Design of Treatment Trials for Functional Gastrointestinal Disorders

E. Jan Irvine,[1,2,*] **Jan Tack**,[3,*] Michael D. Crowell,[4] Kok Ann Gwee,[5] Meiyun Ke,[6] Max J. Schmulson,[7] William E. Whitehead,[8] and Brennan Spiegel[9]

[1]Department of Medicine, University of Toronto, Toronto, Ontario, Canada; [2]Li Ka Shing Knowledge Institute and Department of Medicine, St Michael's Hospital, Toronto, Canada; [3]Departments of Clinical and Experimental Medicine and Gastroenterology, Translational Research Center for Gastrointestinal Disorders, University Hospital KU Leuven, Leuven, Belgium; [4]Division of Gastroenterology and Hepatology, Mayo Clinic, Scottsdale, Arizona; [5]Yong Loo Lin School of Medicine, National University of Singapore, Singapore; [6]Peking Union Medical College Hospital, Center of FGID and MGID, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing, China; [7]Facultad de Medicina, Universidad Nacional Autónoma de México, Laboratorio de Hígado, Páncreas y Motilidad, Unidad de Investigación en Medicina Experimental, Hospital General de México, Mexico City, Mexico; [8]University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; and [9]Cedars-Sinai Health System, Cedars-Sinai Center for Outcomes Research and Education, Los Angeles, California

This article summarizes recent progress and regulatory guidance on design of trials to assess the efficacy of new therapies for functional gastrointestinal disorders (FGIDs). The double-masked, placebo-controlled, parallel-group design remains the accepted standard for evaluating treatment efficacy. A control group is essential, and a detailed description of the randomization process and concealed allocation method must be included in the study report. The control will most often be placebo, but for therapeutic procedures and for behavioral treatment trials, respectively, a sham procedure and control intervention with similar expectation of benefit, but lacking the treatment principle, are recommended. Investigators should be aware of, and attempt to minimize, expectancy effects (placebo, nocebo, precebo). The primary analysis should be based on the proportion of patients in each treatment arm who satisfy a treatment responder definition or a prespecified clinically meaningful change in a patient-reported outcome measure. Data analysis should use the intention-to-treat principle. Reporting of results should follow the Consolidated Standards for Reporting Trials guidelines and include secondary outcome measures to support or explain the primary outcome and an analysis of harms data. Trials should be registered in a public location before initiation and results should be published regardless of outcome.

Keywords: Functional Gastrointestinal Disorders; Controlled Trial; Patient-Reported Outcome Measure; Intention to Treat.

Clinical trial design for functional gastrointestinal disorders (FGIDs) is hampered by several factors, including symptom variability between subjects or groups and within subjects over time and the lack of specific biomarkers. The Rome diagnostic criteria and design recommendations are now routinely applied in clinical treatment trials. Since the publication of the Rome III guidance, there have been substantial advances in several aspects of clinical trial design. The expectations for patient-reported outcome (PRO) measurement have undergone major changes with the dissemination of regulatory guidelines for PROs from the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA).[1–3] Accumulating data also provide new insights for measuring common FGID symptoms, such as abdominal pain, discomfort, diarrhea, urgency, constipation, and bloating, among others. New information about the placebo, "nocebo," and "precebo" responses also challenges researchers to consider the biases inherent in FGID trials. In addition, advances in pragmatic clinical trial (PCT) design offer new approaches to measuring the effectiveness of FGID therapies in the context of everyday clinical practice. This updated Rome IV chapter now addresses each of these new trends, provides guidance for investigators seeking to develop and conduct FGID clinical trials, and emphasizes evolving concepts about how best to test the risks and benefits among the full range of FGID treatments.

Identifying the Hypotheses and Research Questions

The first task is to establish the hypothesis of the putative effect of the studied treatment, based on its expected mechanism of action, which generates the specific research question(s) for the proposed trial. As multiple factors contribute to the pathogenesis of FGIDs, it is likely that no single therapeutic approach will fully abolish all symptoms.

*Authors share co-first authorship.

Abbreviations used in this paper: BSFS, Bristol Stool Form Scale; CONSORT, Consolidated Standards for Reporting Trials; EMA, European Medicines Agency; FDA, US Food and Drug Administration; FGID, functional gastrointestinal disorder; HRQOL, health-related quality of life; IBS, irritable bowel syndrome; IBS-C, irritable bowel syndrome with constipation; IBS-D, irritable bowel syndrome with diarrhea; IBS-M, irritable bowel syndrome with constipation and diarrhea; MCID, minimally clinical important difference; PCT, pragmatic clinical trials; PRO, patient-reported outcome; RCT, randomized controlled trial.

Most current article

0016-5085/$36.00
http://dx.doi.org/10.1053/j.gastro.2016.02.010

Table 1. Goals of a Treatment Trial

To ascertain the ability of the intervention to
Relieve symptoms or decrease symptom severity
Improve functional health status and health-related quality of life
Improve ability to cope with symptoms
Decrease use of health care resources
Avoid harm and be cost-effective

Most FGID intervention studies evaluate the impact of a treatment on the items listed in Table 1, but specific goals can vary widely. Investigators should prioritize their research question(s) pertinent to the specific FGID, develop a hypothesis based on available evidence, and design a study that most effectively answers the research question(s).

In general, the primary question will address whether the study treatment improves FGID symptoms. Consequently, the primary outcome measurement tools must include reporting of the most important symptoms expected to change with the proposed treatment. The secondary questions are best determined by the particular disorder, that is, its specific symptoms and the mechanism of action of the treatment. Pathophysiological factors, while important explanatory parameters, should be considered secondary rather than primary end points.

Defining the Target Condition

Patient Population

A screening log of key variables is mandatory in order for readers to judge the generalizability of the results. The log should include demographic (eg, age, sex, and race) and clinical variables (eg, disease severity, symptom duration, prior treatments for the condition, and the use of concurrent medications) for patients entered and excluded, with reasons for exclusion. Explicit inclusion and exclusion criteria are mandatory for all studies. Most treatment trials in FGIDs have required a minimum severity level for specific symptoms thought to be typical of the condition. Balanced consideration for the potential mechanism of action of the drug must also be given when selecting the study population.

It is advisable to include as broad a spectrum of patients as possible, defined by the Rome- specific FGID criteria. Restricting or modifying the study population must be justified. The EMA requests that early drug development programs include sufficient numbers of both men and women to permit assessment of safety and efficacy for both sexes. The FDA also supports engagement of subjects of different racial backgrounds.[2,3]

Inclusion Criteria

The minimum screening for eligibility should be specified and should adhere to current guidelines. The Rome classification of FGIDs is currently the most comprehensive and well-established diagnostic system, and its use ensures a sufficient degree of standardization of study participants across centers and cultural settings, and allows further exploration for differences in treatment response.

Exclusion Criteria/Appropriate Rule Outs

Important confounding factors to consider for possible exclusion criteria are psychological comorbidities, sociocultural perspectives, and biological variations. Psychological comorbidities are often thought to be predictors of poor response to treatment, but this has not been proven.[4] Other psychologically related influences include the placebo and nocebo effects (see section on placebo and nocebo), and future studies may wish to consider designs that could measure the subject's proneness to these effects.

