Gastroenterology 2026;170:1388–1407

Design of Treatment Trials for Disorders of Gut–Brain Interaction

Gregory S. Sayuk,[1] Darren M. Brenner,[2] Florencia Carbone,[3] Simon R. Knowles,[4] Brian E. Lacy,[5] Heidi M. Staudacher,[6,7] and Alexander C. Ford[8,9]

[1]Division of Gastroenterology, Washington University in St. Louis School of Medicine, Saint Louis, Missouri; [2]Division of Gastroenterology and Hepatology, Northwestern University Feinberg School of Medicine, Chicago, Illinois; [3]Department of Gastroenterology and Hepatology, Leuven University Hospitals, Leuven, Flanders, Belgium; [4]Department of Psychological Sciences, School of Health Sciences, Swinburne University of Technology, Hawthorn, Victoria, Australia; [5]Division of Gastroenterology and Hepatology, Department of Medicine, Mayo Clinic, Jacksonville, Florida; [6]Food & Mood Centre, The Institute for Mental and Physical Health and Clinical Translation (IMPACT), School of Medicine and Barwon Health, Deakin University, Geelong, Victoria, Australia; [7]Department of Medicine, School of Translational Medicine, Monash University, Melbourne, Victoria, Australia; [8]Leeds Institute of Medical Research at St James's, University of Leeds, Leeds, United Kingdom; and [9]Leeds Gastroenterology Institute, Leeds Teaching Hospitals National Health Service Trust, Leeds, United Kingdom

Rigorous treatment trials have resulted in substantial therapeutic advances in disorders of gut–brain interaction (DBGI) over the past several decades. However, much of this work has been conducted in adults with irritable bowel syndrome, and some of the challenges of DGBI trials (eg, symptom-based diagnoses and outcomes) have experienced more limited progress, especially for less common DGBI. Randomized, placebo-controlled trials remain the time-tested gold standard for DGBI, although more recent pragmatic trials have demonstrated real-world efficacy. Moreover, treatment trials have expanded into dietary interventions and brain–gut behavioral therapies, now with greater guidance and recognition of the nuances of these nondrug trials. Development of validated patient-reported outcome measures and adherence to Good Clinical Practice standards will yield novel DGBI therapies and more sophisticated treatment strategies for the management of these complex disorders. This article summarizes best practice, methodology, conduct, and reporting of treatment trials in DGBI as well as regulatory guidance and recent advances in trial design.

Keywords: Abdominal Pain; Bowel Habit; Treatment; Randomized Controlled Trial.

Treatment trials are essential for proving the safety and effectiveness of therapies for disorders of gut–brain interaction (DGBI). During the past 30 years, advances in treatment options, especially for irritable bowel syndrome (IBS), have been supported by rigorous trials, leading to more approved drugs. However, for other DGBI, such as functional dyspepsia (FD), no treatments have received such validation or regulatory approval. For less prevalent DGBI, evidence is often limited to observational or anecdotal reports, leaving uncertainty about treatment effectiveness.

Conducting effective treatment trials in DGBI is challenging due to their variable symptoms and lack of reliable biomarkers. Most existing trials are short-term, despite the chronic and intermittent nature of these disorders. Additionally, few studies explore flexible or periodic treatment approaches, limiting our understanding of long-term management.

Despite these issues, the design of treatment trials in DGBI continues to make substantial progress and is becoming increasingly sophisticated. This chapter highlights the importance of rigorous trials in carefully defined patient groups, considering overlap both between DGBI and with other conditions. Although randomized, placebo-controlled trials remain the gold standard, newer study designs, such as pragmatic trials, may better reflect real-world effectiveness. Progress continues in the use of validated patient-reported outcome measures (PROMs) and tracking of gastrointestinal symptoms. Treatment approaches now include not only medications but also a growing number of dietary interventions, supplements, brain–gut behavioral therapies (BGBTs), and digital therapeutics. We discuss key aspects of trial design, measurement of outcomes, adherence to Good Clinical Practice, and future research priorities in this field.

Abbreviations used in this paper: **AE, adverse event; BGBT, brain–gut behavioral therapy; BSFS, Bristol Stool Form Scale; CC, chronic constipation; CONSORT, Consolidated Standards of Reporting Trials; CSBM, complete spontaneous bowel movement; DGBI, disorders of gut–brain interaction; DSMC, Data Safety and Monitoring Committee; EMA, European Medicines Agency; FD, functional dyspepsia; FDA, Food and Drug Administration; GERD, gastroesophageal reflux disease; 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 mixed bowel habits; IBS-QOL, irritable bowel syndrome quality of life; MCID, minimum clinically important difference; NRS, numeric rating scale; PICOT, Population under study, Intervention being tested, Comparator or control, Outcome of interest, and Time-related aspects; PROMIS, Patient-Reported Outcomes Measurement Information System; PROM, patient-reported outcome measure; QoL, quality of life; RCT, randomized controlled trial; SAE, serious adverse event; SF-36, Medical Outcomes Study 36-Item Short Form Survey.**

Most current article

0016-5085
https://doi.org/10.1053/j.gastro.2025.12.040

Identifying the Hypothesis and Research Question

A proper hypothesis is critical to the success of a treatment trial and should be based on available evidence. The pathophysiology of DGBI is complex, diverse, and incompletely understood. Recognizing this is critical to trial design. As an example, FD may result from any number of mechanistic processes; for example, impaired gastric accommodation, delayed gastric emptying, immune activation, or perturbations of the microbiome. Accordingly, a treatment trial evaluating the efficacy of a new prokinetic agent might only enroll patients with FD and delayed gastric emptying. As the research question is being developed, the investigator should consider what the knowledge deficits are in clinical practice and what needs to be addressed to improve clinical care. The question can be refined using the PICOT (Population under study, Intervention being tested, Comparator or control, Outcome of interest, and Time-related aspects, including treatment duration and timing of assessment of outcomes) format (Table 1). PICOT helps to determine the main objective of the trial as well as the appropriate primary and secondary end points, along with suitable methods to assess these outcomes.

Defining the Target Condition

Patient Population

Critical to the success of a clinical treatment trial is accuracy of the diagnosis of the DGBI under study. Some treatment trials may require a specific subset of patients with a DGBI. In the future, biomarkers may enhance diagnostic certainty or be used as outcome measures, although few examples exist currently.[1] Therefore, most treatment trials in DGBI rely on symptom-based diagnostic criteria to confirm the presence of the condition of interest and PROMs to assess their end points of interest.

Inclusion Criteria

The trial population needs to be defined carefully to provide clarity about which DGBI patients might benefit from the intervention under study and to be as representative as possible of the entire population of interest to allow extrapolation of the findings. Inclusion criteria should specify the DGBI of interest, stating the criteria used to define it, whether a certain subcategory is being studied (eg, IBS with diarrhea [IBS-D]), whether coexisting DGBI are permitted, and whether concomitant long-term therapies are allowed. Although carefully defined inclusion criteria lead to more homogenous study populations, they can also limit generalizability. Biomarkers for DGBI remain elusive, but other features may identify specific DGBI subsets for treatment trials, such as symptom subtypes (eg, bowel habits), pathophysiologic changes (eg, elevated duodenal mast cell counts), or physiological measures (eg, gastrointestinal transit).

Exclusion Criteria

The overarching goal of exclusion criteria is to minimize study design confounders. Areas to consider include comorbidities, prior surgery, medication history, and comorbid medical and psychological conditions. Overly strict exclusion criteria can restrict enrollment and may make a trial cohort less generalizable to the broader population. However, for a novel therapy, the trial population should be as homogenous as possible, the overarching goal

Table 1. Design Elements of Treatment Trials in Disorders of Gut–Brain Interaction According to the PICOT Format

Population under study (P)	What are the characteristics of the disease? How prevalent is the disease under study? How heterogeneous is the population (inter- and intrasubject variability)?
Intervention (I)	What is the active treatment(s) you want to study (drug/dietary intervention/brain–gut behavioral therapy/medical device)? What are the characteristics of the treatments being studied?
Comparator intervention (C)	Current standard of care, placebo or other inactive intervention, or no treatment?
Outcomes (O)	What are the primary study outcomes? What are the secondary and exploratory outcomes? Are there any other considerations (eg, was the intervention adhered to)?
Time-related aspects (T)	How long will the treatment be received? What is the timing of the assessment of outcomes?
Other considerations	What are the ethical considerations (eg, use of a placebo/sham treatment or an active standard therapy)? What is the expected time frame? What are the costs? What team members and expertise (eg, specialist statistician) are required? How complex are the logistics? Should the recruitment be single or multicenter? Will the setting be primary, secondary, tertiary, or quaternary care? Will it use a pragmatic or explanatory approach? How many dropouts would be expected?

being to maximize the chance of detecting a therapeutic signal. Subsequent studies might then examine a broader patient population, applying less restrictive inclusion and exclusion criteria to more closely approximate a "real-world" setting.

Managing Overlap of Disorders of Gut–Brain Interaction, Comorbidities, and Disease Modifiers

Almost all treatment trials limit study enrollment to participants with a single DGBI. Studying a homogeneous population has advantages, but patients often have overlapping DGBI.[2] Treatment trials restricted to a single DGBI limit potential enrollment numbers and lessen applicability of the findings. Future trials should consider enrolling patients with overlapping DGBI to enhance the clinical applicability of the observations.

Nevertheless, it is critical that the DGBI under investigation is the primary symptomatic condition. Thus, a treatment trial in IBS could allow coexisting FD, but the symptoms of IBS must be predominant, and the patient must be able to differentiate between the abdominal pain experienced with FD and the pain experienced with IBS. Similarly, coexisting common mental disorders, such as anxiety or depression, are common in DGBI and generally should not be considered an exclusion criterion, unless the symptoms affect internal validity (eg, the presence of suicidal ideation). Disorders of somatic hypersensitivity, such as fibromyalgia or migraine headaches, are frequently observed in DGBI. These should be acknowledged but should not be an exclusion criterion, unless their symptoms predominate. Trials may plan to stratify patients based on gastrointestinal symptoms, psychiatric disorders, or somatoform conditions.

Clinical Trial Design in Disorders of Gut–Brain Interaction

Unique Challenges for the Design of Treatment Trials for Disorders of Gut–Brain Interaction

There are multiple challenges in the design of treatment trials in DGBI, including factors that must be reconciled to evaluate potential therapies for DGBI for efficacy and safety.[3] Other common challenges stem from high placebo response rates,[4,5] comorbidities (ie, visceral, somatic, or psychological), or treatments that can potentially exacerbate, enhance, or intensify DGBI symptoms. Standardizing the design and end points for treatment trials in DGBI also presents unique challenges. Although the United States Food and Drug Administration (FDA) has developed enrollment and clinical end point thresholds for drug trials in IBS and chronic constipation (CC)[6] such guidance is lacking for many other DGBI and many potential therapies, including dietary interventions, nutraceuticals, BGBT, microbiome-targeted interventions, or medical extended reality. Finally, bias or "systematic error" occurring at any phase from design to publication may negatively affect the trial or result in inaccurate findings (Supplementary Table 1).

Masking Process

Although subtle differences in terminology have been debated, "blinding" and "masking" are often used synonymously and refer to the process whereby individuals involved in clinical trials are made intentionally unaware of treatments being provided or of the study hypothesis to minimize potential biases that may affect trial outcomes (Supplementary Figure 1). Unmasked trials are prone to bias and may lead to larger treatment effects.[7] A single-masked design may be used when the investigator requires knowledge of the treatment assignment or, alternatively, the participant, but not the investigator, knows the assignment. In double-masked studies, patients and investigators are both unaware of treatment allocation. Triple-masked studies involve masking the participants, investigators, and data managers. The greater the degree of masking the lower the risk of bias. Masking participants is more difficult when active drugs cause predictable side effects and in trials of nondrug interventions, such as diet or BGBTs, where use of a placebo is challenging. However, masking participants to the study hypothesis is an alternative.

