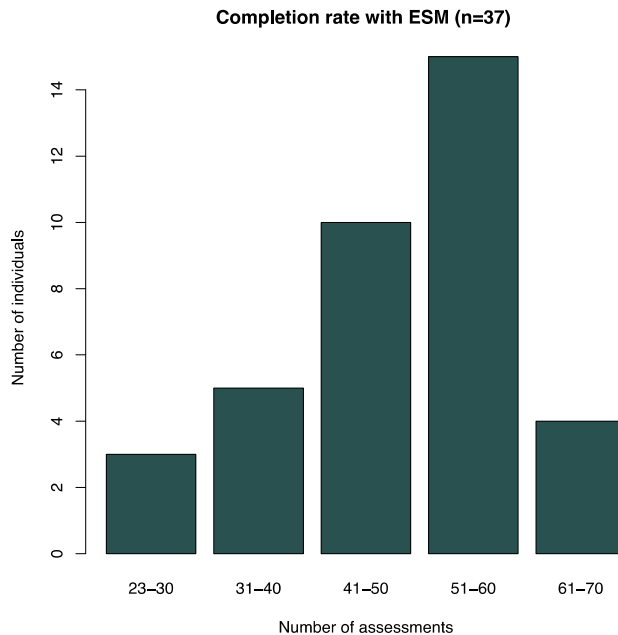


Figure 9.1 Completion rate with ESM; represented by number of individuals (y-axis) per category of completed number of assessments (x-axis).

Concurrent validity

End-of-day scores compared to ESM scores

Gastrointestinal symptoms that were scored in both the ESM and the end-of-day diary were abdominal pain, belching, bloating, flatulence, and urge to defecate. Associations between end-of-day scores and corresponding mean ESM scores were all significant, indicating that both measurement methods assess the same constructs (Table 9.1). Furthermore, ICCs between end-of-day diary scores and mean ESM scores were all above 0.500, demonstrating moderate to good agreement between the two assessment methods (Table 9.1). However, end-of-day scores were all significantly higher than mean ESM-scores (Table 9.2), suggesting that both methods reflect different levels of symptom experience during the day. End-of-day scores for abdominal pain and flatulence did not differ significantly from ESM maximum scores (Table 9.2), indicating peak-level reporting at the end of the day for these symptoms. Figure 9.2 illustrates this concept for abdominal pain.

Table 9.1 Associations and intraclass correlation coefficients between end-of-day diary scores and ESM mean scores.

	Associations		Intraclass Correlations	
	Estimate	SE	ICC	95%-CI
Abdominal pain	0.970 ***	0.093	0.706	0.638; 0.763
Belching	1.030 ***	0.079	0.610	0.526; 0.683
Bloating	0.820 ***	0.056	0.780	0.727; 0.824
Flatulence	0.810 ***	0.085	0.630	0.549; 0.699
Urge to defecate	0.607 ***	0.081	0.567	0.476; 0.645

Significance of associations tested using mixed linear models with end-of-day diary score as the dependent variable and ESM mean scores as the independent variable, corrected for repeated measures (AR1 covariate structure). Estimate indicates the strength and direction of the association.

SE: Standard error; ICC: Intraclass Correlation Coefficient. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Table 9.2 Differences between end-of-day diary scores and ESM mean and maximum scores.

	End-of-day vs. ESM mean		End-of-day vs. ESM max	
	Difference	SE	Difference	SE
Abdominal pain	1.775 ***	0.180	-0.240	0.149
Belching	0.689 ***	0.190	-0.459 *	0.179
Bloating	0.973 ***	0.169	-1.161 ***	0.182
Flatulence	1.637 ***	0.255	-0.486	0.248
Urge to defecate	1.894 ***	0.240	-0.722 **	0.232

A positive difference indicates a higher score in the end-of-day diary compared to ESM. Significance tested using mixed linear models with the delta score (*i.e.* difference between end-of-day diary score and ESM mean/max score) as the dependent variable and corrected for repeated measures (AR1 covariate structure).

SE: Standard error. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

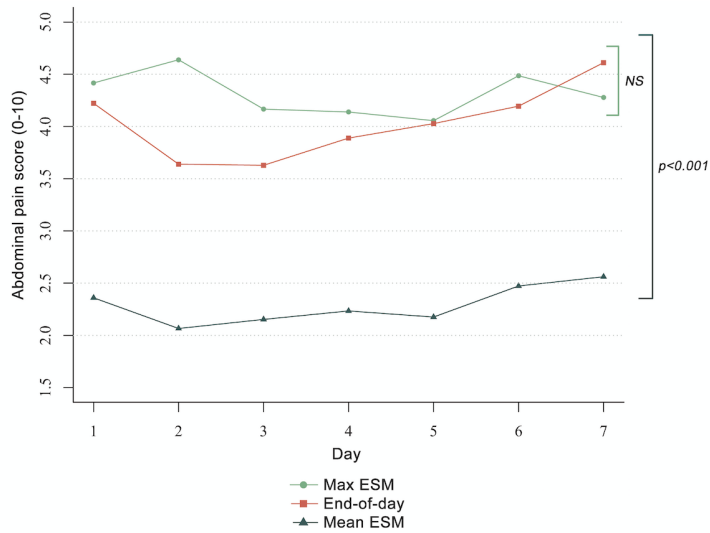


Figure 9.2 Daily end-of-day diary scores and mean and maximum ESM scores for abdominal pain (on an 11-point NRS). Difference tested using mixed linear models. NS=not significant.

In order to visualize the comparison between the ESM and end-of-day reporting into more detail, abdominal pain scores for ESM and the end-of-day diary for one individual are shown in Figure 9.3. This figure depicts a highly fluctuating pattern of abdominal pain during the 7-day study period, when assessed using the ESM, including multiple time-points with no abdominal pain at all (*i.e.* score 0) and a few time-points with higher scores (*i.e.* above 5). When using the diary, the end-of-day scores are extrapolated to the entire day, which results in an abdominal pain score higher than 5 during four of the seven days. This discrepancy between the assessment methods is highlighted in the figure.

End-of-week scores compared to ESM

Table 9.3 shows the comparison and correlation between end-of-week and ESM scores. End-of-week scores, measured by using the GSRS, were significantly higher than mean ESM scores for abdominal pain, belching, bloating, flatulence, urge to defecate, heartburn, and a full sensation. For all symptoms but heartburn, a moderate correlation was found between both measurement methods. Heartburn scores were highly correlated. IBS-SSS scores for abdominal pain and bloating (*i.e.* recall period of 10 days) were significantly higher than corresponding ESM scores, and correlation between the methods was low to moderate.

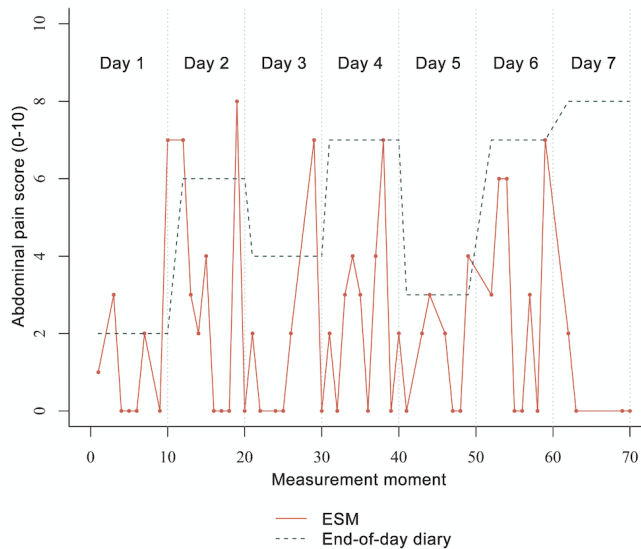


Figure 9.3 Abdominal pain scores (on an 11-point NRS) for ESM and end-of-day diary over the 7-day study period, for one IBS patient. Each day, ten assessments were available for ESM; one assessment was completed at the end of each day in the diary.

Psychological factors measured by using ESM, that corresponded to PHQ-9 or GAD-7, were “feeling low”, “feeling anxious”, “feeling worried”, “feeling irritated”, and “feeling relaxed”. Differences between the answering scales did not allow harmonization of the scores, so mean scores could not be compared between the methods. Correlations between ESM and PHQ-9/GAD-7 were low for all items.

Internal consistency

To determine internal consistency of the ESM-PROM for IBS, items were categorized in five constructs: upper GI symptoms, lower GI symptoms, physical non-GI symptoms, positive affect, and negative affect. Cronbach’s α coefficients for each of these domains are shown in Table 9.4. This demonstrates acceptable internal consistency for physical non-GI symptoms and positive affect, and good internal consistency for lower GI symptoms and negative affect. Internal consistency for upper GI symptoms is relatively low, indicating that these five symptoms might not perfectly reflect the same construct of upper GI symptoms.

Table 9.3 End-of-week scores compared to ESM scores (mean scores on subject-level).

GSRs vs. ESM	GSRs score mean ± SD	ESM score mean ± SD	Pearson correlation
Abdominal pain	4.30 ± 1.22	2.37 ± 0.98 ***	0.439
Belching	2.89 ± 1.68	1.39 ± 0.53 ***	0.642
Bloating	4.84 ± 1.42	3.02 ± 1.19 ***	0.705
Flatulence	4.16 ± 1.66	2.02 ± 0.89 ***	0.647
Urge to defecate	3.76 ± 1.80	1.98 ± 0.78 ***	0.400
Heartburn	1.57 ± 1.12	1.23 ± 0.40 *	0.887
Feeling full	3.99 ± 1.40	3.03 ± 1.15 ***	0.682
IBS-SSS vs. ESM	IBS-SSS score mean ± SD		
Abdominal pain	4.37 ± 1.76	2.26 ± 1.63 ***	0.384
Bloating	5.94 ± 1.87	3.33 ± 1.97 ***	0.534
PHQ-9/GAD-7 vs. ESM			
Feeling low			0.413
Anxious			0.359
Worried			0.497
Irritated			0.424
Relaxed			-0.193

GSRs vs. ESM: ESM scores were rescaled from 11-point to 7-point scale. IBS-SSS vs. ESM: IBS-SSS VAS scores were divided by 10 to match the 11-point ESM scale; ESM scores were not rescaled. PHQ-9/GAD-7 vs ESM: answering scales do not allow comparison between mean scores. Differences tested using paired samples t-test. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Table 9.4 Internal consistency for five symptom-domains within the ESM-PROM, reflected by Cronbach's α coefficient.

	Cronbach's α
Lower GI symptoms	0.750
Abdominal pain	
Bloating	
Flatulence	
Urge to defecate	
Upper GI symptoms	0.545
Rumbling sounds	
Nausea	
Burping	
Heartburn	
Feeling full	
Physical – non-GI	0.635
Palpitations	
Sweating	
Dyspnea	
Dizziness	
Mental – Positive affect	0.623
Good	
Relaxed	
Mental – Negative affect	0.832
Low	
Anxious	
Irritated	
Stressed	
Worried	

Test-retest reliability

Table 9.5 shows mean ESM scores for the first (*i.e.* day 1, 2, 3) and second (*i.e.* day 5, 6, 7) half-week of the study period. Scores did not differ significantly between these two time periods, except for “painful muscles”, “painful joints”, and “urinary urge”, which were all scored higher in the second half-week. ICCs for all symptoms, but palpitations, were above 0.600, indicating moderate to good consistency between the measurements. These findings indicate sufficient test-retest reliability.

Table 9.5 Mean scores for ESM-reported gastrointestinal symptoms, non-gastrointestinal physical symptoms, and mental status, separately for the first half-week and second half-week of the study period. ICC reflects agreement between first and second half-week.

	First half-week Mean score ± SD	Second half-week Mean score ± SD	ICC [95%-CI]
Gastrointestinal symptoms			
Abdominal pain	2.18 ± 1.56	2.36 ± 1.85	0.870 [0.750; 0.933]
Intestinal gas	1.82 ± 1.50	1.61 ± 1.73	0.810 [0.632; 0.902]
Rumbling sounds	2.43 ± 2.08	2.66 ± 2.22	0.915 [0.836; 0.956]
Bloating	3.20 ± 1.99	3.50 ± 2.16	0.887 [0.782; 0.942]
Defecatory urge	1.71 ± 1.55	1.50 ± 1.20	0.769 [0.554; 0.880]
Nausea	0.63 ± 0.92	0.57 ± 0.92	0.811 [0.632; 0.903]
Burping	0.62 ± 0.91	0.69 ± 1.08	0.777 [0.566; 0.885]
Heartburn	0.37 ± 0.73	0.35 ± 0.76	0.817 [0.643; 0.906]
Feeling full	3.22 ± 2.05	3.57 ± 2.04	0.866 [0.740; 0.931]
Physical symptoms (non-gastrointestinal)			
Palpitations	0.14 ± 0.30	0.22 ± 0.83	0.558 [0.141; 0.773]
Sweating	0.88 ± 1.11	0.89 ± 1.47	0.910 [0.824; 0.954]
Dyspnea	0.52 ± 1.15	0.76 ± 1.62	0.909 [0.820; 0.954]
Dizziness	0.54 ± 0.93	0.67 ± 1.21	0.752 [0.520; 0.872]
Painful muscles	1.69 ± 2.20	2.44 ± 2.42 **	0.872 [0.693; 0.940]
Painful joints	1.91 ± 2.49	2.40 ± 2.51 **	0.944 [0.872; 0.973]
Urinary urge	1.33 ± 1.29	1.62 ± 1.54 *	0.888 [0.775; 0.943]
Mental status			
Good	6.41 ± 1.84	6.51 ± 1.92	0.955 [0.914; 0.977]
Low	0.86 ± 1.18	0.73 ± 1.10	0.824 [0.661; 0.909]
Anxious	0.30 ± 0.55	0.37 ± 0.75	0.752 [0.520; 0.872]
Irritated	0.71 ± 1.01	0.89 ± 1.32	0.897 [0.800; 0.947]
Stressed	1.25 ± 1.33	1.29 ± 1.48	0.775 [0.560; 0.884]
Relaxed	6.41 ± 1.27	6.45 ± 1.34	0.894 [0.793; 0.945]
Worried	1.21 ± 1.20	1.19 ± 1.68	0.774 [0.560; 0.884]

First half week represents day 1, 2, and 3; second half week represents day 5, 6, and 7. Significance tested using paired samples t-test. SD: Standard deviation. ICC: Intraclass Correlation Coefficient. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Discussion

The current study evaluated concurrent validity, internal consistency, and test-retest reliability of a previously developed PROM based on the ESM-principle for symptom assessment in IBS populations. Content validity and cross-cultural adaptation of this tool were described previously.¹⁵ We here demonstrate that ESM scores for GI symptoms are significantly associated to corresponding end-of-day scores, with moderate-to-good agreement between the methods. However, end-of-day scores were significantly higher than *mean* ESM scores, and for abdominal pain, the key GI-symptom in IBS, end-of-day scores did not differ significantly from the *maximum* ESM level during the day. When comparing ESM to end-of-week scores, the difference between the methods was even more pronounced and correlations were weaker. Cronbach's α coefficients were good for lower GI symptoms and moderate-to-good for four other domains within the ESM-PROM. Good consistency was shown between scores of the first half-week and scores of the second-half week. Altogether, this indicates adequate concurrent validity, moderate-to-good internal consistency, and good test-retest reliability.

All newly developed PROMs are currently expected to show adequate validity and reliability prior to the use as an outcome in clinical trials or patient health care. Multiple statistical methods that are supposed to test psychometric properties have been described throughout the years, and several guidelines have been published recommending how to accurately evaluate a new PROM.^{27,29} The FDA specifically addressed this issue for IBS and provides reference on how to review PROMs that are suggested for the use in IBS clinical studies.³ We here report on the measures of validity and reliability that are most applicable to our electronic ESM-PROM. An important note to make is that, as opposed to frequently used static measures that describe health status at a specific time-point, the ESM provides dynamic assessments of symptomatology, allowing for the detection of short-term fluctuations. Furthermore, several items are included in the ESM-PROM describing the environmental and social context of a subject at the time of recall, which are not suitable for psychometric evaluation. Therefore, it was expected that solely the statistical values would not reflect perfect validity and reliability, but we would rather be able to give a detailed overview of GI-symptom monitoring using this new electronic tool, as compared to currently available non-momentary PROMs for IBS.

As IBS is a heterogeneous disorder often presenting with psychological disturbances and/or general somatic complaints, also reflected by lower levels of generic mental and physical quality of life,³⁰ items were included on mental state and frequently reported physical complaints in addition to GI symptoms. These were extracted from existing IBS scales and focus groups with IBS patients, as described previously.¹⁵ The assessment of patients' mental state using ESM has been thoroughly described in psychiatric research.³¹⁻³³ The analyses in the current report, however, focused on the GI-part of the PROM as GI-symptoms represent the key symptoms of IBS. Hence, measuring those is the main purpose of the developed tool.

A 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 could have resulted in low compliance, which would obviously not contribute to the accuracy of the measure. However, in the current study, nearly 69% of the total assessments was completed, which is considerably higher than the generally accepted completion rates of at least 33% for ESM in other studies.^{25,26} Since each ESM assessment was only available during the instructed time frame and was marked with an electronic time stamp, reported completion rates are accurate and reliable, which further supports compliance with and adherence to this method.

With regard to concurrent validity, we found significant associations between ESM and end-of-day scores for corresponding GI symptoms, replicating the results of an initial pilot study on the performance of the ESM in an IBS population.¹⁴ Together with moderate-to-good agreement between the assessment methods, this indicates that the ESM-PROM is measuring similar constructs regarding key IBS symptoms. A difference between the methods, with a tendency to report peak levels of GI symptoms when assessed retrospectively at the end of the day, has also been reported previously.^{13,14} The fact that the scores reported in the end-of-day diary did not reflect the last ESM assessments of the day, supports the suggestion of peak-reporting at the end of the day, while confirming that the abdominal symptoms are not worst at that moment per se. Summarizing, the newly developed ESM-PROM has shown here to measure similar constructs, but considerable different levels of GI symptoms, as compared to the conventional end-of-day scores.

As end-of-day reporting is generally accepted and recommended by the regulatory authorities for the use as outcome measures in IBS clinical trials,^{3,4} and the ESM is associated with a higher patient burden, it might be of interest to consider what benefits the ESM provides. One important asset is the option

to give a more detailed view on individual symptom patterns, including symptom formation. With regard to the differences between real-time and end-of-day reporting, we here demonstrated that the ESM is suitable to provide a detailed overview of a person's symptoms over a period of seven days, providing patients and caregivers with insight into fluctuations from assessment to assessment. Apart from reflecting different, possibly more accurate, levels of symptoms, the ESM thus draws a detailed picture of the dynamics of IBS symptoms during daily life. Additionally, it offers the opportunity to analyze interactions with other symptoms, psychological experiences, or environmental factors.^{6,34}

Internal consistency was demonstrated to be good for lower GI symptoms, but relatively poor for upper GI symptoms. This suggests that the five symptoms within the latter category (*i.e.* rumbling sounds, nausea, burping, heartburn, and feeling full) do not perfectly reflect this symptom domain, but should possibly be considered as distinct constructs. However, since these items were selected based on focus group interviews within the target population, we consider those useful for screening of frequently occurring upper GI symptoms in an IBS population. Given the low internal consistency, this part of the tool should be adjusted, for example, when using it in a functional dyspepsia population.

In the current study, the ESM-PROM was completed during one week with continuous repeated measurements. Since the ESM is a real-time assessment tool, test conditions depend on the natural daily life of the subject and are, therefore, never exactly similar between different measurements. Furthermore, IBS symptoms are known to fluctuate over short periods of time. Hence, test-retest reliability is not expected to be perfect for ESM. One could speculate that each subject presents with a symptom pattern that identifies them to a certain extent and that between-subject differences are larger than within-subject differences. Therefore, to expand on test-retest reliability we compared and correlated mean scores of the first half-week to those of the second half-week. For all lower and upper GI symptoms as well as psychological items reflecting positive and negative affect, the scores did not differ significantly between the two time periods, and correlations were moderate to high. For the measured general physical complaints, this was less consistent, with a particularly low correlation for "palpitations". This is possibly explained by the fact that mean scores for this item were very low, with subjects mainly scoring 0 on this item, resulting in a lack of variability between subjects. In order to confidently report on test-retest reliability for the physical non-GI symptoms, analyses should be

performed in a subgroup of IBS patients presenting with these symptoms. However, this is not considered relevant for the current report focusing on the assessment of key GI symptoms.

A strength of our study is that the evaluated tool was previously designed based on specific recommendations on PROM development and IBS outcomes^{3,29}. A potential shortcoming is that our sample size is relatively small. Study populations of at least 50 subjects have been recommended in order to accurately assess validity and reliability of health measures.³⁵ However, a considerable strength of the ESM in this regard is the large number of repeated measures, that significantly increases power of the analyses.