Managing Functional Gastrointestinal Disorders Overlap, Comorbidities, and Disease Modifiers

Overlap disorders, potential disease modifiers, and important comorbidities that might affect treatment response should be assessed and explored. The overlap of FGIDs with other FGIDs and with somatic and psychiatric disorders is a challenge for clinical trail design. First, the accuracy of the FGID diagnosis may be questioned and it is possible that a treatment might improve the symptoms of one disorder while symptoms of the other worsen. Second, the presence of a comorbidity may be associated with increased symptom severity, greater impact on health-related quality of life (HRQOL), and greater psychological distress—all of which could modify the response to treatment. Third, underlying motility or sensory disorders in different parts of the GI tract may interact in ways that could affect the response to specific treatments. The committee recommends that, in most situations, patients with overlapping conditions be included in the trial and the presence of comorbid conditions should be documented.

Role of Biomarkers in Defining Study Population

Continuing research is needed to identify biomarkers that attempt to elucidate disease mechanisms and may facilitate assessment of efficacy of treatments in FGID studies. A biomarker is an indicator of a physiological or pathological state that can be objectively measured and evaluated, in contrast to PROs, which are measured using questionnaires that capture patient perceptions of their illness.[5] A valid and reliable biomarker should optimally distinguish patients with a known clinical syndrome from other conditions, and do so with a high degree of sensitivity and specificity. It may also have predictive value, in that its presence could potentially predict natural history and/or response to specific therapies.[5] While they are not suitable as surrogate end points at this time, they can be used to stratify patients. However, at present, very few biomarkers have been identified that have sufficient sensitivity and specificity.

Clinical Trial Design

Unique Challenges for the Design of Treatment Trials in Functional Gastrointestinal Disorders

There are several challenges to conducting FGID treatment trials, including a high placebo response rate[6]; symptoms that are intermittent and of fluctuating severity[7]; a potential need for multimodal therapy, given the limited efficacy of available treatments or multiple etiological mechanisms affecting the disease process[8]; difficulty maintaining masking of patients and investigators in trials of behavioral interventions[9]; contamination from over-the-counter treatments or medicines taken for other conditions; the necessity of avoiding significant harms[10] in treating non−life-threatening conditions; absence of biomarkers both for diagnosing the disorder in question and for evaluating the treatment response; and absence of acceptable end points for many FGIDs. In addition, clinical trials differ from clinical practice in several ways, including the application of strict inclusion and exclusion criteria, the use of a placebo group, application of a standardized intervention, frequent follow-up visits with extensive data recording, and the use of study coordinators.

The placebo response observed in clinical trials has been attributed in part to the attention given to enrolled subjects, including detailed explanation and reassurance, close monitoring, and ready access to study coordinators, which may in themselves produce a therapeutic effect. Bias, defined as "systematic error" in estimating the treatment effect, may enter a clinical trial at any stage, from design to publication.[11] The major sources of bias are listed in Table 2.

Table 2. Major Sources of Bias in Clinical Trials

Bias type	Comments
Investigator bias	Conscious or unconscious, usually expressed through decisions about eligibility
Patient expectancy (placebo)	Especially a problem where end points are subjective
Ascertainment bias	
Self-selection for treatment	Patients are more likely to respond positively to treatments they prefer and seek out
Changes in subject pool	Publicity or other factors may influence the subject pool over time
Nonspecific effects	
Doctor−patient relationship	Especially important in psychological interventions
Regression to the mean	Patients are usually enrolled when most symptomatic and inevitably improve
Publication bias	Authors are more likely to submit trials with positive results and journals are more likely to publish them

Masking/Blinding Process

It is mandatory to undertake the maximum masking possible, determined by the type of intervention and study design. It is recommended to evaluate and report whether masking was successful. Masking of participants, investigators, and evaluators to treatment assignment is a key feature of a successful controlled trial.[12] Single masking is when only the study subject/patient is unaware of the treatment allocation. Double-masking (both patients and research personnel) is necessary to ensure the highest validity of the primary outcome measurement. Triple-masking includes also masking monitors, data managers, statisticians, and others who interpret outcome tests.[13] Interventions involving procedures such as psychotherapy, hypnotherapy, sphincterotomy, or drug trials in which the active drug causes predictable side effects or rapid symptom changes, are difficult to mask from patients or investigators. Possible solutions include using independent assessors who are unaware of the intervention, or standardized interviewer-administered or self-administered questionnaires.[14] In addition, study investigators are encouraged to ask both patient and interventionist at the end of the trial whether they believe active treatment was administered and to report these data.

Randomization

Investigators must include a detailed description of their randomization process and concealed allocation method in the report of the study. Randomization is the process of assigning subjects to different treatment arms without bias, which can be accomplished either by someone other than an investigator preparing a numbered series of sealed envelopes containing group assignments or use of a computer program for random allocation.[13,15] Critical recommendations to ensure randomized concealed treatment are the randomization code is generated by a noninvestigator (preferably a computer), randomization is done within blocks of variable size (permuted block randomization) or sufficient size to minimize unmasking due to side effects in previously exposed patients, the list of patient treatment assignments should be available only to the medical officer in charge of patient safety, and a record should be kept of patients for whom the mask has been broken. When reporting the trial, the randomization procedure should be described explicitly.[15] Stratified randomization is a variation on randomization that is designed to assure balance on the most important prognostic factors by using a separate randomization sequence for each stratum (eg, male vs female or IBS with constipation [IBS-C] vs IBS with diarrhea] IBS-D])[16].

Selecting a Control Group

A control group is required to establish the true efficacy of a new treatment. As therapies of proven efficacy accumulate, a comparison against an active available treatment can be considered, but this requires higher patient numbers to establish efficacy and may fail to show a statistically significant difference.[17,18] Control groups for therapeutic procedures are equally crucial. In behavioral treatment trials, confirming that the control intervention produces a similar expectation of benefit but does not act on the same physiological or psychological principle is recommended. In trials involving a therapeutic procedure, a sham group is recommended when feasible.

Placebo, Nocebo, and Precebo Responses

Placebo. Placebo (from the Latin "to please") is an intervention that generates the expectation of benefit in the patient but is believed to lack any specific effect to change a particular disorder,[19] or an intervention for which there is no scientific theory explaining its action.[20] When used along with blinding, use of a placebo design may enable investigators to assess side effects of interventions more readily and with less bias. Placebos can be administered as a drug or as a procedural intervention.[20] The placebo effect is well characterized in FGID trials, especially in FD and IBS, with response rates ranging from 6%–72%[21,22] and 0%–84%, respectively.[6] A meta-analysis suggested that the placebo response is larger when a responder is defined by a global improvement in IBS symptoms compared with defining a responder by reduction in abdominal pain.[23,24] A more recent systematic review and meta-analysis found higher placebo rates in European randomized controlled trials (RCTs) compared with those conducted in other continents; in those that used physician-reported outcomes compared with those that used a patient-reported end point; and in RCTs using shorter duration of therapy.[25] Also, pooled placebo response rates were generally higher in RCTs using clinical criteria to define the presence of IBS compared with those using Rome criteria, trials using 3 times daily dosing, trials that assigned patients to placebo or active therapy in a 1:1 ratio, trials of antispasmodics and mixed 5-HT3 antagonists/5-HT4 agonists, and trials of lower scientific quality.[25]

Nocebo. In contrast to placebo, nocebo (from the Latin "I shall harm")[20] is the expectation of distress. The expectation of side effects may increase the frequency with which adverse effects are reported in both the active and control arms of a drug study, and may increase the likelihood that subjects will drop out of the trial.

