Development and Validation of the Rome IV Diagnostic Questionnaire for Adults

Olafur S. Palsson,[1] William E. Whitehead,[1] Miranda A. L. van Tilburg,[1] Lin Chang,[2] William Chey,[3] Michael D. Crowell,[4] Laurie Keefer,[5] Anthony J. Lembo,[6] Henry P. Parkman,[7] Satish S. C. Rao,[8] Ami Sperber,[9] Brennan Spiegel,[10] Jan Tack,[11] Stephen Vanner,[12] Lynn S. Walker,[13] Peter Whorwell,[14] and Yunsheng Yang[15]

[1]University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; [2]Center for Neurobiology of Stress, David Geffen School of Medicine, University of California, Los Angeles, California; [3]Division of Gastroenterology, University of Michigan Health System, Ann Arbor, Michigan; [4]Mayo Clinic, Scottsdale, Arizona; [5]Feinberg School of Medicine, Northwestern University, Chicago, Illinois; [6]Harvard Medical School, Boston, Massachusetts; [7]Lewis Katz School of Medicine, Temple University, Philadelphia, Pennsylvania; [8]Digestive Health Center, Medical College of Georgia, Georgia Regents University, Augusta, Georgia; [9]Kibbutz Farod, D. N. Bikat Bet Hakerem, Israel; [10]Mt. Sinai Medical Center, Los Angeles, California; [11]KU Leuven Center for GI Research, Leuven, Belgium; [12]Division of Gastroenterology, Kingston General Hospital, Kingston, Ontario, Canada; [13]Department of Pediatrics, Adolescent Division, Vanderbilt University Medical Center, Nashville, Tennessee; [14]University Hospital of South Manchester, United Kingdom; and [15]Chinese PLA General Hospital, Beijing, China

The Rome IV Diagnostic Questionnaires were developed to screen for functional gastrointestinal disorders, serve as inclusion criteria in clinical trials, and support epidemiologic surveys. Separate questionnaires were developed for adults, children and adolescents, and infants and toddlers. For the adult questionnaire, we first surveyed 1162 adults without gastrointestinal disorders, and recommended the 90th percentile symptom frequency as the threshold for defining what is abnormal. Diagnostic questions were formulated and verified with clinical experts using a recursive process. The diagnostic sensitivity of the questionnaire was tested in 843 patients from 9 gastroenterology clinics, with a focus on clinical diagnoses of irritable bowel syndrome (IBS), functional constipation (FC), and functional dyspepsia (FD). Sensitivity was 62.7% for IBS, 54.7% for FD, and 32.2% for FC. Specificity, assessed in a population sample of 5931 adults, was 97.1% for IBS, 93.3% for FD, and 93.6% for FC. Excess overlap among IBS, FC, and FD was a major contributor to reduced diagnostic sensitivity, and when overlap of IBS with FC was permitted, sensitivity for FC diagnosis increased to 73.2%. All questions were understandable to at least 90% of individuals, and Rome IV diagnoses were reproducible in three-fourths of patients after 1 month. Validation of the pediatric questionnaires is ongoing.

Keywords: Sensitivity; Specificity; Functional Gastrointestinal Disorder; Irritable Bowel Syndrome; Functional Constipation; Functional Dyspepsia.

The Rome working teams, which are composed of clinical investigators and clinicians who are experts in the functional gastrointestinal disorders (FGIDs) that affect specific regions of the gut, devise diagnostic criteria for these disorders that are intended for use by other clinicians and researchers (not patients). Their goal is to make the criteria as sensitive and specific as possible, and to accomplish this they sometimes develop complex diagnostic algorithms and combine requirements for laboratory evaluations with symptom criteria. However, there is also a need for a questionnaire that translates the diagnostic criteria into questions that are understandable to most patients, in order to enable standardized diagnostic assessment of individuals. Consequently, the Rome Foundation appointed a Questionnaire Development Committee (QDC) of individuals with expertise in test development to develop a patient questionnaire that incorporates the Rome diagnostic criteria. These questions can be used as inclusion criteria in clinical trials, as case definitions in epidemiological surveys, or for clinic screening.

The mandate of the QDC was to develop the Rome IV Diagnostic Questionnaire for Adults based on the new Rome IV criteria and to assess its performance with respect to understandability by patients, test–retest reliability, concordance with independent diagnoses by experienced clinicians, and ability to discriminate patients with the 3 most common FGIDs, which are irritable bowel syndrome (IBS), functional constipation (FC), and functional dyspepsia (FD), from nonpatient controls recruited from the population. Two pediatric questionnaires were also developed by a QDC subcommittee, but testing of these is still underway and will not be described here.

The process of developing and validating the Rome IV Diagnostic Questionnaire for Adults consisted of 5 different project steps, and these will be described in sequence.

VALIDATION

Step 1. Survey The Normal Frequency of Occurrence of Gastrointestinal Symptoms in the General Population of the United States to Guide the Rome IV Working Teams in Defining an Abnormal Symptom Frequency

The symptoms that the Rome working teams identified as diagnostic of FGIDs are not pathognomonic for GI disease; they are symptoms such as abdominal pain, nausea, vomiting, and heartburn that also occur occasionally in healthy nonpatients. These symptoms are only considered clinically significant and indicative of an FGID if they occur at an abnormally high frequency. This observation has 2 important consequences for the design of symptom-based diagnostic questionnaires: (1) the response scales used for questions about the frequency of symptom occurrence must have small enough steps to capture clinically significant differences between individuals, and (2) the thresholds used to define what is an abnormal frequency of occurrence will vary for different symptoms because the normal frequency of occurrence of these symptoms differs.

To address this, the QDC first developed and validated new response scales for the Rome IV diagnostic questionnaire, with more response steps than the Rome III.[1] The committee then conducted a survey of a nationally representative sample of US adults in order to provide the Rome IV working teams with the data needed to set thresholds for identifying meaningful, clinically significant deviations from the normal frequency of occurrence of GI symptoms. A sample of 1665 US adults stratified by sex (50% males), age, race, and ethnicity was recruited by a market research firm, CINT USA, Inc. (Los Angeles, CA) to complete an Internet survey. After inconsistent responders (identified by 3 repeated survey questions) were eliminated, response sets from 1277 individuals were retained for analysis.