Upon completion of trials, the success of masking can be evaluated by asking participants and investigators to guess which treatment was administered, although this approach is controversial.[8] Alternatively, scales to assess masking success objectively, via detection of a low degree of masking, response bias, and differing behavior between trial arms, have been proposed.[9]

Randomization

It is essential that investigators provide detailed and accurate accounts of their randomization and concealment techniques. Randomization is the process by which participants are assigned by chance to different interventions, ensuring objective unbiased assignment to study arms and equal distribution of potential confounding variables that may influence outcomes. Unrestricted randomization simply involves assigning participants without taking known covariates into account. This strategy is potentially problematic, because it can lead to chronological bias and unequal distribution over time. Stratified randomization based on prognostic factors, such as age, sex, or risk factors (eg, postinfection status in IBS) that likely influence treatment outcomes, ensures these covariates are balanced across treatment groups.

Potential imbalances can be minimized further using block randomization with equal block sizes, ensuring equal numbers of participants are assigned to each study arm, because the intervention assignment is randomized within the block. The size of blocks used and their order are concealed to avoid allocation bias. Key to this study arm assignment process is allocation concealment, which prevents the research team from influencing group assignment. Centralized randomization companies or pharmacies,

web-based data collection platforms, or sequentially numbered, opaque, sealed envelopes, can facilitate allocation concealment,[10] although the latter may be less rigorous.

Selecting a Control Group

Drug trials often use a placebo, with the lack of gold standard treatments for a given DGBI used as justification for this choice. However, this may be considered unethical, particularly where effective treatments already exist. In this situation, an alternative approach would be to consider the use of an accepted standard of care or a "rescue" medication on an as-required basis. Considerations specific to selection of control interventions in nondrug trials are discussed below.

Placebo and Nocebo Responses

Placebo response. Placebo response is an improvement in symptoms experienced in response to an inert treatment not attributable to the treatment itself (Figure 1).[11,12] Placebo response rates of 35% to 45% are

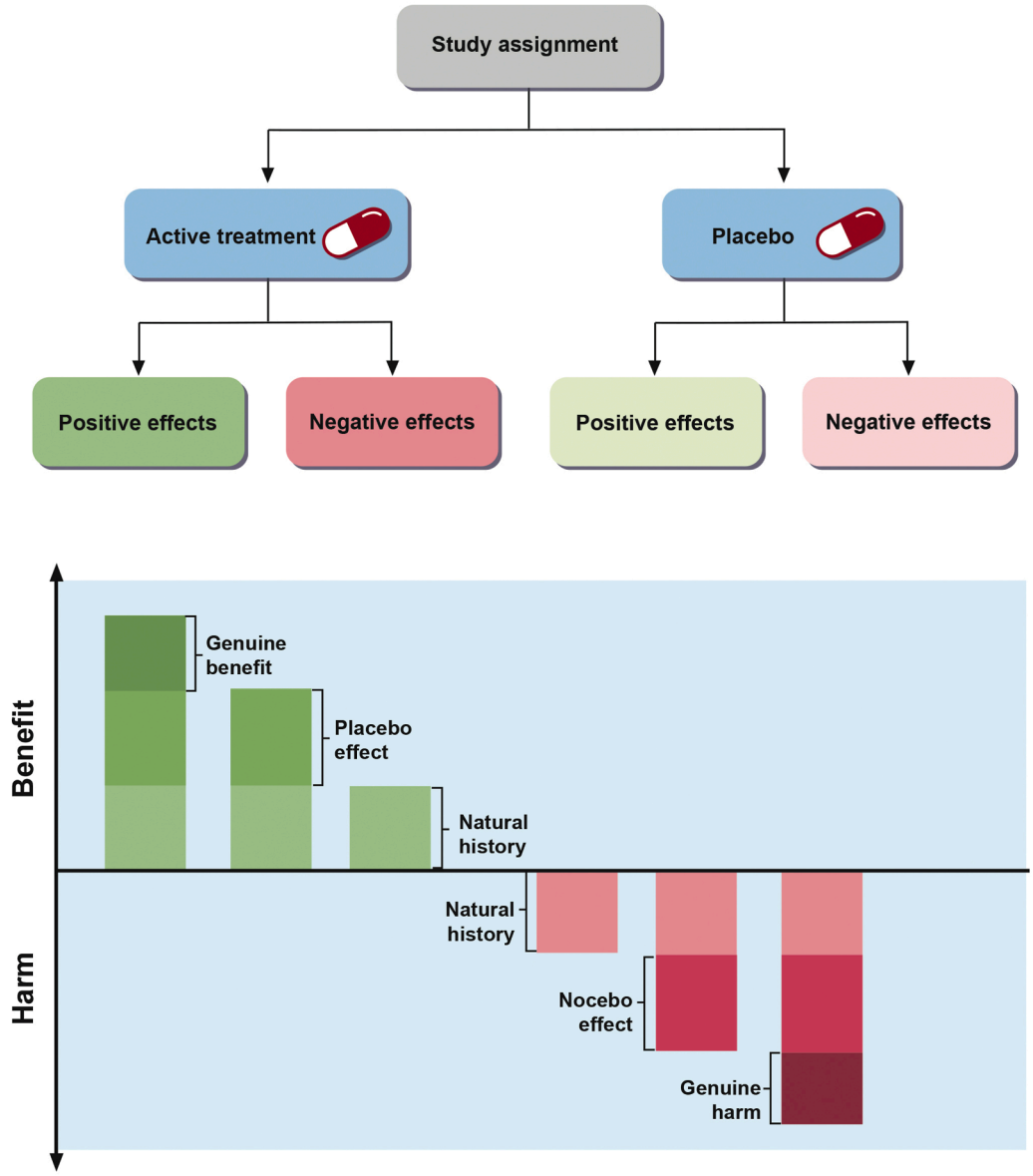

Figure 1. Beneficial and harmful effects of active drug and placebo in a treatment trial. Participants assigned to active treatment and placebo can experience both positive effects ("benefit") and negative effects ("harm"). The components of beneficial changes ("benefit") in patients' clinical symptoms and signs include the genuine treatment effect produced by the active intervention, the placebo effect, and the natural history of untreated disease, together with regression to the mean and a spontaneous remission of symptoms. Conversely, treatment "harm," or the development of AEs or side effect symptoms, relates to combined effects of genuine harm due to undesired pharmacologic or other effects of the active treatment, nocebo effects, and again the natural history of untreated disease with spontaneous worsening of symptoms. Adapted from Özdemir V, Endrenyi L. Rethinking clinical trials and personalized medicine with placebogenomics and placebo dose. OMICS 2021;25:1–12.[12]

reported in DGBI trials,[4,5] although these rates are in line with those observed in other gastrointestinal conditions and other chronic painful disorders.[13] Higher placebo rates may affect power calculations, leading to inadequate sample sizes, study underpowering, and threaten the trial's potential to detect true statistically significant differences between an experimental treatment and a placebo control. Thus, minimizing placebo response in DGBI trials is essential to drug development and determination of the true efficacy of the experimental intervention.

Multiple factors have been proposed as potential mediators of placebo response in trials of drugs in IBS. Use of standardized, more stringent, entry criteria (eg, Rome V criteria), patient-reported outcomes, longer durations of therapy, and more rigorous end points (eg, FDA composite responder definitions in IBS) are associated with significantly lower placebo response rates.[4,14]

Nocebo response. Nocebo responses comprise a worsening of symptoms and are often reported as the proportion of participants experiencing side effects or adverse events (AEs) during an inert intervention (Figure 1).[15] Nocebo effects may be attributable to neurobiological and psychological mechanisms of expectation, usually due to cognitions and worries that an intervention will cause harm. In DGBI, where negative expectations of treatments can be formed faster than positive ones, the nocebo effect may influence clinical outcomes more substantially than placebo effects through reductions in treatment compliance and confidence in therapy, thus biasing trial outcomes.[16,17] The nocebo effect may also increase AE rates in both placebo and treatment arms, confounding accurate assessment of medication safety.

Because DGBI patients often do not respond to initial interventions and cycle through various treatments, previous unsuccessful treatment or negative experiences with specific treatments can hamper positive effects of subsequent treatments.[18] Nocebo effects have also been demonstrated to reduce the efficacy of active treatment.[19]

Baseline Observation Versus Placebo Run-In

Before randomization, run-in periods allow prospective observation of potential participants to determine study eligibility. Run-ins reflect a "bias by design" in that they increase the power to detect intention-to-treat effects. However, this comes at the expense of clinical applicability.[20] Potential benefits of run-in periods include the ability to exclude patients who are noncompliant or do not meet the required symptom severity threshold and to allow participants to reflect further on trial participation.[20,21] However, such prerandomization run-in periods should be for observation only, not for implementing any form of intervention with either placebo or active treatment. Run-ins with placebo to identify and reduce placebo response are problematic because they create a selection bias and are not recommended.[22] Evidence suggests they do not achieve the desired effect, with comparable placebo response rates to trials without a placebo run-in.[22]

Choice of Study Design

The choice of study design is largely determined by the phase of development of the treatment (Supplementary Table 2) and the primary end point of the trial. Designing a treatment trial in DGBI can be facilitated using the aforementioned PICOT approach. The most common and important trial designs are dealt with below, with detail on these and other designs provided in Table 2 and in Figure 2. Mechanistic elements may be incorporated into any trial design to evaluate physiological processes underlying the beneficial effects of the active intervention.

The gold standard for assessing the efficacy of a new treatment is a double-masked, randomized, placebo-controlled, parallel group trial. In this design, patients are divided randomly into parallel groups to receive, in a masked fashion, various treatments, one of which is a placebo. This design is used to demonstrate superiority of an active intervention over a placebo. The design can be modified depending on the aim of the study. For example, some studies will include more than 2 parallel arms in a multiarm trial, whereas others may examine different doses or regimens of the same drug in a dose-ranging trial. Classically, it consists of an initial therapy-free observation period, a standard trial period, and may include a washout period once the trial is completed.

If the aim of the study is to address the effectiveness, convenience, or safety profile of a new therapy against another treatment considered to be the current standard of care, trials of noninferiority may be used. In this case, the study will compare a new therapy with a standard therapy (control) in a parallel group design to show it is at least as effective.

In a randomized crossover trial, each patient receives both active and control treatments for a specific period. The order of the intervention is randomized and masked, with treatment intervals separated by a washout period, whose duration depends upon the presumed longevity of response to the active intervention. Data analysis should account for period-by-treatment and sequence effects and, if sequence effects are identified, should be limited to data collected during the first treatment period.

Factorial trials evaluate 2 or more interventions in various combinations. This design is most appropriate when evaluating the interactive effects of combination therapy that might be desirable or where there may be a plausible mechanistic rationale for combination treatments, such as a dietary intervention and a probiotic. The effects of these different interventions can be measured, and a smaller sample size is required because results within comparator groups can be pooled to examine main effects.