The study is currently further ongoing in four other centers. Here, we report on data of the first center that completed the study. We anticipate that based on data of the five centers, firm conclusions on psychometric properties of the tool can be drawn and additional insight may be obtained into the performance of the measurement in different sociocultural populations. In addition to the present analyses, an important aspect of testing the quality of a PROM is responsiveness or the ability to detect change.^{27,36} We did not include this in the current report, but this should be tested in the future before using the ESM-PROM in the evaluation of treatment efficacy in IBS populations.

In conclusion, the current results support the validity and reliability of the newly developed electronic instrument for measuring GI symptoms in an IBS population, based on the experience sampling methodology. This ESM-PROM has the advantage of a more detailed view on individual symptom patterns, with the option to analyze symptom-symptom and symptom-environment interactions, in order to aid disease monitoring in both clinical practice and research settings. Future research is warranted to consider the responsiveness of the ESM-PROM, in order to determine whether it could also play a role in the evaluation of treatment efficacy.

References

1. Mearin F, Lacy BE. Diagnostic criteria in IBS: useful or not? *Neurogastroenterol Motil* 2012;24(9): 791-801.
2. Mearin F, Lacy BE, Chang L, et al. Bowel Disorders. *Gastroenterology* 2016;150(6): 1393-1407.
3. U.S. Department of Health and Human Services FaDA, Center for Drug Evaluation and Research (CDER). Guidance for Industry Irritable Bowel Syndrome - Clinical Evaluation of Drugs for Treatment. 2012.
4. (EMA) EMA. Guideline on the evaluation of medicinal products for the treatment of irritable bowel syndrome. 2014.
5. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol*. 2008;4:1-32.
6. 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.
7. Houtveen JH, Oei NY. Recall bias in reporting medically unexplained symptoms comes from semantic memory. *J Psychosom Res*. 2007;62(3):277-282.
8. 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.
9. Bos FM, Schoevers RA, aan het Rot M. Experience sampling and ecological momentary assessment studies in psychopharmacology: A systematic review. *Eur Neuropsychopharmacol*. 2015;25(11): 1853-1864.
10. Jamison RN, Raymond SA, Slawsby EA, McHugo GJ, Baird JC. Pain assessment in patients with low back pain: comparison of weekly recall and momentary electronic data. *J Pain*. 2006;7(3):192-199.
11. Stone AA, Broderick JE, Schwartz JE, Shiffman S, Litcher-Kelly L, Calvanese P. Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. *Pain*. 2003;104(1-2):343-351.
12. Friedberg F, Sohl SJ. Memory for fatigue in chronic fatigue syndrome: the relation between weekly recall and momentary ratings. *Int J Behav Med*. 2008;15(1):29-33.
13. 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.
14. 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.
15. 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).
16. Ludidi S, Mujagic Z, Jonkers D, et al. Markers for visceral hypersensitivity in patients with irritable bowel syndrome. *Neurogastroenterol Motil*. 2014;26(8):1104-1111.
17. Thijssen AY, Mujagic Z, Jonkers DM, et al. Alterations in serotonin metabolism in the irritable bowel syndrome. *Aliment Pharmacol Ther*. 2016;43(2):272-282.
18. Mujagic Z, Ludidi S, Keszthelyi D, et al. Small intestinal permeability is increased in diarrhoea predominant IBS, while alterations in gastroduodenal permeability in all IBS subtypes are largely attributable to confounders. *Aliment Pharmacol Ther*. 2014;40(3):288-297.

19. Palsson O, Whitehead W, van Tilburg MA, et al. Development and Validation of the Rome IV Diagnostic Questionnaire for Adults. *Gastroenterology*. 2016;150:1481-1491.
20. Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit time. *Scand J Gastroenterol*. 1997;32(9):920-924.
21. Wiklund IK, Fullerton S, Hawkey CJ, et al. An irritable bowel syndrome-specific symptom questionnaire: development and validation. *Scand J Gastroenterol*. 2003;38(9):947-954.
22. Francis CY, Morris J, Whorwell PJ. The irritable bowel severity scoring system: a simple method of monitoring irritable bowel syndrome and its progress. *Aliment Pharmacol Ther*. 1997;11(2):395-402.
23. 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.
24. 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.
25. 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.
26. Delespaul P. Assessing schizophrenia in daily life the experience sampling method.: UPM, Universitaire Pers Maastricht, Maastricht University; 1995.
27. Alrubaiy L, Hutchings HA, Williams JG. Assessing patient reported outcome measures: A practical guide for gastroenterologists. *United Eur Gastroenterol J*. 2014;2(6):463-470.
28. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155-163.
29. 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.
30. Spiegel BM. The burden of IBS: looking at metrics. *Curr Gastroenterol Rep*. 2009;11(4):265-269.
31. 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.
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 Psychiatr Scand*. 2003;107(2):124-131.
33. 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. *Eur Psychiatry*. 2015;30(4):441-447.
34. Verhagen SJ, Hasmi L, Drukker M, van Os J, Delespaul PA. Use of the experience sampling method in the context of clinical trials. *Evid-Based Ment Health*. 2016;19(3):86-89.
35. 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.
36. 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.



PART III
THE EXPERIENCE SAMPLING METHOD IN IBS
APPLICATIONS IN CLINICAL RESEARCH

Chapter 10

Patient specific stress-abdominal pain interaction in irritable bowel syndrome: An experience sampling method study

Lisa Vork, Daniel Keszthelyi, Sander M.J. van Kuijk, Emilio G. Quetglas, Hans Törnblom, Magnus Simrén, Qasim Aziz, Maura Corsetti, Jan Tack, Zlatan Mujagic, Carsten Leue, Joanna W. Kruimel, Ad A.M. Masclee

Clin Transl Gastroenterol. Jul;11(7)

Abstract

INTRODUCTION Gastrointestinal symptoms in irritable bowel syndrome (IBS) have been correlated to psychological factors using retrospective symptom assessment. However, real-time symptom assessment might reveal the interplay between abdominal and affective symptoms more reliably in a longitudinal perspective. The aim of this study was to evaluate the association between stress and abdominal pain, using the Experience Sampling Method (ESM) as a real-time and repeated measurement method.

METHODS Thirty-seven IBS patients (26 female; mean age 36.7) and 36 healthy subjects (HC; 24 female; mean age 31.1) completed an electronic ESM during seven consecutive days. Abdominal pain and stress were scored on an 11-point Numeric Rating Scale at a maximum of ten random moments each day.

RESULTS Abdominal pain scores were 2.21 points higher in IBS compared to HC ($P < 0.001$), whereas stress levels did not differ significantly ($B: 0.250$, $P = 0.406$). In IBS, a 1-point increase in stress was associated with, on average, 0.10 points increase in abdominal pain ($P = 0.017$). In HC, this was only 0.02 ($P = 0.002$). Stress levels at $t = -1$ were not a significant predictor for abdominal pain at $t = 0$ in both groups, and vice versa.

CONCLUSIONS Our results demonstrate a positive association between real-time stress and abdominal pain scores, and indicate a difference in response to stress and not a difference in experienced stress per se. Furthermore, an in-the-moment rather than a longitudinal association is suggested. This study underlines the importance of considering the individual flow of daily life and supports the use of real-time measurement when interpreting potential influencers of abdominal symptoms in IBS.

Introduction

Irritable bowel syndrome (IBS) is a disorder of gut-brain interaction, in which psychological factors play a role in both symptom origination and perpetuation.¹ Associations between anxiety and depression and IBS have extensively been studied, but also daily stress and life hassles have been associated with gastrointestinal (GI) symptoms in IBS.^{2,3} However, whether those psychological features precede or follow the occurrence of GI symptoms remains to be elucidated.^{4,5} Furthermore, it is likely that the order of occurrence of these symptoms and the magnitude of the association between both, differ between subjects. Further unraveling this association by considering the heterogeneity between IBS patients could give insight in individual interactions between psychological and GI symptoms, leading the way to a more personalized approach of IBS.

The majority of previous studies on this issue are either cross-sectional studies that concurrently assessed both psychological and GI symptoms for between-subject analyses, or longitudinal studies that report within-subject analyses based on repeated assessments on daily or weekly basis.⁶⁻⁸ Although the latter methodology offers a better opportunity to analyze the relationship between both symptoms over time, it does not take into account the individual daily life flow of symptom interactions.

IBS has been described to present as a relapsing-remitting disorder, with symptoms varying over time, where highly fluctuating symptom patterns are often even reported within one day.^{9,10} This underlines the importance of considering this daily life symptom variability when evaluating associations with possible triggering or concomitant factors, such as stress. In addition, as is true for the assessment of abdominal pain^{11,12}, end-of-week or end-of-day reports of stress are likely influenced by recall bias, since stress (*i.e.* in terms of daily life hassles) is a complex process subject to within-day fluctuations.^{13,14} This further emphasizes the need for studying the temporal interplay between stress and GI symptoms.

For this purpose, we previously developed an electronic smartphone-based patient-reported outcome measure (PROM) specifically for the use of the Experience Sampling Methodology (ESM) in IBS populations.¹⁵ The ESM is a momentary assessment method collecting repeated measurements randomly during the day, that concern the current status and natural environment of the subjects. These repeated in-the-moment assessments result in an extensive individual pattern of symptoms, and provide insight into the complex interplay

between this longitudinal symptom formation and possibly associated daily life factors.

Chan *et al.* recently reported results of an ESM study evaluating the temporal relationship between psychological factors and bowel symptoms in IBS patients with predominant diarrhea (IBS-D) and healthy volunteers.¹⁶ An interesting finding was a negative association between abdominal pain and preceding stress levels, whereas abdominal pain was predictive for the occurrence of daily life stress and negative affect. In this report, a potential association between concurrent abdominal and psychological symptoms was not evaluated, while Blanchard *et al.* previously pointed towards an important association as such using daily reports.⁶

Therefore, the aim of the current study was to evaluate the association between stress and abdominal pain, using concurrent as well as time-lagged assessments, in patients with irritable bowel syndrome and healthy subjects, using the Experience Sampling Method. A detailed interpretation of the association between daily life stress and abdominal pain on the subject-level, taking into account between-subject heterogeneity, is provided.

Materials and methods

This prospective observational study was executed as part of a larger international, multicenter project. The current report presents data of one center, the Maastricht University Medical Center+ (Maastricht UMC+), Maastricht, the Netherlands. The study protocol has been approved by the Maastricht UMC+ Committee of Ethics in November 2016 and was executed according to the revised Declaration of Helsinki (64th WMA General Assembly, Fortaleza, Brazil; October 2013). The study has been registered in the US National Library of Medicine (<http://www.clinicaltrials.gov>, NCT02880722).

Study participants

Recruitment of IBS patients, aged between 18 and 70 years, took place at the outpatient clinic of Gastroenterology-Hepatology of Maastricht UMC+, a secondary/tertiary referral center. In addition, subjects that previously participated in the Maastricht IBS Cohort¹⁷⁻¹⁹ were contacted to participate in the current study. IBS, including subtype assignment, was diagnosed according to the Rome IV criteria^{20,21}, which were evaluated by a trained clinical researcher in a face-to-face interview. Reason for exclusion was abdominal

surgery in the past (except for uncomplicated appendectomy, cholecystectomy, and/or hysterectomy).

Healthy subjects, also aged between 18 and 70 years, were recruited via advertisements in Maastricht UMC+. Subjects were eligible if they did not have a past or present diagnosis of any GI disorder and did not fulfill Rome IV criteria for IBS. All study subjects could only participate if they could understand the Dutch language and were able to use the smartphone application. Subjects that did not have a smartphone able to run the application, were provided a suitable device for the duration of the study period. Participants were not allowed to change their medication use, or start any non-pharmacological treatment within one month before the start of study participation until the end of the study period. All subjects gave written informed consent prior to participation.

Data collection

All study participants completed the ESM for seven consecutive days during their regular daily life. A digital application (Maastricht Electronic Abdominal Symptom REcording; MEASuRE), that was specifically developed for the use of ESM in IBS populations¹⁵, was downloaded on the participants' smartphones and was activated for the course of the study period. Subjects were instructed to carry their smartphone with them during this week and to complete the real-time questionnaires as often as possible. The MEASuRE application was set to send out an auditory and written signal ten times a day at random moments between 07:30 a.m. and 10:30 p.m., with at least 15 minutes and a maximum of 3 hours between subsequent signals. Every 10 minutes following a signal, the ESM questionnaire was available. Questionnaires that were not completed within 10 minutes following a signal were considered missing data. Therefore, subjects were instructed to complete as many questionnaires as possible each day, as soon as possible following each signal, but to skip questionnaires when considering completing impossible at that moment (for example when driving a car). Previous studies using this specific ESM algorithm have shown reasonably high completion rates.^{11,22}

The development of this ESM questionnaire for the momentary assessment of GI symptoms, the affective state, and environmental factors, was described previously.¹⁵ This questionnaire was repeated in the same order at all measurement moments and questions were scored on an 11-point Numeric Rating Scale (0=*not at all* to 10=*very severely*), according to the FDA and EMA recommendations.^{23,24} Figure 10.1 shows a screenshot of the MEASuRE

application for the items “abdominal pain” and “stress”, that were used in the current analyses.

At the end of the 7-day study period, the Patient Health Questionnaire-9 (PHQ-9; 0-3 scale; calculates a total composite score for severity of depressive symptoms; recall period of two weeks)²⁵, Generalized Anxiety Disorder-7 (GAD-7; 0-3 scale; calculates a total composite score for severity of anxious symptoms; recall period of two weeks)²⁶, and the Visceral Sensitivity Index (VSI) were completed.^{27,28} The 15-item VSI assesses GI-specific anxiety on a 6-point Likert scale. A composite sum score of all items is calculated; higher scores indicate more GI-specific anxiety. The item “In stressful situations, my belly bothers me a lot” was here used as the self-reported association between stress and abdominal symptoms.

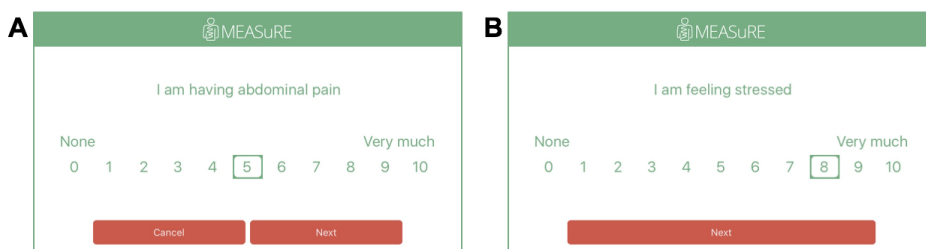


Figure 10.1 Screenshots of ESM questions for the assessment of abdominal pain (A) and stress (B), respectively. Questions were phrased in Dutch; here translated into English.

Statistical analysis

All analyses were performed using R version 3.5.1 (R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria). Continuous outcomes are presented as mean \pm standard deviation (SD), and differences between groups were tested using the independent samples t-test. Proportions for categorical variables were tested using the χ^2 -test. A *P*-value of <0.05 was considered statistically significant. Subjects who completed at least 1/3 of the total number of assessments (*i.e.* 23 out of 70) were included in the analyses regarding ESM data.^{29,30}

Levels of abdominal pain and stress over the 7-day study period were compared between IBS patients and healthy subjects using linear mixed-effects models, since ESM data are hierarchical with repeated measures (level 1) nested within subjects (level 2). ESM scores (*i.e.* abdominal pain or stress) were used as the

dependent variable and disease status (*i.e.* IBS or healthy) as the predictor variable. Models were corrected for repeated measures and autocorrelation, using an autoregressive covariate (AR1) structure.

In order to evaluate associations between stress and abdominal pain, similar analyses were performed. First, the association between concurrent stress and abdominal pain was evaluated, using abdominal pain scores as dependent and stress scores as independent variables. A first model including the total study population included the interaction between stress and disease status as a predictor variable. When finding a significant association between this interaction term and abdominal pain scores, models were performed separately for IBS and healthy subjects, to further evaluate the association between stress and abdominal pain in each group. Since it was *a priori* hypothesized that stress and abdominal pain would particularly be related in subgroups of IBS patients, the model was performed separately for each IBS subject, in order to obtain individual regression coefficients.

Secondly, we assessed whether stress scores at one point in time could predict subsequent abdominal pain scores, by using lagged scores (*i.e.* $t=-1$) of stress as the predictor variable. Similarly, lagged scores for abdominal pain were tested as predictor variables in models with stress scores as the dependent variable, to assess a possible association between stress scores and preceding abdominal pain scores. In all models, random slopes for predictor variables were tested, but the models with the best model-fit (*i.e.* based on Akaike Information Criterion) are here reported.

Results

Study population

Thirty-seven IBS patients and 36 healthy subjects (HC) were enrolled and completed the study. Female sex was predominant in both groups (70% in IBS and 67% in HC) and mean age was 36.7 ± 13.6 years in IBS and 31.2 ± 17.7 years in healthy subjects. These characteristics were not significantly different between the groups. In the IBS population, IBS subtypes were represented as follows: 14 diarrhea predominant (IBS-D), 10 constipation predominant (IBS-C), 8 mixed stool pattern (IBS-M), and 5 undefined predominant stool pattern (IBS-U). In both IBS and HC, the majority of subjects scored only minimal or mild for both depressive (PHQ-9) and anxious (GAD-7) symptoms, *i.e.* total PHQ-9 or GAD-7 score of 0-9. These scores were not significantly different between the

groups. Completion rate of ESM was 68.6% in the IBS and 69.4% in the control group. All subjects completed at least 1/3 of the total number of assessments (70), and were therefore all included in the analyses.

Levels of abdominal pain and stress

Overall, mean abdominal pain scores from day to day were significantly higher in the IBS group than pain scores reported by the healthy subjects. On average, IBS subjects reported abdominal pain scores 2.21 points higher compared to HC (SE: 0.27, $p < 0.001$). However, mean stress levels were not significantly different between the groups (mean difference: 0.25, SE: 0.30, $P = 0.406$) (Figure 10.2).

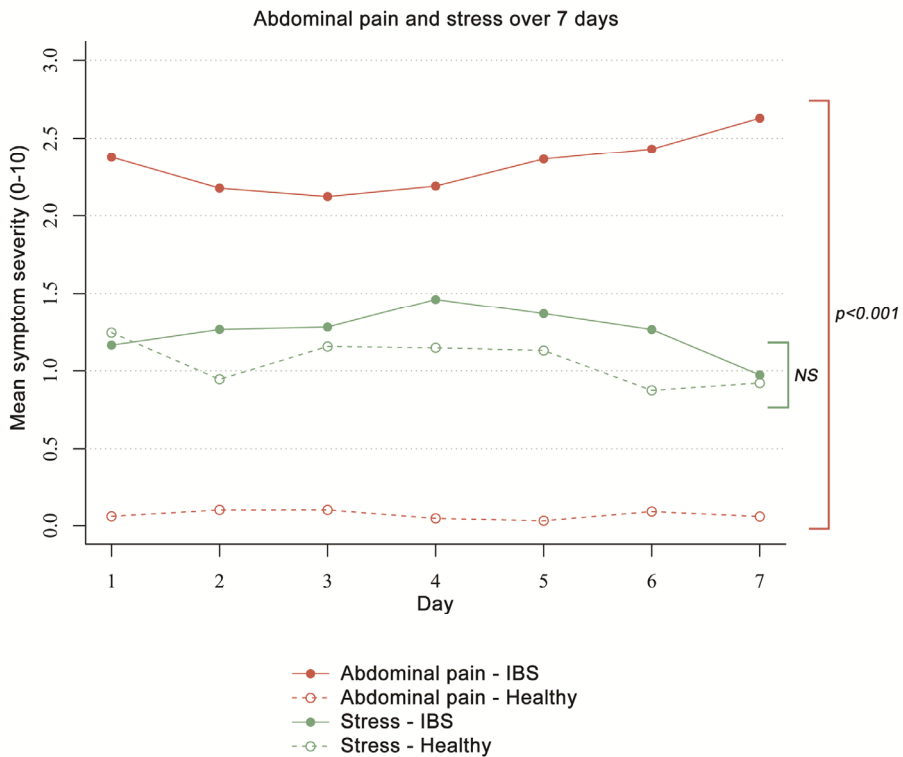


Figure 10.2 Symptom severity for abdominal pain and stress over the 7-day study period, separately for IBS and healthy subjects. Symptom scores are measured using ESM over 7 days; mean scores per day are presented.

Association between stress and abdominal pain scores

In the IBS population, a significant association was found between concurrent stress and abdominal pain scores. On average, a 1-point increase in stress was associated with 0.10 points increase in abdominal pain (SE: 0.04, $P=0.017$). In healthy subjects, this was only 0.02 (SE: 0.01, $P=0.002$). Figure 10.3 visualizes the repeated measures for abdominal pain and stress over the 7-day study period, separately for three of the IBS patients, which were selected to illustrate between-subject differences in the association between stress and abdominal pain. The plots show that the interplay between stress and abdominal pain is different for each of the three individuals. This is confirmed when performing the above described linear mixed-effects model, separately for each IBS patient: the coefficients for the association between stress and abdominal pain range from -1.92 to 1.56. This means that in one subject a 1-point increase in stress corresponded to a *decrease* in abdominal pain of 1.92 points, whereas in another this was associated with 1.56 points *increase* in abdominal pain.