Precebo. The term *precebo* was coined to describe the effect that influences placebo even before the study begins.[26] The precebo effect refers to the potential for a drug benefit during a clinical trial to be influenced by preconceived notions or by communications about the trial contained in advertisements and consent forms.

Baseline Observation vs Placebo Run-In

A period of prospective baseline measurement before treatment is useful to evaluate patient eligibility. This also limits recall and reporting biases and ensures that patients are currently symptomatic. It allows comparison of patients in the active and placebo groups, as well as evaluation of a clinically important change in health status.

Older studies have used a placebo run-in period where all patients received placebo for a specified period and their responses were assessed using the study outcome measures. Patients who significantly improved were excluded. Although acceptable to regulatory agencies, placebo run-in can underestimate the overall effect size.[27]

Choice of Study Design

The double-masked, randomized, placebo-controlled, parallel-group trial is the gold standard for testing the efficacy of new treatments. Variations of this basic design include different groups receiving different doses of the active treatment (dose-ranging, in phase 2), more than one control treatment, multiarm trials, a baseline period of no treatment, and a washout period after treatment is completed.

As there is no universally effective treatment for any FGID, the standard approach is to test a new therapy against placebo to prove its superiority. Occasionally, trials of equivalence and noninferiority are performed where a new therapy is more convenient or less expensive.[18]

Crossover designs have been popular in FGID treatment trials.[6] Subjects receive both treatments during distinct time periods, usually separated by a washout phase, in randomized order, with the aim of comparing the treatments. Theoretically, a crossover design can increase sensitivity to detect change, allowing a smaller sample size for the desired statistical power. However, there are down sides: patient dropout and missing data have a greater impact than in a parallel-group design, carryover effects that occur when the first treatment influences the response to the second treatment, and there is a higher risk of unmasking due to side effects. Therefore, crossover trials seems most applicable in physiological studies where end points are objectively measured.

A factorial design is appropriate when evaluating combination treatments, which may be desirable in patients with severe FGID symptoms.[28] This requires a control group for each intervention. The withdrawal trial is an "enrichment design" in which all subjects receive the active treatment and, at a predefined time point, are classified as responders or nonresponders. The latter are then excluded and responders are randomly assigned to continue with treatment or placebo. The efficacy assessment is based only on the second part of the trial. Potential carryover effects from active treatment are the major drawback.[29]

Design of Trials for Behavioral, Surgical, and Complementary and Alternative Medicine Interventions

In trials evaluating the efficacy of behavioral, surgical, or many types of complementary and alternative medicine interventions, it is not possible to mask the intervention from the therapist (the individual implementing the intervention) or from the patient. Expectation of benefit is the most important variable to balance across intervention arms. Some published trials of behavioral interventions have compared symptom improvement in the active treatment group to symptom changes in people who remain on a waiting list to receive the intervention or who continue to receive “standard medical care.” However, both of them create a negative expectancy of improvement and therefore have potential to overestimate the efficacy of the investigational treatment. A better approach is to identify an alternative, active treatment that generates a similar expectation of benefit and is assumed to be less effective. Investigators have also tried to balance the amount of contact time with the therapist and other characteristics across treatment arms.[30,31] The expectation of benefit should be measured in both groups to confirm that the treatment arms are balanced.[14]

A number of steps are recommended to minimize the impact of investigator bias in behavioral trials: (1) randomize patients to the treatment arms only after they have been screened and found eligible; (2) have an expert develop a detailed treatment manual for all treatment arms; (3) use multiple well-trained interventionists and test whether outcomes differ across interventionists; and (4) use patient-completed outcome questionnaires or outcome assessors who are blind to the treatment assignment of each subject.

Duration of Treatment

Prior recommendations for treatment durations of trials of 8−12 weeks were based on experience together with considerations of cost and ability to retain patients. The EMA guidelines differentiate between trials to establish short-term efficacy, for which a treatment duration of 4 weeks or longer would be acceptable vs trials intended to establish long-term efficacy, for which a minimum of 6 months is recommended.[3] Extended patient follow-up should be considered to determine the treatment durability and should also relate to the presumed treatment mechanism and periodicity of symptoms. Recent long-term studies of 6 months duration in parallel design[32] or on demand have now been undertaken in FGID patients.[33]

Adherence to Treatment and Study Protocol

During a clinical trial of FGID, adherence to medication is critical in interpreting the results and efficacy of treatment. Most trials accept 80% as a reasonable level of adherence that allows valid assessment of the treatment intervention. Measuring adherence can be achieved by measuring a metabolite of a drug treatment, counting unused medication, use of electronic or paper diaries, prescription purchase monitoring, patient interview, or physician impression. Strategies that appear to enhance patient adherence during clinical trials include short-term vs longer trials, clear written instruction and education before and during the trial, reminders to take medication, recording symptoms or attending appointments, self-monitoring, severity of condition or symptoms, efficacy of the treatment, and patient education and understanding of the importance of adherence.[34]

Considerations for Dietary Interventions

Major challenges in dietary trials include masking the intervention and innovations in the choice of control diets. Applying any standardized diet, such as the average national diet, is likely to be an intervention, but it does control diet in a standardized fashion. Recently, the methodological rigor of dietary intervention trials has improved, with studies in which all meals were provided in a masked fashion to the patients for the duration of the trial.[35,36]

Considerations for Probiotic Trials

There is a rapidly increasing interest in using probiotics and prebiotics for the treatment of FGIDs, but interpreting the trials to date has been hampered by suboptimal trial design, small sample sizes, and the wide variety of probiotic strains and formulations that have been used.[37] A minimum requirement for probiotic trials is to demonstrate that the test organisms are present in stools or in the lumen of the gut in a representative subset of exposed subjects. Whether for registration as drugs, or as food supplements or functional foods, they require the same rigorous criteria, design, and end points as classical pharmacological efficacy studies.[38]

Considerations for Pediatric Trials

Compared with adults, there are far fewer published clinical trials on FGID in the pediatric population. Primary end points in children vary widely among studies. Developmental limitations make it difficult to obtain reliable PROs in young children. In trials involving infants, toddlers, and younger children, reports of symptoms are based on parental observation, so-called “observer-reported outcomes.” Unfortunately, minimally clinical important differences (MCIDs) and factors determining the magnitude of placebo responses have not been established in pediatrics.

Pragmatic Clinical Trials

PCTs focus on the risks, benefits, and costs of competing therapies within the context of usual practice settings.[39] Whereas explanatory RCTs restrict variability in treatment delivery between sites and between treatment arms, PCTs aim to understand health outcomes between competing management approaches for a common clinical dilemma, may include a broad range of patients from diverse settings, and may even allow for different patterns of care within a study arm to emulate clinical reality. However, as PCTs do not tightly control all aspects of study design, it can be difficult to untangle mechanisms of action, or to isolate key prognostic variables, and they are more susceptible to effects of investigator bias, patient expectancy, and ascertainment bias related to self-selection for different treatments.

Registering With ClinicalTrials.gov

All clinical trials should be registered before initiation on a dedicated, publically accessible website. One example is ClinicalTrials.gov, established by the National Institutes of Health. US law now not only mandates reporting of clinical trials, but also establishes penalties for noncompliance.