For the purpose of setting frequency thresholds for defining what should be considered abnormal, we identified the 90th percentile for all questions (males and females combined) and reported these to the committees as recommended thresholds. We reasoned that using these thresholds for diagnoses would result in no more than 10% of healthy subjects being misclassified as patients, and when these 90th percentile thresholds were combined for multiple symptoms used to diagnose a disorder, the specificity could be expected to be in excess of 90%. In the calculation of these cutoffs, we excluded subjects who reported having a prior medical diagnosis of upper GI diagnoses from analysis of upper GI symptom thresholds, and conversely excluded those reporting lower GI diagnoses from analysis of lower GI symptoms.

Figure 1 illustrates how these data were analyzed: The cardinal symptom defining IBS is the frequency of abdominal pain, and the figure shows a histogram of the frequency of all responses to the question on abdominal pain. If the threshold for clinical significance is set at once a week as the Rome IV Bowel Committee recommended, the proportion of these population controls who might be misclassified as IBS is only 6.7%, indicating that the specificity of this symptom criterion for IBS diagnosis is 93.3%. Figure 1 also shows how the specificity of the diagnosis would be impacted if an alternative threshold for clinical significance were selected, and it shows that abdominal pain or discomfort is reported significantly more frequently by females than by males. Because males have a lower prevalence of abdominal pain and discomfort than females, the 90th percentile for males could be set at 2 to 3 days per month rather than once a week, and the 90th percentile for females occurs once a week, as is the case for the combined sample. For simplicity, the QDC recommended that the frequency thresholds for IBS diagnosis should be based on the 90th percentile for women and men combined.

A summary report[2] on the distribution of symptom occurrence rates for all the Rome III symptoms was distributed to the Rome IV working teams together with recommendations for selecting frequency thresholds for diagnosis. This report is available as a supplement to this article. Most of the working teams adopted these suggestions.

Step 2. Development of the Rome IV Diagnostic Questionnaire for Adults

The 6 working teams that were tasked with updating the Rome diagnostic criteria for the FGIDs in each region of the GI tract were appointed in 2013 and were requested to complete draft documents, including revised diagnostic criteria by May 2014. These committees worked by e-mail and conference calls, and met together for the first time at a satellite meeting held in conjunction with Digestive Disease Week 2014. Previous to this, the QDC completed its survey of the base rates of symptom occurrence in the population and also revised and validated new, more sensitive response scales[1] for patients to use when reporting their symptoms. The QDC developed a draft of the Rome IV Diagnostic Questionnaire (R4DQ) and distributed the questions appropriate to each region of the GI tract to the chair and co-chair of the working teams immediately before 3-day meetings of all the committees in December 2014. The chair and co-chair of the QDC met with each committee to discuss how well the draft diagnostic questions embodied the revised diagnostic criteria.

The QDC's interactions with the working teams were iterative: When the working teams revised their diagnostic criteria based on feedback from the Rome Foundation's Editorial Committee, or the critiques submitted by outside reviewers, the QDC revised the draft diagnostic questions. After each revision to either the diagnostic criteria or the diagnostic questions, the QDC asked the committees to again review the diagnostic questions for consistency with the diagnostic criteria.

Translatability Assessment

When the QDC was confident that the diagnostic questionnaire was in a near-final form, it was submitted to the professional translation company, Transperfect Inc. (New

How Often Did You Have Discomfort or Pain Anywhere in Your Abdomen?

Figure 1. This figure shows a histogram of the frequency of occurrence of "discomfort or abdominal pain" in the general population after excluding subjects with physician-diagnosed lower gastrointestinal disorders. The *dotted vertical line* shows the recommended threshold that is ≥90th percentile for the combined male and female sample. MT, shows the 90th percentile for males; FT, the 90th percentile for females.
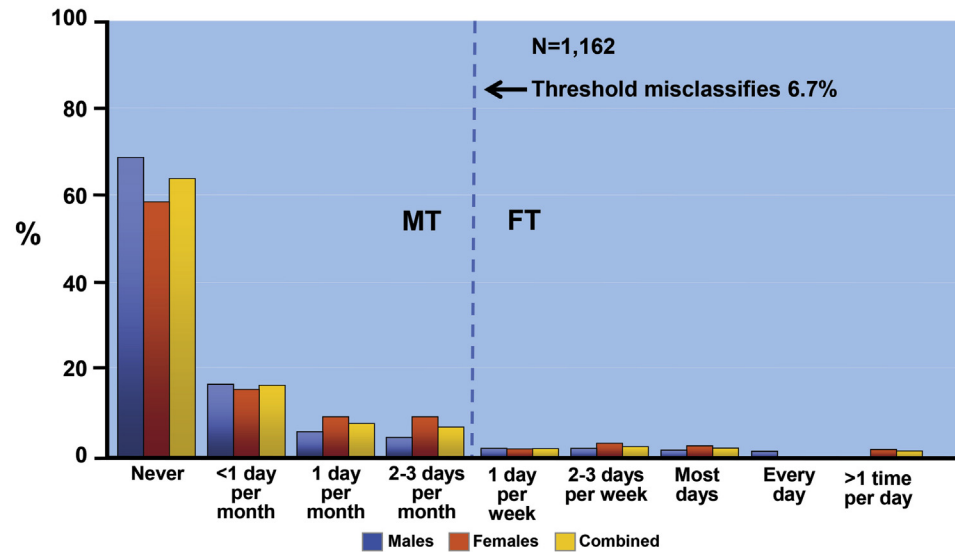
York, NY), for an assessment of the translatability into the following major languages: Arabic, French, German, Hindi, Italian, Japanese, Korean, Russian, and simplified Chinese. This translatability assessment resulted in no changes to the English version of the questionnaire, but did provide guidance that will be useful to the translators when the Rome IV questionnaire is translated into these and other languages.

Step 3: Understandability Assessment

In order to assess the understandability of the R4DQ, the QDC recruited a new US general population sample of 589 adult subjects stratified by sex, age, race, ethnicity, and years of education to complete an Internet survey. Each subject was asked to evaluate one-third (approximately 28) of the total questions. Subjects first answered each symptom question and were then asked (1) whether the question was difficult to understand (yes or no) and, if the question was difficult, (2) what about the question made it difficult to understand and (3) to give suggestions for alternative more understandable wording. On average, 1.5 questions of 28 were rated as difficult to understand, and neither older age nor fewer years of education were associated with the number of questions rated difficult to understand. Seven questions were rated difficult by 10% or more subjects, and these were revised to enhance their simplicity and clarity based on the suggestions of the subjects. The QDC concluded that all questions in the final form of the R4DQ are understandable to >90% of US adults, and that understandability is not influenced by age or education.