Adaptive trials allow for several different treatments to be evaluated concurrently and for data accumulated during the trial to be used to modify the trial prospectively, increasing efficiency. Such adaptations include discontinuing an arm for futility or safety concerns, closing recruitment to an arm early if clinical efficacy is achieved, or altering sample size or allocation of enrollees based on early efficacy signals in 1 or more arms of the trial. In these

Table 2. Summary of Available Trial Designs With Advantages and Disadvantages of Each

Type of trial	Summary of design	Advantages	Disadvantages
Double-masked, randomized, placebo-controlled, parallel group trial	Patients are divided randomly into parallel groups to receive, in a masked fashion, various treatments, one of which is a placebo.	Classical, gold standard approach. Appropriate when concerns about carryover effects are present. Beneficial when disease variability may occur during the trial.	May require a large sample size.
Noninferiority trial	As for parallel group trial, with the aim being to confirm noninferiority of a new therapy against another treatment currently considered to be the standard of care.	May be easier to recruit to, given that all participants receive an active intervention. Allows alternative interventions to be studied in people with the condition, which may be less costly, have fewer side effects, or may be less cumbersome to apply.	Larger sample sizes are necessary to prove noninferiority or equivalence, because the noninferiority margin is smaller than the effect size in superiority trials. Use of a wide noninferiority margin may lead to false-positive results. May be difficult to recruit an appropriate sample size based on the prevalence of the underlying condition.
Crossover trial	Each patient receives both treatments (a comparator and control) for a specific period with the order of the intervention randomized and masked and treatment intervals separated by a washout period.	Trials require smaller sample sizes. Each patient serves as her or his own control, increasing the sensitivity to detect changes (less variability in measurements). May facilitate recruitment because if one treatment is perceived to be "better" by the participants, everyone will receive the perceived better treatment.	Carryover effect of treatments. In drug trials, there is a higher likelihood of unmasking due to side effects of the drug. Studies take longer, increasing the risk of missing data and patient drop out, potentially compromising the study power and complicating data analysis and interpretation.
N-of-1 trial	A variation of crossover trials, with similar benefits and limitations. This trial design helps overcome logistic and statistical limitations in rarer DGBI, with each patient acting as their own experimental and control group and receiving each treatment in a random order a repeated number of times during a trial. Washout periods between interventions may be necessary.	Can help assess treatments that elicit heterogenous responses in different subjects (eg, subgroups of patients within one DGBI).	As for crossover trials and require complex statistical analysis.
Factorial trial	Evaluate $\geq$2 interventions in various combinations; eg, in a classic 2$\times$2 design with placebo, patients are randomized into 4 groups: (1) treatment A plus placebo B; (2) treatment B plus placebo A; (3) both treatment A and treatment B; (4) placebo A and placebo B.	Smaller sample size. Allows for greater statistical powering. May be easier to recruit to because there is a higher chance for participants to receive an active intervention. Permits studying synergistic effects of 2 treatment modalities.	Complex statistical analysis. If an interaction exists between treatments, a loss of power is possible.

Table 2. Continued

Type of trial	Summary of design	Advantages	Disadvantages
Randomized withdrawal trial	All patients initially receive an open-label intervention and, at a predefined time point, are classified as responders or nonresponders. Responders are subsequently randomized to either continue to receive the intervention or a placebo during the second phase of the trial. Response data are assessed only during the withdrawal part of the trial, when active treatment is compared with placebo.	Reduced time on placebo, since only responders are randomized to placebo, thereby giving an ethical advantage. Increased acceptability to potential participants and, hence, facilitates recruitment.	Carryover effect from the enrichment phase. The treatment effect may be overestimated, because only responders are included.
Adaptive trial	Allows for several different treatments to be evaluated concurrently and for data accumulated during the trial course to be used to make modifications prospectively during the study. These include discontinuing an arm for futility or safety concerns, closing recruitment to an arm, in the event that clinical efficacy is achieved, or altering sample size or allocation of enrollees based on early efficacy signals in $\geq$1 arms of the trial.	Reduced trial time and cost by tackling multiple research questions simultaneously under a single protocol. May be more efficient and ethical because interim adaptations are allowed. Participants are more likely to be assigned to active treatment. May be a better reflection of real clinical practice.	Possible dismissal of a potentially beneficial treatment. Complex statistical analysis is required.
Proof-of-concept trial	Exploratory trials that are typically done in the early clinical development phase of a drug to provide evidence that a certain dosage may be efficacious for either the condition, or the subset of patients, of interest. They may be used to determine whether a specific dose of a drug, or a nondrug intervention, is taken on to phase 3 trials.	Requires a smaller sample size. Usually allows greater latitude for statistical significance testing.	May be underpowered for the end point of interest or for the study of a suboptimal dosage of a drug, resulting in the abandoned development of a potentially efficacious drug. Lack of generalizability of findings to broader patient populations.
Pragmatic trial	Takes place in real-world clinical practice settings with limited control of variables. Hence, they usually relax eligibility criteria to correspond with everyday clinical care.	Provides real-world evidence. Outcome measures considered from the patient's perspective and intention-to-treat analysis.	Increased number of variables requiring statistical correction. Greater study population heterogeneity.

trials, all experimental treatments are evaluated at planned time points to control for predefined outcomes, facilitating decisions to continue or terminate treatments. The FDA has provided core principle guidance for such trials.[23]

Whereas traditional randomized controlled trials (RCTs) are designed to limit confounding and biases, the results obtained are not always practical or applicable to the real world. Pragmatic trials, in contrast, take place in real-world settings with limited control of variables. Such trials usually relax eligibility criteria to correspond with everyday clinical care.[24] This study design increases the likelihood that the trial outcomes are relevant to more patients and that processes are similar to routine practice. Developing a fully pragmatic trial is challenging, and hence, such studies are instead often designed with multiple pragmatic features. Tools exist to assist with pragmatic trial design.[25]

Design of Trials for Nondrug Interventions in Disorders of Gut–Brain Interaction

Challenges unique to designing nondrug trials include selecting appropriate controls and masking participants. Recommendations for placebo (sham) controls exist for

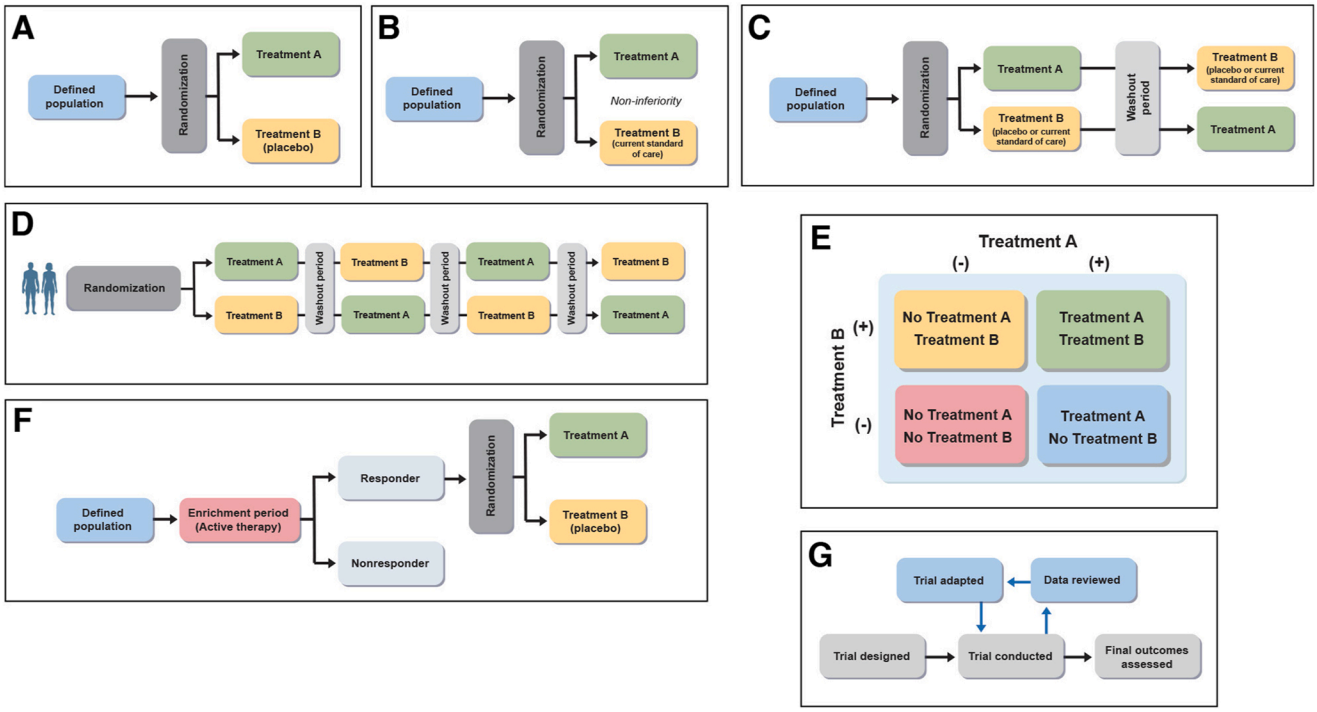

Figure 2. Representation of different treatment trial designs depicting (*A*) double-masked, randomized, placebo-controlled, parallel-group trial; (*B*) noninferiority/comparative effectiveness trial; (*C*) crossover trial; (*D*) N-of-1 trial; (*E*) factorial (2×2) trial; (*F*) randomized withdrawal trial; and (*G*) adaptive trial.

trials of whole-diet interventions.[26] For example, a sham diet must appear as a special dietary intervention and alter intake of some dietary components, but not the component of interest or others that might influence the outcome of interest. Recommendations for the development of sham interventions are lacking for studies assessing procedures, devices, BGBTs, or digital therapeutics. Use of a waiting list or no treatment controls is common in nondrug trials. These are limited by an inability to mask adequately and are most appropriate in proof-of-concept or feasibility trials.[27] These can then be followed by comparative trials that incorporate an active control, trials with higher-risk sham interventions (eg, endoscopic procedures), or trials with placebos requiring intensive development.

Ease of access to information on dietary, lifestyle, or psychological management of health through the media or scientific literature may inform certain beliefs about the therapeutic potential of a particular treatment. Therefore, expectations of benefit may be stronger in nondrug trials. Where placebo or masking is not possible, control participants may have strong negative expectations of improvement, which will overestimate efficacy of the intervention. Choosing a control that is equally attractive or is likely to have treatment effects allows both intervention and control to be presented as potentially beneficial and may neutralize participant expectations. Masking the intent of the research may also be ethically appropriate, without violating principles of informed consent, to reduce expectation bias.

Investigators may also have preconceived beliefs about the benefit of a treatment, and these beliefs can, consciously or unconsciously, influence patients. This investigator bias is especially problematic in nondrug trials in the absence of double-masking. To minimize this, details of the interventions for all treatment arms should be documented before the trial is initiated, and treatment fidelity should be assessed. Nondrug trials can be influenced by the training and experience of the researcher applying the intervention. This may explain variability in trial outcomes. Interventions should be manualized, with detailed documentation of their components, delivery procedures, and the training and qualifications of those administering them. Standardized trial-specific training should be provided, with ongoing training sessions available, if necessary. When multiple people are delivering the intervention, precision or proficiency in implementation and differences in therapeutic alliance can be assessed in secondary analyses.

Just as in drug trials, adherence to treatment should be assessed. Studies in which treatments are self-administered, such as dietary interventions or digital cognitive behavioral therapy, should report this objectively, where possible, with a threshold informed by previous literature, established a priori, to classify adherence and non-adherence.