Stress levels at $t=-1$ (*i.e.* lagged scores) were not a significant predictor for abdominal pain at $t=0$ in both IBS patients and healthy subjects. Likewise, abdominal pain scores at $t=-1$ could not predict stress levels at $t=0$ in both groups.

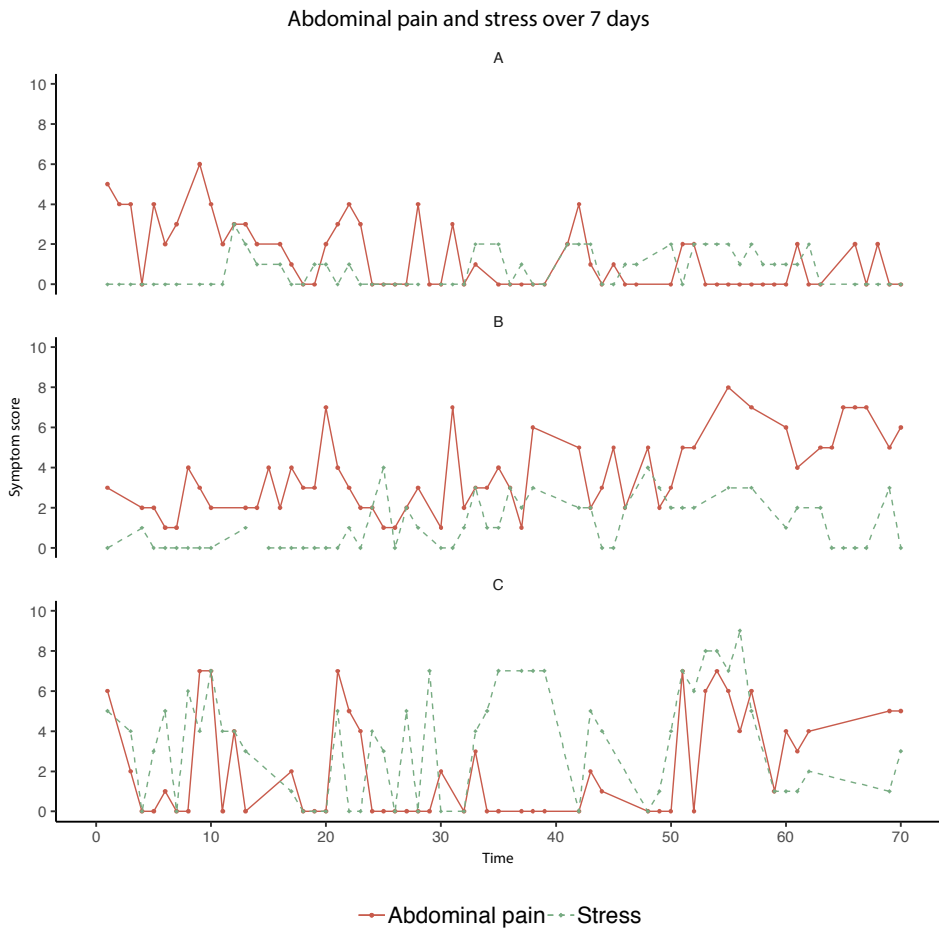


Figure 10.3 Symptom severity scores for abdominal pain and stress over seven days, separately for three IBS subjects. The heterogeneity between IBS subjects is shown by the different patterns of stress and abdominal pain; also indicated by the difference in corresponding regression coefficients for **A)** -0.23, **B)** 0.07, and **C)** 0.28.

Agreement with self-reported association between stress and IBS symptoms

The Visceral Sensitivity Index item “In stressful situations, my belly bothers me a lot” can be interpreted as a representation of the subjectively experienced association between stress and GI symptoms (“subjective association”). If this corresponds well with the association as found in the current real-time ESM data (“objective association”), this would indicate that subjects are able to

retrospectively report an association between stress and abdominal pain reliably.

In total, 81.1% of the subjects reported to have more GI symptoms during stressful situations to some extent (*i.e.* answering options “mildly agree”, “moderately agree”, or “strongly agree”). More than half of the study population (54.1%) strongly agreed with this statement. As described above, we found large heterogeneity in the magnitude of the association between stress and abdominal pain, when assessing the real-time ESM data (“objective association”). Together with a correlation coefficient of only 0.125 between the subjective (*i.e.* VSI) and the objective (*i.e.* ESM) association, this suggests that IBS patients retrospectively report an association between stress and abdominal pain as more important compared to when momentary assessed during daily life.

Discussion

The current ESM study demonstrates that concurrent levels of stress and abdominal pain are associated in IBS patients, but not in healthy subjects. High heterogeneity in the magnitude of this association was found, pointing towards different levels of stress-sensitivity within the IBS population. Whereas average abdominal pain scores over seven days were significantly higher in IBS compared to healthy subjects, stress levels did not differ between the groups. An association between stress and preceding abdominal pain scores could not be confirmed, and neither were stress levels predictive for subsequent abdominal pain scores. In addition, a subjectively reported relationship between stress and GI symptoms, represented by the Visceral Sensitivity Index, was not in agreement with the association as found within the ESM data, suggesting overestimation of the role of stress in retrospective self-reported questionnaires.

As expected, reported abdominal pain was on average significantly lower in the healthy subjects compared to individuals diagnosed with IBS. In contrast, momentary stress levels were not significantly different over time between IBS patients and their non-IBS counterparts. Both findings are consistent with a recent ESM study of Chan *et al.*¹⁶, in an IBS-D population. In addition to this we could demonstrate a significant and clinically relevant association between stress levels and concurrent abdominal pain scores in the IBS but not the healthy group, which suggests that IBS subjects do not experience more stress

per se, but appear to be more susceptible to a greater impact of feeling stressed during daily life. This has been described previously in the context of (major) life events.^{31,32} It has been suggested that the association could work the other way around as well; in that case abdominal pain would lead to (increasing) stress^{2,7,33}, but this would have been reflected by higher stress levels in IBS subjects compared to the control population. The current ESM analyses do not allow us to draw conclusions in terms of a causal relation, but based on the above-mentioned considerations, we assume that stress triggers the onset or worsening of abdominal pain in IBS patients.

An interaction between concurrent stress and abdominal pain as such was previously demonstrated in a report by Blanchard and colleagues⁶, in which weekly averages of daily symptom scores were analyzed. The current study adds evidence to this by exploring short-term associations in more detail, that is within the day. Our results underline that even within-day changes in stress can be associated with daily fluctuations in abdominal pain, of which the pattern could be overlooked when only analyzing on week-level. Furthermore, we demonstrate high heterogeneity in the magnitude of the association between subjects, which points towards different levels of stress-sensitivity in terms of a stress-abdominal pain interaction within the IBS population. This emphasizes the need for an individualized approach in both research and clinical management regarding IBS.

Daily life stress did not predict subsequent abdominal pain scores, or vice versa, contradicting the findings of Chan *et al.*¹⁶ Whereas in this recent study daily life stress was represented by the stressfulness of the major activity that took place since the previous entry, we used the scoring of to what extent a subject felt stressed at the current entry. Likewise, the reporting of abdominal pain was different: Chan and colleagues analyzed to what extent abdominal pain between the current and previous timepoint had affected the subject's daily activity, whereas our participants scored their abdominal pain on each assessment moment using an 11-point NRS. These methodological differences in definitions of daily life stress and abdominal pain as well as the time span to which the assessments refer (*i.e.* "between the previous and current assessments" vs. "at the current assessment") might therefore be a reason for differences in results. Furthermore, slightly different timing of the assessments (*i.e.* 8 times a day for 14 consecutive days vs. 10 times a day for 7 consecutive days) combined with unavoidable differences in missing data, might lead to alterations in the time-lags. Future research might benefit from taking into account the exact length of time between assessments, herewith providing a

more reliable comparison between studies. Even though we did not find any significant associations in time-lagged analyses, this does not fully reject the potential presence of interactions as such, possibly only in a subgroup of IBS patients.

The disagreement between a self-reported relationship (using the VSI) between stress and abdominal pain and the association between these factors in our momentary data (using ESM) suggests that subjects do not reliably report this interplay between symptoms in retrospective questionnaires. This could be due to recall bias, in which our autobiographical memory is not able to correctly recall the exact interaction between experiences. The reporting of peak rather than average levels of abdominal pain in IBS populations has been linked to retrospective assessments in previous studies^{10,11}, which may also hold true for the stress-abdominal pain association. Being aware of the fact that patients may overreport an effect of stress on their abdominal symptoms is relevant for clinicians, since subjects showing a considerable stress-related pain response might benefit from different management strategies than others.³ In addition, patients might benefit from education in terms of providing insight into their individual, moment-to-moment flow of both psychological and abdominal symptoms, herewith creating awareness for their disease course.

The current study demonstrates that ESM is suitable to more deeply characterize individual symptom patterns using repeated, momentary assessments and taking into account psychosocial factors in IBS patients. The inclusion of a healthy control group puts extra confidence to the conclusion that the presented results are reflecting the IBS population and by considering all four IBS subtypes the results become generalizable to the heterogeneous IBS population as a whole. Data was collected using a previously validated IBS-specific ESM tool, that was developed according to FDA guidelines for the development of Patient-Reported Outcome Measures (PROM)¹⁵. All statistical models were corrected for autocorrelation, of which the importance in this specific context was previously underlined by Blanchard *et al.*⁶

An inevitable limitation of using the ESM is the occurrence of missing data. However, this is accounted for as much as possible by only including subjects that complete at least 1/3 of the total number of assessments, and by using advanced statistical modeling correcting for repeated measures and within-subject autocorrelation. The inclusion of subjects in a secondary/tertiary center might result in a specific study population not fully generalizable to the entire IBS population. However, subjects were previously also recruited via general practitioners' practices (*i.e.* for the Maastricht IBS Cohort), so the current study

population most likely represents a heterogeneous IBS population. Furthermore, selection bias could have arisen by the fact that ESM might be experienced as burdensome. Since the current report only includes data from one of the five intended inclusion centers, sample size is relatively small. However, due to multilevel statistics using repeated measures this should not affect the reliability of the identified association. Future analyses including more subjects, from different centers, could give additional insight into the heterogeneity in a stress-abdominal pain interaction between subjects.

In conclusion, using real-time symptom assessment, we demonstrated that stress levels are positively associated with concurrent abdominal pain scores, with high heterogeneity in the magnitude of this association between individual IBS patients, pointing towards different levels of stress-sensitivity regarding stress-abdominal pain interaction in IBS. Whereas IBS patients and healthy subjects showed comparable levels of stress, this association was stronger in IBS patients compared to healthy subjects, indicating a difference in response to stress rather than a difference in experienced stress per se. In addition, abdominal pain scores could not be predicted by preceding stress levels, and vice versa, suggesting an in-the-moment rather than a longitudinal association. This study underlines the importance of taking into account the individual flow of daily life when evaluating symptom patterns in IBS populations and supports the use of real-time measurement when interpreting potential influencers of abdominal symptoms.

References

1. Drossman DA. Functional Gastrointestinal Disorders: History, Pathophysiology, Clinical Features and Rome IV. *Gastroenterology*. 2016.
2. Van Oudenhove L, Crowell MD, Drossman DA, et al. Biopsychosocial Aspects of Functional Gastrointestinal Disorders. *Gastroenterology*. 2016.
3. Palsson OS, Drossman DA. Psychiatric and psychological dysfunction in irritable bowel syndrome and the role of psychological treatments. *Gastroenterol Clin North Am*. 2005;34(2):281-303.
4. Koloski NA, Jones M, Kalantar J, Weltman M, Zaguirre J, Talley NJ. The brain-gut pathway in functional gastrointestinal disorders is bidirectional: a 12-year prospective population-based study. *Gut*. 2012;61(9):1284-1290.
5. Jones MP, Tack J, Van Oudenhove L, et al. Mood and Anxiety Disorders Precede Development of Functional Gastrointestinal Disorders in Patients but Not in the Population. *Clin Gastroenterol Hepatol*. 2017;15(7):1014-1020.e1014.
6. Blanchard EB, Lackner JM, Jaccard J, et al. The role of stress in symptom exacerbation among IBS patients. *J Psychosom Res*. 2008;64(2):119-128.
7. Dancy CP, Taghavi M, Fox RJ. The relationship between daily stress and symptoms of irritable bowel: a time-series approach. *J Psychosom Res*. 1998;44(5):537-545.
8. Levy RL, Cain KC, Jarrett M, Heitkemper MM. The relationship between daily life stress and gastrointestinal symptoms in women with irritable bowel syndrome. *J Behav Med*. 1997;20(2):177-193.
9. Mearin F, Baro E, Roset M, Badia X, Zarate N, Perez I. Clinical patterns over time in irritable bowel syndrome: symptom instability and severity variability. *Am J Gastroenterol*. 2004;99(1):113-121.
10. 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.
11. 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.
12. Stone AA, Schwartz JE, Broderick JE, Shiffman SS. Variability of momentary pain predicts recall of weekly pain: a consequence of the peak (or salience) memory heuristic. *Pers Soc Psychol Bull*. 2005;31(10):1340-1346.
13. Havermans R, Nicolson NA, Devries MW. Daily hassles, uplifts, and time use in individuals with bipolar disorder in remission. *J Nerv Ment Dis*. 2007;195(9):745-751.
14. DeLongis A, Folkman S, Lazarus RS. The impact of daily stress on health and mood: psychological and social resources as mediators. *J Pers Soc Psychol*. 1988;54(3):486-495.
15. Vork L, Keszhelyi 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* 2017.
16. 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*. 2018:e13514.
17. Ludidi S, Mujagic Z, Jonkers D, et al. Markers for visceral hypersensitivity in patients with irritable bowel syndrome. *Neurogastroenterol Motil*. 2014;26(8):1104-1111.
18. Thijssen AY, Mujagic Z, Jonkers DM, et al. Alterations in serotonin metabolism in the irritable bowel syndrome. *Aliment Pharmacol Ther*. 2016;43(2):272-282.

19. Mujagic Z, Ludidi S, Keszthelyi D, et al. Small intestinal permeability is increased in diarrhoea predominant IBS, while alterations in gastroduodenal permeability in all IBS subtypes are largely attributable to confounders. *Aliment Pharmacol Ther.* 2014;40(3):288-297.
20. Palsson O, Whitehead W, van Tilburg MA, et al. Development and Validation of the Rome IV Diagnostic Questionnaire for Adults. *Gastroenterology.* 2016;150:1481-1491.
21. Mearin F, Lacy BE, Chang L, et al. Bowel Disorders. *Gastroenterology.* 2016.
22. Vork L, Mujagic Z, Drukker M, et al. The Experience Sampling Method-Evaluation of treatment effect of escitalopram in IBS with comorbid panic disorder. *Neurogastroenterol Motil.* 2018:e13515.
23. U.S. Department of Health and Human Services FaDA, Center for Drug Evaluation and Research (CDER). Guidance for Industry Irritable Bowel Syndrome - Clinical Evaluation of Drugs for Treatment. 2012.
24. (EMA) EMA. Guideline on the evaluation of medicinal products for the treatment of irritable bowel syndrome. 2014.
25. 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.
26. Spitzer RL, Kroenke K, Williams JB, Lowe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern CMed.* 2006;166(10):1092-1097.
27. Labus JS, Bolus R, Chang L, et al. The Visceral Sensitivity Index: development and validation of a gastrointestinal symptom-specific anxiety scale. *Aliment Pharmacol Ther.* 2004;20(1): 89-97.
28. Labus JS, Mayer EA, Chang L, Bolus R, Naliboff BD. The central role of gastrointestinal-specific anxiety in irritable bowel syndrome: further validation of the visceral sensitivity index. *Psychosom Med.* 2007;69(1):89-98.
29. 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.
30. Delespaul P. Assessing schizophrenia in daily life the experience sampling method.: UPM, Universitaire Pers Maastricht, Maastricht University; 1995.
31. Mayer EA, Craske M, Naliboff BD. Depression, anxiety, and the gastrointestinal system. *J Clin Psychiatry.* 2001;62 Suppl 8:28-36; discussion 37.
32. Whitehead WE, Crowell MD, Robinson JC, Heller BR, Schuster MM. Effects of stressful life events on bowel symptoms: subjects with irritable bowel syndrome compared with subjects without bowel dysfunction. *Gut.* 1992;33(6):825-830.
33. Craske MG, Wolitzky-Taylor KB, Labus J, et al. A cognitive-behavioral treatment for irritable bowel syndrome using interoceptive exposure to visceral sensations. *Behav Res Ther.* 2011;49(6-7):413-421.

Chapter 11

The experience sampling method - Evaluation of treatment effect of escitalopram in IBS with comorbid panic disorder

Lisa Vork, Zlatan Mujagic, Marjan Drukker, Daniel Keszthelyi, José M. Conchillo, Martine A.M. Hesselink, Jim van Os, Ad A.M. Masclee, Carsten Leue, Joanna W. Kruimel

Neurogastroenterology & Motility. 2019;31(1).

Abstract

INTRODUCTION Confirming treatment response in clinical trials for irritable bowel syndrome (IBS) is challenging, due to the lack of biomarkers and limitations of the currently available symptom assessment tools. The Experience Sampling Method (ESM) might overcome these limitations by collecting digital assessments randomly and repeatedly during daily life. This study evaluated differences in change in abdominal pain between real-time (*i.e.* ESM) and retrospective (*i.e.* Gastrointestinal Symptom Rating Scale (GSRS) and an end-of-day symptom diary) measurements, using data of an RCT on escitalopram versus placebo in patients with IBS and comorbid panic disorder.

METHODS Twenty-nine IBS patients with comorbid panic disorder were included in a 6-month RCT. The GSRS, diary, and ESM were completed at baseline (t=0) and after 3 (t=3) and 6 months (t=6). Linear mixed models were used.

RESULTS ESM analyses revealed a significant interaction between escitalopram and time, and ESM abdominal pain scores were 1.4 points lower in the escitalopram group compared to placebo at t=6 (on a 1-to-7 scale; $P=0.021$). When including the interaction with momentary anxiety, the reduction in abdominal pain scores in escitalopram versus placebo was even more pronounced for higher levels of anxiety. Average GSRS- and end-of-day-abdominal pain scores were not significantly different between escitalopram and placebo at t=3 and t=6.

CONCLUSIONS Real-time ESM has the potential to capture treatment response more sensitively compared to a retrospective end-of-day gastrointestinal symptom diary and the GSRS, by taking into account day-to-day symptom variability

Introduction

The main challenge in the management of Irritable Bowel Syndrome (IBS) is the selection of a suitable therapy for the individual patient. Overall, the efficacy of most available treatments is modest, the patient population is heterogeneous, and the pathophysiology of IBS is multifactorial and incompletely understood, limiting the development of new treatments. Furthermore, due to the lack of biomarkers and limitations of the currently available symptom assessment tools, confirming treatment response for new and existing therapies in clinical trials remains challenging.

Sensitive tools for symptom assessment are crucial in the evaluation of treatment effect in IBS populations.¹ Drug regulatory authorities, both in USA and Europe, recommend the use of end-of-day symptom scores for evaluation of treatment efficacy in IBS. However, it is well-known that such retrospective (*i.e.* end-of-day), self-reported measurements have important limitations, such as recall and ecological biases.^{2,3} We have previously demonstrated possible advantages of real-time symptom assessment, using the Experience Sampling Method (ESM) compared to end-of-day reporting.⁴ The ESM is an electronic assessment method, which collects data repeatedly at random moments over the day, during several consecutive days. The assessments focus on the in-the-moment physical and mental state of the subject, in the natural environment.^{3,5} Considerable differences between real-time and end-of-day abdominal pain scores were demonstrated in a cross-sectional study, raising the question of whether real-time symptom assessment may be superior to retrospective methods in the detection of treatment response in IBS patients.⁶ It is expected that ESM shows an increased sensitivity to detect within-subject change over time, as a result of, among other things, repeated data collection.⁷⁻⁹

IBS is defined as a disorder of gut-brain interaction¹⁰ with alterations in serotonin metabolism as a presumed key pathophysiological factor.¹¹⁻¹³ Antidepressants have been utilized in the treatment of IBS for many years. With regard to the efficacy of antidepressants in IBS, global symptom improvement and increased quality of life as well as a reduction in abdominal pain have been demonstrated for tricyclic antidepressants (TCAs), but not for selective serotonin reuptake inhibitors (SSRIs).^{14,15} Furthermore, overlap between IBS and psychiatric disorders is common in daily practice¹⁶⁻¹⁸ and IBS patients with psychological comorbidities may constitute the IBS subgroup most prone to respond to neuromodulating therapy.