Final Reconciliation Between the Questionnaire and Working Team Criteria

After completing the understandability and translatability analyses and making minor changes to the

questionnaire, the chair and co-chair of the QDC committee again compared the questions with the Rome IV diagnostic criteria and developed scoring criteria for assigning provisional diagnoses based on questionnaire responses. This review identified a need for the working teams to make minor changes to the wording of the diagnostic criteria to ensure agreement with the questionnaire, and these suggestions were communicated to the working teams and incorporated into the Rome IV diagnostic criteria.

Step 4. Clinical Validation of the English Version of the Diagnostic Questionnaire

The preferred method for assessing the accuracy of a new diagnostic test is to compare classifications based on the new test to classifications based on an objectively measured biological marker that is known to be related to the pathophysiology of the disease. However, this is not possible for the FGIDs because there is no consensus on specific pathophysiological mechanisms for them, and no biomarkers exist that can identify the FGIDs with acceptable precision.[3] Another barrier to validation of symptom criteria for the FGIDs is that similar symptoms (eg, abdominal pain, nausea, vomiting, constipation, and diarrhea) are present in other diseases; examples are inflammatory bowel disease, gastroesophageal reflux disease, celiac disease, and GI cancers. Many view the exclusion of other disease explanations for symptoms as necessary for the diagnosis of the FGIDs.

In the absence of an objective reference standard for diagnosis of FGIDs, 2 models of validation have been used[4]: In the first model, the reference standard is a negative endoscopy, alternative imaging study, or laboratory test to exclude other diseases that could cause the symptoms, and the measure of diagnostic accuracy is whether the symptom criteria correctly identify patients who do not have an

alternative basis for their symptoms.[5,6] The second model uses positive diagnoses made by experienced clinicians as the reference standard, and the measure of diagnostic accuracy is whether the symptom criteria are concordant with the clinical diagnosis.[7,8] A third model is a hybrid of the first 2; for example, Vanner et al[9] first excluded from consideration any patient with alarm signs or symptoms suggestive of possible structural disease (eg, blood in stools or family history of GI cancer) and then examined the agreement of the Rome II symptom criteria with clinical diagnosis; he found that the positive predictive value of the Rome II criteria was 100%.

The QDC utilized a hybrid model for validation of the criteria for IBS and FD: To be included as a reference case of IBS, patients must have had a negative endoscopy within the last 5 years as well as a positive clinical diagnosis; and to be included as a reference case of FD, patients must have had a negative upper endoscopy within the last 5 years plus a positive clinical diagnosis. However, red flag signs or alarm symptoms, such as blood in stools or a family history of colon cancer, did not result in exclusion of the patient.

Aims of the Clinical Validation Study

The primary goal of the validation study was to test the sensitivity of the R4DQ for identifying patients who were diagnosed IBS, FC, or FD by clinicians. These reference clinical diagnoses were made by experienced clinicians before the patients completing the R4DQ, thus guaranteeing that the reference diagnoses were independent. In addition, a diagnosis of IBS required that the patient have a colonoscopy within the last 5 years, and a diagnosis of FD required that the patient have an upper endoscopy within the same time frame. Additional goals of this study were to assess the test−retest reliability of the R4DQ by having approximately 30 patients at each site complete the R4DQ a second time after a 30-day interval, and to examine the overlap between FGIDs diagnosed by the R4DQ.

Methods

The study was conducted at 9 clinical sites managed by academic gastroenterologists who are familiar with the FGIDs and with previous versions of the Rome criteria. Each site was asked to recruit 100−150 patients, including 25% with a clinical diagnosis of IBS, 25% with a clinical diagnosis of FC, 25% with a clinical diagnosis of FD, and 25% with other FGID diagnoses. The purpose of concentrating recruitment on IBS, FC, and FD was to insure that sufficient numbers of patients could be recruited to have adequate statistical power to test sensitivity. These are the most common FGID diagnoses made in clinical practice.

Inclusion/Exclusion Criteria

To participate, subjects were required to have been diagnosed with an FGID in a medical clinic, have personal access to an Internet-connected computer or tablet and be able to use it to answer questionnaires online; be able read and write English fluently; be at least 18 years old; have had GI symptoms for at least 6 months; and for the subsample of patients with IBS or FD clinical diagnosis, have had an endoscopy (colonoscopy for IBS, upper GI endoscopy for FD) with negative findings within the past 5 years. Individuals who had been diagnosed with the following organic health problems likely to affect GI symptoms were excluded from participation in the study: inflammatory bowel disease (Crohn's disease or ulcerative colitis), cancer anywhere in the GI tract, current infection of the GI tract, celiac disease, diabetes mellitus, and/or an eating disorder. Individuals who had undergone bariatric surgery or resection of any part of their bowels except appendix or gallbladder operations were also excluded from participation.

Enrollment and Questionnaire Completion

The research coordinator at each site reviewed the medical records for the past 3−6 months to retrospectively identify all patients assigned a clinical diagnosis of any FGID. At sites where this retrospective medical record review did not yield enough patients in all categories, prospective enrollment of new patients was permitted. Enrollment and data collection occurred between April 1 and September 15, 2015.

Research coordinators at the clinical sites sent e-mails or letters to eligible patients explaining the purpose of the study. Subjects were given the web address where they could complete the questionnaire and, to ensure that the database contained no information that could be used to identify them, they were assigned a randomly selected ID and a unique password to enter when they took the questionnaire. When potential study participants accessed the study website, they first reviewed a study description and recorded their consent before completing the online study questionnaire.

The questionnaire included the Rome IV Diagnostic Questionnaire (from 26 to 86 questions depending on skip patterns); the Rome III Diagnostic Questionnaire modules for IBS, FC, and FD (up to 27 questions); 6 demographic questions; and 6 questions about the frequency and types of medications used for GI symptoms. They were also asked about excluded diagnoses to confirm eligibility. Completion of the online questionnaires required 15 to 20 minutes, and patients received $25 for participation.

The research coordinator at each site had access on the study website to a password-protected study-management interface to view a list of the IDs of all patients from their site who had completed the questionnaire. The coordinators were encouraged to contact all subjects who had not completed the questionnaires within 2 weeks to remind them to do so.