Considerations for Pediatric Trials in Disorders of Gut–Brain Interaction

Many treatments effective in treating DGBI in adults have proven to be ineffective in pediatric studies. Of the drugs available in adults with DGBI, to date, only linaclotide has been approved by the FDA for use in children with chronic constipation.[28] Although many topics presented in

this article are applicable to treatment trials irrespective of patient age, there are unique design and ethical considerations when performing trials in children with DGBI. Some challenges unique to pediatric DGBI trials include a paucity of pediatric PROMs and complexity in their development, difficulty in symptom reporting, and reluctance to enroll children in clinical trials (Supplementary Table 3).[29] Working groups have been convened to establish standardized criteria for pediatric DGBI trials, and guidelines have been published by the Rome Foundation.[30,31]

Patient-Reported Outcome Measures

Definition, importance, and selection of patient-reported outcome measures. The FDA defines a PROM as "any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else."[32] PROMs measure a "construct," or defining concept, such as pain,[33] and may be generic or disease-specific. Generic PROMs are useful to compare outcomes with non-DGBI disease groups (eg, vs inflammatory bowel disease or when assessing nondisease variables). For example, health-related quality of life (HRQoL) can be measured using the generic Medical Outcomes Study 36-Item Short Form Survey (SF-36),[34] or DGBI-specific tools such as the IBS Quality of Life (IBS-QOL) questionnaire.[35]

PROM selection depends on a variety of factors, including reliability and validity, psychometric performance, change sensitivity, instrument conceptual framework, the intended condition and study population, the recall period, and the respondent burden.[32] An instrument that is intended for use in one specific population may not be fit for purpose across other target populations. Consequently, use and comparison of unique PROMs across multiple conditions or samples is potentially problematic. This section is intended to help guide the assessment and selection of PROMs for DGBI treatment trials.

Classification of outcome measures. DGBI symptoms can be measured across various attributes, including their frequency, severity, and bothersomeness.[36] Symptoms can also be measured as a combined symptom cluster (eg, an integrative constipation PROM assessing straining, sense of incomplete evacuation, stool form, and stool frequency). There is increasing recognition of the importance of psychological, emotional, and social function, and HRQoL in treatment trials in DGBI. PROMs are also available to assess dietary outcomes, work productivity, and resource utilization in DGBI (Supplementary Figure 2).

Regulatory agency guidance for clinical trial outcome measures. In 2012, the FDA published its first set of recommendations for the design of drug trials in IBS.[6] In 2015, the European Medicines Agency (EMA) released a similar IBS-specific document, with concordant outcome recommendations (Supplementary Table 4).[37] For IBS trial enrollment, the FDA endorsed a weekly average of worst daily abdominal pain $\geq$3 on a 0- to 10-point numeric rating scale (NRS) for study inclusion, and recommended the Bristol Stool Form Scale (BSFS) to characterize stool form.[6] Although this guidance limited the potential for negative trial outcomes due to floor effect, it also resulted in the exclusion of individuals with milder symptoms. Further, this guidance presumed that IBS study participants would have daily symptoms, despite the recognition that individuals often experience much less frequent symptoms.[38] The FDA did not establish responder criteria for IBS with mixed bowel habits (IBS-M) or IBS unclassified, although IBS-M responder criteria have been proposed by the EMA (Supplementary Table 4).[37]

The IBS Working Group section of the Critical Path Institute's Patient-Reported Outcome Consortium developed the Diary for IBS Symptoms-Constipation, -Diarrhea, and -Mixed to identify future patient-centered end points for IBS treatment trials.[39] The Diary for IBS Symptoms-Constipation was used to establish 3 key abdominal sensory symptoms as the most meaningful to patients: pain, discomfort, and bloating. These were combined to establish an abdominal score, with each measured on a 0- to 10-point NRS.[40]

PROMs used as primary end points leading to FDA and EMA drug approval for constipation (both CC and opioid-induced) have measured the number of spontaneous bowel movements or complete spontaneous bowel movements (CSBMs) on daily diaries.[41,42] Stool form is reported using the BSFS[43] or Likert scales, because stool frequency alone appears to be insufficiently reliable to evaluate constipation.[44]

For FD, the Leuven Postprandial Distress Scale,[45] the FD Symptom Diary,[46] and the experience sampling method-PROM[47] were developed in line with FDA guidance.[23] The EMA has expressed support for using the Leuven Postprandial Distress Scale in trials assessing postprandial distress syndrome.[48] However, the FDA has not endorsed any specific instrument for FD, to date.

Binary outcome measures. Many treatment trials in DGBI have used a binary PROM as an end point, such as "adequate relief," or "satisfactory relief," which provides a dichotomous responder status ("yes/no").[49] Binary end points are easy to administer, straightforward to interpret, and have construct validity compared against other DGBI end points.[50,51] Pragmatic trials, therefore, may consider binary outcome measures as a primary end point.

The FDA discourages binary end point use in clinical registration trials due to several limitations: (1) they do not capture whether benefit is achieved across multiple signs and symptoms; (2) responses may depend on baseline severity; (3) they do not capture the magnitude or direction of symptom changes; and (4) binary end points were not derived from patient focus groups as the gold standard approach.[52]

Integrated symptom questionnaires. Integrated symptom questionnaires incorporate measures of multiple symptoms, often combining them into a single score (eg, the IBS Severity Scoring System), thereby providing a global assessment of severity across a range of symptoms. Integrated symptom outcomes could be used as secondary end points in drug treatment trials.[53]

Pictograms. Written DGBI symptom descriptors require an adequate level of literacy and comprehension

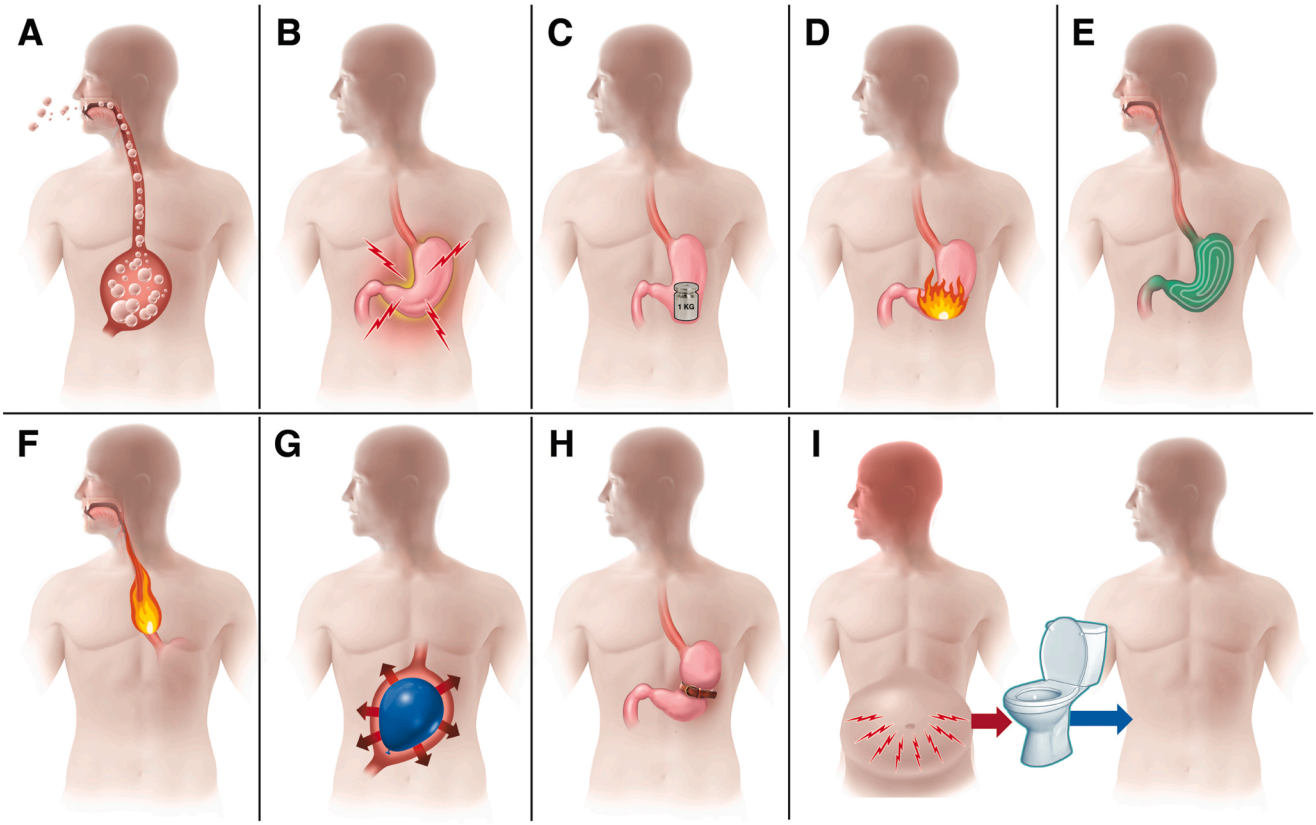

Figure 3. Pictograms for the assessment of DGBI symptoms depicting (*A*) belching; (*B*) epigastric pain; (*C*) postprandial fullness; (*D*) epigastric burning; (*E*) nausea; (*F*) retrosternal burning; (*G*) bloating; (*H*) early satiation; and (*I*) lower abdominal discomfort or pain relieved by defecation.

as well as abstract thinking. Pictograms overcome some of the limitations of written questions by conveying messages through pictorial or symbolic representation (Figure 3). An example of a DGBI pictogram is the BSFS.[43] In FD, validated pictograms used in conjunction with a questionnaire lead to better concordance between patient-reported symptom severity and frequency and a physician's evaluation.[54,55] Pictograms measuring the presence and severity of symptoms can be used in isolation (ie, without verbal descriptors), a useful method when language lacks specific terms for symptoms.

Measuring disorders of gut–brain interaction symptoms. This section will review the measurement of some of the main gastrointestinal symptoms in DGBI and the essential considerations and challenges in their assessment. The National Institutes of Health Patient-Reported Outcomes Measurement Information System (PROMIS)-Gastrointestinal Symptom Scales cover a range of gastrointestinal symptoms,[56] but other questionnaires are also discussed.