In our tertiary referral centre, the prevalence of comorbid panic disorder in IBS patients is high (up to 50% of this population).¹⁹ According to current guidelines for management of anxiety disorders, the first-choice treatment for panic disorder are SSRIs.^{20,21} However, it is not known whether IBS patients with panic disorder benefit from SSRIs with regard to their IBS symptoms and whether anxiety might moderate treatment response. Therefore, the current randomised controlled trial was developed to examine the effect of the SSRI escitalopram on abdominal pain in IBS patients, diagnosed with comorbid panic disorder. However, due to slow patient inclusion rates the study has been prematurely ended. Since the originally intended sample size was not achieved, a firm conclusion on the efficacy of SSRIs in IBS with panic disorder cannot be drawn. Nevertheless, as both real-time and retrospective patient-reported symptom scores were collected before, during, and after the treatment period, the data are suitable to compare treatment response measurements between ESM and retrospective symptom reports.

The primary objective of the current study was to evaluate differences in change in abdominal pain by use of real-time (*i.e.* ESM) and retrospective (*i.e.* Gastrointestinal Symptom Rating Scale (GSRS) and an end-of-day symptom diary) symptom assessment methods, in the escitalopram treatment group versus the placebo group after respectively 3 and 6 months of treatment, in patients with IBS and comorbid panic disorder. Secondary, we aimed to evaluate the modulating effects of psychological comorbidity, *i.e.* anxiety, on the effect of escitalopram on abdominal pain, by using ESM.

Materials and methods

Study design

The current study uses data of a single-centre, double-blind, parallel-group, randomized controlled trial of escitalopram versus placebo for patients with IBS and comorbid panic disorder. The study protocol was approved in 2011 by the medical-ethics committee of Maastricht University Medical Centre (MUMC), Maastricht, the Netherlands, and was registered in the US National Library of Medicine (<http://www.clinicaltrials.gov>, NCT01551225). The study was executed according to the principles of the revised Declaration of Helsinki (59th WMA General Assembly, Seoul, Republic of Korea, October 2008). All subjects gave their written informed consent prior to participation.

Study participants and medication

Study participants were recruited between February 2012 and June 2016, via the MUMC's outpatient clinics of Gastroenterology and the Med-Psych-Centre. The latter is a secondary/tertiary outpatient clinic for the combined consultation of a gastroenterologist and a hospital-psychiatrist, delivering integrated care. Subjects aged 18-70 years, clinically diagnosed with IBS and comorbid panic disorder, were considered eligible. IBS was diagnosed by a gastroenterologist (JK) using the Rome III criteria^{22,23} and, if indicated, GI endoscopy, abdominal imaging, and/or blood, breath or faecal analyses were performed to exclude organic GI diseases. The presence of a comorbid panic disorder was diagnosed by a psychiatrist (CL) using the DSM-IV-TR criteria and the Mini-International-Neuropsychiatric-Interview.²⁴ Details on exclusion criteria and subject monitoring during the trial are described in Supporting Information 1.

Included subjects were randomly assigned to one of the two treatment groups (*i.e.* escitalopram or placebo) in a 1:1 ratio and started with an initial dose of study medication of 5 mg once a day. This was increased to 10 mg once daily after the first week and could be gradually increased to a maximum of 20 mg once daily, depending on the individual response to treatment in terms of self-reported symptom improvement. Further information on randomization procedures and study medication can be found in Supporting Information 1.

Study procedures

Three standardized measurement moments, *i.e.* baseline ($t=0$), after three months ($t=3$), and after 6 months ($t=6$) were implemented in the six-month study period. At baseline, questionnaires were completed on demographics, lifestyle, and medication use. At baseline and the two follow-up moments, validated retrospective symptom questionnaires, a 14-day end-of-day symptom diary, and a 7-day Experience Sampling Method (ESM) period were completed.

Symptom questionnaires

Validated questionnaires that were completed at baseline and after three and six months included the Gastrointestinal Symptom Rating Scale (GSRS) and the Rand 36-item Short-Form Health Survey (SF-36). The GSRS is a 15-item self-reported scale, evaluating GI symptom severity based on a single measurement, with a recall period of one week. All individual items are scored on a 7-point Likert scale (1=*not at all* to 7=*extremely*) and are subsequently clustered into

five domains (abdominal pain, reflux, indigestion, diarrhoea, and constipation); higher scores indicate more severe symptoms.^{25,26} The SF-36 was used to evaluate general quality of life, resulting in a score for physical (PCS) as well as mental (MCS) quality of life (*i.e.* higher scores indicate better quality of life).^{27,28}

Symptom diary

A 14-day end-of-day symptom diary was used to evaluate GI symptom severity on a daily basis over 14 consecutive days. GI symptoms (*i.e.* abdominal pain, abdominal discomfort, abdominal bloating, flatulence, constipation, and diarrhoea) were scored using a 5-point Likert scale (1=*not at all* to 5=*extremely*) at the end of each day, so considering a recall period of one day.

Experience Sampling Method (ESM)

On seven consecutive days during the 14-day diary period, the electronic ESM was used for momentary assessment of GI symptoms and the in-the-moment affective state. ESM assessments were performed on a palmtop computer; the device sent out an auditory signal ('beep') at 10 random moments between 7.30 a.m. and 10.30 p.m., and subjects completed an electronic questionnaire following each beep. The ESM-questionnaire was repeated in the same order at all measurement moments and all questions were scored by using a 7-point Likert scale (1=*not at all* to 7=*extremely*). In contrast to the retrospective symptom questionnaires and end-of-day diary, there is no recall but ESM measures real-time experiences.

Statistical analyses

All analyses were performed using Stata, version 13.1 (Stata Corporation, College Station, Texas, 2009). Baseline demographical characteristics and symptom severity scores are presented as medians (Q1; Q3) for continuous data and as proportions for categorical data. The primary outcome measures were abdominal pain severity, as measured by using the ESM, the end-of-day symptom diary, and the abdominal pain-domain of the GSRS (GSRS-AP). ESM and end-of-day diary data had a three-level structure (level 1: repeated measurements at each ESM assessment (ESM) and at the end of each day (diary), respectively; level 2: time (t=0; t=3; t=6); level 3: subject); GSRS data had a two-level structure (level 1: time (t=0; t=3; t=6); level 2: subject). The primary ESM analysis included all ESM data of the subjects that completed three valid ESM periods (*i.e.* at least 23 out of 70 (1/3) assessments per period). This rule of

thumb for validity of ESM periods has been common standard in ESM protocols, in order to take into account the fact that including data of subjects that complete too few assessments, may distort overall study findings.^{29,30} An ESM sensitivity analysis included all subjects that collected ESM-data at all three measurement moments. All analyses with end-of-day diary or GSRs as the outcome measure, included all subjects that collected data at the three measurement moments (*i.e.* excluding two drop-outs). A *P*-value of 0.05 was considered statistically significant.

Change in abdominal pain - To analyse whether the change in abdominal pain over time was different in escitalopram compared to placebo, mixed linear models with unstructured covariance, random slopes for treatment and time, and including the treatment*time interaction were used. Separate models were used for the three different assessment methods, with abdominal pain scores (*i.e.* ESM, end-of-day diary, and GSRs) as the dependent variable. All main effects were included as a random slope, however, in some models the random slope for time and/or treatment had to be removed because the model did not reach convergence. All non-significant interaction terms ($\alpha=0.05$) were removed from the models top-down. Regression coefficients in each strata were obtained by using the Stata `lincom` post-estimation command in the final models.

In order to evaluate a possible modulating effect of anxiety, anxiety scores were additionally included in the model with ESM abdominal pain scores as the dependent variable. Both two-way interactions anxiety*escitalopram, anxiety*time, and escitalopram*time, and the three-way interaction anxiety*escitalopram*time as well as the main effect of anxiety were included. Momentary anxiety scores assessed using ESM (on a 1-to-7 point scale) were used. Using the Stata post-estimation command `lincom`, regression coefficients are shown for scores 1 (no anxiety), 2 (low anxiety), 3 (mild anxiety), and 4 (moderate anxiety), as subjects did not frequently score anxiety higher than 4.

Sample size

We originally intended to include 80 subjects (*i.e.* 40 in each group) in this RCT, aiming to evaluate the effect of escitalopram versus placebo on abdominal pain in IBS patients with comorbid panic disorder. However, patient inclusion rate remained lower than expected and the study was prematurely ended with group sizes of 15 and 14 for escitalopram and placebo respectively. Therefore, we cannot draw a firm conclusion on the efficacy of escitalopram, but we here

exploratory describe differences in the measurement of treatment response between real-time and retrospective assessment methods.

Results

Study population

During the enrolment period, 215 IBS patients with comorbid panic disorder were screened for eligibility, of whom twenty-nine (13.5%) were included in the study (see Figure 11.1). Twenty-one (72.4%) were female, and the median age was 37 years [IQR: 24-49]. IBS subtype distribution was as follows: 12 diarrhoea-predominant (IBS-D), 5 constipation-predominant (IBS-C) and 12 mixed type (IBS-M).

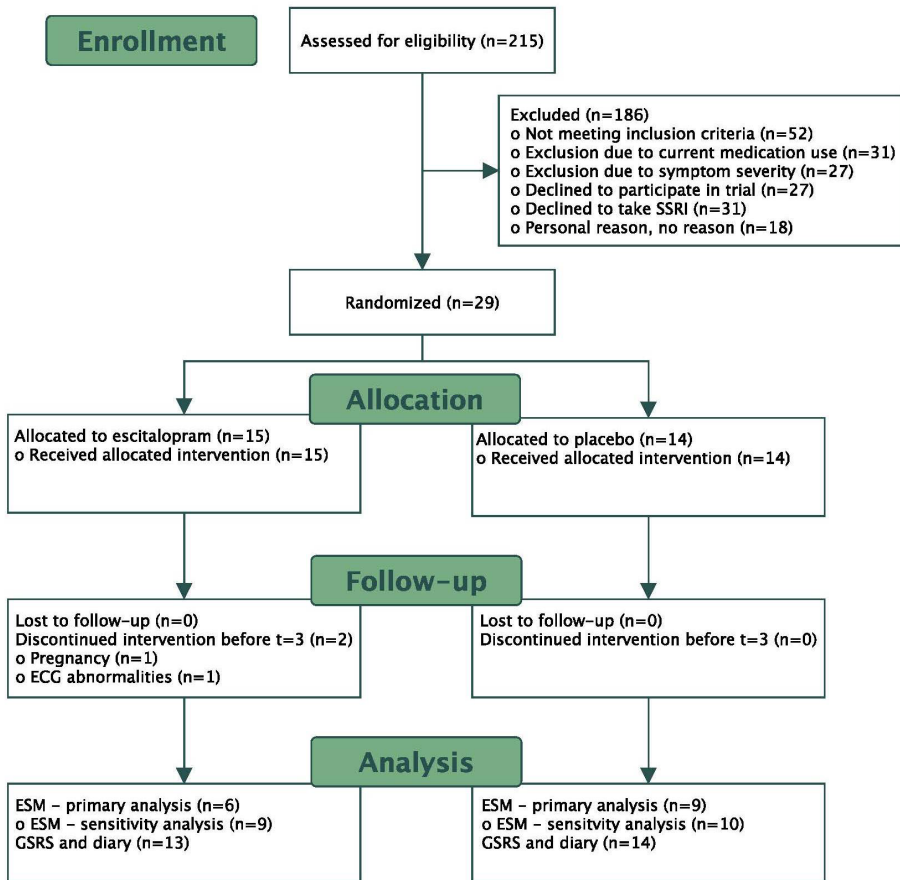


Figure 11.1 Flow-chart of subject enrolment.

Fifteen subjects were randomized to the escitalopram group (12 female (80.0%); median age 39 years) and 14 to the placebo group (9 female (64.3%); median age 25.5 years). Abdominal symptom scores at baseline, according to the GSRS and end-of-day symptom diary, were similar in both groups, just as the SF-36 scores. Baseline characteristics are shown in Table 11.1. Two subjects (6.9%; both from the escitalopram group) dropped out before t=3 months, for reasons of unplanned pregnancy (n=1) and development of ECG abnormalities on routine ECG before dose elevation of escitalopram (n=1). Since the primary aim of the current study was not to evaluate the effect of escitalopram versus placebo, but to compare outcomes of different methods for treatment response measurement, these two subjects with incomplete data were not included in the analyses.

Table 11.1 Demographical and clinical patient characteristics for escitalopram and placebo treatment groups, at baseline (t=0).

	Baseline, t=0 months	
	Escitalopram (n=15)	Placebo (n=14)
Female sex , n (%)	12 (80.0)	9 (64.3)
Age , median [Q1; Q3]	39.00 [25.00; 55.00]	25.50 [23.00; 46.00]
BMI , median [Q1; Q3]	26.93 [25.99; 29.40]	25.40 [23.53; 28.04]
IBS subtype , n (%)		
IBS-D	6 (40.0)	6 (42.9)
IBS-C	5 (33.3)	0 (0.0)
IBS-M	4 (26.7)	8 (57.1)
IBS-U	0 (0.0)	0 (0.0)
GSRS , median [Q1; Q3]		
Abdominal pain	3.33 [2.33; 4.33]	3.67 [2.67; 4.00]
Reflux syndrome	2.50 [1.00; 3.00]	1.50 [1.00; 4.00]
Diarrhea syndrome	3.67 [3.00; 5.67]	2.67 [2.33; 4.00]
Constipation syndrome	4.00 [2.00; 4.00]	3.33 [2.33; 4.33]
Indigestion syndrome	3.75 [2.75; 4.50]	3.38 [2.25; 4.25]
End-of-day diary , median [Q1; Q3]		
Abdominal pain	2.71 [1.93; 2.86]	2.39 [2.21; 3.50]
Discomfort	2.71 [2.31; 3.36]	2.75 [2.43; 3.57]
Bloating	2.57 [1.64; 3.08]	2.71 [1.78; 3.29]
Diarrhea	1.29 [1.00; 1.69]	1.15 [1.00; 1.50]
Constipation	1.15 [1.00; 1.69]	1.46 [1.14; 2.23]
SF-36 , median [Q1; Q3]		
Mental Composite Score (MCS)	46.84 [41.10; 53.85]	43.30 [30.75; 51.41]
Physical Composite Score (PCS)	41.39 [32.66; 47.84]	40.86 [34.27; 46.17]

GSRS scores are presented for five subscales. End-of-day diary scores are presented as average scores over 14 days, based on the sum of individual scores per day. SF-36 scores are presented for two subscales.

Of the total study population, fifteen subjects (51.7%; 6 escitalopram, 9 placebo) completed a valid number of ESM assessments (*i.e.* at least 23 out of 70 assessments per period) at all three time points. The data of these subjects were used for the primary ESM analysis. Baseline characteristics of this subgroup were comparable to those of the total study population. Information on side effects during the trial can be found in Supporting Information 1.

Evaluation of treatment response using ESM

Twenty-one subjects (72.4%) agreed to complete the ESM questionnaires, of whom 15 (71.4%) collected a valid number of assessments (*i.e.* at least 23 out of 70 assessments per period) at all three time points. In the primary ESM analysis including ESM data of these 15 subjects, a significant two-way interaction escitalopram*time was found (χ^2 : 101.22, *df*=2, $P<0.001$), revealing that abdominal pain decreased more over time in the escitalopram group than the placebo group. Furthermore, the lincom post-estimation command showed that abdominal pain scores were significantly lower in the escitalopram group compared to placebo at *t*=6 ($B=-1.425$, $P=0.021$; on a 1-to-7 scale). In the ESM sensitivity analysis, including data from all 19 subjects that completed ESM, the two-way interaction was significant as well (χ^2 : 31.08, *df*=2, $P<0.001$), but abdominal pain scores were not significantly different between escitalopram and placebo at *t*=3 ($B=-0.280$, $P=0.631$) and *t*=6 months ($B=-0.783$, $P=0.181$) (Table 11.2). Mean ESM abdominal pain scores for escitalopram and placebo at all three time points are shown in Figure 11.2. The differences in change in abdominal pain scores between escitalopram and placebo at *t*=3 and *t*=6 months are shown in Table 11.2.

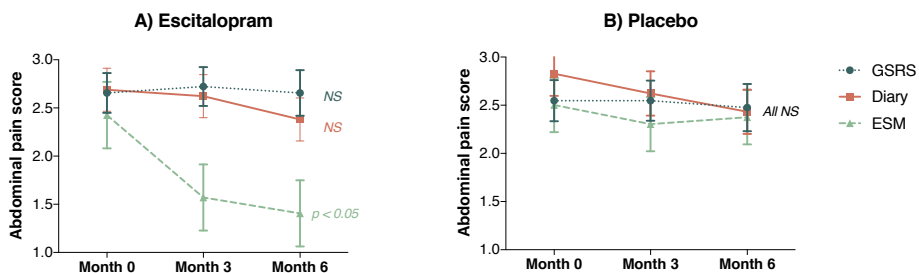


Figure 11.2 Mean abdominal pain scores at *t*=0 months (baseline), *t*=3 months, and *t*=6 months for the escitalopram group (A) and placebo group (B), separately for GRS, end-of-day diary, and ESM (on a 1-to-5 point scale). Change in abdominal pain scores over time tested using linear mixed models. NS: No significant change in abdominal pain scores after 3 and 6 months. $P<0.05$: Significant change in abdominal pain scores after 6 months.

Interaction with anxiety - Additionally, anxiety levels were also included in this model. The three-way interaction anxiety*escitalopram*time was revealed not statistically significant (χ^2 : 4.45, df=2, $P=0.108$). However, since we *a priori* expected a limited power to show this interaction, we performed the lincom post-estimation command to obtain the regression coefficients for the different strata despite the p-value above alpha. At t=6 months, average abdominal pain score was 1.414 point lower ($P=0.040$) for escitalopram compared to placebo when there was no anxiety (1) and this difference increased to 1.601 ($P=0.024$) for low anxiety (2), 1.788 ($P=0.026$) for mild anxiety (3) and 1.975 ($P=0.036$) for moderate anxiety (4). Results are shown in Table 11.3.

Evaluation of treatment response using GSRS and end-of-day diary

In models assessing GSRS-AP and the end-of-day abdominal pain score, we did not find a significant two-way interaction between escitalopram and time, showing that the association between escitalopram and abdominal pain was similar at all time points using retrospective assessments (Table 11.2). Mean abdominal pain scores for escitalopram and placebo at all three time points are shown in Figure 11.2, for both methods.

Table 11.2 Difference in abdominal pain scores between escitalopram and placebo for outcome measures GSRs, end-of-day diary, and ESM, tested using linear mixed models. Results shown for t=0 months, t=3 months and t=6 months.

Abdominal pain	Month 0 (baseline)			Month 3			Month 6			
	Interaction term escitalopram*time χ^2 (df), p-value	Estimate [95% CI]	SE	P-value	Estimate [95% CI]	SE	P-value	Estimate [95% CI]	SE	P-value
GSRs ^{a,d}	0.21 (2), p=0.900	0.183 [-0.574; 0.940]	0.386	0.635						
End-of-day diary ^{a,d}	1.14 (2), p=0.565	-0.199 [-0.816; 0.418]	0.315	0.528						
ESM	Primary ^b	101.22 (2), p<0.001	-0.108 [-1.315; 1.098]	0.616	0.860	0.615	0.084	-1.425 [-2.632; -0.217]	0.616	0.021
	Sensitivity ^c	31.08 (2), p<0.001	-0.044 [-1.184; 1.097]	0.582	0.940	0.583	0.631	-0.783 [-1.930; 0.364]	0.585	0.181

Estimate indicates the difference between abdominal pain scores (i.e., GSRs, end-of-day diary, and ESM) in the escitalopram group versus the placebo group. A negative estimate indicates a lower abdominal pain score in the escitalopram group compared to placebo. SE: standard error. ^a n=27; ^b n=15; ^c n=19. ^d Estimate, SE, and P-value at month 0 reflect results for all time points.

Table 11.3 Difference in ESM abdominal pain scores between escitalopram and placebo for ESM-anxiety levels of 1 (no anxiety), 2 (low anxiety), 3 (mild anxiety), and 4 (moderate anxiety) (on a 1-to-7 Likert scale) respectively, tested using linear mixed models. Results shown for t=0 months, t=3 months and t=6 months.