Patient-Reported Outcomes Measurement

Definition and uses of patient-reported outcomes. The patient report is of primary importance in evaluating effectiveness of FGID therapies. PROs are designed to capture the patients' illness experience in a structured format and may help bridge the gap between patients and providers, while providing outcome targets for clinical trials. The FDA, EMA, and other regulatory agencies consider the patient report in drug approval, and have developed guidance for development and use of PROs in clinical trials.[40,41] The National Institutes of Health has also supported a major PRO initiative, called the Patient Reported Outcome Measurement Information System (PROMIS; www.nihpromis.org), designed to develop and evaluate several PRO domains.[42]

Classification of Outcomes Measures

Individual symptoms can be measured across various attributes, including frequency, severity, bothersomeness, and predictability, among other factors.[43] Individual symptoms and their attributes can be combined into symptom clusters. FGID symptoms, in turn, may impact physical, social, and emotional function, measured in terms of HRQOL. Finally, the broader illness experience of FGIDs, as measured by HRQOL instruments, may have downstream impacts on FGID-related resource utilization and work productivity.[44] Health utility measures, like the EQ-5D, are a specialized form of PROs that inform cost−utility analyses—another key resource utilization outcome.[45]

Developing Patient-Reported Outcomes for Clinical Trials: Guidance From Regulatory Agencies

Regulatory agencies have developed detailed guidance for how to validate and document developmental steps for a PRO. Examples include the FDA guidance on Patient-Reported Outcome Measures[40] and guidance for IBS registration trials.[2]

PRO development should begin with a systematic review of the literature to build a conceptual framework for developing a new PRO. The framework should be expanded based on direct input from representative patients of the target population. Both the FDA and EMA emphasize rigorous qualitative methods in conducting patient focus groups or interviews. PRO developers next focus on creating the individual items for the instrument, in easily understood language. The FDA and EMA generally recommend no more than 1-day recall periods for a PRO. Investigators must next conduct quantitative empirical testing of the instrument in a representative sample of patients, including both cross-sectional and longitudinal psychometric testing. The provisional PRO should undergo exploratory factor analysis to evaluate the quantitative structure of the instrument, analysis of convergent validity by evaluating the relationship with legacy instrument of relevance, and also measurement of internal consistency and reliability of each PRO subscale. Finally, the investigators must calculate the MCID for each scale in the PRO.

US Food and Drug Administration Interim Guidance for Irritable Bowel Syndrome Clinical Trial Outcomes

Recognizing that it would take time before an FDA-qualified IBS PRO could be developed, the FDA agreed to allow interim end points for registration trials until a final PRO is developed. The interim guidance, published in May 2012, suggests the following co-primary end points: abdominal pain and abnormal defecation. The standard 11-point numeric rating scale should be used to measure abdominal pain in IBS. For abnormal defecation, the FDA recommends measuring stool frequency for IBS-C trials, and stool form measured using the Bristol Stool Form Scale (BSFS) for IBS-D trials. Using the tools of the numeric rating scale and the BSFS, the FDA recommends specific IBS responder definitions, which is largely followed by the EMA.[3]

The FDA interim guidance for IBS trials has significant limitations: Different inclusion criteria and different end points are recommended for trials of IBS-C vs IBS-D, and neither inclusion criteria nor end points are specified for patients with IBS with constipation and diarrhea (IBS-M). These limitations appear to restrict the target populations that can be studied and the likely indications for drugs that could be approved.

Binary Outcomes Measures

Many high-quality FGID RCTs have employed a binary PRO end point, such as "adequate relief," "satisfactory relief," or "considerable relief,"[46] which provide a dichotomous responder status (yes/no relief). Binary end points are easy to administer and straightforward to interpret.[46,47] Previous systematic reviews and the Rome III guidance supported the use of binary end points as a standard for IBS and FGID clinical trials.[46–48] However, the FDA currently discourages the use of these binary end points in clinical registration trials, based on concerns about the possibility that a clinical response with a binary end point may depend on baseline severity, may not detect MCIDs, and may lack capacity to track key illness domains or discriminate between clinical disease subgroups. Nevertheless, a working party, which analyzed patient-level data from 12 existing clinical trials in 10,000 patients, showed that the binary response demonstrated excellent construct validity across a wide range of variables and was able to detect MCIDs in key bowel symptoms.[49]

Integrative Symptom Questionnaires

The IBS Severity Scoring System[50] and the Functional Bowel Disease Symptom Index[51] are 2 well-validated symptom questionnaires that integrate several components of IBS symptoms into a single score. The IBS Severity Scoring System instrument, a widely used PRO for IBS clinical research studies, incorporates pain, abdominal distention, bowel dysfunction, and HRQOL to estimate overall patient illness severity. The Functional Bowel Disease Symptom Index[51] also includes the resource utilization variable—number of physician visits. However, neither meets current psychometric EMA or FDA standards for a qualified PRO for clinical trials. Functional dyspepsia trials have used several outcome measures, with varying degrees of validation, but none meets all the FDA and EMA end point criteria for registration trials.[52]

Pictograms

Verbal symptom descriptors are used in most PRO instruments to evaluate symptom patterns and severity in FGIDs.[2,36,52] Pictorial representations have been proven to be effective in improving comprehension and recall of new information.[53] The use of pictograms to evaluate FGID is being examined and in functional dyspepsia shows potential to improve concordance between the clinician's and patient's evaluation of symptom pattern and severity (Figure 1).[54] The impact of cultural context on interpretation and acceptability of pictograms needs evaluation.

Measuring Abdominal Pain vs Discomfort

The Rome criteria have historically combined "pain" and "discomfort" into the same symptom complexes. Abdominal pain is a defining characteristic for many FGIDS and an important driver of symptom severity, HRQOL decrements, and health care resource utilization.[55] FGID pain is typically measured as severity, using the previously discussed numeric rating scale, but less is known about the impact of other pain attributes, including frequency, constancy, recency, duration, predominance, predictability, speed of onset, and its relation to bowel movements. Although past clinical trials have used PROs that group pain and discomfort together,[46,49] the separation between pain and discomfort is inconsistent across patient groups, and discomfort often refers to a range of symptoms, such as bloating, gas, fullness, flatulence, sensation of incomplete evacuation, and urgency. It is recommended that pain be measured separately from discomfort, and that the type of nonpainful symptom be specified.

Measuring Bloating

Bloating is reported by up to 31% of the general population,[56,57] and is also very common among patients with FGIDs. The feeling of bloating should be distinguished from visible abdominal distention, as they do not always overlap.[43] The use of pictograms has been recommended to express bloating more accurately.

Measuring Stool Frequency and Form

Stool form and frequency are now part of FDA and EMA end points for IBS clinical trials. The BSFS is used as a validated measure of stool form that also correlates with intestinal transit time.[57] The BSFS has reasonable face validity for patients and has improved the reliability of patient reports on stool consistency. However, there are important limitations of the BSFS, including occurrence of multiple BSFS forms within the same bowel movement and the difficulty to determine the "start" and "end" of a bowel movement.

In current constipation trials, and supported by FDA, stool frequency is measured in a diary using terms of (1) bowel movement, (2) spontaneous bowel movement (without need for manual maneuvers or rescue laxatives), and complete spontaneous bowel movement (adds a sensory aspect in that the patient must experience a full evacuation without the residual feeling of retained stool).