Research coordinators entered the primary FGID clinical diagnosis (reference diagnosis) into the website interface for all subjects who completed the questionnaire. The coordinators also reviewed the medical records of questionnaire completers and abstracted data on all other GI diagnoses and relevant medical tests, such as endoscopy, transit study, gastric emptying study, esophageal pH study, esophageal manometry, barium enema or magnetic

resonance imaging of the colon, and fecal stool test for ova/parasites.

The research coordinators randomly selected a subset of questionnaire completers with clinical diagnoses of IBS, FD, and functional constipation and invited them to complete the questionnaire a second time 30 days after the first completion, for an additional payment of $25. Reminders to complete the repeat test were sent by e-mail as needed.

Data Analysis

For the primary analysis of diagnostic test sensitivity, the data from all clinical sites were pooled. Sensitivity and 95% confidence intervals for sensitivity were computed for each diagnosis (IBS, FC, and FD) separately. Sensitivity was defined as the proportion of all patients with a primary clinical diagnosis of the index disorder who fulfilled Rome IV criteria for the same diagnosis based on questionnaire responses. Patients receiving a secondary clinical diagnosis of the index disorder were excluded from the analysis due to a concern that clinicians may vary in the amount of time and effort devoted to secondary diagnoses in a busy clinic and may make secondary diagnoses of IBS or FD without endoscopy, thus causing secondary diagnoses to be less reliable.

Specificity, which is defined as the proportion of all patients without the index clinical diagnosis who also do not fulfill Rome IV questionnaire criteria for the index diagnosis, was not a goal of the analysis of data collected from the clinical sites for the following reasons: (1) The ability to discriminate patients with the index FGID from patients with other FGIDs is less relevant to the performance of the diagnostic questionnaire than is the ability to discriminate true cases from general population controls because it is known that there is a high degree of overlap between FGIDs in specialty medical clinics. (2) The design of the study was not appropriate for estimating the ability to discriminate the index disorder from other FGIDs even in the medical clinic because the patient sample was not representative of the clinic population; instead, patients were selectively recruited based on having established clinical diagnoses of specific FGIDs. Consequently, we estimated specificity of the Rome IV criteria from a large, representative population sample (see Three Country Population Survey), which was recruited specifically for this purpose.

Secondary sensitivity analyses were performed to assess the influence of symptom frequency thresholds and other specifics of the diagnostic criteria on test accuracy. Additionally, the degree of overlap between FGID diagnoses based on Rome IV criteria was calculated. In this analysis, all FGID diagnoses (primary and secondary) recorded in the medical record were included. Test−retest reliability was assessed for each of the 3 key FGID diagnoses in the study by computing percent agreement and κ statistics between 2 administrations of the questionnaire to the same subjects approximately 1 month apart, for each of the 3 primary diagnoses separately.

Results

A total of 881 patients enrolled and completed the study questionnaire across the 9 sites, but 38 were disqualified because they were found not to meet study criteria after enrollment, leaving 843 for analysis. Females comprised 76.3% of the 843 evaluable patients, and age ranged from 18 to 81 years (mean 43.6 years). Race/ethnicity distribution was 85.3% white, 4.3% black, 4.5% Asian, and 6.0% other or not disclosed. Among US patients, 8.2% reported Hispanic ethnicity. Table 1 shows the frequency of all FGID clinical diagnoses (not Rome IV diagnoses based on the questionnaire) found in the medical records of the 843 patients in the analysis sample.

Sensitivity of the Rome IV Criteria

Columns 1−3 in Table 2 show the sensitivity statistics for the 3 most prevalent FGIDS: IBS, FC, and FD. (The data on specificity in columns 4−5 come from the Three Country Population Survey and will be discussed later.)

Irritable bowel syndrome: sensitivity analysis. The sensitivity of the Rome IV criteria for IBS is comparable with levels previously reported for Rome III,[4] but is suboptimal. The QDC examined the impact on diagnostic sensitivity of each of the 3 changes made to the Rome III criteria:

1. The threshold frequency of abdominal pain was made more stringent, changing from 2−3 days per month in Rome III to at least once a week in Rome IV. If the frequency of abdominal pain was relaxed to 2−3 days

Table 1. Frequency of All Clinical Functional Gastrointestinal Disorder Diagnoses in the Sample of 843 Gastrointestinal Patients, Including Multiple Diagnoses for the Same Patient

FGID diagnosis	No. of cases	% of Total sample
Irritable bowel syndrome	442	52.4
Functional constipation	235	27.9
Functional dyspepsia	153	18.1
Functional bloating	96	11.4
Functional diarrhea	79	9.4
Functional abdominal pain	71	8.4
Functional fecal incontinence	55	6.5
Chronic idiopathic nausea	34	4
Dysphagia	13	1.5
Globus	11	1.3
Functional heartburn	11	1.3
Functional vomiting	9	1.1
Unspecified belching	7	0.8
Cyclic vomiting	7	0.8
Functional chest pain	5	0.6
Proctalgia fugax	3	0.4
Rumination syndrome	2	0.2
Sphincter of Oddi dysfunction	2	0.2
Chronic proctalgia	2	0.2
Aerophagia	1	0.1
Levator ani	1	0.1

VALIDATION

Table 2. Sensitivity and Specificity of the 3 Most Common Functional Gastrointestinal Disorders

Diagnosis	Clinical validation sample		Population control sample	
	Sensitivity, %	95% CI	Specificity, %	95% CI
IBS	62.7	57.8–67.6	97.1	96.6–97.6
FD	54.7	46.3–63.1	93.3	92.5–94.0
FC excluding OIC and IBS	33.9	27.0–40.8	94.5	93.9–95.2
FC including OIC and IBS	70.5	63.8–77.2	93.1	92.5–93.9

CI, confidence interval.

per month and all other Rome IV criteria remained the same, this would increase the sensitivity of Rome IV to 69.9%.

2. "Discomfort" was dropped from the key qualifying symptom, "abdominal discomfort or pain"; Rome IV requires "abdominal pain." However, this more restrictive phrasing did not reduce diagnostic sensitivity: The proportion of patients with a primary clinical diagnosis of IBS who reported "abdominal discomfort or pain" at least weekly on the Rome III pain question was similar to the proportion who reported "abdominal pain" at least weekly on the equivalent Rome IV question (74.9% vs 78.4% respectively).