Abdominal pain	Month 0 (baseline)			Month 3			Month 6		
	Estimate [95% CI]	SE	P-value	Estimate [95% CI]	SE	P-value	Estimate [95% CI]	SE	P-value
No anxiety (1)	-0.117 [-1.467; 1.233]	0.689	0.865	-1.179 [-2.528; 0.170]	0.688	0.087	-1.414 [-2.762; -0.066]	0.688	0.040
Low anxiety (2)	0.085 [-1.268; 1.437]	0.690	0.902	-0.909 [-2.265; 0.447]	0.692	0.189	-1.601 [-2.994; -0.208]	0.711	0.024
Mild anxiety (3)	0.286 [-1.130; 1.702]	0.722	0.692	-0.640 [-2.070; 0.791]	0.730	0.381	-1.788 [-3.359; -0.216]	0.802	0.026
Moderate anxiety (4)	0.488 [-1.044; 2.019]	0.781	0.533	-0.370 [-1.934; 1.194]	0.798	0.643	-1.975 [-3.820; -0.129]	0.942	0.036

Estimate indicates the difference between ESM abdominal pain scores in the escitalopram group versus the placebo group. A negative estimate indicates a lower abdominal pain score in the escitalopram group compared to placebo. SE: standard error. The model included a three-way interaction term for anxiety*escitalopram*time; χ^2 : 4.45 (2), P=0.108.

Discussion

In the present exploratory study, we evaluated the possible advantages of real-time symptom assessment using ESM, in quantifying treatment response, when compared to retrospective symptom measurements using an end-of-day diary and the GSRS. Hereto, data from a prematurely ended randomised controlled trial, developed to evaluate the effect of escitalopram vs. placebo on abdominal pain in IBS patients with comorbid panic disorder, were used. Hence, no beneficial effect of escitalopram over placebo in reducing abdominal pain in subjects with IBS and comorbid panic disorder was found, when evaluated by conventional, retrospective symptom assessment methods (*i.e.* GSRS and end-of-day abdominal pain scores). However, when momentarily assessed by the real-time electronic ESM, a significant decrease in abdominal pain was observed in escitalopram as compared to placebo, after six months of treatment. Due to the exploratory nature of the used method, we cannot draw a firm conclusion on the effect of SSRIs on abdominal pain in IBS patients based on the presented results, however, the current study underlines the difference in outcome between retrospective and repeated in-the-moment measurements. Furthermore, these results show that with the high number of repeated measures and the real-time assessment of symptoms, ESM may have the potential to detect within-subject changes over time with higher sensitivity, yielding higher analytical power in small numbers of study participants, when compared to conventional retrospective questionnaires.

Differences between retrospective and real-time assessments have previously been demonstrated in IBS populations. Reporting of peak abdominal pain rather than average scores has been described, for weekly or daily retrospective measurement compared to momentary abdominal pain assessment.^{4,31} This phenomenon can be attributed to recall bias, suggesting that real-time assessments might be more accurate with regard to abdominal pain evaluation in a cross-sectional study design. The present study is the first to evaluate differences in outcome of treatment effect over time, between retrospective and momentary assessments of abdominal pain. Apart from eliminating recall bias, ESM has the advantage of capturing within-subject variability over time, as a result of repeated data collection. In the context of evaluating treatment effect, this is beneficial as it leads to an increased sensitivity to detect within-subject change over time.⁷⁻⁹ Regarding the ESM analyses in the current study, it should be noted that a significant decrease in abdominal pain was only found in the primary analysis which included the ESM data of subjects that completed

three valid ESM periods, *i.e.* at least 23 out of 70 assessments per period, underlining the importance of collecting sufficient repeated assessments when using ESM. Our findings confirm results from earlier trans-diagnostic ESM data obtained from patients with functional or unexplained somatic symptoms and psychiatric comorbidity treated at an academic hospital-psychiatry outpatient ward with regard to ESM's sensitivity to detect change.³²

Given the recognized efficacy of escitalopram in panic disorder treatment,^{33,34} we hypothesized an interaction between anxiety, escitalopram, and time in the model evaluating the change in abdominal pain. Even though no significant three-way interaction term could be found, a more pronounced decrease in abdominal pain in the escitalopram group could be demonstrated for subjects with higher levels of momentary anxiety as compared to subjects with lower levels of anxiety. These results demonstrate the potential of ESM to evaluate the effect of momentary factors possibly modulating symptom formation and/or treatment response and adds value to the field by integrating somatic and psychological assessment at the same time.

The current study is based on data of a prematurely ended RCT, which was set up to evaluate the effect of an SSRI on abdominal pain in IBS patients with comorbid panic disorder. The intended number of subjects to be enrolled was 80, while after over four years only 29 subjects were included. This was 13.5% of all eligible patients approached for participation within this period. Given the availability of the drug (SSRI) in regular clinical care and the anticipation of a possible placebo treatment for 6 months, many patients decided not to participate in the trial. This illustrates the difficulty of performing a high-quality placebo-controlled RCT for the evaluation of drug efficacy in IBS patients, or any other chronic condition, especially in cases of comorbidities. Furthermore, well conducted RCTs are costly and time-consuming, and may put participants at risk, due to the unknown nature of treatment effects. Therefore, methods that may reduce the number of required study participants in clinical trials are highly needed. ESM provides the opportunity to identify significant treatment effects in a lower number of study participants, compared to the currently available retrospective questionnaires, by gathering a high number of repeated measures and more reliable data that are not affected by recall and ecological biases.

A strength of the current study is that symptoms were concurrently reported using ESM, an end-of-day diary, and the GRSR during an RCT, offering the opportunity to directly compare outcome of treatment effect between real-time and retrospective measurements. Furthermore, the inclusion of a well-characterized tertiary IBS population with pronounced comorbid anxiety, in

which both IBS and anxiety symptoms were repeatedly measured using ESM, enabled us to evaluate the potential of ESM to take into account momentary factors modulating symptom formation and treatment effect.

A limitation of this study is that not all included subjects agreed to complete ESM, which decreased the sample size in the ESM analyses and might have induced selection bias. As discussed previously, the use of ESM might overcome this limitation concerning sample size by taking into account multiple repeated measurements in a single patient and thereby increasing the power to detect change over time. With regard to possible selection bias, we found no differences in demographical characteristics and gastrointestinal and psychological symptom severity between subjects that completed ESM and subjects that did not. A drawback of ESM can be the fact that the method is time-consuming, which might lead to a lack of compliance. On the other hand, due to the repeating character, it is expected that ESM results are not easily influenced by missing data. Furthermore, we took into account the risk of low compliance by excluding subjects that completed too few ESM assessments from the primary analyses. With regard to the study participants, IBS subtype distribution is different between the treatment groups in that no subjects with the constipation predominant subtype were present in the placebo group. However, as this study focuses on abdominal pain as the main outcome measure and subtype distribution is the same for all measurement methods, our primary results are most likely not affected by this difference in subtype distribution. Nevertheless, given the exploratory nature of this study and the fact that ESM has not yet been validated as an evaluation tool in the context of clinical trials evaluating abdominal pain, our results need replication in order to draw a firmer conclusion on the advantages of ESM over retrospective methods in quantifying treatment response in IBS patients.

In conclusion, the current study demonstrates a potential benefit of repeated in-the-moment measurements as compared to retrospective assessments, in the evaluation of treatment effect in IBS patients. Given the exploratory nature of the study, no firm conclusion can be drawn with regard to the effect of SSRIs on abdominal pain in IBS patients with comorbid panic disorder. Nevertheless, by taking into account day-to-day symptom variability as well as momentary factors that might moderate treatment effect, such as psychological symptoms, ESM has shown the potential to capture treatment response more sensitively compared to a retrospective end-of-day GI-symptom diary and the GSRS.

References

1. Mujagic Z, Keszthelyi D, Aziz Q, et al. Systematic review: instruments to assess abdominal pain in irritable bowel syndrome. *Aliment Pharmacol Ther.* 2015;42(9):1064-1081.
2. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol.* 2008;4:1-32.
3. 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.
4. 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.
5. 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.* 2017.
6. 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.
7. Verhagen SJ, Hasmi L, Drukker M, van Os J, Delespaul PA. Use of the experience sampling method in the context of clinical trials. *Evid Based Ment Health.* 2016;19(3):86-89.
8. Bolger N, Laurenceau JP. *Intensive longitudinal methods: An introduction to diary and experience sampling research.* New York: NY: Guilford Press; 2013.
9. Bolger N, Davis A, Rafaeli E. *Diary methods: capturing life as it is lived.* *Annu Rev Psychol.* 2003;54:579-616.
10. Drossman DA, Tack J, Ford AC, Szigethy E, Tornblom H, Van Oudenhove L. *Neuromodulators for Functional Gastrointestinal Disorders (Disorders of Gut-Brain Interaction): A Rome Foundation Working Team Report.* *Gastroenterology.* 2018;154(4):1140-1171.e1141.
11. Lesurtel M, Soll C, Graf R, Clavien PA. Role of serotonin in the hepato-gastrointestinal tract: an old molecule for new perspectives. *Cell Mol Life Sci.* 2008;65(6):940-952.
12. Cremon C, Carini G, Wang B, et al. Intestinal serotonin release, sensory neuron activation, and abdominal pain in irritable bowel syndrome. *Am J Gastroenterol.* 2011;106(7):1290-1298.
13. Thijssen AY, Mujagic Z, Jonkers DM, et al. Alterations in serotonin metabolism in the irritable bowel syndrome. *Aliment Pharmacol Ther.* 2016;43(2):272-282.
14. Ford AC, Quigley EM, Lacy BE, et al. Effect of antidepressants and psychological therapies, including hypnotherapy, in irritable bowel syndrome: systematic review and meta-analysis. *Am J Gastroenterol.* 2014;109(9):1350-1365.
15. Ruepert L, Quartero AO, de Wit NJ, van der Heijden GJ, Rubin G, Muris JW. Bulking agents, antispasmodics and antidepressants for the treatment of irritable bowel syndrome. *The Cochrane database of systematic reviews.* 2011(8):Cd003460.
16. Drossman DA, Creed FH, Olden KW, Svedlund J, Toner BB, Whitehead WE. Psychosocial aspects of the functional gastrointestinal disorders. *Gut.* 1999;45 Suppl 2:li25-30.
17. Palsson OS, Drossman DA. Psychiatric and psychological dysfunction in irritable bowel syndrome and the role of psychological treatments. *Gastroenterol Clin North Am.* 2005;34(2):281-303.
18. Ballou S, Keefer L. The impact of irritable bowel syndrome on daily functioning: Characterizing and understanding daily consequences of IBS. *Neurogastroenterol Motil.* 2017;29(4).

19. Kruiemel J, Leue C, Winkens B, et al. Integrated medical-psychiatric outpatient care in functional gastrointestinal disorders improves outcome: a pilot study. *Eur J Gastroenterol Hepatol.* 2015;27(6):721-727.
20. Balkom van ALJM, Vliet van IM, Emmelkamp PMG, et al. Multidisciplinaire Richtlijn Angststoornissen. 2013; Versie 1.0 (3e revisie):<https://assets-sites.trimbos.nl/docs/cc5b0f4e-9398-452d-8792-1453a38bf0a8.pdf>, 2013.
21. Association AP. Practice guideline for the treatment of patients with panic disorder. Second edition ed: American Psychiatric Association; 2010.
22. Longstreth GF, Thompson WG, Chey WD, Houghton LA, Mearin F, Spiller RC. Functional bowel disorders. *Gastroenterology.* 2006;130(5):1480-1491.
23. Drossman DA. Rome III: the new criteria. *Chin J Dig Dis.* 2006;7(4):181-185.
24. Diagnostic and statistical manual of mental disorders (DSM)-IV-TR. 4th ed. Washington, DC: American Psychiatric Association; 2000.
25. Svedlund J, Sjodin I, Dotevall G. GSRS--a clinical rating scale for gastrointestinal symptoms in patients with irritable bowel syndrome and peptic ulcer disease. *Dig Dis Sci.* 1988;33(2): 129-134.
26. Revicki DA, Wood M, Wiklund I, Crawley J. Reliability and validity of the Gastrointestinal Symptom Rating Scale in patients with gastroesophageal reflux disease. *Qual Life Res.* 1998;7(1):75-83.
27. Farivar SS, Cunningham WE, Hays RD. Correlated physical and mental health summary scores for the SF-36 and SF-12 Health Survey, V.I. *Health Qual Life Outcomes.* 2007;5:54.
28. McHorney CA, Ware JE, Jr., Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care.* 1993;31(3):247-263.
29. Palmier-Claus JE, Myin-Germeyns I, Barkus E, et al. Experience sampling research in individuals with mental illness: reflections and guidance. *Acta Psychiatr Scand.* 2011;123(1):12-20.
30. Delespaul P. Assessing schizophrenia in daily life the experience sampling method.: UPM, Universitaire Pers Maastricht, Maastricht University; 1995.
31. 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.
32. 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.
33. Bakker A, van Balkom AJ, Spinhoven P. SSRIs vs. TCAs in the treatment of panic disorder: a meta-analysis. *Acta Psychiatr Scand.* 2002;106(3):163-167.
34. Stahl SM, Gergel I, Li D. Escitalopram in the treatment of panic disorder: a randomized, double-blind, placebo-controlled trial. *J Clin Psychiatry.* 2003;64(11):1322-1327.

Supporting Information 11.1

Materials and methods

Exclusion criteria

Exclusion criteria were a history of abdominal surgery (except for uncomplicated appendectomy, cholecystectomy, or hysterectomy) and current use of (and inability to stop) medication that could influence GI motility, GI perception, or serotonin metabolism. Further exclusion criteria were pregnancy/lactation, current medication use specific for IBS or anxiety, administration of investigational drugs 180 days prior to study participation, excessive alcohol consumption (>20 alcoholic consumptions per week) or drug abuse, known QT-interval prolongation, and other cardiac disease or use of medication that prolongs QT-interval.

Randomization procedures and study medication

Included subjects were randomly assigned to one of the two treatment groups (*i.e.* escitalopram or placebo) in a 1:1 ratio, with stratification by gender, using a web-based randomization system (ALEA Software for randomization in clinical trials). Randomization was performed by an independent institution, the Clinical Trial Centre Maastricht (CTCM), and both the study participants and the researchers were blinded for treatment allocation throughout the study.

All study medication was supplied by Lundbeck B.V. Nederland, in identical blister packs, and tablets were similar with regard to appearance, in order to assure blinding conditions. Escitalopram and placebo tablets consisted of a core (microcrystalline cellulose (E460), croscarmellose sodium (E468), colloidal silica and magnesium stearate (E470b)) and a shell (hypromellose (E464), titanium dioxide (E171) and macrogol 400), with or without escitalopram respectively. Subjects were instructed to return all blister packs in order to monitor compliance.

In case of any side effects during the first two weeks of treatment or in the first two weeks after dosage augmentation, Alprazolam 0.25 mg once or twice a day was provided.¹

Subject monitoring

At baseline (t=0), one week after treatment initiation and thereafter at week 4, 8, 12 (t=3), 18, and 24 (t=6), standardized visits to the gastroenterologist (JK)

and psychiatrist (CL) were scheduled, in order to closely monitor patient safety. During those visits, symptoms and possible side effects were evaluated and a general physical examination was executed. Furthermore, at t=0, t=3, t=6, and before each dose elevation, an electrocardiogram (ECG) was performed to check for QT-prolongation. Also, at t=0, t=3, and t=6, blood analysis was performed to check for any abnormalities, especially changes in liver function, as a result of escitalopram treatment. Based on these clinical investigations, study treatment was immediately discontinued if necessary.

Results

Side effects

Adverse events were reported by eighteen subjects. One serious adverse event, not related to the study medication, was an appendicitis treated with appendectomy. Further side effects were headache (33.3% escitalopram, 14.3% placebo), epigastric pain or heartburn (6.7% escitalopram, 21.4% placebo), dizziness, sweating and palpitations (13.3% escitalopram, 14.3% placebo), fatigue (6.7% escitalopram), hematoma (7.1% placebo), and experiencing nightmares (7.1% placebo). One subject in the escitalopram group presented with slightly abnormal serum liver enzymes, but those normalized within one month without any action undertaken. Furthermore, an increase in stool frequency was reported by one subject in the escitalopram group.

Sinus bradycardia occurred in two subjects (both escitalopram), which normalized in one of them after a dose reduction of the study medication. The other patient presented with a worsening of pre-existent (asymptomatic) sinus bradycardia, with ECGs during safety monitoring visits showing a decrease of the sinus rhythm from 56/min to 49/min after initiation of the treatment period. A CYP450 polymorphism was found in this subject, possibly explaining the reduction in heart rate due to a slow escitalopram metabolism. It was decided to discontinue the study medication and to define the subject as a dropout. After discontinuation, the sinus rhythm normalized to the pre-existent level of 55/min.

References

1. Balkom van ALJM, Vliet van IM, Emmelkamp PMG, et al. Multidisciplinaire Richtlijn Angststoornissen. 2013; Versie 1.0 (3e revisie):<https://assets-sites.trimbos.nl/docs/cc5b0f4e-9398-452d-8792-1453a38bf0a8.pdf>, 2013.

Chapter 12

General discussion

IBS diagnosis evolving over time

Irritable bowel syndrome (IBS) has for many years been labeled as a diagnosis *per exclusionem*, since it often mimics the symptoms of a number of gastrointestinal (GI) diseases with an organic origin or substrate.¹ However, throughout the past decades more knowledge about the phenotypical presentation of IBS has been gained and it has been claimed that IBS should be addressed as a symptom-based disorder.^{2,3} The condition has originally been defined as a “functional GI disorder”. Based on results of studies that focused on pathophysiological alterations in the GI tract, IBS has more recently been described as a “disorder of gut-brain interaction”.³ In this current concept, the disorder is assumed to result from a multifactorial pathophysiology with a central role for the gut-brain axis.^{4,5} This alternating conceptualization of IBS has led to an ongoing process of development and revision of diagnostic criteria for IBS, which inevitably results in variable prevalence estimates throughout the years.^{6,7}

For this purpose, we compared IBS prevalence between the previous (Rome III; 2006⁸) and updated (Rome IV; 2016^{3,5}) version of the diagnostic Rome criteria in a well-characterized Dutch IBS population (**Chapter 2**). It was found that approximately 61-87% of Rome III IBS patients is likely to also fulfill the more stringent Rome IV criteria and that the Rome IV IBS population is represented by younger females with higher overall GI symptom severity, as well as comorbid psychological symptoms, and lower quality of life compared to the Rome III IBS population. These findings were more recently confirmed by a comparable study in a Swedish Rome III IBS population⁹ and underline the importance of harmonizing IBS diagnostic criteria between studies worldwide. However, this discrepancy might question the validity and accuracy of such symptom-based diagnostic criteria, since a subgroup of patients previously diagnosed with IBS according to the Rome III criteria suddenly “lost” this diagnosis when Rome IV criteria were implemented. This issue has also been raised in an expert review, concluding that the latest revision of the Rome criteria likely results in the inclusion of more homogeneous groups of IBS patients in clinical trials.¹⁰ Indeed, for scientific purposes, adherence to the most recently published diagnostic criteria for study inclusion should become the standard, in order to be able to compare results between studies. However, as the use of the Rome criteria in clinical practice is still experienced as burdensome,¹⁰ clinicians might pragmatically manage patients with IBS-like symptoms whether or not they strictly fulfill diagnostic criteria and are still in

need for new methods to reliably diagnose IBS and measure symptom severity and symptom fluctuation over time. Meanwhile, the search for biological markers to diagnose IBS and distinguish between IBS and organic counterparts will be ongoing in order to improve the correct classification and management of these patients.^{11,12}

The heterogeneity of IBS: can we predict individual disease course?

As a result of the multifactorial pathophysiology, several mechanistic pathways may underlie IBS symptoms, leading to variability in the clinical presentation between IBS subjects.^{3,13} This heterogeneity has hampered overall treatment efficacy and the development of new management strategies throughout the years.¹⁴ In addition to this between-subject variation, the disorder is generally of a chronic relapsing-remitting nature, with high variability in symptom presentation and severity within subjects over time.^{15,16} This further impedes the success of available treatment strategies in the long-term treatment of IBS patients. The choice of treatment is generally based on the patient-specific key presenting symptoms, for example abdominal pain or altered stool habits. To aid the management of the latter, IBS populations are often subdivided into subtypes based on their predominant bowel habit: diarrhea (IBS-D), constipation (IBS-C), mixed diarrhea and constipation (IBS-M), or unspecified predominant bowel habit (IBS-U).³ However, the variability in presentation over time also holds true for these subtypes, with shifts between predominant bowel habit in up to 75% of IBS patients.¹⁷ Therefore, in order to determine the best treatment option for each individual patient, it would be helpful if we could better predict the natural disease course over time.

Several studies have looked into this by searching for clinical or biological predictors for IBS outcomes over time.¹⁸⁻²¹ In a prospective study including 161 Rome III IBS patients we demonstrated that, after a follow-up period of 5 years, 30.4% of patients did no longer fulfill the Rome III criteria (**Chapter 3**), which is in line with previous studies.^{18,22} We could not find any independent predictors of whether or not meeting Rome III criteria at the time of follow-up. Our results did indicate that comorbid psychological symptoms, rather than GI symptom severity, might be associated with long-term (*i.e.* after 5 years) generic quality of life (QoL) in the IBS population. In contrast, Clevers *et al.* demonstrated that IBS-specific QoL was significantly higher in IBS patients with a decrease in GI

symptom severity after a follow-up period of 1 year.²¹ Altogether, these findings highlight the complexity of IBS, and suggest that future research should focus not only on the GI-part of the disorder, but also look from a multidisciplinary perspective, by concurrently taking into account psychological and other somatic comorbidities. Furthermore, the importance of considering the typical episodic course of IBS when evaluating disease severity, instead of drawing conclusions based on a single timepoint, is herewith underlined.