Measuring functional heartburn. There are few functional heartburn treatment trials, and their outcome measures are often extrapolated from the gastroesophageal reflux disease (GERD) literature rather than being developed specifically for functional heartburn. GERD instruments applied to functional heartburn include ReQuest,[57] the GerdQ,[58] the Reflux Disease Questionnaire,[59] the Gastrointestinal Short Form Questionnaire,[60] and the GERD Analyzer (GERDyzer) tool.[61]

Measuring satiation and fullness. Symptoms of satiation and fullness can be assessed with previously described questionnaires, such as the Leuven Postprandial Distress Scale or FD Symptom Diary.[45,46] In addition, the Dyspepsia Symptom Severity Index, Leeds Dyspepsia Questionnaire, and Glasgow Dyspepsia Severity Score have been used to assess the severity of gastroduodenal symptoms.[62–64]

Measuring abdominal pain. Chronic, recurrent, abdominal pain is a cardinal symptom of IBS and a feature of many other DGBI. Reporting of pain is a *sine qua non* to diagnose IBS, centrally mediated abdominal pain syndrome, and other DGBI. A challenge to the objective assessment of abdominal pain is the recognition that considerable fluctuations in pain can be experienced throughout the day in patients both with and without a DGBI (Figure 4).[65,66]

Established measures of abdominal pain focus primarily on pain intensity. Several simple, quick, methods of abdominal pain intensity assessment have been used in studies, including visual analog scales, NRS, and verbal rating scales.[65] Assessing abdominal pain symptoms in real time can be conducted using ecological momentary assessment. This uses digital devices carried by the patient to record current symptoms throughout the day following audio prompts, addressing both daily variations in pain and the potential for recall bias that may result in overestimates

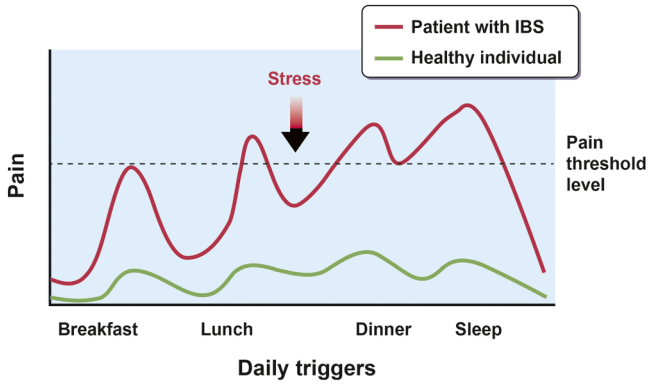

Figure 4. Theoretical representation of daily variations in abdominal pain symptoms in patients with IBS and healthy individuals. Adapted from Mujagic Z, Keszthelyi D, Aziz Q, et al. Systematic review: instruments to assess abdominal pain in irritable bowel syndrome. Aliment Pharmacol Ther 2015;42:1064–1081.[65]

of abdominal pain.[66,67] Baseline assessment of pain has important implications on pain responses in DGBI treatment trials, and a minimum 7-day diary is the standard approach.[68]

Measuring abdominal discomfort. Inconsistencies in the theoretical and operational definitions of pain and discomfort are highlighted in the literature.[69] Individual cognitive interviews have suggested that nearly 90% of patients with IBS consider pain and discomfort as different.[70] Together, these data support the measurement of discomfort as a nonspecific abdominal symptom separate from the pain construct. It can be measured similarly to abdominal pain.

Measuring abdominal bloating and distension. Abdominal bloating (a perception of increased abdominal fullness, tightness, or swelling) and distension (a physical manifestation of objective increases in abdominal girth) often coexist in patients but may be experienced individually.[71] The presence of abdominal bloating is regarded as a supporting symptom in IBS and is part of the criteria for functional abdominal bloating. Although bloating is often a therapeutic priority clinically, treatment trials historically have relegated bloating and distension to secondary end points or symptoms considered as part of a composite score.[72]

Open-ended qualitative interviews have suggested that patients with IBS categorize bloating by how it looks (eg, "full of air") or how it feels (eg, "tightness," "fullness").[36] Although some patients identify with looking "distended," many patients perceive this term as "very technical," or "overly clinical." Bloating and distension frequently coexist, but only weakly correlate.[73] The Mayo Bloating Questionnaire has been developed as an easy to use instrument for implementation in clinical settings and research studies.[74] Abdominal distension (girth) can be measured accurately and objectively using abdominal inductance plethysmography.[75]

Measuring stool frequency and form. As with abdominal pain, abnormalities of stool form or frequency are cardinal symptoms of IBS.[38] These symptoms are also required to meet Rome criteria for CC and functional diarrhea.[38] Although there are several validated questionnaires for use in CC, there are none for functional diarrhea. The most widely used instrument in treatment trials of CC is the validated Patient Assessment of Constipation-Symptoms questionnaire.[76] However, the questionnaire does not measure stool frequency, and current FDA end points for CC treatment trials require assessment of self-reported CSBMs per week, with response defined as a minimum increase (eg, of at least 1 CSBM per week) above baseline. A Bowel Function Diary has been validated for use in opioid-induced constipation treatment trials.[77]

The IBS Severity Scoring System does not assess stool form or frequency,[78] but instead asks patients to report how satisfied they are with their bowel habit. Similarly, the Gastrointestinal Symptom Rating Scale-IBS does not assess stool form or frequency,[79] although it does gauge bothersomeness of constipation, diarrhea, loose stools, and hard stools. Other validated questionnaires for IBS, such as the Bowel Symptom Severity Scale, only capture stool form and frequency.[80]

The FDA and EMA have made identical recommendations for the assessment of stool form in IBS-D and stool frequency in IBS with constipation (IBS-C), the former by a measure of stool form only, and the latter by a measure of stool frequency only.[6,81,82] For IBS-D this consists of a pictorial and verbal description of stool form by the patient via the BSFS.[43] For IBS-C, this consists of assessing patient-reported stool frequency measured by the number of CSBMs per week, as is done for CC.

Measuring urgency. Urgency is an unpleasant sensation of needing to rush to the toilet to prevent soiling,[83] which is most commonly reported in IBS-D, although it can also be experienced in IBS-C or IBS-M. Usually regarded as a key secondary end point, urgency has been included as a primary end point in some IBS trials.[84] Urgency questions are incorporated into PROMs for other non-DGBI, such as fecal incontinence,[85] but validated questionnaires measuring urgency are lacking in IBS. The PROMIS-Diarrhea Scale includes 3 Likert-style response items relating specifically to urgency. Despite this, there is a clear lack of urgency-specific PROMs for IBS, and consequently, the FDA considers urgency to be an exploratory end point.[6]

Measuring other disorders of gut–brain interaction symptoms. Although there are few validated PROMs to capture symptoms in DGBI other than IBS and FD, the PROMIS- Gastrointestinal Symptom Scales cover a range of other gastrointestinal symptoms, including incontinence, nausea and vomiting, urgency, and reflux.[56] Other specific PROMs have been developed for dysphagia, globus, supragastric belching, and dyssynergic defecation.[86–90] When PROMs for less common DGBI are available, but have not been subjected to extensive psychometric testing, they are recommended to be used in conjunction with other generic validated PROMs or PROMs that have been validated in other conditions.

Measuring health-related quality of life. Quality of life (QoL), a measure of overall perceived biopsychosocial well-being, is a frequently used PROM in DGBI treatment trials. Differentiation between QoL and HRQoL is complex,

but at its simplest, HRQoL scales reflect more on the patient's pathology and associated treatment.[91]

Reflecting this concept of pathologic conceptualization of QoL, the FDA defines HRQoL as "a multidomain concept that represents the patient's general perception of the effect of illness and treatment on physical, psychological, and social aspects of life."[92] HRQoL instruments can be further defined as generic (not condition or disease-focused) or DGBI-specific.

The SF-36 is a well-established generic HRQoL scale[93] that assesses overall physical and mental functioning across 8 domains. Due to its strong psychometric properties and established norms, the SF-36 is a suitable PROM for trials in DGBI.[93,94] The PROMIS Global-10 assesses HRQoL across physical and mental health domains and, like the SF-36, has strong psychometric properties and established norms.[95,96] A recent study undertook a comprehensive evaluation of QoL across all DGBI using the PROMIS.[97,98] A series of confirmatory factor analyses across different DGBI (both individual and combined by anatomical region) found that the scale overall lacked construct validity, although further analysis indicated that a shorter version of the PROMIS demonstrated strong construct validity and DGBI specific norms.[99] Accordingly, the brief PROMIS is recommended as a HRQoL PROM for trials in DGBI.[100]

One of the most well-used DGBI-specific HRQoL instruments is the IBS-QOL.[35] Based on its use and evidence of strong psychometric properties across multiple studies, the IBS-QOL is, arguably, the best HRQoL scale available for IBS.[101] Further recommendations relating to QoL measures and gastrointestinal-specific HRQoL measures are provided elsewhere.[102,103]

Measuring diet and diet-related outcomes. Dietary intervention trials in DGBI can be nutraceutical or nutrient supplementation trials (eg, fiber supplementation),[104] food supplementation trials (eg, prunes),[105] or whole-diet trials (eg, gluten-free diet), which usually involve exclusion of 1 or more dietary components.[106] In nutrient and food supplementation trials, assessment of overall dietary intake should be undertaken to confirm that background dietary intake has not changed. In whole-diet intervention trials, dietary intake should be measured to confirm adherence to the intervention.

Measurement should occur for short periods (ie, 3 to 7 days) at baseline and at the end of the intervention to minimize burden and recording fatigue.[107] Selecting a measurement method for diet depends on trial design, the dietary constituent(s) of interest, and the resources and staffing available. Research dietitians are best suited to collect, analyze, and interpret dietary data and should be involved early in the trial design.

Diet is challenging to measure because it is multidimensional (ie, we eat many foods, in a variety of combinations and preparations), and there is high temporal variation in dietary intake. Hence, the number of recording days for high precision can be considerable[108] and must be weighed against the participant burden associated with extensive diet recording. Finally, dietary assessment relies on self-report, which is subject to reporting error due to the Hawthorne effect (ie, behavior modification in response to an awareness of being observed), response bias, and underreporting.[107]

Dietary trials often use food records to measure current dietary intake. This involves recording all food and beverages in real time, on paper or digitally, usually for 3 to 7 days. The 24-hour recall method involves a structured interview by a trained interviewer to collect participant data on food and fluid intake in the preceding day. In feeding trials, all food is provided to participants, allowing for tight control of dietary intake,[109] and adherence can be measured using completion logs detailing consumption of provided meals.

Best practice in collecting patient-reported outcome measures data. The selection of a PROM depends on the research question and the trial design. For example, simple binary adequate relief end points may be appropriate for pragmatic trials. Other considerations should include the PROM's psychometric properties and the required completion time. The recall period, which should be sufficiently long enough to capture an "average" symptom experience in a particular DGBI, should also be considered. The recording duration should balance the number of days needed for optimal data precision with a recording period that limits participant burden. In the case of abdominal pain in IBS, for example, daily recordings are recommended; for stool form and frequency, a recording duration of 7 days may be sufficient,[110] although several weeks may be needed for optimal precision.[111] Real-time momentary assessments, such as experience sampling method,[47] may reduce recall bias. The mode of data capture, whether electronic or hard copy, in-person or remote, should also be considered. Guidelines for design and reporting of these methods are available.[112]

Adverse events and safety monitoring. The accurate acquisition and reporting of safety information are crucial for the ongoing evaluation and management of known and potential risks to trial participants. Good Clinical Practice guidelines provide general principles and recommendations for safety recording and reporting in clinical trials.[113–115] Supplementary Table 5 summarizes the definitions for AEs, adverse reactions, and serious AEs (SAEs). It is recommended that the collection of safety information is centralized and risk-based to take appropriate measures in accordance with the level of expected risks. The extent to which AEs are recorded depends on (1) the study aims and purpose of the data; (2) whether the drug is used in accordance with the licensed indication ("on label"); and (3) the known safety profile of the drug. SAEs should always be collected and promptly reported.[116] The minimum information collected should include the description, timing, severity, seriousness, causality (in relation to the intervention), and outcome of the event. For scientific journals, the Consolidated Standards of Reporting Trials (CONSORT) statement also provides several recommendations and a checklist for proper reporting of "harm" (ie, "the totality of possible adverse consequences of an intervention; the direct opposite of benefits, against which they must be compared").[117]

Recording adverse events in nondrug trials. Because safety guidelines have mainly been developed for drug trials, they may not always be appropriate for nondrug trials, potentially leading to underreporting of AEs and SAEs. Although BGBTs are generally considered to be safe, AEs and SAEs are rarely reported adequately in clinical trials testing them.[118] AEs occurring in BGBT trials should be based not only on medical grounds (eg, worsening of symptoms or self-harm) but also on emotional (eg, increased aggression or anxiety) and social (eg, increased family strains) aspects. A stepwise process to define AEs and adverse reactions in BGBT trials has been proposed (Supplementary Figure 3).[119,120]