3. Rome III required that pain or discomfort improve after defecation and that changes in stool frequency and/or consistency occur "when the pain started." By contrast, Rome IV requires only that pain and defecation are "associated" in time or that changes in stool consistency and frequency are "associated" with abdominal pain. However, this change did not have a significant impact on the sensitivity of the Rome criteria.

To understand the causes of misclassifications in IBS, we examined which Rome IV diagnoses were assigned by the questionnaire to the 140 patients who received a clinical diagnosis of IBS but did not meet Rome IV criteria for IBS. The most common Rome IV diagnoses were FC (35 of 140), FD (34 of 140), functional diarrhea (22 of 140), and levator ani syndrome (20 of 140). We also tabulated the clinical diagnoses assigned to 192 patients who met the Rome IV criteria for IBS but did not receive a clinical diagnosis of IBS by their physician. The most common clinical diagnoses (either primary or secondary diagnosis) were FC (95 of 192), FD (58 of 192), functional bloating (27 of 192), functional abdominal pain (25 of 192), and functional diarrhea (17 of 192). These data show that the core symptoms that combine to diagnose IBS are the ones accounting for most misclassifications, and they suggest that clinicians likely base their clinical diagnoses on the predominant or most bothersome symptom when patients present with the constellation of constipation, diarrhea, abdominal pain, and bloating. A significant number of misclassifications also result from the overlap of FD with IBS.

Functional dyspepsia: sensitivity analysis. The Rome IV criteria for unspecified FD had a sensitivity of 54.7% (Table 2). Sensitivity was defined as the proportion of patients with a physician primary diagnosis of FD who fulfilled Rome IV criteria for unspecified FD, a Rome IV diagnosis that requires that the patient meet symptom criteria for either postprandial distress syndrome or epigastric pain syndrome. Only patients who had an upper endoscopy were included in this analysis. It was not possible to examine the sensitivity of the questionnaire for diagnosing postprandial distress syndrome and epigastric pain syndrome because clinicians usually did not distinguish between these subtypes of FD.

To understand the causes of misclassifications in FD, we examined the other Rome IV diagnoses assigned to 63 clinic patients who received a clinical diagnosis of FD, but did not meet Rome IV criteria for FD. The most common Rome IV diagnoses were unspecified functional bowel disorder (20 of 63) and IBS (19 of 63). To further explore the causes of misclassification, we examined the medical diagnoses assigned to 308 patients who met the Rome IV criteria for FD, but did not receive a clinical diagnosis of FD by their physician. The most common clinical diagnoses (either primary or secondary diagnosis) were IBS (172 of 308), FC (102 of 308), functional bloating (37 of 308), and functional abdominal pain (32 of 308). These analyses suggest that patients in whom FD symptoms overlap the symptoms of IBS or constipation are the ones more likely to be misclassified.

Functional constipation: sensitivity analysis. The Rome IV criteria for FC require that patients meeting symptom criteria for opioid-induced constipation (OIC) and IBS be excluded from the diagnosis of FC. When this is done, the sensitivity of the Rome IV criteria for identifying clinically diagnosed FC is only 33.9%, which indicates that clinicians are classifying many more patients as FC than are being classified FC by the Rome IV criteria. However, when the Rome IV criteria were relaxed to permit patients with OIC and IBS to be classified as FC if they met the other FC criteria (ie, when comorbid diagnoses were permitted), the sensitivity increased to 70.5% (Table 2).

To understand the causes of misclassifications in FC, we examined which Rome IV diagnoses were assigned by the questionnaire to the 121 patients who received a clinical diagnosis of FC but did not meet Rome IV criteria for FC. The most common Rome IV diagnoses were IBS (84 of 121), FD

(67 of 121), proctalgia fugax (28 of 121), and levator ani syndrome (21 of 121). We also tabulated the clinical diagnoses assigned to 55 patients who met the Rome IV criteria for FC but did not receive a clinical diagnosis of FC by their physician. The most common clinical diagnoses (either primary or secondary diagnosis) were IBS (39 of 55), FD (12 of 55), and functional bloating (10 of 55). These data show that two-thirds of the misclassifications are due to the overlap of IBS with symptoms of FC, and the requirement that a patient meeting criteria for IBS cannot be assigned a diagnosis of FC.

We infer from this that the clinicians in our study were, for the most part, not behaving as if a diagnosis of IBS or OIC rules out a clinical diagnosis of FC. Such behavior is consistent with statements in the Rome IV Bowel Disorders chapter that FC and constipation-predominant IBS frequently overlap and should be seen as parts of a spectrum rather than as distinct disorders. This view is also supported by recent studies showing that the symptoms of FC and constipation-predominant IBS frequently overlap[10,11] and, over time, many patients transition back and forth between these 2 diagnostic categories.[11] Therefore, a major cause of low sensitivity for the FC criteria is that the Rome IV criteria do not permit FC to overlap with IBS-C and OIC.

Overlap of Rome IV Diagnoses of Irritable Bowel Syndrome, Functional Dyspepsia, and Functional Constipation

Table 3 shows the overlap between Rome IV diagnoses in the clinical validation study. This is a complex table because it shows the prevalence of each diagnosis in patients with each of the other 2 diagnoses and enables the reader to compare this overlap with the prevalence of the separate diagnoses. For example, the cell in the second row of the third column shows that 66.9% (257 of 384) patients with FD have IBS, which is greater than the prevalence of IBS in the whole sample (50.7% [427 of 843 patients]). Conversely, the cell in the third row of the second column

shows that 60.2% (257 of 427) of patients with IBS also have FD, which is greater than the prevalence of FD in the total sample (45.6% [384 of 843 patients]). The degree of overlap among the other combinations of diagnoses shows a similar excess overlap. This excess overlap in the symptoms of IBS, FD, and FC is consistent with the sensitivity analyses shown previously that identified the overlap among the symptoms of these disorders as a factor that consistently contributes to lower sensitivity for the diagnostic criteria.