Measuring Bowel Urgency

Bowel urgency is also a bothersome symptom that can undermine HRQOL in patients with FGIDs like IBS[55] and fecal incontinence.[56,58] The term *urgency* in IBS-D is

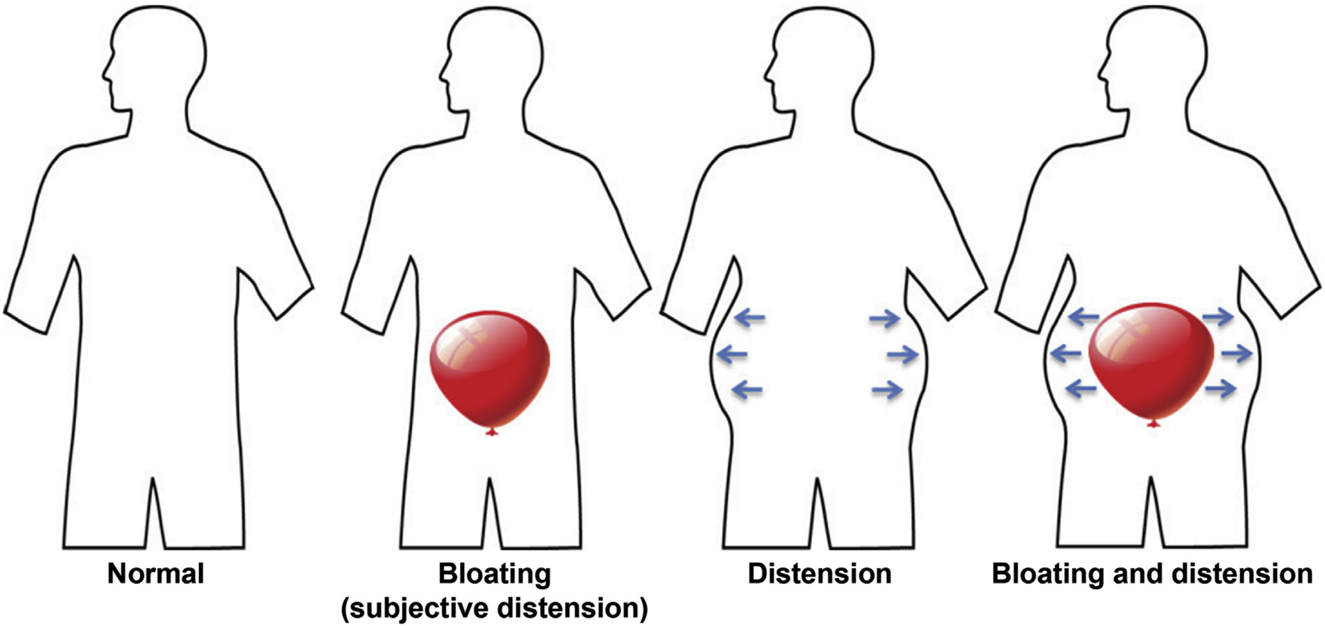


Figure 1. Three pictograms depicting the patient experience of abdominal bloating and distension.

multifaceted, and includes attributes like frequency, intensity, interference, and fluctuation, which may need to be taken into account when creating a PRO.

Health-Related Quality of Life Questionnaires

HRQOL is a type of PRO that captures biopsychosocial health rather than individual symptoms. HRQOL is usually measured with patient questionnaires or instruments that collect data across several areas of health, including physical, psychological, and social functioning. HRQOL instruments are generally classified as either "generic," which measure HRQOL across many different conditions or populations; or "disease-targeted," when they measure HRQOL in 1 or more specific conditions.[59] Examples of the former include the Short Form-36 Health Survey[60] and the Sickness Impact Profile.[61] Disease-targeted HRQOL instruments appear to be more responsive than generic HRQOL instruments to treatments for specific diseases. More than 110 disease-targeted HRQOL instruments were developed in gastroenterology and cover a range of FGIDs.[62] Of the multiple disease-targeted instruments in IBS, the Irritable Bowel Syndrome Quality of Life questionnaire[63] has the most extensive data supporting its validity.

Data Collection Strategies

Both the FDA and the EMA recommend the use of daily diaries to assess the interim primary symptom outcomes in IBS, while specific PROs are being developed. Most likely, this recommendation can also be extended to other FGIDs for which no guidelines have been produced to date. The use of daily assessment serves to minimize recall bias and avoids influence by the presence of the investigator. Symptom ratings can be performed at a fixed time (eg, bedtime), or at the time when symptoms actually occur. The FDA suggests using an interactive voice response or personal digital assistant to assure accurate data collection because concern has been raised that paper diaries may be retrospectively completed just prior to a visit.[64,65] A number of secondary outcome variables, such as QOL questionnaires, use a longer retrospective recall period, and need to be collected only intermittently during the course of a trial. More recently, real-time capturing of symptom occurrence, using electronic applications on hand-held devices, has the potential to provide more accurate information on actual onset of symptoms and their time course. However, patient compliance is crucial and the gain over daily diaries completed at a fixed time of day needs to be proven.

Adverse Events and Safety Monitoring

It is important to report all adverse events found in treatment trials, as new or unexpected side effects may be encountered when testing new therapies. An extension of the Consolidated Standards of Reporting Trials (CONSORT) statement[66,67] encouraged the use of the term *harms* rather than *safety*. If the collection of harms data is a key trial objective, this fact should be reflected in the title and abstract, as well as the body of the article. The methods section should also clearly define how adverse events were measured. Details of individual adverse events, including impact and severity, are needed to allow pooling across trials to allow calculation not only of number needed to treat, but also of number needed to harm.

Analysis and Data Reporting

Consolidated Standards of Reporting Trials Guidelines

The original CONSORT statement, first published in 1996, was developed to improve the quality of reporting of 2-group, parallel RCTs.[66,67] The latest iteration, the CONSORT 2010 Statement,[68] re-emphasized the importance of clearly and transparently reporting the reason the study was undertaken, and how it was carried out and analyzed. It includes a 22-item checklist, including key elements of statistical reporting to which investigators should adhere,[67] and a flow diagram with 4 sections (ie, enrollment, allocation, follow-up, and analysis). Many journals now require that manuscripts describing clinical trials conform to the CONSORT guidelines.

Primary Efficacy Analysis

Defining minimally clinical important differences and responder definitions. In general, a study should have 1 and no more than 2 primary outcome measures.[67] The FDA has published guidance on trial design, end points, and responder definitions for the treatment of IBS.[2] The most recent update of the EMA guidance follows the same principles.[3] The primary statistical analysis should focus on the chosen primary outcome measure(s), and the result of this planned analysis determines whether or not the study has a positive result in support of a new treatment.

Although the main outcome often is reported as a comparison between the end of treatment and baseline observations, it is also important that data are provided describing how patients changed throughout the course of the study. When 2 primary outcome variables are included in the trial, the investigators should specify in advance whether the trial is positive if only one of the outcome measures is significant, or if they require that both be significant. If significance of any primary outcome will provide evidence for efficacy of the treatment, the analysis should adjust for multiple comparisons.[69] For all outcome measures, the estimated effect of the intervention (difference between active and placebo treatment) and a 95% two-sided confidence interval should be included.[70] Statistically significant differences between study groups can also be expressed using a *P* value (actual values). The reciprocal of the therapeutic gain can also allow computation of the number of patients needed to treat to encounter a patient who will experience a clinical benefit.[71]

The statistical analysis should be based on an intention-to-treat principle, which includes all patients randomized to treatment.[72] Dropouts can be considered treatment nonresponders, or the last observation of the primary outcome variable that was available can be carried forward. Both approaches should be examined to test for differences in results. Many studies also report a per-protocol (all patients who followed the protocol) or an all-patients-treated (all patients who received treatment after randomization) analysis. These analyses may provide insight as to whether a treatment works under optimal conditions but cannot replace the intention-to-treat analysis.