Safety considerations for dietary interventions involve many aspects that might differ depending on the type of trial. Weight loss, limited dietary diversity, nutritional insufficiency, and alterations in the gut microbial composition all are potential AEs with a dietary intervention.[121] Moreover, abdominal symptoms may be aggravated by a dietary intervention (eg, bloating with a high-fiber diet). Best practices guidance has been published for dietary intervention trials.[107,122]

Participant withdrawal from study participation. Reasons for participant withdrawal from the trial may not always be due to a lack of improvement, and include poor adherence to the study or even the occurrence of an important AE. Clear criteria for participant withdrawal are necessary, not only to ensure the protection of the participants but also to minimize loss of critical study data.[123]

Breaking the mask and early study termination. Breaking the mask of a clinical trial for an individual trial participant should be undertaken only if relevant for the safety of the trial participants (eg, an AE plausibly related to the intervention), and the decision to proceed should involve the principal investigator, study sponsor, and Data Safety and Monitoring Committee (DSMC). Early termination or suspension of studies should be considered to avoid harm to the participants when a concerning or unexpected number of AEs have been reported. These decisions, made by the study sponsor after discussion with the principal investigator and with the involvement of the DSMC, should be taken carefully as they may lead to statistical and methodological difficulties and inappropriate interpretation of the data.[124]

Analysis and Data Reporting

Guidelines for Reporting Clinical Trial Results

Quality, clear, and transparent communication of clinical trial findings in medical journals is fundamental for readers to understand the design and results of studies appropriately as well as to assess their validity.[125] However, a review found that detailed information was often lacking or that the available information was not properly described.[126] Although investigator competition and the drive for "productivity" may compel researchers to report positive trial results preferentially, publication of negative results, such as nonconfirmatory, unexpected, or inconclusive data, is equally important.[127]

The CONSORT statement proposes key elements that clinical trialists should adhere to, with the aim of improving the quality of reporting results of parallel RCTs.[128] The CONSORT statement includes a 25-item checklist and a flow diagram, including (1) the enrollment section, reporting the number of participants assessed for eligibility, and a description of patients who were excluded or refused participation; (2) the randomization and allocation section, which indicates the total number of patients randomized and how many patients were allocated per study arm; (3) the follow-up section, which displays the number of individuals who received the intervention and those who did not (lost to follow-up); and (4) the analysis section, which describes the number of individuals who were analyzed and the number excluded from the analysis with the reasons for exclusion (Supplementary Figure 4).[128–130] A recent CONSORT extension for outcome reporting has been published.[131]

Primary and Secondary Outcome Measures

The outcome of greatest importance associated with a trial is identified as the primary outcome, with others being defined as secondary outcomes.[132,133] Investigators should consider the priority and number of outcome measures and specify no more than 2 primary outcome measures in a trial to minimize issues associated with multiplicity and interpretation.[132] Secondary outcome measures can be useful to include in a trial to assess other potential associated effects, including those relating to AEs or side effects. Investigators must identify, define, and describe each outcome clearly in their trial protocol, including their provenance and psychometric properties.[132,133]

The selection of primary and secondary outcomes needs to include clear, preplanned methodological details relating to data management (eg, evaluation of data integrity and security, and screening), statistical procedures undertaken (eg, how missing data will be addressed, and primary and secondary data analyses), and the rationale for their selection and interpretation. The selection of statistical methods and any conclusions regarding the efficacy of an intervention (ie, evidence for or against the intervention) should be based on the primary outcome measure and its estimated treatment effect (ie, the difference between the intervention and control arms). Trials involving 2 primary outcomes need to identify the preplanned determinations, including the rationale in determining how intervention success is defined (eg, do both primary outcomes need to show a significant positive effect or only 1 of the 2 outcomes?) and all statistical adjustments undertaken (eg, Bonferroni's correction).

Minimum clinically important difference. Although effect size identifies the magnitude of effect, the minimum clinically important difference (MCID) refers to the minimum amount of change important to a participant. Based on validated PROMs, MCIDs have become increasingly important in trials to determine both effect sizes and variability scores used in sample size calculations.[134,135] Benefits of using MCIDs include (1) the ability to use patient-centered

measures that ensure the change being measured is meaningful to the participant; (2) a reduction in bias; (3) provision of a therapeutic threshold to define success or failure of a trial; and (4) the ability to guide clinical practice and improve the interpretation of trial outcomes.[134–137] Despite the benefits of using MCIDs in trials, they also present several challenges, including (1) the limited number of validated PROMs available with established MCIDs; (2) the methodological debate regarding how MCIDs should be derived (eg, anchor- vs distribution-based approaches); and (3) the heterogeneity and potential limited generalizability of MCIDs.[136]

Intention-to-treat and per-protocol analysis. The CONSORT statement recommends that the assessment of both primary and secondary outcomes in RCTs be based on an intention-to-treat analysis.[132] An intention-to-treat approach minimizes bias involving the nonrandom loss of data and is generally conservative for superiority hypotheses.[138] Intention-to-treat requires a determination of the method by which missing data are treated (eg, last observation carried forward, use of multiple imputation regression–based methods).[139,140]

Limitations of the intention-to-treat approach are that it may include participants who were randomized but did not receive any treatment and may be imprecise or biased, or both, due to the heterogeneity of including dropouts unrelated to treatment effects and noncompliant participants.[138] Recognizing the potential for underestimation of the treatment effect, CONSORT also recommends the consideration of secondary analyses (eg, a per-protocol approach).[132] A per-protocol approach only considers participants who did not deviate from the trial protocol and adhered to the intervention.[141] It tends to overestimate treatment differences, and comparability is more problematic (ie, uncertainty whether benefits of randomization are maintained).[142] RCTs should use an intention-to-treat approach for the primary analysis and a per-protocol approach for supportive analyses.[132] If the findings from both intention-to-treat and per-protocol analyses are similar, investigators can have more confidence in the results.[138,143]

Sample Size and Power Calculations

A trial protocol must include a clear statement about what the minimum required sample size was and how this was determined.[132,133] The trial should plan to enroll the minimum number of participants to provide high probability (power) of detecting a prespecified MCID, where it exists.[132,144]

Interim Analysis and Stopping Rules

Although the CONSORT statement highlights the importance of trials including predefined interim analyses, associated stopping rules, and engagement with a DSMC,[132,145] research indicates that many trials do not adhere to these recommendations.[146] Interim analyses can be an important tool for investigators conducting RCTs to monitor progress and assess efficacy and safety-related aspects. However, the relevance of interim analyses to treatment trials in DGBI is more limited, particularly because (1) there is a lack of mortality associated with DGBI; (2) any SAE will trigger a safety committee review; and (3) interim analysis may inadvertently influence a trial, especially if it is unmasked. Stopping rules are an important component of trials, especially when interventions involve potential mortality or serious complications.

Ethical Issues

All trials must have a favorable opinion from a local ethics committee. Where an investigational product is used, regulatory agency approval is required before the trial can start. All participants need to provide written informed consent. It is important that participants understand the objectives, risks, and inconveniences of the trial. Participants should be informed of their right to withdraw at any time without it affecting their future care.

A balance should be maintained between scientific ambitions and the patients' interest in receiving therapy without being exposed to risks. Invasive procedures should be kept to the minimum that is clinically and scientifically required. To maintain confidentiality, all trial information needs to be properly recorded, handled, and stored.

Recommendations for Future Research

Future research priorities in DGBI treatment trials include:

1. A need for PROMs to assess the efficacy of treatments for the majority of DGBI that have been recognized.
2. Additional treatment trials in pediatric patients, and patients with DGBI other than IBS (eg, functional abdominal bloating).
3. Establishing MCIDs for symptoms other than abdominal pain.
4. Treatment trials evaluating abdominal pain response as the primary end point where an identified pharmacotherapy is targeted at relieving pain alone, without an anticipated effect on bowel habits.
5. Trials to assess the efficacy of treatments over durations exceeding 3 months.
6. Trials using on-demand therapy or intermittent dosing strategies.
7. Pragmatic trials in terms of end points used to define response and patient populations studied (eg, studying efficacy of a drug for constipation in both IBS-C and CC or for diarrhea in both IBS-D and functional diarrhea).[3]
8. Further assessment of what best constitutes control therapy in nondrug trials (eg, trials of dietary interventions or BGBTs).

9. Treatment trials comparing efficacious drugs with each other, combining approaches of more than 1 drug, a drug and a dietary intervention, or a drug and a BGBT, or comparing dietary interventions with BGBTs.
10. Establishing biomarkers that can be used to personalize the treatment to the patient with a DGBI, based on their likely response to a drug or other intervention.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at https://dx.doi.org/10.1053/j.gastro.2025.12.040

References

1. Camilleri M, Chedid V. Actionable biomarkers: the key to resolving disorders of gastrointestinal function. Gut 2020;69:1730–1737.
2. Fairlie T, Shah A, Talley NJ, et al. Overlap of disorders of gut-brain interaction: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol 2023;8:646–659.
3. Black CJ, Ford AC. Irritable bowel syndrome: a spotlight on future research needs. Lancet Gastroenterol Hepatol 2021;6:419–422.
4. Bosman M, Elsenbruch S, Corsetti M, et al. The placebo response rate in pharmacological trials in patients with irritable bowel syndrome: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol 2021;6:459–473.
5. Bosman M, Smeets F, Elsenbruch S, et al. Placebo response in pharmacological trials in patients with functional dyspepsia—a systematic review and meta-analysis. Neurogastroenterol Motil 2023;35:e14474.
6. Food and Drug Administration. Guidance for Industry: Irritable Bowel Syndrome—Clinical Evaluation of Drugs for Treatment. May 2022. Available at: https://www.fda.gov/media/78622/download. Accessed September 13, 2025.
7. Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273:408–412.
8. Miller LE, Stewart ME. The blind leading the blind: use and misuse of blinding in randomized controlled trials. Contemp Clin Trials 2011;32:240–243.
9. Bang H, Ni L, Davis CE. Assessment of blinding in clinical trials. Control Clin Trials 2004;25:143–156.
10. Altman DG, Schulz KF. Statistics notes: Concealing treatment allocation in randomised trials. BMJ 2001;323:446–447.
11. Lembo AJ. Understanding the placebo and nocebo effects in patients with irritable bowel syndrome. Gastroenterol Hepatol (N Y) 2020;16:374–376.
12. Özdemir V, Endrenyi L. Rethinking clinical trials and personalized medicine with placebogenomics and placebo dose. OMICS 2021;25:1–12.
13. Elsenbruch S, Enck P. Placebo effects and their determinants in gastrointestinal disorders. Nat Rev Gastroenterol Hepatol 2015;12:472–485.
14. Ford AC, Moayyedi P. Meta-analysis: factors affecting placebo response rate in the irritable bowel syndrome. Aliment Pharmacol Ther 2010;32:144–158.
15. Ballou S, Iturrino J, Rangan V, et al. Improving medication tolerance: a pilot study in disorders of gut-brain interaction treated with tricyclic antidepressants. J Clin Gastroenterol 2022;56:452–456.
16. Li R, Chen F, He X, et al. Nocebo response intensity and influencing factors in the randomized clinical trials of irritable bowel syndrome: a systematic review and meta-analysis. Front Med (Lausanne) 2022;9:1018713.
17. Glissen Brown JR, Sanayei A, Proctor S, et al. Examining the nocebo effect in trials of neuromodulators for use in disorders of gut-brain interaction. Am J Gastroenterol 2023;118:692–701.
18. Kessner S, Wiech K, Forkmann K, et al. The effect of treatment history on therapeutic outcome: an experimental approach. JAMA Intern Med 2013;173:1468–1469.
19. Bingel U, Wanigasekera V, Wiech K, et al. The effect of treatment expectation on drug efficacy: imaging the analgesic benefit of the opioid remifentanil. Sci Transl Med 2011;3:70ra14.
20. Laursen DRT, Paludan-Müller AS, Hróbjartsson A. Randomized clinical trials with run-in periods: frequency, characteristics and reporting. Clin Epidemiol 2019;11:169–184.
21. Rees JR, Mott LA, Barry EL, et al. Randomized controlled trials: who fails run-in? Trials 2016;17:374.
22. Bosman MHMA, Keszthelyi D. Letter: placebo run-in for IBS clinical trials-is it useful? Aliment Pharmacol Ther 2020;52:1237–1238.
23. Food and Drug Administration. Adaptive Designs for Clinical Trials of Drugs and Biologics: Guidance for Industry. November 2019. Available at: https://www.fda.gov/media/78495/download. Accessed September 13, 2025.
24. Ford AC, Wright-Hughes A, Alderson SL, et al. Amitriptyline at Low-Dose and Titrated for Irritable Bowel Syndrome as Second-Line Treatment in primary care (ATLANTIS): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet 2023;402:1773–1785.
25. Health Informatics Centre (University of Dundee). PRECIS2—PRagmatic Explanatory Continuum Indicator Summary. Available at: https://www.precis-2.org/. Accessed September 13, 2025.
26. Staudacher HM, Irving PM, Lomer MCE, et al. The challenges of control groups, placebos and blinding in clinical trials of dietary interventions. Proc Nutr Soc 2017;76:628.
27. Keefer L, Ballou SK, Drossman DA, et al. A Rome Working Team Report on brain-gut behavior therapies for disorders of gut-brain interaction. Gastroenterology 2022;162:300–315.
28. Di Lorenzo C, Khlevner J, Rodriguez-Araujo G, et al. Efficacy and safety of linaclotide in treating functional