Test−Retest Reliability of the Rome IV Diagnostic Questionnaire for Adults

A subsample of the clinic patients completed retests with the diagnostic questionnaire approximately 1 month after the first administration. Percentage agreement between the 2 administrations was assessed with regard to the 3 key diagnoses. Only data from the 140 patients who completed the questionnaire the second time within a range of 20−40 days after the first one were included in the analysis. Mean time interval between administrations was 31.1 days. The agreement between the first and second questionnaire administration for the diagnoses was 75.7% for IBS ($\kappa = 0.51$), 76.4% ($\kappa = 0.53$) for FD, and 79.2% ($\kappa = 0.44$) for FC (κ values in the range from 0.41 to 0.60 are considered indicative of moderate agreement).[12]

Step 5. Assessment of Specificity, Prevalence, and Overlap of Functional Gastrointestinal Disorder Diagnoses in the Three Country Population Survey

Study Aims

The primary aims of this survey were to estimate the specificity of the Rome IV criteria for distinguishing patients with the 3 most common FGID diagnoses (IBS, FC, and FD) from healthy population controls, to estimate the prevalence in the population of the FGIDs based on Rome IV criteria,

Table 3. Overlap Between Rome IV Diagnoses of Irritable Bowel Syndrome, Functional Dyspepsia, and Functional Constipation in the Clinic Validation Study (Irritable Bowel Syndrome and Opioid-Induced Constipation Were Not Excluded From Functional Constipation in This Table)

	IBS (n = 427)	FD (n = 384)	FC (n = 287)	Total sample (N = 843)
IBS	—	66.9[a] (257/384)	56.0[a] (161/287)	50.7[a] (427/843)
FD	60.2[b] (257/427)	—	52.3[a] (150/287)	45.6[a] (384/843)
FC	37.7[b] (161/427)	39.0[b] (150/384)	—	34.0[a] (287/843)
Total sample	50.7[b] (427/843)	45.6[b] (384/843)	34.0[b] (287/843)	843

Note: Data are presented as % (n/N).

[a]Percent of patients with the diagnosis in the column heading who also have the diagnosis in the row heading, for example, in the second row, third column, 66.9% (257/384) patients with FD also had IBS. This compares with an overall prevalence of IBS in the total clinical validation sample (column 5) of 50.7%.

[b]Percent of patients with the diagnosis in the row heading who also have the diagnosis in the column heading, for example, in the third row, second column, 60.2% (257/427) patients with IBS also had FD, compared with a prevalence of FD in the whole sample of 45.6% (row 5).

and to assess the overlap of Rome IV diagnoses of IBS, FC, and FD.

Methods

Qualtrics Inc. (Provo, UT), a global survey research company, was commissioned to identify nationally representative general population samples of adults based on specific demographic quotas in the United States, English-speaking Canada, and the United Kingdom. Recruitment was stratified based on sex, age, years of education, race, and, for the US sample, Hispanic ethnicity. Qualtrics directed interested subjects to the study website at the University of North Carolina at Chapel Hill for survey completion.

The survey included up to 86 R4DQ questions (depending on skip pattern); up to 27 questions from the IBS, FC, and FD modules of the Rome III Diagnostic Questionnaire; 15 questions from the Physical Health Questionnaire[13]; 8 questions from the SF-8 Health Survey[14]; 15 demographic questions; 8 questions on medical history; 4 questions on access to health care; 1 question on diet; and 3 questions on psychological distress related to GI symptoms. The median completion time for the survey was 19 minutes.

A total of 6300 respondents completed the survey in the 3 countries, 2100 in each country. Subjects who self-reported on the questionnaire that they had been diagnosed with inflammatory bowel disease, celiac disease, or GI cancer, or who had undergone GI resection except gallbladder or appendix, were excluded from analysis. An additional 369 (5.9%) who failed data-quality checks were also excluded, leaving 5931 valid response sets for statistical analysis. The demographic characteristics of this analysis sample are presented in Table 4.

Data Analysis

A primary aim for the Three Country Population Survey was to estimate the specificity of the Rome IV criteria for distinguishing patients with the 3 primary FGID diagnoses (IBS, FC, and FD) from healthy population controls. Those results are presented in Table 2. For each index diagnosis of an upper GI disorder, we first eliminated subjects who reported a relevant upper GI medical diagnosis, and then calculated the proportion of these nonpatient population controls who did not fulfill the Rome IV diagnostic criteria for the index disorder. Similarly, for each index diagnosis of a lower GI disorder, we eliminated subjects who reported a relevant lower GI medical diagnosis and then calculated the proportion of these nonpatient population controls who did not fulfill the Rome IV diagnostic criteria for the index lower GI FGID. The prevalence estimates (as distinct from specificity estimates) were performed on all available subjects, including those with self-reported medical GI diagnoses. These estimates of specificity and prevalence were performed on the pooled samples from all three countries. A third aim of the Three Country Population Survey was to estimate the overlap between the 3 most common FGIDs (IBS, FD, and FC) in the general population. For these analyses, subjects with IBS could also be diagnosed with FC.

Table 4. Demographic characteristics of the analysis sample from the Rome IV Three Country Population Survey

Characteristic	United States	Canada	United Kingdom	Overall
Sample size (valid responders)	1949	1988	1994	5931
Sex, %				
Female	49.4	49.3	48.9	49.2
Male	50.6	50.7	51.1	50.8
Age groups, %				
18–39 y	38.7	39.2	38.9	38.9
40–64 y	40.5	40.1	40.4	40.4
65+ y	20.9	20.6	20.7	20.7
Years of education, %				
<13	34.6	33.8	37.1	35.2
13–16	49.2	44.6	42.0	45.2
>16	16.2	21.6	20.9	19.6
Race/ethnicity, %				
Asian	1.9	9.1	3.1	4.7
Black	20.2	1.8	1.2	6.6
White	56.7	75.3	83.8	72.0
Hispanic (United States only)	19.1			
Other/mixed	2.1	4.0	1.8	2.6
Missing race information[a]	0.0	9.9[a]	10.2[a]	6.7[a]

[a]Race/ethnicity data were collected in the Canadian and UK samples only after the initial pilot sampling that amounted to 10% of the total final sample.

Results

Estimates for the population prevalence of the FGIDs and the estimated specificity of a Rome IV diagnosis of each FGID are given in Table 5.

Overlap Between Irritable Bowel Syndrome, Functional Constipation, and Functional Dyspepsia

Table 6 shows the overlap among the 3 most common FGIDs. The overlap among all pairs of FGID diagnoses is 3–6 times the prevalence of these diagnoses in the population.