Putting IBS research in a time-varying perspective

Chapter 4 provides an example of considering the episodic changes in IBS presentation, as described above, in clinical research. In this chapter we further evaluated a previously demonstrated correlation between stool consistency and the fecal microbiota composition,²³⁻²⁵ by taking into account the day-to-day variation in stool consistency per individual over seven consecutive days. We confirmed the association between stool consistency and the microbial composition, and even though we found moderate to high variability in stool consistency from day to day in both IBS and healthy subjects, we could not demonstrate a significant effect of time on the association between stool consistency and the microbiota. As this indicates that the microbial profile of a fecal sample at one timepoint is representative for an individual's microbiota during short periods of time (*i.e.* at least seven days), even in case of highly fluctuating stool consistency, this has important implications for clinical microbiota research. This approach supports the use of repeated measurements and multilevel statistics, correcting for intra-individual correlations, when studying IBS outcomes that are known to fluctuate over time.

Symptom assessment in IBS: towards daily life monitoring

In the light of the incompletely elucidated pathophysiology and the resulting lack of reliable biomarkers, accurate symptom registration appears essential as it currently is the only clinical read-out for IBS diagnosis, disease monitoring, and treatment evaluation. As discussed above, given the heterogeneity in symptom presentation between subjects as well as within subjects over time, an individual approach that considers the temporal fluctuations, might be an

important step towards unraveling the often inconceivable and difficult-to-manage symptom patterns.

In **chapter 5** and **11** we exploratively evaluated the experience sampling method (ESM) as a novel method for symptom assessment (**chapter 5**) and treatment evaluation (**chapter 11**) in small IBS populations. We demonstrated that real-time symptom scores (*i.e.* ten times a day) over seven consecutive days were significantly different from end-of-day scores, which are the current standard as outcome measure in IBS clinical trials. A tendency to report peak day-levels of abdominal pain at the end of the day was demonstrated, and hereby we confirmed the findings of a previous study by Weinland *et al.*²⁶ In addition, the research performed in **chapter 11** suggests that ESM has the potential to capture treatment response more sensitively compared to retrospective tools, by taking into account day-to-day symptom variability as well as momentary factors that might moderate treatment effect.

In order to further elaborate on these exploratory studies, a patient-reported outcome measure (PROM) specific for the use of ESM in IBS patients, was developed (**Chapter 6**) and the performance of this tool was evaluated in terms of psychometric properties in a Dutch Rome IV-IBS population (**Chapter 9**). In addition to supporting the validity and reliability of the developed ESM-tool for symptom assessment in IBS, this study highlights some important advantages of ESM as compared to currently used retrospective questionnaires. Firstly, we replicated the finding of peak-level reporting in end-of-day and end-of-week reports, suggesting that real-time assessment reflects GI symptoms more accurately than retrospective methods. Furthermore, we illustrated that the use of repeated measures by ESM offers the opportunity to look into individual symptom patterns very thoroughly, herewith visualizing short-term symptom fluctuations during the day.

An additional strength of momentary symptom registration is elucidated in **chapter 10**, in which a possible interplay between abdominal pain as the key-symptom of IBS and stress, a largely considered trigger of GI symptoms in IBS patients,²⁷ was evaluated. We demonstrated that concurrent levels of stress and abdominal pain are associated in IBS patients, but not in healthy subjects. High heterogeneity in the magnitude of this association was found, pointing towards different levels of stress-sensitivity within the IBS population. The fact that a subjectively reported relationship between stress and GI symptoms, represented by the Visceral Sensitivity Index, did not agree with the association as found within the ESM data, suggests overestimation of the role of stress in retrospective self-reported questionnaires. This highlights the potency of ESM

not only to register individual moment-to-moment symptom patterns, but also to evaluate symptom-symptom or symptom- environment interactions during daily life.²⁸

A multidisciplinary approach of IBS

IBS is not only a multifaceted disorder in terms of the multifactorial pathophysiology and variable presentation of GI symptoms, it is also associated with comorbid multisystem symptoms or diagnoses.^{29,30} This was already mentioned in **chapter 3**, where an important interference of psychological comorbidities with IBS presentation and implications for generic quality of life were observed. However, the occurrence and severity of these comorbidities are again subject to significant heterogeneity between IBS subjects. The simultaneous measurement of those symptoms in a single patient may give a better insight into the nature of complex overlapping symptom systems in IBS in general and may have important implications for individual patient treatment.

In **chapter 7** and **8** we describe the development of ESM-based PROMs for symptom assessment in functional dyspepsia and overactive bladder syndrome, respectively; two other functional disorders in gastroenterology and urology. Both have shown to overlap with IBS, in subgroups of patients. Future studies will have to focus on the psychometric evaluation of these PROMs in the target populations and thereafter, the three tools together (or combinations thereof) could be used to evaluate overlap in symptoms and multisystem symptom-symptom interactions during daily life in individual patients. Finally, this could further aid the multidisciplinary approach of IBS and other/comorbid functional disorders.

Implications for clinical care and future research

It is often stated that the first step in IBS treatment is education and reassurance about the nature and disease course of the disorder.³¹ ESM could herein be of benefit by 1) giving patients and their healthcare providers direct insight into the presenting symptoms during day-to-day flow; 2) conceptualizing the interplay between GI symptoms and psychological experiences, contextual factors, and otherwise presumed influencers during daily life; and 3) elucidating links between IBS and comorbid multisystem symptoms. This might lead to a

better understanding of symptom occurrence and could aid in determining adequate and tailored personal life style interventions or the decision-making process regarding pharmacological treatment options targeting the predominant symptom and/or triggering concomitant factors. Ideally, ESM could also become a helpful instrument in the evaluation of treatment efficacy in clinical practice as well as in clinical trials, but hereto future research should focus first on the responsiveness of the developed PROM(s).

References

1. Spiegel BM, Farid M, Esrailian E, Talley J, Chang L. Is irritable bowel syndrome a diagnosis of exclusion?: a survey of primary care providers, gastroenterologists, and IBS experts. *Am J Gastroenterol.* 2010;105(4):848-858.
2. Spiller R, Aziz Q, Creed F, et al. Guidelines on the irritable bowel syndrome: mechanisms and practical management. *Gut.* 2007;56(12):1770-1798.
3. Mearin F, Lacy BE, Chang L, et al. Bowel Disorders. *Gastroenterology.* 2016.
4. Holtmann GJ, Ford AC, Talley NJ. Pathophysiology of irritable bowel syndrome. *Lancet Gastroenterol Hepatol.* 2016;1(2):133-146.
5. Drossman DA. Functional Gastrointestinal Disorders: History, Pathophysiology, Clinical Features and Rome IV. *Gastroenterology.* 2016.
6. Boyce PM, Koloski NA, Talley NJ. Irritable bowel syndrome according to varying diagnostic criteria: are the new Rome II criteria unnecessarily restrictive for research and practice? *Am J Gastroenterol.* 2000;95(11):3176-3183.
7. Park DW, Lee OY, Shim SG, et al. The Differences in Prevalence and Sociodemographic Characteristics of Irritable Bowel Syndrome According to Rome II and Rome III. *J Neurogastroenterol Motil.* 2010;16(2):186-193.
8. Foundation R. Appendix A: Rome III Diagnostic Criteria for FGIDs. 885-897.
9. Aziz I, Tornblom H, Palsson OS, Whitehead WE, Simren M. How the Change in IBS Criteria From Rome III to Rome IV Impacts on Clinical Characteristics and Key Pathophysiological Factors. *Am J Gastroenterol.* 2018.
10. Whitehead WE, Palsson OS, Simren M. Irritable bowel syndrome: what do the new Rome IV diagnostic guidelines mean for patient management? *Expert Rev Gastroenterol Hepatol.* 2017;11(4):281-283.
11. Sood R, Gracie DJ, Law GR, Ford AC. Systematic review with meta-analysis: the accuracy of diagnosing irritable bowel syndrome with symptoms, biomarkers and/or psychological markers. *Aliment Pharmacol Ther.* 2015;42(5):491-503.
12. Mujagic Z, Tigchelaar EF, Zhernakova A, et al. A novel biomarker panel for irritable bowel syndrome and the application in the general population. *Sci Rep.* 2016;6:26420.
13. Enck P, Aziz Q, Barbara G, et al. Irritable bowel syndrome. *Nat Rev Dis Primers.* 2016;2:16014.
14. Corsetti M, Whorwell P. Novel pharmacological therapies for irritable bowel syndrome. *Expert Rev Gastroenterol Hepatol.* 2016.
15. Palsson OS, Baggish J, Whitehead WE. Episodic nature of symptoms in irritable bowel syndrome. *Am J Gastroenterol.* 2014;109(9):1450-1460.
16. Mearin F, Baro E, Roset M, Badia X, Zarate N, Perez I. Clinical patterns over time in irritable bowel syndrome: symptom instability and severity variability. *Am J Gastroenterol.* 2004;99(1):113-121.
17. Drossman DA, Morris CB, Hu Y, et al. A prospective assessment of bowel habit in irritable bowel syndrome in women: defining an alternator. *Gastroenterology.* 2005;128(3):580-589.
18. Ford AC, Forman D, Bailey AG, Axon AT, Moayyedi P. Fluctuation of gastrointestinal symptoms in the community: a 10-year longitudinal follow-up study. *Aliment Pharmacol Ther.* 2008;28(8):1013-1020.
19. Mearin F, Badia X, Balboa A, et al. Predictive factors of irritable bowel syndrome improvement: 1-year prospective evaluation in 400 patients. *Aliment Pharmacol Ther.* 2006;23(6):815-826.

20. El-Serag HB, Pilgrim P, Schoenfeld P. Systemic review: Natural history of irritable bowel syndrome. *Aliment Pharmacol Ther.* 2004;19(8):861-870.
21. Clevers E, Tack J, Tornblom H, et al. Development of Irritable Bowel Syndrome Features Over a 5-year Period. *Clin Gastroenterol Hepatol.* 2018;16(8):1244-1251.e1.
22. Card T, Enck P, Barbara G, et al. Post-infectious IBS: Defining its clinical features and prognosis using an internet-based survey. *United Eur Gastroenterol J.* 2018;6(8):1245-1253.
23. Vandeputte D, Falony G, Vieira-Silva S, Tito RY, Joossens M, Raes J. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut.* 2016;65(1):57-62.
24. Falony G, Joossens M, Vieira-Silva S, et al. Population-level analysis of gut microbiome variation. *Science.* 2016;352(6285):560-564.
25. Tigchelaar EF, Bonder MJ, Jankipersadsing SA, Fu J, Wijmenga C, Zhernakova A. Gut microbiota composition associated with stool consistency. *Gut.* 2016;65(3):540-542.
26. 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.
27. Van Oudenhove L, Crowell MD, Drossman DA, et al. Biopsychosocial Aspects of Functional Gastrointestinal Disorders. *Gastroenterology.* 2016.
28. 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.
29. Clevers E, Vaes B, Henrard S, et al. Health problems associated with irritable bowel syndrome: analysis of a primary care registry. *Aliment Pharmacol Ther.* 2018.
30. Whitehead WE, Palsson O, Jones KR. Systematic review of the comorbidity of irritable bowel syndrome with other disorders: what are the causes and implications? *Gastroenterology.* 2002;122(4):1140-1156.
31. Corsetti M, Whorwell P. The global impact of IBS: time to think about IBS-specific models of care? *Therap Adv Gastroenterol.* 2017;10(9):727-736.



ADDENDUM

Addendum

Summary

Nederlandse samenvatting

Impact paragraph

Dankwoord

List of publications

Curriculum vitae

Summary

Irritable bowel syndrome (IBS) is a prevalent functional gastrointestinal disorder (FGID), currently considered “disorders of gut-brain interaction”, and is defined as a combination of abdominal pain and altered bowel habits. IBS affects 5-11% of the Western population and leads to considerable (health care) costs for society, due to high disease burden. The underlying pathophysiology of IBS is considered multifactorial, but incompletely understood, which hampers the identification of reliable biological markers and development of effective treatments. Hitherto, diagnosis, follow-up of disease severity over time, and evaluation of treatment efficacy all rely on patient-reported symptoms.

The current thesis evaluates different methods for symptom registration in IBS and focuses on the development of a novel real-time symptom assessment method, the experience sampling method (ESM). Consisting of three parts, the thesis focuses on diagnosing IBS and following symptom severity over time (**part I**), the development and validation of disease-specific ESM-tools (**part II**), and the potential benefit of ESM in evaluating symptom patterns, symptom-symptom interactions, and treatment efficacy (**part III**).

Part I – Diagnosis and symptoms in IBS

IBS is a symptom-based diagnosis, for which the Rome IV criteria are the latest updated version of the diagnostic criteria. Core symptoms are abdominal pain and altered bowel habits, *i.e.* constipation, diarrhea, or a combination of both. However, symptom presentation can vary widely, leading to a heterogeneous patient population. Furthermore, due to the continuous process of adaptations to the diagnostic criteria since 1978, characteristics of the IBS population might vary depending on the criteria as used.

The Rome IV criteria were published in 2016 and in **chapter 2** we compare IBS prevalence between Rome III and Rome IV criteria. Furthermore, clinical characteristics of Rome III IBS patients and subjects that also fulfill the Rome IV criteria were compared. Of 404 Rome III-diagnosed IBS patients, between 62% and 87% was likely to also fulfill the more stringent Rome IV criteria. Rome IV-positive subjects reflected a subgroup of Rome III patients with more severe gastrointestinal (GI) symptoms and psychological comorbidities, and lower quality of life. Although IBS will be considered the same entity, it is of importance to be aware of these differences when evaluating IBS populations in clinical settings or comparing results from multiple studies.

In addition to cross-sectional differences between subjects, it is also known that symptoms can vary widely over time within subjects. In **chapter 3**, the evolution of symptoms over time is elucidated in 161 Rome III IBS patients. After a mean follow-up duration of 4.7 years, 30% did not fulfill Rome III criteria any longer and showed significantly lower levels of GI symptoms and GI-specific anxiety, whereas psychological symptoms (*i.e.* depression and anxiety) and quality of life were comparable between subjects that did and did not fulfill IBS diagnostic criteria at follow-up. These results imply that long-term quality of life and general well-being might depend on concurrent psychological symptoms, rather than GI symptom improvement over time.

Chapter 4 also considers symptom variability over time, focusing on bowel habits in relation to gut microbiota, one of the supposed pathophysiological mechanisms of IBS. In this chapter, we aimed to evaluate fluctuations in stool consistency over seven days and relate this to possible changes in fecal microbial composition. A significant association between stool consistency and the microbial composition was demonstrated, but we found rather stable microbiota within subjects over seven days, even in cases of highly fluctuating stool consistency from day to day. This implies that taking into account stool consistency in future microbiota analyses is important, but in the context of between-subject comparisons, a single fecal sample seems representative for the individual microbial picture over a short time-period.

Part I of the thesis herewith underlines the multifactorial characteristic of IBS, showing high variability in symptom presentation between subjects, but also within-subject changes over time. In order to capture the entire subject-specific picture of the disorder, there is a need for accurate and valid assessment tools that not only evaluate GI symptoms, but also consider comorbid psychological features and general quality of life, and take into account changes over time.

Part II – The experience sampling method – Development and validation

The second part of this thesis focuses on different methods for symptom registration in irritable bowel syndrome and other functional disorders. The risk for several biases (*i.e.* recall and ecological bias) and a lack of compliance come with the currently used retrospective, and often paper-and-pencil methods. Therefore, there is a need for validated methods that overcome these

limitations. The experience sampling method (ESM) is here proposed as a novel, potentially more accurate, symptom assessment method for such symptom-based disorders. The digital ESM is based on random and repeated measures, concerning the in-the-moment state and natural environment of a subject.

In **chapter 5**, an exploratory study describes the comparison between ESM, end-of-day, and end-of-week reports in IBS. Abdominal pain and flatulence scores reported at the end of the day were higher than the daily mean of repeated ESM scores, and a preference for reporting peak-levels at the end of the day was found. These differences were even more pronounced when comparing mean ESM scores over a one-week period to conventional end-of-week questionnaires. This suggests that ESM measures IBS symptoms more accurately by eliminating recall and ecological biases. In order to use the ESM in future studies or clinical care, however, disease-specific tools should be developed and validated.

In **chapter 6**, we developed a patient-reported outcome measure specific for the use of ESM in irritable bowel syndrome populations. According to current guidelines, this process included focus group interviews and cognitive interviewing with subjects from the target population, as well as expert opinions from international experts in the field of neurogastroenterology. Also, cross-cultural validation took place for the purpose of using the tool in different countries and languages. **Chapter 7** and **8** describe this development process of similar tools for functional dyspepsia and functional urologic complaints, respectively. In **chapter 9**, the further validation of the IBS-specific tool is elaborated. A completion rate of 68.6% was found, indicating that subjects completed 48 measurements over seven days on average. Adequate concurrent validity was demonstrated by comparing ESM scores to corresponding end-of-day and end-of-week scores; moderate-to-good agreement between the different methods was found. Cronbach's α coefficients showed moderate-to-good internal consistency for five constructs within the ESM-instrument (*i.e.* upper GI symptoms, lower GI symptoms, physical non-GI symptoms, positive affect, and negative affect), and test-retest reliability was confirmed by good consistency between scores of the first half-week and scores of the second half-week. These results support the validity and reliability of the IBS-specific ESM-tool.

Part III – The experience sampling method in IBS – Applications in clinical research

The last part of this thesis elaborates on applications of the experience sampling method in IBS clinical research. **Chapter 10** presents a thorough evaluation of the association between abdominal pain and stress in IBS and healthy subjects, using repeated measures by the experience sampling method over seven days. A positive association between stress levels and concurrent abdominal pain scores was found, but with high heterogeneity in the magnitude of this association between individual IBS patients. The results pointed towards a difference in response to stress between IBS and healthy subjects, rather than a difference in experienced stress during daily life. Furthermore, the results suggested an in-the-moment but not a longitudinal association, since abdominal pain scores could not be predicted by preceding stress levels, and vice versa. This study describes the heterogeneity and complexity of the interaction between stress and abdominal pain in IBS, and illustrates the additive value of repeated measures using the experience sampling method, when evaluating this complex interplay between symptoms.

In **chapter 11**, the use of ESM as a method of treatment evaluation is described. Since no biological markers for IBS are currently available, evaluation of treatment efficacy of current and newly developed therapies highly relies on symptom improvement. Using data of a randomized controlled trial on escitalopram versus placebo in patients with IBS and comorbid panic disorder, differences in treatment outcome between momentary and retrospective assessment methods were examined. When assessing treatment effect using the ESM, a significant decrease in abdominal pain was observed in the escitalopram as compared to placebo group, after six months of treatment. However, when assessing this using end-of-day or end-of-week reports, no beneficial effect of escitalopram over placebo was found. The difference in treatment outcome between the measures indicates a potential benefit of ESM over conventional symptom assessment methods by detecting within-subject changes over time with higher sensitivity, due to the high number of repeated measures. Furthermore, the potential of ESM to capture treatment response more sensitively than retrospective reports, by taking into account symptom fluctuations during the day, but also momentary factors that might moderate treatment effect (for example psychological symptoms), has been postulated.

Altogether, this thesis illustrates the challenges in capturing the entire picture of IBS symptomatology including heterogeneity between subjects and changes over time within subjects, as well as comorbid psychological and environmental factors potentially interacting with GI complaints. It serves as a foundation for the use of ESM in IBS and other symptom-based disorders. First steps in development and validation of disease-specific ESM tools are described and examples of the potential use in scientific research and clinical care have been given. Future studies could benefit from this by using the developed ESM tools to further unravel the complex association between key symptoms and influencing factors during daily life, potentially leading to more effective patient-tailored treatment strategies.

Addendum

Summary

Nederlandse samenvatting

Impact paragraph

Dankwoord

List of publications

Curriculum vitae

Nederlandse samenvatting

Het prikkelbare darm syndroom (PDS), irritable bowel syndrome (IBS) in het Engels, is een veel voorkomende functionele maag-darmaandoening (FGID; functional gastrointestinal disorder), die momenteel wordt omschreven als een "aandoening betreffende de hersen-darmas", en wordt gedefinieerd als een combinatie van buikpijn en veranderd ontlastingspatroon. PDS komt voor bij ongeveer 5-11% van de westerse bevolking en leidt tot aanzienlijke kosten voor de samenleving, waaronder gezondheidszorgkosten, ten gevolge van de hoge ziektelast. De onderliggende, multifactoriële pathofysiologie van PDS is nog niet volledig bekend, wat de identificatie van betrouwbare biologische markers en de ontwikkeling van effectieve behandelingen hindert. Daardoor zijn tot nu toe de diagnose, de follow-up van het natuurlijke ziektebeloop en de evaluatie van de effectiviteit van behandelingen afhankelijk van door de patiënt gerapporteerde klachten.