In prespecified analyses, the effect of potential modifiers such as sex, age, duration, or severity of disease, and presence of psychological stress can be assessed using a logistic regression analysis, where the binary dependent variable represents the a priori specified definition of a responder.[73]

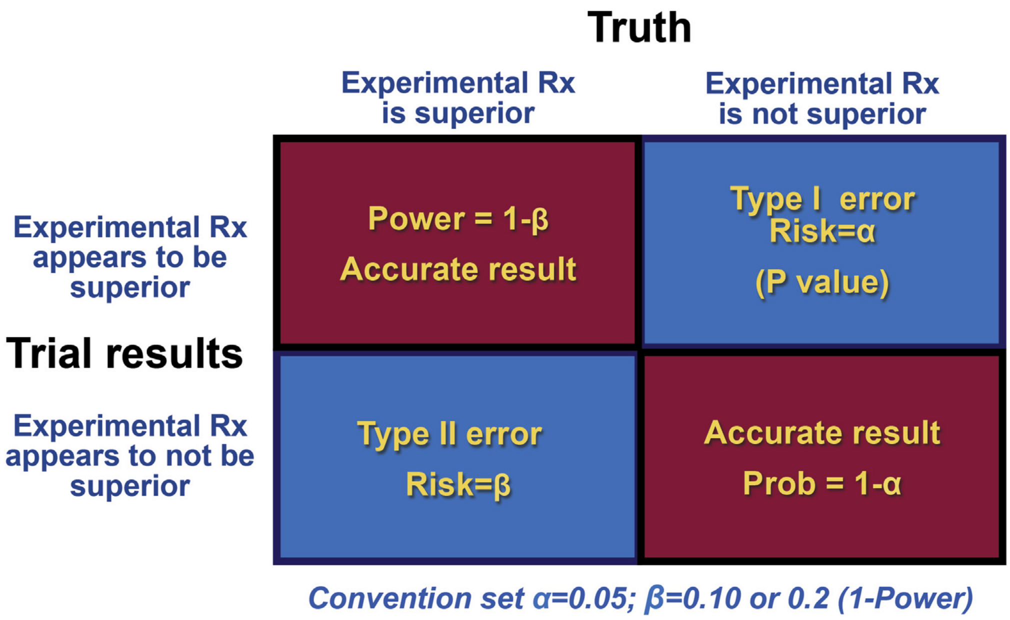
Analysis of secondary outcome measures and subgroups. It is recommended that changes in each of the symptoms that comprise the entry criteria be analyzed by intention to treat and reported. This may support (or refute) the direction and magnitude of the interventional effect on the primary outcome measure. Investigators sometimes include a large number of secondary variables to identify predictors of response or to explore other possible effects of the intervention unrelated to the primary hypothesis. In such cases, adjustment for repeated testing is needed, or the intestiagors may use descriptive rather than inferential statistics, or they may choose a conservative α level (eg, .01) to protect against type I error without unduly inflating the type II error rate.

Exploratory subgroup analyses are commonly performed in trials addressing the effectiveness of therapies for patients with FGIDs. This practice is controversial and some researchers question its validity,[67,74–76] particularly when undertaken after initial evaluation of the dataset (post-hoc subgroup analysis). The test of interaction is the most appropriate type of subgroup analysis.[76] Specific plans to present and analyze harms data should be clearly described, and reported as actual incidence rates and 95% confidence intervals.[77]

Sample size and power calculations. The protocol should clearly specify the assumptions upon which the sample size calculation was based.[78] This includes the minimum effect size that the trial is designed to detect, α (type I) error level, the statistical power or β (type II) error level, and when evaluating changes (differences) in continuous outcomes, the standard deviation of the difference (Figure 2). Trials have generally been powered to detect differences between 10% and 15%, at a power of 80% and an α level of 5%, using a 2-sided test. An allowance for dropouts should also be made in determining the appropriate sample size, and the number and timing of the dropouts should be reported.

For responder analyses, the protocol should clearly state what constitutes a patient responder. The study must have sufficient power to detect a clinically important difference in the proportion of responders.[79,80]

Interim analysis and stopping rules. Plans for interim analyses should be clearly prespecified in the study protocol, but usually there is no compelling reason to incorporate specific interim analyses for interventions in the FGIDs. Unplanned preliminary analyses should be avoided because premature presentation of results can affect the further conduct of the trial and can lead to the reporting of inaccurate observations.[81,82] On the other hand, adaptive clinical trial designs that explicitly allow for study design modifications based on interim analyses have become increasingly attractive due to their flexibility and efficiency in pharmaceutical/clinical development.[81–84] Limitations of these designs include control of type I error rate, minimization of statistical and operational bias on the estimates of treatment effects, and the interpretability of the results.[84] The FDA provided draft guidance on adaptive design clinical trials for drug and biologic development.[85]



Trial results	Truth: Experimental Rx is superior	Truth: Experimental Rx is not superior
Experimental Rx appears to be superior	Power = 1-β Accurate result	Type I error Risk=α (P value)
Experimental Rx appears to not be superior	Type II error Risk=β	Accurate result Prob = 1-α

Convention set α=0.05; β=0.10 or 0.2 (1-Power)

Figure 2. Elements of a sample size calculation including Type I and Type II error.

Interim analyses to assess the futility of continuing a trial should be overseen by a Data and Safety Monitoring Board that is independent from the investigators, should test for equivalence rather than superiority of one treatment relative to the other, and should use a priori−defined liberal equivalence margins for the effect size.

Noninferiority trials. The design and methods for the statistical analysis of equivalence and noninferiority trials can be complex and have not been developed to the extent seen in superiority trials. The most important design and analysis aspects are estimating the effect size of the active comparator based on previous RCTs, selection of appropriate noninferiority margins, and determining the sample size.[86] The FDA has published Guidance for Industry Non-Inferiority Clinical Trials.[87]

Ethical Issues

All investigational products should receive approval from local regulatory agencies and all clinical trials should be approved by local ethical committees before their start. All patients should sign informed consent before any study-related procedure, and all personnel involved in clinical trials should adhere to Good Clinical Practice guidelines and have recent (last 2 years) evidence of qualification. Most hospitals and research institutions are also mandating manuals documenting evidence of training in standard operating procedures. Clinical trials should maintain a balance between scientific ambitions and the patients' interest in receiving effective therapy without being exposed to risk or unnecessary evaluations. Adverse events must be monitored and registered throughout the study, and serious adverse events need to be rapidly reported and evaluated for their relationship to the investigational treatment. For the placebo-controlled study phases in which patients might receive placebo or a treatment of uncertain therapeutic value for longer periods, rescue medication should be offered whenever possible. Patient benefit for participating in trials may increase when an open-label follow-up treatment can be offered after the controlled study phase.

Many studies offer financial compensation to participating patients for the time lost due to study visits. These incentives, and other coverage of expenses (travel, food), should be of reasonable magnitude to prevent patients from enrolling in a study purely for financial reasons. Similarly, financial compensation to physicians or institutions for participating in clinical trials should also be reasonable and concordant with standards of reimbursement based on resource utilization or time commitment.