Development and Validation of the Pediatric Rome IV Diagnostic Questionnaires

In contrast to the adult questionnaire, where one version is used to assess all patients, several different assessments are available for the pediatric population. The Questionnaire on Pediatric Gastrointestinal Symptoms was first developed to measure Rome II criteria in children and adolescents (aged 4–18 years), and updates were made for Rome III. Questionnaires for both parents and children (aged 10 years and older) were developed, and evidence exists for their reliability and validity.[15–18] In addition, an infant/toddler version was developed and tested recently, based on parental report of infant/toddler symptoms.[19] The mandate of the pediatric subcommittee of the QDC was to update

Table 5. Population Prevalence of the Rome IV Functional Gastrointestinal Disorders in the Three Country General Population Survey

Diagnosis	Subjects, n	Prevalence, %	Specificity, % (95% CI)
Functional dysphagia	266	4.5	97.1 (96.7–97.6)
Functional heartburn	107	1.8	99.0 (98.7–99.3)
Reflux hypersensitivity	85	1.4	99.3 (99.1–99.6)
Globus sensation	61	1.0	99.1 (98.8–99.3)
Functional chest pain	59	1.0	99.2 (99.0–99.5)
Excessive belching	50	0.8	99.6 (99.4–99.8)
FD	551	9.3	93.3 (92.5–94.0)
Postprandial distress syndrome	454	7.7	94.2 (93.5–94.9)
Epigastric pain syndrome	212	3.6	98.0 (97.6–98.4)
Rumination	202	3.4	97.6 (97.2–98.1)
Chronic nausea and vomiting	80	1.3	99.0 (98.8–99.3)
Cyclic vomiting	71	1.2	99.2 (98.9–99.4)
Cannabinoid hyperemesis syndrome	10	0.2	99.9 (99.8–100.0)
Functional biliary pain	12	0.2	99.9 (99.8–100.0)
Central abdominal pain syndrome	1	0	—
Functional bowel disorder unspecified	752	12.7	88.4 (87.9–89.3)
FC excluding IBS and OIC	374	6.3	94.5 (93.9–95.2)
FC with IBS and OIC	525	8.9	93.2 (92.5–93.9)
IBS	341	5.7	97.1 (96.6–97.6)
OIC	88	1.5	99.2 (98.9–99.5)
FD	323	5.4	95.3 (94.7–95.9)
Functional bloating/distention	51	0.9	99.3 (99.1–99.6)
Proctalgia fugax	318	5.4	96.3 (95.8–96.9)
Levator ani syndrome	101	1.7	99.0 (98.7–99.3)
Fecal incontinence	196	3.3	98.0 (97.6–98.4)

CI, confidence interval.

these questionnaires based on the Rome IV criteria. Reliability and validity testing is currently ongoing, therefore, the current article will only describe the development of the questionnaire itself.

The committee followed procedures similar to those described here for the Rome IV Diagnostic Questionnaire for Adults, but with important exceptions: First, the new response scales developed for the adult questionnaire were not used in the pediatric questionnaires because they are not appropriate for the cognitive developmental level of children. For example, percentages can be difficult to understand and apply by children as young as age 10 years old. Although the new response scales would be appropriate for parental report, the committee desired to keep response

Table 6. Overlap Between Irritable Bowel Syndrome, Functional Dyspepsia, and Functional Constipation in the General Population (Irritable Bowel Syndrome and Opioid-Induced Constipation Not Excluded From Functional Constipation in This Table)

	IBS (n = 341)	FD (n = 551)	FC (n = 525)	Total sample (N = 5931)
IBS	—	31.6[a] (174/551)	20.0[a] (105/525)	5.7[a] (341/5931)
FD	51.0[b] (174/341)	—	23.0[a] (121/525)	9.3[a] (551/5931)
FC	30.8[b] (105/341)	22.0[b] (121/525)	—	8.9[a] (525/5931)
Total sample	5.7[b] (341/5931)	9.3[b] (551/5931)	8.9[b] (525/5931)	5931

Note: Data are presented as % (n/N).
[a]Percent of patients with the diagnosis in the column heading who also have the diagnosis in the row heading, for example, in the second row, third column, 31.6% (174/551) patients with FD also had IBS. This compares with an overall prevalence of IBS in the total sample (column 5) of 5.7%.
[b]Percent of patients with the diagnosis in the row heading who also have the diagnosis in the column heading, for example, in the third row, second column, 51.0% (174/341) patients with IBS also have FD, compared to a prevalence of FD in the whole sample of 9.3% (row 5).

VALIDATION

scales similar for parents and children so that they can be compared in research and clinical practice. Secondly, because fewer resources were available for the validation of the 2 pediatric questionnaires, current validation efforts focus on comparison of questionnaire diagnoses with clinical diagnoses, and the prevalence of Rome IV symptoms in a normal sample of US adults. These will be presented in another publication as data become available.

Discussion

The studies presented here show that the Rome IV Diagnostic Questionnaire for Adults has adequate sensitivity and excellent specificity for IBS and FD diagnoses: for IBS, sensitivity was 62.7% and specificity was 94.5%; and for FD, sensitivity was 54.7% and specificity was 93.3%. For FC, on the other hand, the sensitivity of the diagnostic questionnaire was inadequate: sensitivity was 33.9% and specificity was 93.6%. The poor performance of the diagnostic questionnaire for FC was related to the Rome IV requirement that patients meeting criteria for IBS or OIC be excluded from the diagnosis of FC. When these exclusions were removed, 70.5% of the patients given a clinical diagnosis of FC were identified by the questionnaire.

We also showed that the questions on this diagnostic questionnaire are understandable to at least 90% of US adults and that comprehension of the questions is not significantly affected by older age or by limited education. Diagnoses of IBS, FC, and FD based on the questionnaire were reliable for three-fourths of clinic patients during a 30-day interval.

A unique strength of this series of studies was the inclusion of secondary analyses to identify the sources of misclassifications by the diagnostic questionnaire: A major contributor to reduced sensitivity was revealed to be the overlap of symptoms related to IBS, FC, and FD. In particular, these sensitivity analyses suggest that FC and constipation-predominant IBS are on a continuum. This is consistent with recent research[10,11] and with the opinions expressed by the Rome IV Bowel Committee, although it is not reflected in the Rome IV diagnostic criteria for IBS and FC.