Dit proefschrift evalueert verschillende methoden voor klachtenregistratie bij PDS en richt zich op de ontwikkeling van een nieuwe "real-time" methode, de Experience Sampling Method (ESM). Het proefschrift bestaat uit drie delen en richt zich op het diagnosticeren van PDS en het natuurlijke beloop van de ernst van klachten (**deel I**), de ontwikkeling en validatie van ESM-instrumenten voor specifieke aandoeningen (**deel II**) en potentiële voordelen van ESM in het evalueren van klachtenpatronen, interacties tussen verschillende klachten en de werkzaamheid van behandelingen (**deel III**).

Deel I – Diagnose en klachten bij PDS

De diagnose PDS wordt gesteld op basis van klachten, waarbij gebruik gemaakt wordt van de meest recente diagnostische criteria, de Rome IV-criteria. Buikpijn en veranderd ontlastingspatroon, d.w.z. obstipatie, diarree of een combinatie van beiden, staan hierin centraal als klachten. De presentatie van klachten kan echter sterk variëren tussen personen, waardoor PDS wordt gekenmerkt door een heterogene patiëntenpopulatie. Doordat er sinds 1978 meermaals aanpassingen aan de diagnostische criteria zijn gedaan, kunnen de specifieke kenmerken van PDS-patiënten ook variëren afhankelijk van de gebruikte criteria.

In 2016 werden de Rome IV-criteria gepubliceerd en in **hoofdstuk 2** vergelijken we de PDS-prevalentie tussen Rome III- en Rome IV-criteria. Verder worden in dit hoofdstuk de klinische kenmerken vergeleken tussen Rome III-PDS-patiënten

en proefpersonen die ook aan de Rome IV-criteria voldoen. Van 404 Rome III-PDS-patiënten was het waarschijnlijk dat 62% tot 87% ook aan de strengere Rome IV-criteria voldeed. Rome IV-positieve proefpersonen weerspiegelden een subgroep van de Rome III-patiënten met ernstigere maag-darmklachten en psychologische comorbiditeiten, en een lagere kwaliteit van leven. Hoewel PDS door de jaren heen nog steeds als dezelfde aandoening wordt beschouwd, is het van belang om deze verschillen mee te wegen wanneer men PDS-patiënten evalueert in de kliniek of wanneer resultaten van verschillende onderzoeken met elkaar worden vergeleken.

Naast verschillen tussen patiënten op een bepaald moment, is het ook bekend dat klachten per persoon sterk kunnen variëren van moment tot moment. In **hoofdstuk 3** wordt deze variatie van klachten over de tijd beschreven bij 161 PDS-patiënten (Rome III). Na een gemiddelde follow-upduur van 4,7 jaar voldeed 30% niet meer aan de Rome III-criteria en rapporteerde significant lagere scores voor maag-darmklachten en hieraan gerelateerde angstscores. Psychologische klachten, zoals depressie en angst, en kwaliteit van leven waren echter vergelijkbaar tussen proefpersonen die op het moment van follow-up wel, en zij die niet, voldeden aan de diagnostische criteria voor PDS. Deze resultaten impliceren dat de kwaliteit van leven en het algemeen welbevinden op de lange termijn mogelijk niet afhankelijk zijn van verbetering van de maag-darmklachten na verloop van tijd, maar eerder van de aanwezigheid van psychische klachten.

In **hoofdstuk 4** wordt ook gekeken naar schommelingen in klachten over een bepaalde tijdsperiode, maar met de nadruk op het ontlastingspatroon in relatie tot de darmflora, één van de veronderstelde pathofysiologische mechanismen van PDS. In dit hoofdstuk worden schommelingen in de ontlastingsconsistentie over zeven dagen geëvalueerd en wordt onderzocht of deze samenhangen met veranderingen in de microbiële samenstelling in de ontlasting. Er werd een significante associatie tussen de ontlastingsconsistentie en de microbiota aangetoond. De microbiële samenstelling was echter per persoon vrij stabiel gedurende zeven dagen, zelfs in het geval van sterk wisselende ontlastingsconsistentie van dag tot dag. Dit impliceert dat het belangrijk is om bij toekomstige microbiota-analyses rekening te houden met de ontlastingsconsistentie, maar bij vergelijkingen van de microbiota tussen proefpersonen lijkt een enkel ontlastingsmonster representatief voor het microbiële beeld van een specifieke persoon over een korte tijdsperiode.

Deel I van het proefschrift onderschrijft hiermee het multifactoriële aspect van PDS, met een hoge variatie in de klachtenpresentatie tussen proefpersonen, maar ook veranderingen van klachten van moment tot moment per persoon. Om het individuele klachtenpatroon van de aandoening volledig in beeld te kunnen brengen, is er behoefte aan nauwkeurige en valide instrumenten voor klachtenregistratie, die niet alleen maag-darmklachten, maar ook comorbide psychologische klachten en kwaliteit van leven vastleggen, en rekening houden met veranderingen over de tijd.

Deel II – De experience sampling method – Ontwikkeling en validatie

Het tweede deel van dit proefschrift richt zich op verschillende methoden voor klachtenregistratie bij het prikkelbare darm syndroom en andere functionele aandoeningen. Tot nu toe worden hiervoor met name retrospectieve, papieren methoden gebruikt, zoals dagboeken aan het einde van de dag of vragenlijsten aan het einde van een week. Deze brengen echter het risico op verschillende vormen van vertekening met zich mee, zoals het zich verkeerd herinneren van klachten gedurende een bepaalde periode in het verleden (“recall bias”) of de beïnvloeding van antwoorden door omgevingsfactoren op het moment van invullen (“ecological bias”). Ook worden met name papieren instrumenten vaak niet op het juiste moment ingevuld. Zodoende is er behoefte aan gevalideerde methoden, die deze beperkingen niet hebben. De experience sampling method (ESM) wordt in dit proefschrift aangedragen als een nieuwe, potentieel nauwkeurigere methode voor het meten van symptomen voor dergelijke aandoeningen. De digitale ESM maakt gebruik van herhaalde metingen op willekeurige momenten, die betrekking hebben op de huidige staat en de natuurlijke omgeving van een persoon.

In **hoofdstuk 5** worden in een exploratieve studie klachtenregistraties van PDS-patiënten, verzameld middels ESM-, einde-van-de-dag- en einde-van-de-weekrapportages, met elkaar vergeleken. Scores voor buikpijn en winderigheid werden aan het einde van de dag hoger gerapporteerd dan gemiddeld over de dag middels herhaalde ESM-scores. Tevens werd een voorkeur voor het rapporteren van piek-scores aan het eind van de dag gezien. Deze verschillen waren nog meer uitgesproken wanneer gemiddelde ESM-scores over een periode van een week werden vergeleken met conventionele vragenlijsten aan het einde van de week. Dit suggereert dat ESM PDS-klachten nauwkeuriger meet door recall en ecological bias te elimineren. Om de ESM in toekomstige

studies of klinische zorg te kunnen gebruiken, moeten echter instrumenten specifiek voor bepaalde aandoeningen worden ontwikkeld en gevalideerd.

In **hoofdstuk 6** hebben we een patiënt-gerapporteerde uitkomstmaat ontwikkeld, specifiek voor het gebruik van ESM door patiënten met het prikkelbare darm syndroom. Op basis van de huidige richtlijnen bestond dit ontwikkelingsproces uit onder andere focusgroep- en cognitieve interviews met personen uit de doelgroep. Tevens zijn de meningen van internationale experts op het gebied van neurogastro-enterologie hierbij betrokken en vond culturele validatie plaats om het instrument in verschillende landen en talen te kunnen gebruiken. **Hoofdstuk 7** en **8** beschrijven eenzelfde ontwikkelingsproces van vergelijkbare instrumenten voor respectievelijk functionele dyspepsie en functionele urologische klachten.

In **hoofdstuk 9** wordt de verdere validatie van de PDS-specifieke tool uitgewerkt. Een invul-percentag (‘‘compliance’’) van 68,6% werd gevonden, wat aangeeft dat proefpersonen gemiddeld 48 metingen over zeven dagen voltooiden. Adequate ‘‘concurrent validity’’ werd aangetoond door ESM-scores te vergelijken met overeenkomstige einde-van-de-dag- en einde-van-de-weekscores; er werd een matige tot goede overeenkomst tussen de verschillende methoden gevonden. Cronbach's α -coëfficiënten toonden een matige tot goede ‘‘internal consistency’’ voor vijf constructen binnen het ESM-instrument (d.w.z. bovenste maag-darmklachten, lagere maag-darmklachten, fysieke niet-maag-darmklachten, positieve emoties en negatieve emoties), en ‘‘test-retest reliability’’ werd bevestigd door goede overeenkomst tussen scores van de eerste helft van de week en scores van de tweede helft van de week. Deze resultaten ondersteunen de validiteit en betrouwbaarheid van het PDS-specifieke ESM-instrument.

Deel III – De experience sampling method voor PDS – Toepassingen voor klinisch onderzoek

Het laatste deel van dit proefschrift gaat in op toepassingen van de ESM voor klinisch onderzoek naar PDS. **Hoofdstuk 10** beschrijft een uitgebreide evaluatie van de associatie tussen buikpijn en stress bij PDS-patiënten en gezonde proefpersonen, op basis van herhaalde metingen middels ESM gedurende zeven dagen. Een positieve associatie tussen stress- en gelijktijdige buikpijnscores werd gevonden, maar met een hoge variatie in de mate van deze associatie tussen individuele PDS-patiënten. De resultaten wijzen op een

verschil in reactie op stress tussen PDS-patiënten en gezonde proefpersonen, maar niet op een verschil in ervaren stress tijdens het dagelijks leven. Bovendien suggereerden de resultaten een cross-sectionele (op hetzelfde moment) maar geen longitudinale (op verschillende momenten) associatie, aangezien buikpijnscores niet konden worden voorspeld door voorgaande stress-levels, en vice versa. Deze studie beschrijft de heterogeniteit en complexiteit van de interactie tussen stress en buikpijn bij PDS, en illustreert de toegevoegde waarde van herhaalde metingen met behulp van de ESM bij het evalueren van dit complexe samenspel aan klachten.

In **hoofdstuk 11** wordt het gebruik van de ESM voor evaluatie van behandelingen beschreven. Aangezien er momenteel geen biologische markers voor PDS beschikbaar zijn, is de beoordeling van de effectiviteit van huidige en nieuw-ontwikkelde therapieën sterk afhankelijk van de mate van symptoomverlichting. Met behulp van gegevens van een gerandomiseerde en gecontroleerde studie (RCT) over escitalopram versus placebo bij patiënten met PDS en comorbide paniekstoornis, werden verschillen in effectiviteit van behandeling tussen real-time en retrospectieve methoden voor klachtenregistratie onderzocht. Bij het beoordelen van het behandelings-effect met behulp van de ESM werd na zes maanden behandeling een significante afname van buikpijn waargenomen in de escitalopram- vergeleken met de placebogroep. Bij de beoordeling hiervan middels einde-van-de-dag- of einde-van-de-week-methoden werd echter geen gunstig effect van escitalopram ten opzichte van placebo gevonden. Het verschil in gevonden behandelings-effectiviteit tussen de methoden wijst op een potentieel voordeel van de ESM ten opzichte van conventionele methoden voor het meten van symptomen. Vanwege het hoge aantal herhaalde metingen is de ESM in staat om veranderingen over de tijd binnen patiënten met een grotere gevoeligheid vast te leggen. Bovendien heeft de ESM de potentie om schommelingen in klachten gedurende de dag te registreren, maar ook rekening te houden met factoren die het behandelings-effect zouden kunnen beïnvloeden (bijvoorbeeld psychologische symptomen), waardoor de respons op behandeling gevoeliger geregistreerd kan worden dan met retrospectieve methoden.

In dit proefschrift worden de uitdagingen van het volledig in kaart brengen van PDS-klachten beschreven, inclusief het vastleggen van heterogeniteit tussen proefpersonen, veranderingen over een tijdsperiode binnen proefpersonen en comorbide psychologische en omgevingsfactoren die mogelijk een

wisselwerking hebben met maag-darmklachten. Het dient hiermee als basis voor het gebruik van de ESM bij PDS en andere functionele aandoeningen. De eerste stappen in de ontwikkeling en validatie van ESM-instrumenten voor een aantal specifieke aandoeningen worden beschreven, en er worden voorbeelden gegeven van mogelijke toepassingen van de ESM in wetenschappelijk onderzoek en klinische zorg. Toekomstige studies zouden hiervan kunnen profiteren door de ontwikkelde ESM-instrumenten te gebruiken om de complexe associatie tussen de belangrijkste symptomen en beïnvloedende factoren in het dagelijks leven verder te ontrafelen, wat mogelijk kan leiden tot effectievere patiënt-specifieke behandelstrategieën.

Addendum

Summary

Nederlandse samenvatting

Impact paragraph

Dankwoord

List of publications

Curriculum vitae

Impact paragraph

Irritable bowel syndrome (IBS) is a highly prevalent chronic gastrointestinal (GI) disorder, presenting with variable abdominal complaints, but defined as abdominal pain accompanied by altered bowel habits (*i.e.* constipation, diarrhea, or a combination of both). These symptoms lead to a high patient burden with impaired quality of life, work absenteeism and presenteeism, and increased health care consumption. Altogether, this results in decreased quality of life for the affected individual and substantial direct and indirect costs for society. IBS therefore not only influences physical and mental wellbeing on an individual level, but also has an important societal impact. The underlying pathophysiology of IBS is hitherto incompletely understood, but is believed to be multifactorial. Resulting, reliable biological markers for diagnosis and follow-up of disease course have yet to be developed, and we mostly rely on patient-reported symptoms to evaluate disease severity and treatment efficacy. Currently available methods for symptom registration have, however, important limitations, and might therefore be suboptimal to rely on in scientific as well as clinical settings.

The main aim of the current thesis is to develop a novel symptom assessment method, the Experience Sampling Method (ESM), specific for IBS populations, and to evaluate its advantages over currently used methods. ESM is an electronic, real-time symptom assessment method that uses random and repeated measurements in order to create a full picture of a subject's symptoms during daily life. The purpose of the research described in this thesis is to pave the way towards the use of ESM for symptom registration in IBS, in both scientific research and patient care, in order to improve understanding of the pathophysiology as well as available treatment options in the field of IBS.

In the first part of this thesis, the currently used symptom-based diagnostic criteria for IBS, *i.e.* Rome criteria, and symptom variability over time are evaluated. Differences in disease prevalence and patient characteristics, when using the current Rome IV criteria compared to the previous Rome III criteria, underline the importance of accurate symptom assessment when diagnosing and evaluating IBS. Furthermore, the need for symptom assessment methods that take into account both between-subject differences and within-subject variability over time, as well as concurrent psychological symptoms and environmental factors, is emphasized.

Subsequently, the second and third part of this thesis describe the development and validation of disease-specific patient-reported outcome measures (PROMs) for the use of ESM, and give examples of the advantages of applying ESM in the evaluation of symptoms and treatment efficacy in IBS. Important differences in symptom scores were shown between ESM and retrospective end-of-day and end-of-week assessments, suggesting that ESM reflects symptom patterns more accurately than the commonly used retrospective methods. More specific, the retrospective methods showed a tendency to report peak symptom scores rather than average symptom scores during the day. Additionally, differences in the measurement of treatment response were demonstrated between ESM and retrospective methods. An IBS-specific ESM-tool was then developed according to the current international guidelines on PROM-development, and adequate validity was shown. Lastly, this thesis describes the use of ESM in a thorough evaluation of the association between abdominal pain and stress in IBS patients and healthy control subjects. In this evaluation, ESM data elegantly present high variability in both abdominal pain and stress between as well as within subjects, and demonstrates heterogeneity in the association between these symptoms in IBS patients.

Altogether, the current thesis emphasizes the need for considering both between- and within-subject differences in symptom reports when evaluating IBS. The results presented in this thesis are mainly of interest from a scientific point of view, at least on the short term. The developed IBS-specific ESM-tool is readily available and can easily be used to evaluate symptom-related research questions in IBS study populations. As we have shown, it is not only of use in the evaluation of specific GI symptoms, but also in the evaluation of interactions between several GI symptoms, between GI symptoms and psychological symptoms, or between symptoms and life style factors that are assumed to be of influence on IBS complaints. It is therefore that the current thesis could form a basis for an important shift in the way that the clinical picture of IBS is presented in scientific literature, and offers the opportunity of looking into symptom patterns with more detail and on an individual level.

On the somewhat longer term, this thesis could also have a considerable impact from a societal point of view, since it could have an important role in the diagnostic and therapeutic processes in clinical care. Once the described IBS-specific ESM-tool has been further developed, it could be used during patients' clinical consultation of a gastroenterologist or general practitioner. Ideally, the

data collected would be transformed to a clear overview of a patient's symptom pattern during the past week, including the most important triggers of symptoms during daily life. Such an overview could enable patients themselves to get more insight into triggers, like life style factors, situations, or behavior that could be of influence on symptom formation. This is useful in diagnosing the disorder and has an educative aspect by informing subjects on the nature of their disease and the development of symptoms during daily life. Ultimately, this person-tailored overview of symptoms could be used as a basis for management of the disorder. Subjects could discuss the most significant disease determinants according to the ESM data and opt to target these by life style interventions, dietary changes, psychological treatment, or drug interventions. Since many therapeutic options for IBS have been suggested, but all show only moderate efficacy at best in therapeutic trials on group level, this approach would be a step towards patient-tailored personalized care. Therefore, a future step will be to make ESM readily available for individual care givers and their patients, in order to explore all the possibilities of this new and promising tool for diagnosing and treating IBS, as well as create self-awareness and self-management in this patient group.

Addendum

Summary

Nederlandse samenvatting

Impact paragraph

Dankwoord

List of publications

Curriculum vitae

Dankwoord

Na 4,5 jaar fulltime en bijna 2 jaar naast het klinisch werk te hebben besteed aan “mijn” onderzoek, is dit proefschrift de kers op de taart. Dit was er natuurlijk nooit gekomen zonder de hulp en steun van zoveel mensen. Ik kan onmogelijk iedereen persoonlijk benoemen, maar wil in dit hoofdstuk toch een aantal mensen nadrukkelijk bedanken voor de afgelopen jaren.

Mijn paranimfen.

Zsa Zsa, lieve Zsas, ik kan me mijn PhD niet voorstellen zonder jou. We hebben als “IBS twins” 4,5 jaar zij aan zij doorgebracht op de UNS 50 en konden hierdoor altijd bij elkaar terecht om te sparren over studie-opzet, analyses of manuscripten, maar ook om ons te beklagen over alles (en iedereen) als er iets tegenzat. Dank je wel dat je mijn gevloek en tirades, meestal tegen de computer of R, altijd accepteerde (of tactvol negeerde). Ik zal vaak terugdenken aan alle leuke momenten die we samen hebben meegemaakt; de simpele lachbuien op ons kantoor, feestjes met collega’s en de vele binnen- en buitenlandse reizen die we hebben mogen maken. Ik ben je eeuwig dankbaar voor de mentale support (vaak met chocola) die zo belangrijk was; gelukkig weten we elkaar nog steeds bijna dagelijks digitaal te vinden!

Maxime, lieve Max, na ruim 18 jaar vriendschap weten wij precies wat we aan elkaar hebben. Het plan om samen in Amsterdam Geneeskunde te gaan studeren werd overhoopgegooid, maar onze vriendschap is er niet minder hecht door geworden. En hoe bijzonder is het dat we los van elkaar precies hetzelfde pad hebben gekozen. Zo hebben we op afstand toch alle PhD-struggles met elkaar kunnen delen en geldt dit inmiddels ook voor de kliniek. In 2017 samen naar de UEGW in Barcelona was toch “zuks gezelligs”; ik kijk nu al uit naar alle UEGW’s en DDW’s die nog gaan volgen! En wie weet vinden we die 2-onder-1-kap ook nog wel ;)

Mijn promotieteam.

Prof. Masclee, beste Ad, dank voor het vertrouwen om mij onder uw supervisie mijn PhD te laten uitvoeren. Tijdens mijn sollicitatiegesprek benoemde u dat het prikkelbare darm syndroom niet altijd gezien wordt als het meest “sexy” onderwerp om onderzoek naar te doen, maar uw werkdrive en passie voor het vak, en soms ook strenge woorden, hebben mij gestimuleerd als onderzoeker en daardoor mijn proefschrift mede gevormd tot het huidige eindresultaat.

Dank voor uw geduld zodat ik mijn pad binnen de MDL op mijn eigen tempo kan bewandelen.