constipation in paediatric patients: a randomised, double-blind, placebo-controlled, multicentre, phase 3 trial. Lancet Gastroenterol Hepatol 2024;9:238–250.
29. Saps M, Nurko S, Benninga MA, et al. A collaborative effort to advance drug development in pediatric constipation and irritable bowel syndrome. J Pediatr Gastroenterol Nutr 2021;73:145–149.
30. Koppen IJN, Saps M, Lavigne JV, et al. Recommendations for pharmacological clinical trials in children with functional constipation: the Rome Foundation Pediatric Subcommittee on Clinical Trials. Neurogastroenterol Motil 2018;30:e13294.
31. Saps M, van Tilburg MA, Lavigne JV, et al. Recommendations for pharmacological clinical trials in children with irritable bowel syndrome: the Rome Foundation Pediatric Subcommittee on Clinical Trials. Neurogastroenterol Motil 2016;28:1619–1631.
32. Food and Drug Administration. Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. December 2009. Available at: https://www.fda.gov/media/77832/download. Accessed September 13, 2025.
33. Johnson B, Patrick DL, Devji T, et al. Patient-reported outcomes. In: Higgins J, Thomas J, Chandler J, eds. Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021.
34. Brazier JE, Harper R, Jones NM, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. BMJ 1992;305:160–164.
35. Patrick DL, Drossman DA, Frederick IO, et al. Quality of life in persons with irritable bowel syndrome: development and validation of a new measure. Dig Dis Sci 1998;43:400–411.
36. Spiegel BM, Bolus R, Agarwal N, et al. Measuring symptoms in the irritable bowel syndrome: development of a framework for clinical trials. Aliment Pharmacol Ther 2010;32:1275–1291.
37. European Medicines Agency. Guideline on the evaluation of medicinal products for the treatment of chronic constipation (including opioid induced constipation) and for bowel cleansing. June 25, 2015. Available at: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-evaluation-medicinal-products-treatment-chronic-constipation-including-opioid-induced-constipation-and-bowel-cleansing_en.pdf. Accessed September 13, 2025.
38. Mearin F, Lacy BE, Chang L, et al. Bowel disorders. Gastroenterology 2016;150:1393–1407.
39. Critical Path Institute. Patient-Reported Outcome Consortium. Available at: https://c-path.org/program/patient-reported-outcome-consortium/. Accessed September 13, 2025.
40. Fehnel SE, Ervin CM, Carson RT, et al. Development of the diary for irritable bowel syndrome symptoms to assess treatment benefit in clinical trials: foundational qualitative research. Value Health 2017;20:618–626.
41. Camilleri M, Kerstens R, Rykx A, et al. A placebo-controlled trial of prucalopride for severe chronic constipation. N Engl J Med 2008;358:2344–2354.
42. Tack J, van Outryve M, Beyens G, et al. Prucalopride (Resolor) in the treatment of severe chronic constipation in patients dissatisfied with laxatives. Gut 2009;58:357–365.
43. Heaton KW, Ghosh S, Braddon FE. How bad are the symptoms and bowel dysfunction of patients with the irritable bowel syndrome? A prospective, controlled study with emphasis on stool form. Gut 1991;32:73–79.
44. Ashraf W, Park F, Lof J, et al. An examination of the reliability of reported stool frequency in the diagnosis of idiopathic constipation. Am J Gastroenterol 1996;91:26–32.
45. Carbone F, Vandenberghe A, Holvoet L, et al. Validation of the Leuven Postprandial Distress Scale, a questionnaire for symptom assessment in the functional dyspepsia/postprandial distress syndrome. Aliment Pharmacol Ther 2016;44:989–1001.
46. Taylor F, Higgins S, Carson RT, et al. Development of a symptom-focused patient-reported outcome measure for functional dyspepsia: the Functional Dyspepsia Symptom Diary (FDSD). Am J Gastroenterol 2018;113:39–48.
47. Smeets FGM, Keszthelyi D, Vork L, et al. Development of a real-time patient-reported outcome measure for symptom assessment in patients with functional dyspepsia using the experience sampling method. Neurogastroenterol Motil 2019;31:e13496.
48. European Medicines Agency. Letter of support for Leuven Postprandial Distress Scale (LPDS) as PRO in Postprandial Distress Syndrome (PDS). December 20, 2015. Available at: https://www.ema.europa.eu/en/documents/other/letter-support-leuven-postprandial-distress-scale-lpds-pro-postprandial-distress-syndrome-pds_en.pdf. Accessed September 13, 2025.
49. 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.
50. 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.
51. 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.
52. 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.
53. Tack J, Carbone F, Chang L, et al. Patient-reported outcomes in disorders of gut-brain interaction. Gastroenterology 2024;166:572–587.
54. Goelen N, Scheepers J, Carbone F, et al. The Waiting Room Questionnaire, a patient-reported outcome instrument including pictograms for facilitating the diagnosis of functional gastrointestinal disorders. Neurogastroenterol Motil 2025;37:e70019.
55. 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.
56. Spiegel BM, Hays RD, Bolus R, et al. Development of the NIH Patient-Reported Outcomes Measurement Information System (PROMIS) gastrointestinal symptom scales. Am J Gastroenterol 2014;109:1804–1814.
57. Bardhan KD, Stanghellini V, Armstrong D, et al. Evaluation of GERD symptoms during therapy. Part I. Development of the new GERD questionnaire ReQuest. Digestion 2004;69:229–237.
58. Jones R, Junghard O, Dent J, et al. Development of the GerdQ, a tool for the diagnosis and management of gastro-oesophageal reflux disease in primary care. Aliment Pharmacol Ther 2009;30:1030–1038.
59. Shaw M, Dent J, Beebe T, et al. The Reflux Disease Questionnaire: a measure for assessment of treatment response in clinical trials. Health Qual Life Outcomes 2008;6:31.
60. Pare P, Meyer F, Armstrong D, et al. Validation of the GSFQ, a self-administered symptom frequency questionnaire for patients with gastroesophageal reflux disease. Can J Gastroenterol 2003;17:307–312.
61. Holtmann G, Chassany O, Devault KR, et al. International validation of a health-related quality of life questionnaire in patients with erosive gastro-oesophageal reflux disease. Aliment Pharmacol Ther 2009; 29:615–625.
62. Leidy NK, Farup C, Rentz AM, et al. Patient-based assessment in dyspepsia: development and validation of Dyspepsia Symptom Severity Index (DSSI). Dig Dis Sci 2000;45:1172–1179.
63. Moayyedi P, Duffett S, Braunholtz D, et al. The Leeds Dyspepsia Questionnaire: a valid tool for measuring the presence and severity of dyspepsia. Aliment Pharmacol Ther 1998;12:1257–1262.
64. el-Omar EM, Banerjee S, Wirz A, et al. The Glasgow Dyspepsia Severity Score—a tool for the global measurement of dyspepsia. Eur J Gastroenterol Hepatol 1996;8:967–971.
65. Mujagic Z, Keszthelyi D, Aziz Q, et al. Systematic review: instruments to assess abdominal pain in irritable bowel syndrome. Aliment Pharmacol Ther 2015;42:1064–1081.
66. Mujagic Z, Leue C, Vork L, et al. The Experience Sampling Method—a new digital tool for momentary symptom assessment in IBS: an exploratory study. Neurogastroenterol Motil 2015;27:1295–1302.
67. Weinland SR, Morris CB, Hu Y, et al. Characterization of episodes of irritable bowel syndrome using ecological momentary assessment. Am J Gastroenterol 2011; 106:1813–1820.
68. Farrar JT, Troxel AB, Haynes K, et al. Effect of variability in the 7-day baseline pain diary on the assay sensitivity of neuropathic pain randomized clinical trials: an ACTTION study. Pain 2014;155:1622–1631.
69. Ashkenazy S, DeKeyser Ganz F. The differentiation between pain and discomfort: a concept analysis of discomfort. Pain Manag Nurs 2019;20:556–562.
70. Sach J, Bolus R, Fitzgerald L, et al. Is there a difference between abdominal pain and discomfort in moderate to severe IBS patients? Am J Gastroenterol 2002; 97:3131–3138.
71. Chang L, Lee OY, Naliboff B, et al. Sensation of bloating and visible abdominal distension in patients with irritable bowel syndrome. Am J Gastroenterol 2001; 96:3341–3347.
72. Schmulson M, Chang L. Review article: the treatment of functional abdominal bloating and distension. Aliment Pharmacol Ther 2011;33:1071–1086.
73. Houghton LA, Lea R, Agrawal A, et al. Relationship of abdominal bloating to distention in irritable bowel syndrome and effect of bowel habit. Gastroenterology 2006;131:1003–1010.
74. Lacy BE, Cangemi DJ, Wise JL, et al. Development and validation of a novel scoring system for bloating and distension: the Mayo Bloating Questionnaire. Neurogastroenterol Motil 2022;34:e14330.
75. Lewis MJ, Reilly B, Houghton LA, et al. Ambulatory abdominal inductance plethysmography: towards objective assessment of abdominal distension in irritable bowel syndrome. Gut 2001;48:216–220.
76. Yiannakou Y, Tack J, Piessevaux H, et al. The PACSYM questionnaire for chronic constipation: defining the minimal important difference. Aliment Pharmacol Ther 2017;46:1103–1111.
77. Camilleri M, Rothman M, Ho KF, et al. Validation of a bowel function diary for assessing opioid-induced constipation. Am J Gastroenterol 2011;106:497–506.
78. 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.
79. Wiklund IK, Fullerton S, Hawkey CJ, et al. An irritable bowel syndrome-specific symptom questionnaire: development and validation. Scand J Gastroenterol 2003;38:947–954.
80. Talley NJ, Boyce PM, Owen BK, et al. Initial validation of a bowel symptom questionnaire and measurement of chronic gastrointestinal symptoms in Australians. Aust N Z J Med 1995;25:302–308.
81. European Medicines Agency. Guideline on the evaluation of medicinal products for the treatment of irritable bowel syndrome. September 25, 2014. Available at: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-evaluation-medicinal-products-treatment-irritable-bowel-syndrome-revision-1_en.pdf. Accessed September 13, 2025.
82. 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.
83. Mangel AW, Wang J, Sherrill B, et al. Urgency as an endpoint in irritable bowel syndrome. Gastroenterology Res 2011;4:9–12.
84. Lembo AJ, Olden KW, Ameen VZ, et al. Effect of alosetron on bowel urgency and global symptoms in women with severe, diarrhea-predominant irritable bowel syndrome: analysis of two controlled trials. Clin Gastroenterol Hepatol 2004;2:675–682.