The sensitivity of the Rome IV Diagnostic Questionnaire for Adults for identifying patients with IBS was also influenced by the threshold frequency of abdominal pain required to make the diagnosis. The Rome IV criteria for IBS appear less sensitive than the Rome III criteria because Rome IV requires abdominal pain at least once a week, and Rome III required abdominal discomfort or pain only 2–3 days per month.

This series of studies has several limitations:

1. The symptoms of the FGIDs fluctuate over time,[20] and there was a lapse of unknown duration between the clinical diagnosis and the time the questionnaire was completed. Treatment occurred during this interval, and might have reduced the patient's symptoms. Test–retest agreement over 1 month was about 75%, which could be regarded as the upper limit of

concordance between the clinical diagnosis and the diagnostic questionnaire.

2. Another reason perfect agreement between clinical diagnosis and a questionnaire-based diagnosis would not be expected is that clinicians are able to consider additional sources of information, such as physiological tests, medical history, and family history, while the questionnaire-based diagnosis relies exclusively on the patient's self-reported symptoms.

3. No effort was made to define the minimum diagnostic workup at the participating clinical sites except for the requirement that upper endoscopy was required to diagnose FD and colonoscopy was required to diagnose IBS. Variations in practices between centers might have resulted in heterogeneity of the reference diagnostic groups.

4. Some Rome IV diagnoses now require physiological tests (eg, disordered defecation), exclusion of structural disease (eg, functional esophageal disorders), physical examination findings (eg, tenderness in levator ani syndrome), or physician judgment (eg, whether abdominal pain has a central or a peripheral cause in central abdominal pain syndrome) in addition to symptoms; these disorders cannot be diagnosed by self-reported symptoms on a questionnaire. Thus, not all of the Rome IV classification system can be validated by the study design used here.

Despite these limitations, the studies described here show that the R4DQ for adults has adequate sensitivity and excellent specificity for the diagnosis of IBS, FD, and many of the other FGIDs. The R4DQ is understandable to 90% of patients, and diagnoses based on the questionnaire show good test–retest reliability. The Diagnostic Questionnaire is also translatable to other languages, and this should make it possible to carry out cross-cultural and global epidemiologic studies. Future studies should allow, rather than arbitrarily excluding, the overlap of symptoms of IBS, FC, and FD and should explore the implications of this overlap for diagnostic classification and treatment.

References

1. Whitehead WE, Palsson OS, Cascade E. Validation of Response Scales for ROME Diagnostic Questionnaire. Gastroenterology 2013;144(Suppl 1). S-916.

2. Whitehead WE, Palsson OS. Report on the Rome III Normative Gastrointestinal Symptom Survey. Chapel Hill, NC: University of North Carolina, 2013.

3. Spiller RC. Potential biomarkers. Gastroenterol Clin North Am 2011;40:121–139.

4. Whitehead WE, Drossman DA. Validation of symptom-based diagnostic criteria for irritable bowel syndrome: a critical review. Am J Gastroenterol 2010;105:814–820; quiz 813, 821.

5. Ford AC, Talley NJ, Veldhuyzen van Zanten SJ, et al. Will the history and physical examination help establish that

irritable bowel syndrome is causing this patient's lower gastrointestinal tract symptoms? JAMA 2008;300: 1793–1805.

6. Sood R, Gracie DJ, Law GR, et al. Systematic review with meta-analysis: the accuracy of diagnosing irritable bowel syndrome with symptoms, biomarkers and/or psychological markers. Aliment Pharmacol Ther 2015;42:491–503.

7. Whitehead WE, Palsson OS, Feld AD, et al. Utility of red flag symptom exclusions in the diagnosis of irritable bowel syndrome. Aliment Pharmacol Ther 2006;24:137–146.

8. Whitehead WE. Development and validation of the Rome III diagnostic questionnaire. In: Drossman DA, Corazziari E, Delvaux M, et al, eds. Rome III: The Functional Gastrointestinal Disorders. McLean, VA: Degnon Associates, 2006:835–853.

9. Vanner SJ, Depew WT, Paterson WG, et al. Predictive value of the Rome criteria for diagnosing the irritable bowel syndrome. AmJ Gastroenterol 1999;94:2912–2917.

10. Heidelbaugh JJ, Stelwagon M, Miller SA, et al. The spectrum of constipation-predominant irritable bowel syndrome and chronic idiopathic constipation: US survey assessing symptoms, care seeking, and disease burden. Am J Gastroenterol 2015;110:580–587.

11. Wong RK, Palsson OS, Turner MJ, et al. Inability of the Rome III criteria to distinguish functional constipation from constipation-subtype irritable bowel syndrome. Am J Gastroenterol 2010;105:2228–2234.

12. Viera AJ, Garret JM. Understanding interobserver agreement: the Kappa statistic. Fam Med 2005;37:360–363.

13. Kroenke K, Spitzer RL, Williams JB. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. Psychosom Med 2002;64:258–266.

14. Ware JE, Kosinski M, Dewey JE, et al. How to Score and interpret Single-Item Health Status Measures: A Manual for Users of the SF-8 Health Survey. Lincoln, RI: QualityMetric Inc., 2001.

15. van Tilburg MA, Squires M, Blois-Martin N, et al. Test of the child/adolescent Rome III criteria: agreement with physician diagnosis and daily symptoms. Neurogastroenterol Motil 2013;25. 302–e246.

16. Baber KF, Anderson J, Puzanovova M, et al. Rome II versus Rome III classification of functional gastrointestinal disorders in pediatric chronic abdominal pain. J Pediatr Gastroenterol Nutr 2008;47:299–302.

17. Chogle A, Dhroove G, Sztainberg M, et al. How reliable are the Rome III criteria for the assessment of functional gastrointestinal disorders in children? Am J Gastroenterol 2010;105:2697–2701.

18. Helgeland H, Flagstad G, Grotta J, et al. Diagnosing pediatric functional abdominal pain in children (4-15 years old) according to the Rome III Criteria: results from a Norwegian prospective study. J Pediatr Gastroenterol Nutr 2009;49:309–315.

19. van Tilburg MA, Rouster A, Silver D, et al. Development and validation of a Rome III questionnaire for infant and toddler functional gastrointestinal disorders. J Pediatr Gastroenterol Nutr 2016;62:384–386.

20. Palsson OS, Baggish J, Whitehead WE. Episodic nature of symptoms in irritable bowel syndrome. Am J Gastroenterol 2014;109:1450–1460.

VALIDATION