Daniel, zonder jou was dit proefschrift nooit geworden wat het nu is. Ondanks je eeuwige pessimisme en “stabiel slecht” als antwoord op bijna elke vraag, hebben jouw werkdrive en enthousiasme voor het onderzoek me altijd gestimuleerd. Jouw vermogen om altijd in oplossingen te denken heeft ervoor gezorgd dat ik bij jou terecht kon als ik het even niet meer overzag. Daarnaast zeer veel dank dat ik altijd binnen kon komen razen als ik even mijn frustratie kwijt moest. Ik hoop dat we nog lang mogen samenwerken.

Joanna, dankzij jou ben ik bij het IBS-team in Maastricht terecht gekomen, doordat ik in 2013 solliciteerde voor de WESP. Jouw passie voor zowel het klinische als het wetenschappelijke MPU-werk hebben mij enthousiast gemaakt voor dit vakgebied. Dank voor jouw steun en vertrouwen om mijzelf als arts en onderzoeker verder binnen de MDL te kunnen ontwikkelen.

Het IBS-team.

Daisy en Martine, jullie hebben als vaste leden van het IBS-team van begin af aan een belangrijke rol gespeeld in mijn PhD. Ik kon altijd bij jullie terecht met vragen over het IBS-cohort, frustraties over METC-protocollen of struggles met manuscripten. Daisy, jouw kritische blik en wetenschappelijke kennis hebben een grote bijdrage geleverd aan mijn proefschrift, waarvoor veel dank. Daarnaast is de gezelligheidsfactor, die jullie als buurvrouwen op de UNS 50 door de jaren heen hebben geleverd, natuurlijk ook van onschatbare waarde.

Zlatan, jij vroeg mij aan het eind van mijn WESP of ik, als ik dan toch huisarts wilde worden, niet een gepromoveerde huisarts wilde worden. Bedankt voor dit subtiele zetje richting de MDL! Jouw vermogen om overal het positieve van in te zien, is meer dan eens van pas gekomen wanneer ik vooral beren op de weg zag. Dank voor jouw hulp bij het verder uitrollen van de ESM en je kritische blik op manuscripten, en natuurlijk ook voor de gezelligheid tijdens feestjes en congressen.

De leescommissie.

Prof. dr. Rutten, prof. dr. Smout, prof. dr. Van Oudenhove, prof. dr. Cals en dr. Breukink; hartelijk dank voor jullie moeite en tijd om mijn proefschrift te evalueren.

Patiënten.

Heel veel dank aan alle patiënten die zich vrijwillig beschikbaar hebben gesteld voor deelname aan de studies, die de basis hebben gevormd voor dit proefschrift.

Collega's UM.

Natuurlijk was mijn PhD-tijd nooit zo'n mooie periode geweest zonder de collega's op de UNS 40 en 50. Ik kwam als beginnende promovenda terecht in een warm bad waar er altijd iemand bereikbaar was voor logistieke of PhD-technische vragen. Het is door de jaren heen een komen en gaan van collega's geweest, maar de warmte binnen de groep is altijd gebleven. Ik kan dan ook alleen maar met een brede glimlach terugkijken op alle gezellige lunch- en vloaimomenten, congresbezoeken in binnen- en buitenland, kerst- en promotiefeesten, VRIJMIBO's en wintersportweekenden. Ik wil iedereen, van dinosaurus tot jonge bloem, bedanken voor deze mooie jaren: Annick, Anke, Ankie, Bas, Bouke, Bram, Chantal, Corinne, Danyta, Dion, Elhaseen, Ellen, Evelien, Fabiënne, Fedde, Gonny, Hao Ran, Heike, Kirsten, Laura, Lonne, Marin, Mark, Marlijne, Montserrat, Pan, Pauline, Philippe, Quirine, Roel, Roy, Selwyn, Steven, Tim H., Tim K., Toon, Vince, Wenke, Wiesje en Yala.

Daarnaast wil ik alle "collega's" van de interne geneeskunde en chirurgie bedanken voor de gezelligheid op de gang! Jean, heel veel dank voor het ontwerpen van de cover voor mijn proefschrift. En natuurlijk ook voor de gezellige jaren op UM! Mirjam, bedankt voor de gezellige aanvulling die jij was op het 2-persoons "IBS"-huishouden in de UNS 50 gedurende jouw stage, en voor de vriendschap sindsdien.

Studenten.

Gedurende mijn PhD zijn er vele studenten geweest die mij, in de vorm van een WESP, WIP of stage anderszins, hebben geholpen bij het uitvoeren van de studies die uiteindelijk dit proefschrift vormen, waarvoor veel dank!

MDL MUMC+.

Ik wil graag alle mensen uit de kliniek bedanken voor de fijne samenwerking door de jaren heen. Ton en de dames van de functiekamer; Anja, Nicole en Rachele, waar ik altijd terecht kon als ik een onderzoekskamer nodig had, vragenlijsten kwijt was of gewoon om even te kletsen. Natuurlijk ook de dames van het secretariaat; Elly, Mietsie, Nienke en Tiny, die altijd bereikbaar waren

voor logistieke of organisatorische zaken. Dank aan alle MDL-artsen en -assistenten die door de jaren heen hebben bijgedragen aan dit proefschrift.

MEMIC.

Het MEMIC-team; Dirk, Donovan, Jacqueline, Koert, Luc en Paula, dank jullie wel voor het enthousiasme waarmee jullie met ons in het MEASuRE-project zijn gestapt. We hebben heel wat uren doorgebracht om de logistiek en inhoud van LDOT, Castor en de MEASuRE-app te bespreken. Gaandeweg zijn we ontelbare struikelblokken tegengekomen, maar gezamenlijk vonden we overal een oplossing voor. Dank voor jullie inzet en betrokkenheid bij de projecten voor mijn proefschrift.

Statistische ondersteuning.

Marjan Drukker, Richel Lousberg en Sander van Kuijk, dank voor jullie ondersteuning bij de statistiek. Het heeft me aardig wat frustratie opgeleverd, maar jullie hebben mij van een SPSS-newbie, via Stata, uit laten groeien tot redelijk zelfstandige R-gebruiker. Jullie hebben een onmisbare bijdrage aan dit proefschrift geleverd door mij de linear mixed-effects models te leren ontcijferen.

Co-auteurs/collaborators.

Naast mijn officiële promotieteam en alle collega's binnen de MDL, wil ik ook graag een aantal mensen in binnen- en buitenland danken voor hun bijdrage aan dit proefschrift.

Carsten Leue en John Penders, dank voor het delen van jullie expertise met betrekking tot respectievelijk de psychosomatiek en intestinale microbiota, en bijdrage aan manuscripten.

Esther van Barneveld, Alexandra Herrewegh, Daisy Vrijens en prof. Gommert van Koeveringe, dank voor de interessante samenwerking, die tot een aantal mooie manuscripten heeft geleid en waardoor we de Experience Sampling Method weer een stapje verder hebben kunnen ontwikkelen.

Prof. Magnus Simrén, prof. Qasim Aziz, prof. Jan Tack, Ceca Bojic, Maura Corsetti, Joanna Jalanka, Rabia Lalani, Chays Manichanh, Irina Midenfjord, Marta Pozuelo, Mirjana Rajilic-Stojanovic, and Hans Törnblom, thank you all very much for the fruitful collaboration, for sharing your expertise, and for welcoming me at your sites during my PhD-years.

John Bothmer and Emilio Quetglas, thank you very much for making our collaboration with Grünenthal GmbH work.

Pélerin commissie.

Tijdens mijn PhD heb ik een aantal jaren het Pélerin symposium mogen organiseren. Ik wil dan ook Audrey, John, Sophie en Yvon heel erg bedanken voor alle vergaderingen aan het eind van een werkdag, die hebben geleid tot een aantal succesvolle symposia, maar mij ook vooral veel energie gaven. Daarnaast dank voor de beste variant van onze vergaderingen, namelijk de gezellige avonden met wijn en goed eten!

Máxima Medisch Centrum.

De overgang van arts-onderzoekster naar ANIOS interne geneeskunde was in maart 2019 een flinke omschakeling, maar de geweldige groep assistenten heeft er zeker aan bijgedragen dat dit is gelukt en ik me nu helemaal thuis voel in het MMC. Lieve collega's, heel veel dank voor de fijne samenwerking, maar ook voor de gezellige cappuccino-momenten en borrels!

Alle internisten in het MMC, en in het bijzonder de opleider Louis Lieverse, wil ik bedanken voor hun begeleiding, samenwerking en steun de afgelopen twee jaar. Alle MDL-artsen in het MMC, en Aniek, dank voor de maanden op de MDL, waardoor ik mijn keuze voor de opleiding MDL definitief heb kunnen maken.

Mijn vriendinnen.

Een wijze vrouw schreef ooit "beter een verre vriend dan een goede buur"; en niets is minder waar. Want al vind ik de afstand niet altijd makkelijk, het is steeds weer een feest om met jullie samen te zijn.

De Fijne Vrouwen; lieve Max, Char en Kat, al ruim 15 jaar gaat onze vriendschap mee. Van wafels met kersen en slagroom tijdens de tussenuren, en poolen in het Centrum, tot serieuze gesprekken over het leven met een wijntje op de vrije vrijdagavond. Ondanks dat we allemaal veranderd zijn de afgelopen jaren en ieder een eigen weg is ingeslagen, en hoe lang we elkaar ook niet gezien hebben door onze drukke levens, het voelt altijd als thuiskomen bij jullie. Ik ben jullie dankbaar voor jullie onvoorwaardelijke steun; ik kan alleen maar hopen dat we uiteindelijk gezamenlijk dansend achter de rollators eindigen!

De Moeders; Gus, Mik, Mi en Kim, zonder jullie had ik het nooit volgehouden in Maastricht. En ondanks dat we nu letterlijk het hele land van Noord naar Zuid en van Oost naar West moeten doorreizen om elkaar te zien, om nog maar van alle tijdelijke emigraties te zwijgen, is het altijd als vanouds vertrouwd om jullie te zien. Ik hoop dat ik nog lang van jullie vriendschap mag genieten!

De Psycho's; lieve Nien en Eem, wij zijn pas later door het lot samengebracht tijdens ons blok "Psychomedisch". En eigenlijk is de vriendschap pas na de

studie hechter geworden. Jullie hebben jaren uitgekeken naar mijn promotie-feest en ik vrees dat ik jullie toch nog wat langer moet laten wachten.. Maar ik hoop op nog vele feestelijke jaren met jullie in het Brabantse!

Anne, Lotte en Pieternel, dank voor de vele etentjes, drankjes en fietstochten in en om Eindhoven, die zo belangrijk waren om even af te schakelen.

Lieve meiden, ik ben jullie allemaal onwijs dankbaar voor jullie betrokkenheid, begrip en vooral ook gezelligheid tijdens de afgelopen jaren <3

Rob, jij hebt mijn promotietijd, zowel de leuke als de stressvolle momenten, van dichtbij meegemaakt. Dank je wel dat je er voor me was als ik er doorheen zat of er een publicatie gevierd moest worden.

Jan en Monique, Esther en Ramon, Erik, Marloes, Eefje, Tom en Vik, heel veel dank voor jullie interesse, steun en gezelligheid tijdens alle verjaardagen, weekendjes Eifel en kerstavonden.

Familie.

Heel veel dank aan al mijn familieleden voor de steun op afstand. Ik vind het jammer dat jullie waarschijnlijk niet allemaal bij mijn verdediging kunnen zijn vanwege de COVID-maatregelen, maar hoop dat ik jullie online toch kan laten zien wat mijn promotie heeft ingehouden.

Lieve oma, ik vind het fantastisch dat u op 91-jarige leeftijd mijn promotie kan meemaken!

Lieve Tom, de Buck, ook voor jou is het niet makkelijk geweest dat ik zo plots naar de andere kant van het land verhuisde en is het een ver-van-je-bed-show wat je kleine zusje allemaal uitvoert in het zuiden. Dank je wel dat je, hoe lang ik ook ben weg geweest, altijd met wijd open armen klaar staat om me een dikke knuffel te geven als ik weer thuiskom.

Lieve pap en mam, zonder jullie was dit boekje er niet geweest. Ik ben jullie dankbaar dat jullie me hebben losgelaten om in Maastricht te gaan studeren en me altijd steunen in de keuzes die ik maak. Dank jullie wel dat jullie er altijd voor me zijn. Ik hoop dat jullie trots zijn op dit boekje.

Addendum

Summary

Nederlandse samenvatting

Impact paragraph

Dankwoord

List of publications

Curriculum vitae

List of publications

Kreiter D, Drukker M, Mujagic Z, **Vork L**, Rutten BPF, van Os J, Masclee AAM, Kruiemel JW, Leue C. 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 Feb; 141:110351.

Drukker M, Peters JCH, **Vork L**, Mujagic Z, Rutten BPF, van Os J, Masclee AAM, Kruiemel JW, Leue C. 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 Dec; 139:110261.

Weerts ZZRM, Heinen KGE, Masclee AAM, Quanjel ABA, Winkens B, **Vork L**, Rinkens PELM, Jonkers DMAE, Kesthelyi D. Smart Data Collection for the Assessment of Treatment Effects in Irritable Bowel Syndrome: Observational Study. *JMIR Mhealth Uhealth.* 2020 Nov 2;8(11):e19696.

Vork L, Keszthelyi D, van Kuijk SMJ, Quetglas EG, Törnblom H, Simrén M, Aziz Q, Corsetti M, Tack J, Mujagic Z, Leue C, Kruiemel JW, Masclee AAM. Patient specific stress-abdominal pain interaction in Irritable Bowel Syndrome: An exploratory Experience Sampling Method study. *Clin Transl Gastroenterol.* 2020 Jul;11(7):e00209.

Beckers AB, **Vork L**, Fikree A, de Ridder R, Aziz Q, Masclee AAM, Keszthelyi D. Colonoscopy is safe and not associated with higher pain scores in patients with Hypermobility Spectrum Disorder: Results from an exploratory prospective study. *Therap Adv Gastroenterol.* 2020 Jul 17;13:1756284820927310.

Mujagic Z, Weerts ZZRM, **Vork L**, Leue C, Kruiemel JW, Hesselink MAM, Muris JWM, Jonkers DMAE, Masclee AAM, Keszthelyi D. Quality of life in irritable bowel syndrome: Authors' reply. *Neurogastroenterol Motil.* 2020 Jan;32(1):e13729.

Vork L[#], Wilms E[#], Penders J, Jonkers DMAE. Stool consistency: looking beyond the Bristol Stool Form Scale. *J Neurogastroenterol Motil.* 2019 Oct 30;25(4):625.

Vork L[#], van Avesaat MHP[#], van Hoboken EA, Kruiemel JW, Conchillo JM, Keszthelyi D, Masclee AAM. Normal Colonic Transit Time predicts the outcome of Colonic Manometry in patients with Chronic Constipation – An exploratory study. *Int J Colorectal Dis.* 2019 Oct;34(10):1819-1822.

Vork L[#], Weerts ZZRM[#], Mujagic Z, Keszthelyi D, Hesselink MAM, Kruiemel JW, Leue C, Muris JWM, Jonkers DMAE, Masclee AAM. Reduction in IBS symptom severity is not paralleled by improvement in quality of life in patients with irritable bowel syndrome. *Neurogastroenterol Motil.* 2019 Aug;31(8):e13629.

Vork L, Mujagic Z, Drukker M, Keszthelyi D, Conchillo JM, Hesselink MAM, van Os J, Masclee AAM, Leue C, Kruiemel JW. The Experience Sampling Method - Evaluation of treatment effect of escitalopram in IBS with comorbid panic disorder. *Neurogastroenterol Motil.* 2019 Jan; 31(1):e13515.

Smeets FGM, Keszthelyi D, **Vork L**, Tack J, Talley NJ, Simrén M, Aziz Q, Ford AC, Conchillo JM, Kruiemel JW, Van Os J, Masclee AAM. 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 Feb; 31(2):e13496.

Weerts ZZRM, Keszthelyi D, **Vork L**, Aendekerk NCP, Frijlink HW, Brouwers JRBJ, Neef C, Jonkers DMAE, Masclee AAM. A novel ileo-colonic release peppermint oil soft capsule for the treatment of Irritable Bowel Syndrome: a pharmacokinetic phase I study in healthy volunteers. *Adv Ther.* 2018 Nov; 35(11): 1965-1978.

Herrewegh A, **Vork L**, Eurlings E, Leue C, Kruiemel JW, van Koeveringe G, Vrijens D. 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 Nov; 37(8): 2893-2903.

Vork L, Keszthelyi D, Mujagic Z, Kruiemel JW, Leue C, Pontén I, Törnblom H, Simrén M, Albu-Soda A, Aziz Q, Corsetti M, Holvoet L, Tack J, Rao SS, van Os J, Quetglas EG, Drossman DA, Masclee AAM. 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 Mar; 30(3).

Vork L[#], Weerts ZZRM[#], Mujagic Z, Kruiemel JW, Hesselink MAM, Muris JWM, Keszthelyi D, Jonkers DMAE, Masclee AAM. Rome III vs Rome IV criteria for irritable bowel syndrome: A comparison of clinical characteristics in a large cohort study. *Neurogastroenterol Motil.* 2018 Feb; 30(2).

Beckers AB, Keszthelyi D, Fikree A, **Vork L**, Masclee AAM, Farmer AD, Aziz Q. Gastrointestinal disorders in joint hypermobility syndrome/Ehlers-Danlos syndrome hypermobility type: A review for the gastroenterologist. *Neurogastroenterol Motil.* 2017 Aug; 29(8).

Imhann F, Bonder MJ, Vich Vila A, Fu J, Mujagic Z, **Vork L**, Tigchelaar EF, Jankipersadsing SA, Cenit MC, Harmsen HJ, Dijkstra G, Franke L, Xavier RJ, Jonkers DMAE, Wijmenga C, Weersma RK, Zhernakova A. Proton pump inhibitors affect the gut microbiome. *Gut.* 2016 May; 65(5):740-8.

Mujagic Z, **Vork L**[#], Leue C[#], Lousberg R, Jonkers DMAE, Keszthelyi D, Hesselink MAM, van Schagen TJ, van Os J, Masclee AAM, Kruiemel JW. The Experience Sampling Method--a new digital tool for momentary symptom assessment in IBS: an exploratory study. *Neurogastroenterol Motil.* 2015 Sep; 27(9):1295-302.

Submitted for publication

Vork L[#], Penders J[#], Jalanka J, Bojic C, van Kuijk SMJ, Salonen A, de Vos WM, Rajilic-Stojanovic M, Masclee AAM, Pozuelo M, Manichanh C, Jonkers DMAE. *Does day-to-day variability in stool consistency link to the fecal microbiota composition?*

Barneveld E, **Vork L**, Herrewegh A, van Poll M, Manders J, van Osch F, Spaans W, van Koevinge G, Vrijens D, Kruiemel JW, van Hanegem N, Lim A, Bongers M, Leue C. *The development of a patient-reported outcome measure for real-time symptom assessment in patients with endometriosis.*

Snijkers J, van den Oever W, Weerts Z, **Vork L**, Mujagic Z, Leue C, Hesselink M, Kruiemel J, Muris J, Bogie R, Masclee A, Jonkers D, Keszthelyi D. *Examining the optimal cut-off values of HADS, PHQ-9 and GAD-7 as screening instruments for depression and anxiety in Irritable Bowel Syndrome.*

Klaassen T, **Vork L**, Smeets FGM, Troost FJ, Kruimel JW, Leue C, Masclee AAM, Keszthelyi D. *The interplay between stress and fullness in functional dyspepsia and healthy controls: an exploratory experience sampling method study.*

Klaassen T, Smeets FGM, **Vork L**, Tack J, Talley NJ, Simren M, Aziz Q, Ford AC, Kruimel JW, Conchillo JM, Leue C, Masclee AAM, Keszthelyi D. *Psychometric evaluation of an experience sampling method-based Patient-Reported Outcome Measure in Functional Dyspepsia.*

Addendum

Summary

Nederlandse samenvatting

Impact paragraph

Dankwoord

List of publications

Curriculum vitae

Curriculum vitae

Lisa Vork was born on the 31st of March 1990 in Heemskerk, the Netherlands. She attended primary school in West-Knollendam and finished secondary school (VWO) at the Bertrand Russell College in Krommenie. In 2008, she started medical school at Maastricht University. After obtaining her medical degree in 2014, she started her PhD at the division of Gastroenterology-Hepatology of Maastricht University Medical Center, under the supervision of prof. dr. A.A.M. Masclee, dr. D. Keszthelyi, and dr. J.W. Kruimel.



The research during the PhD is presented in this thesis and was performed within the NUTRIM School of Nutrition and Translational Research in Metabolism. In March 2019, Lisa started working as a resident (ANIOS) at the department of Internal Medicine and Gastroenterology-Hepatology in Máxima Medical Center in Veldhoven. In October 2020, she started her residency (AIOS) Gastroenterology-Hepatology under supervision of prof. dr. A.A.M. Masclee and dr. J.W. Kruimel.