85. Vaizey CJ, Carapeti E, Cahill JA, et al. Prospective comparison of faecal incontinence grading systems. Gut 1999;44:77–80.
86. Hemmink GJ, Bredenoord AJ, Weusten BL, et al. Supragastric belching in patients with reflux symptoms. Am J Gastroenterol 2009;104:1992–1997.
87. Glasinovic E, Wynter E, Arguero J, et al. Treatment of supragastric belching with cognitive behavioral therapy improves quality of life and reduces acid gastroesophageal reflux. Am J Gastroenterol 2018;113:539–547.
88. Parker CH, Tomlinson G, Correia A, et al. The use of standardized questions in identifying patients with dyssynergic defecation. J Can Assoc Gastroenterol 2018;1:60–66.
89. Deary IJ, Wilson JA, Harris MB, et al. Globus pharyngis: development of a symptom assessment scale. J Psychosom Res 1995;39:203–213.
90. Grudell AB, Alexander JA, Enders FB, et al. Validation of the Mayo Dysphagia Questionnaire. Dis Esophagus 2007;20:202–205.
91. Eisen GM, Locke GR, Provenzale D. Health-related quality of life: a primer for gastroenterologists. Am J Gastroenterol 1999;94:2017–2021.
92. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research. U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research. U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes 2006;4:79.
93. Ware JE, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–483.
94. Jenkinson C, Coulter A, Wright L. Short form 36 (SF36) health survey questionnaire: normative data for adults of working age. BMJ 1993;306:1437–1440.
95. Hays RD, Bjorner JB, Revicki DA, et al. Development of physical and mental health summary scores from the Patient-Reported Outcomes Measurement Information System (PROMIS) global items. Qual Life Res 2009;18:873–880.
96. HealthMeasures. PROMIS® Reference Populations. Updated April 21, 2025. Available at: https://www.healthmeasures.net/score-and-interpret/interpret-scores/promis/reference-populations. Accessed September 13, 2025.
97. Knowles SR, Skvarc D, Ford AC, et al. Negative impact of disorders of gut-brain interaction on health-related quality of life: results from the Rome Foundation Global Epidemiology Survey. Gastroenterology 2023;164:655–668.
98. Sperber AD, Bangdiwala SI, Drossman DA, et al. Worldwide prevalence and burden of functional gastrointestinal disorders, results of Rome Foundation Global Study. Gastroenterology 2021;160:99–114.e3.
99. Apputhurai P, Palsson OS, Bangdiwala SI, et al. Validation of the PROMIS Global Physical and Mental Health Scale for Gastrointestinal Disorders: a large-scale cross-sectional survey. Neurogastroenterol Motil 2025e70059.
100. Apputhurai P, Palsson OS, Bangdiwala SI, et al. Confirmatory validation of the patient health questionnaire—4 (PHQ-4) for gastrointestinal disorders: a large-scale cross-sectional survey. J Psychosom Res 2024;180:111654.
101. Lee J, Lee EH, Moon SH. A systematic review of measurement properties of the instruments measuring health-related quality of life in patients with irritable bowel syndrome. Qual Life Res 2016;25:2985–2995.
102. 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.
103. Bowling A. Measuring Health: A Review of Quality of Life Measurement Scales. New York: Open University Press, 2005.
104. Remes-Troche JM, Taboada-Liceaga H, Gill S, et al. Nopal fiber (*Opuntia ficus-indica*) improves symptoms in irritable bowel syndrome in the short term: a randomized controlled trial. Neurogastroenterol Motil 2021;33:e13986.
105. Chey SW, Chey WD, Jackson K, et al. Exploratory comparative effectiveness trial of green kiwifruit, psyllium, or prunes in US Patients With Chronic Constipation. Am J Gastroenterol 2021;116:1304–1312.
106. Algera JP, Magnusson MK, Öhman L, et al. Randomised controlled trial: effects of gluten-free diet on symptoms and the gut microenvironment in irritable bowel syndrome. Aliment Pharmacol Ther 2022;56:1318–1327.
107. Staudacher HM, Yao CK, Chey WD, et al. Optimal design of clinical trials of dietary interventions in disorders of gut-brain interaction. Am J Gastroenterol 2022;117:973–984.
108. Nybacka S, Störsrud S, Liljebo T, et al. Within- and between-subject variation in dietary intake of fermentable oligo-, di-, monosaccharides, and polyols among patients with irritable bowel syndrome. Curr Dev Nutr 2019;3:nzy101.
109. 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.
110. Lackner JM, Jaccard J, Keefer L, et al. The accuracy of patient-reported measures for GI symptoms: a comparison of real time and retrospective reports. Neurogastroenterol Motil 2014;26:1802–1811.
111. Palsson OS, Baggish JS, Turner MJ, et al. IBS patients show frequent fluctuations between loose/watery and hard/lumpy stools: implications for treatment. Am J Gastroenterol 2012;107:286–295.
112. Stone AA, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med 2002;24:236–243.
113. European Medicines Agency. ICH E2A Clinical safety data management: definitions and standards for expedited reporting—Scientific guideline. 1995. Available at: https://www.ema.europa.eu/en/ich-e2a-clinical-safety-

data-management-definitions-standards-expedited-reporting-scientific-guideline. Accessed September 13, 2025.
114. Food and Drug Administration. Guidance for Industry and Investigators. Safety Reporting Requirements for INDs and BA/BE Studies. December 2012. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/safety-reporting-requirements-inds-and-babe-studies. Accessed September 13, 2025.
115. Therapeutic Goods Association. Notes for guidance on clinical safety data management: definitions and standards for expedited reporting (CPMP/ICH/377/95);1995. Available at: https://www.tga.gov.au/sites/default/files/ich37795.pdf. Accessed September 13, 2025.
116. Irving E, van den Bor R, Welsing P, et al. Series: Pragmatic trials and real world evidence: Paper 7. Safety, quality and monitoring. J Clin Epidemiol 2017;91:6–12.
117. Ioannidis JP, Evans SJ, Gøtzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med 2004;141:781–788.
118. Duggan C, Parry G, McMurran M, et al. The recording of adverse events from psychological treatments in clinical trials: evidence from a review of NIHR-funded trials. Trials 2014;15:335.
119. Linden M, Schermuly-Haupt ML. Definition, assessment and rate of psychotherapy side effects. World Psychiatry 2014;13:306–309.
120. Linden M. How to define, find and classify side effects in psychotherapy: from unwanted events to adverse treatment reactions. Clin Psychol Psychother 2013;20:286–296.
121. Lacy BE. Review article: an analysis of safety profiles of treatments for diarrhoea-predominant irritable bowel syndrome. Aliment Pharmacol Ther 2018;48:817–830.
122. Weaver CM, Fukagawa NK, Liska D, et al. Perspective: US documentation and regulation of human nutrition randomized controlled trials. Adv Nutr 2021;12:21–45.
123. European Medicines Agency. ICH E6 Good clinical practice—Scientific guideline; 2025. Available at: https://www.ema.europa.eu/en/ich-e6-good-clinical-practice-scientific-guideline. Accessed September 13, 2025.
124. Briel M, Bassler D, Wang AT, et al. The dangers of stopping a trial too early. J Bone Joint Surg Am 2012;94(Suppl 1):56–60.
125. Hopewell S, Dutton S, Yu LM, et al. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. BMJ 2010;340:c723.
126. Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. Lancet 2005;365:1159–1162.
127. Bespalov A, Steckler T, Skolnick P. Be positive about negatives-recommendations for the publication of negative (or null) results. Eur Neuropsychopharmacol 2019;29:1312–1320.
128. Schulz KF, Altman DG, Moher D, et al. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. PLoS Med 2010;7:e1000251.
129. Turjeman A, Poran I, Daitch V, et al. Inadequate reporting of participants eligible for randomized controlled trials—a systematic review and meta-analysis. J Clin Epidemiol 2021;140:125–134.
130. Falci SG, Marques LS. CONSORT: when and how to use it. Dental Press J Orthod 2015;20:13–15.
131. Butcher NJ, Monsour A, Mew EJ, et al. Guidelines for reporting outcomes in trial reports: the CONSORT-Outcomes 2022 Extension. JAMA 2022;328:2252–2264.
132. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001;134:663–694.
133. Moher D, Schulz KF, Altman D, et al. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. JAMA 2001;285:1987–1991.
134. Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. J Clin Epidemiol 2017;82:128–136.
135. Draak THP, de Greef BTA, Faber CG, et al. The minimum clinically important difference: which direction to take. Eur J Neurol 2019;26:850–855.
136. Wright A, Hannon J, Hegedus EJ, et al. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). J Man Manip Ther 2012;20:160–166.
137. Wright JG. The minimal important difference: who's to say what is important? J Clin Epidemiol 1996;49:1221–1222.
138. Gupta SK. Intention-to-treat concept: a review. Perspect Clin Res 2011;2:109–112.
139. Nich C, Carroll KM. Intention-to-treat meets missing data: implications of alternate strategies for analyzing clinical trials data. Drug Alcohol Depend 2002;68:121–130.
140. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. Clin Trials 2004;1:368–376.
141. Abraha I, Montedori A. Modified intention to treat reporting in randomised controlled trials: systematic review. BMJ 2010;340:c2697.
142. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: intention-to-treat versus per-protocol analysis. Perspect Clin Res 2016;7:144–146.
143. Tripepi G, Chesnaye NC, Dekker FW, et al. Intention to treat and per protocol analysis in clinical trials. Nephrology (Carlton) 2020;25:513–517.
144. Noordzij M, Tripepi G, Dekker FW, et al. Sample size calculations: basic principles and common pitfalls. Nephrol Dial Transplant 2010;25:1388–1393.
145. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surg 2012;10:28–55.
146. Stegert M, Kasenda B, von Elm E, et al. An analysis of protocols and publications suggested that most

discontinuations of clinical trials were not based on preplanned interim analyses or stopping rules. J Clin Epidemiol 2016;69:152–160.

Received October 14, 2025. Accepted December 18, 2025.

Correspondence
Address correspondence to Gregory S. Sayuk, MD, MPH, Division of Gastroenterology, Washington University in St. Louis School of Medicine, 660 South Euclid Avenue, Saint Louis, Missouri 63110-1010. e-mail: gregsayuk@wustl.edu.

Conflicts of interest
This author discloses the following: Gregory S. Sayuk reports speaker fees and consultancy fees from AbbVie, Salix, Ardelyx, and Sanofi/Regeneron, and is a council member of the American Neurogastroenterology and Motility Society. The remaining authors disclose no conflicts.

Funding
The Rome Foundation provided organizational and editorial support.