Recommendations for Future Research

Several issues that should be addressed with future research include:

1. Examining and minimizing the magnitude of placebo effects
2. Creating and validating measures of expectancy for FGIDs (placebo, nocebo, and precebo)
3. Assessing the natural history of FGID to determine the appropriate duration of acute and long-term RCTs
4. Extending follow-up after RCTs to determine the durability of interventions.
5. Further elucidating the characteristics of IBS-M
6. Developing PROs for IBS-M
7. Developing PROs, observer-reported outcomes and pictograms for pediatric trials in FGID
8. Establishing normal bowel habit ranges in children by age and sex
9. Assessing whether a 1-day recall period for outcomes is appropriate
10. Assessing whether the current FDA and EMA guidance for primary outcomes (co-primary outcomes) for IBS are optimal or other outcomes (including binary outcomes) are also valid

Supplementary Material

Note: The first 50 references associated with this article are available below in print. The remaining references accompanying this article are available online only with the electronic version of the article. Visit the online version of *Gastroenterology* at www.gastrojournal.org, and at http://dx.doi.org/10.1053/j.gastro.2016.02.010.

References

1. Corsetti M, Tack J. FDA and EMA end points: which outcome end points should we use in clinical trials in patients with irritable bowel syndrome? Neurogastroenterol Motil 2013;25:453–457.
2. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER). Guidance for Industry: Irritable Bowel Syndrome−Clinical Evaluation of Drugs for Treatment. Available at: http://www.fda.gov/downloads/Drugs/Guidances/UCM205269.pdf. Published May 2012. Accessed January 10, 2015.
3. European Medicines Agency. Guideline on the Evaluation of Medicinal Products for the Treatment of Irritable Bowel Syndrome. Vol 2014. London, UK: EMA, 2013.
4. Holtmann G, Kutscher SU, Haag S, et al. Clinical presentation and personality factors are predictors of the response to treatment in patients with functional dyspepsia; a randomized, double-blind placebo-controlled crossover study. Dig Dis Sci 2004;49:672–679.
5. Spiller RC. Potential biomarkers. Gastroenterol Clin N Am 2011;40:121–139.
6. Spiller RC. Problems and challenges in the design of irritable bowel syndrome clinical trials: experience from published trials. Am J Med 1999;107:91S–97S.
7. Palsson OS, Baggish J, Whitehead WE. Episodic nature of symptoms in irritable bowel syndrome. Am J Gastroenterol 2014;109:1450–1460.

8. Drossman DA, Thompson WG. The irritable bowel syndrome: review and a graduated multicomponent treatment approach. Ann Intern Med 1992;116:1009–1016.
9. Whitehead WE. Control groups appropriate for behavioral interventions. Gastroenterology 2004;126 (Suppl 1):S159–S163.
10. Camilleri M. Safety concerns about alosetron. Arch Intern Med 2002;162:100–101.
11. Sackett DL. Bias in analytic research. J Chronic Dis 1979; 32:51–63.
12. Jamshidian F, Hubbard AE, Jewell NP. Accounting for perception, placebo and unmasking effects in estimating treatment effects in randomised clinical trials. Stat Methods Med Res 2011;23:293–307.
13. Spilker B. External influences on protocol design. Epilepsy Res Suppl 1993;10:115–124.
14. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. J Behav Ther Exp Psychiatry 2000;31:73–86.
15. Altman DG. Randomisation. BMJ 1991;302:1481–1482.
16. Altman DG. Comparability of randomised groups. Statistician 1985:125–136.
17. Tinmouth JM, Steele LS, Tomlinson G, et al. Are claims of equivalency in digestive diseases trials supported by the evidence? Gastroenterology 2004;126:1700–1710.
18. Temple RJ. When are clinical trials of a given agent vs. placebo no longer appropriate or feasible? Control Clin Trials 1997;18:613–620; discussion 661−666.
19. Thompson WG. Placebos: a review of the placebo response. Am J Gastroenterol 2000;95:1637–1643.
20. Bernstein CN. Placebos in medicine. Semin Gastrointest Dis 1999;10:3–7.
21. Veldhuyzen van Zanten SJ, Cleary C, Talley NJ, et al. Drug treatment of functional dyspepsia: a systematic analysis of trial methodology with recommendations for design of future trials. Am J Gastroenterol 1996;91:660–673.
22. Musial F, Klosterhalfen S, Enck P. Placebo responses in patients with gastrointestinal disorders. World J Gastroenterol 2007;13:3425–3429.
23. Vase L, Robinson ME, Verne GN, et al. The contributions of suggestion, desire, and expectation to placebo effects in irritable bowel syndrome patients. An empirical investigation. Pain 2003;105:17–25.
24. Pitz M, Cheang M, Bernstein CN. Defining the predictors of the placebo response in irritable bowel syndrome. Clin Gastroenterol Hepatol 2005;3:237–247.
25. Ford AC, Moayyedi P. Meta-analysis: factors affecting placebo response rate in the irritable bowel syndrome. Aliment Pharmacol Ther 2010;32:144–158.
26. Kim SE, Kubomoto S, Chua K, et al. "Pre-cebo": an unrecognized issue in the interpretation of adequate relief during irritable bowel syndrome drug trials. J Clin Gastroenterol 2012;46:686–690.
27. Berger VW, Rezvani A, Makarewicz VA. Direct effect on validity of response run-in selection in clinical trials. Control Clin Trials 2003;24:156–166.
28. Drossman DA, Camilleri M, Mayer EA, et al. AGA technical review on irritable bowel syndrome. Gastroenterology 2002;123:2108–2131.
29. Silvers D, Kipnes M, Broadstone V, et al. Domperidone in the management of symptoms of diabetic gastroparesis: efficacy, tolerability, and quality-of-life outcomes in a multicenter controlled trial. DOM-USA-5 Study Group. Clin Ther 1998;20:438–453.
30. Heymen S, Scarlett Y, Jones K, et al. Randomized controlled trial shows biofeedback to be superior to pelvic floor exercises for fecal incontinence. Dis Colon Rectum 2009;52:1730–1737.
31. Rao SS, Seaton K, Miller M, et al. Randomized controlled trial of biofeedback, sham feedback, and standard therapy for dyssynergic defecation. Clin Gastroenterol Hepatol 2007;5:331–338.
32. Chey WD, Lembo AJ, Lavins BJ, et al. Linaclotide for irritable bowel syndrome with constipation: a 26-week, randomized, double-blind, placebo-controlled trial to evaluate efficacy and safety. Am J Gastroenterol 2012;107:1702–1712.
33. Ducrotte P, Grimaud JC, Dapoigny M, et al. On-demand treatment with alverine citrate/simeticone compared with standard treatments for irritable bowel syndrome: results of a randomised pragmatic study. Int J Clin Pract 2014; 68:245–254.
34. Stone AA, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med 2002;24:236–243.
35. Yao CK, Gibson PR, Shepherd SJ. Design of clinical trials evaluating dietary interventions in patients with functional gastrointestinal disorders. Am J Gastroenterol 2013; 108:748–758.
36. Halmos EP, Power VA, Shepherd SJ, et al. A diet low in FODMAPs reduces symptoms of irritable bowel syndrome. Gastroenterology 2014;146:67–75 e5.
37. Passariello A, Agricole P, Malfertheiner P. A critical appraisal of probiotics (as drugs or food supplements) in gastrointestinal diseases. Curr Med Res Opin 2014; 30:1055–1064.
38. Haller D, Antoine JM, Bengmark S, et al. Guidance for substantiating the evidence for beneficial effects of probiotics: probiotics in chronic inflammatory bowel disease and the functional disorder irritable bowel syndrome. J Nutr 2010;140:690S–697S.
39. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. JAMA 2003; 290:1624–1632.
40. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health. Guidance for Industry. Patient-Reported Outcomes Measures: Use in Medical Product Development to Support Labeling Claims. Available at: http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf. Published December 2009. Accessed January 10, 2015.
41. Burke LB, Kennedy DL, Miskala PH, et al. The use of patient-reported outcome measures in the evaluation of medical products for regulatory approval. Clin Pharmacol Ther 2008;84:281–283.
42. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric evaluation and calibration of health-related quality of life

item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Med Care 2007;45(Suppl 1):S22–S31.

43. Spiegel B, Bolus R, Agarwal N, et al. Measuring symptoms in irritable bowel syndrome: development of a framework for clinical trials. Aliment Pharmacol Ther 2010;32:1192–1202.

44. Dean BB, Aguilar D, Barghout V, et al. Impairment in work productivity and health-related quality of life in patients with IBS. Am J Manag Care 2005;11(Suppl):S17–S26.

45. 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.

46. Camilleri M, Mangel AW, Fehnel SE, et al. Primary endpoints for irritable bowel syndrome trials: a review of performance of endpoints. Clin Gastroenterol Hepatol 2007;5:534–540.

47. Bijkerk CJ, de Wit NJ, Muris JW, et al. Outcome measures in irritable bowel syndrome: comparison of psychometric and methodological characteristics. Am J Gastroenterol 2003;98:122–127.

48. Mangel AW, Hahn BA, Heath AT, et al. Adequate relief as an endpoint in clinical trials in irritable bowel syndrome. J Int Med Res 1998;26:76–81.

49. Spiegel B, Camilleri M, Bolus R, et al. Psychometric evaluation of patient reported outcomes in IBS randomized controlled trials: a Rome Foundation report. Gastroenterology 2009;137:1944–1953; e1–e3.

50. 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;11:395–402.

Reprint requests

Address requests for reprints to: Jan Tack, TARGID, University of Leuven, Herestraat 49, 3000 Leuven, Belgium. e-mail: jan.tack@med.kuleuven.be.

Conflicts of interest

The authors disclose no conflicts.

Supplementary References

51. Drossman DA, Li Z, Toner BB, et al. Functional bowel disorders. A multicenter comparison of health status and development of illness severity index. Dig Dis Sci 1995; 40:986–995.
52. Ang D, Talley NJ, Simren M, et al. Review article: endpoints used in functional dyspepsia drug therapy trials. Aliment Pharmacol Ther 2011;33:634–649.
53. Mansoor LE, Dowse R. Effect of pictograms on readability of patient information materials. Ann Pharmacother 2003;37:1003–1009.
54. Tack J, Carbone F, Holvoet L, et al. The use of pictograms improves symptom evaluation by patients with functional dyspepsia. Aliment Pharmacol Ther 2014; 40:523–530.
55. Spiegel BM, Gralnek IM, Bolus R, et al. Clinical determinants of health-related quality of life in patients with irritable bowel syndrome. Arch Intern Med 2004; 164:1773–1780.
56. Sandler RS, Stewart WF, Liberman JN, et al. Abdominal pain, bloating, and diarrhea in the United States: prevalence and impact. Digest Dis Sci 2000;45:1166–1171.
57. Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit time. Scand J Gastroenterol 1997; 32:920–924.
58. Bharucha AE, Zinsmeister AR, Locke GR, et al. Risk factors for fecal incontinence: a population-based study in women. Am J Gastroenterol 2006;101:1305–1312.
59. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med 1993;118:622–629.
60. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–483.
61. Bergner M, Bobbitt RA, Carter WB, et al. The Sickness Impact Profile: development and final revision of a health status measure. Med Care 1981;19:787–805.
62. Khanna P, Agarwal N, Khanna D, et al. Development of an online library of patient-reported outcome measures in gastroenterology: the GI-PRO Database. Am J Gastroenterol 2014;109:234–248.
63. Patrick DL, Drossman DA, Frederick IO, et al. Quality of life in persons with irritable bowel syndrome: development and validation of a new measure. Diges Dis Sci 1998;43:400–411.
64. Koop A, Mosges R. The use of handheld computers in clinical trials. Control Clin Trials 2002;23:469–480.
65. Stone AA, Shiffman S, Schwartz JE, et al. Patient compliance with paper and electronic diaries. Control Clin Trials 2003;24:182–199.
66. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996;276:637–639.
67. Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. JAMA 2001;285:1987–1991.
68. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010;152:726–732.
69. Perneger TV. What's wrong with Bonferroni adjustments. BMJ 1998;316:1236–1238.
70. Guyatt G, Jaeschke R, Heddle N, et al. Basic statistics for clinicians: 2. Interpreting study results: confidence intervals. CMAJ 1995;152:169–173.
71. Nuovo J, Melnikow J, Chang D. Reporting number needed to treat and absolute risk reduction in randomized controlled trials. JAMA 2002;287:2813–2814.
72. Gore SM, Jones G, Thompson SG. The Lancet's statistical review process: areas for improvement by authors. Lancet 1992;340:100–102.
73. Katz MH. Multivariable analysis: a primer for readers of medical research. Ann Intern Med 2003;138:644–650.
74. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992;116:78–84.
75. Smith DG, Clemens J, Crede W, et al. Impact of multiple comparisons in randomized clinical trials. Am J Med 1987;83:545–550.
76. Assmann SF, Pocock SJ, Enos LE, et al. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet 2000;355:1064–1069.
77. Ioannidis JP, Evans SJ, Gotzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med 2004;141:781–788.
78. Campbell MJ, Julious SA, Altman DG. Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons. BMJ 1995; 311:1145–1148.
79. Detsky AS, Sackett DL. When was a "negative" clinical trial big enough? How many patients you needed depends on what you found. Archives of internal medicine 1985;145:709–712.
80. Makuch RW, Johnson MF. Some issues in the design and interpretation of 'negative' clinical studies. Arch Intern Med 1986;146:986–989.
81. Berry DA. Interim analyses in clinical trials: classical vs. Bayesian approaches. Stat Med 1985;4:521–526.
82. Geller NL, Pocock SJ. Interim analyses in randomized clinical trials: ramifications and guidelines for practitioners. Biometrics 1987;43:213–223.
83. Chow SC. Adaptive clinical trial design. Annu Rev Med 2014;65:405–415.
84. Pocock SJ. When to stop a clinical trial. BMJ 1992; 305:235–240.
85. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. Guidance for Industry Adaptive Design Clinical Trials for Drugs and Biologics. Available at: http://www.fda.gov/downloads/Drugs/.../Guidances/UCM201790.pdf. Published February 2010. Accessed January 10, 2015.
86. Pocock SJ. The pros and cons of noninferiority trials. Fundam Clin Pharmacol 2003;17:483–490.
87. US Department of Health and Human Services, Food and Drug Administration Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. Guidance for Industry Non-Inferiority Clinical Trials. Available at: http://www.fda.gov/downloads/Drugs/Guidances/UCM202140.pdf. Published March 2010. Accessed January 10, 2015.