Multidomain Patient-Reported Outcomes of Irritable Bowel Syndrome: Exploring Person-Centered Perspectives to Better Understand Symptom Severity Scores

Jeffrey M. Lackner, PsyD1*, James Jaccard, PhD2, Charles Baum, MD3

1Division of Gastroenterology, Department of Medicine, University at Buffalo School of Medicine, SUNY, Buffalo, NY, USA; 2Silver School of Social Work, New York University, New York, NY, USA; 3Takeda Pharmaceuticals International, Deerfield, Illinois, USA.

ABSTRACT

Objectives: Patient-reported outcomes assessing multiple gastrointestinal symptoms are central to characterizing the therapeutic benefit of novel agents for irritable bowel syndrome (IBS). Common approaches that sum or average responses across different illness components must be unidimensional and have small unique variances to avoid aggregation bias and misinterpretation of clinical data. This study sought to evaluate the unidimensionality of the IBS Symptom Severity Scale (IBS-SSS) and to explore person-centered cluster analytic methods for characterizing multivariate-based patient profiles. Methods: Ninety-eight Rome-diagnosed patients with IBS completed the IBS-SSS and a single, global item of symptom severity (UCLA Symptom Severity Scale) at pretreatment baseline of a clinical trial funded by the National Institutes of Health. k-means cluster analyses were performed on participants’ symptom severity scores.

Results: The IBS-SSS was not unidimensional. Exploratory cluster analyses revealed four common symptom profiles across five items of the IBS-SSS. One cluster of patients (25%) had elevated scores on pain frequency and bowel dissatisfaction, with less elevated but still high scores on life interference and low pain severity ratings. A second cluster (19%) was characterized by intermediate scores on both pain dimensions but more elevated scores on bowel dissatisfaction. A third cluster (16%) had elevated scores across all IBS-SSS subcomponents. The fourth and the most common cluster (37%) had relatively low scores on all dimensions except bowel dissatisfaction and life interference due to IBS symptoms.

Conclusions: Patient-reported outcome endpoints and research on IBS more generally relying on multicomponent assessments of symptom severity should take into account the multidimensional structure of symptoms to avoid aggregation bias and to optimize the sensitivity of detecting treatment effects.

Keywords: disease severity, global assessment, health status indicators, outcome research, psychometric properties, questionnaire development, rating scale.

Copyright © 2013, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.
items reflecting those symptoms and establish their psychometric properties; and culminates in the production of a PRO measure that reflects therapeutic benefit from the patient’s perspective. Historically, attempts to develop patient-reported end points have gravitated toward the construct of perceived severity of symptoms as a metric for gauging both illness status and the benefit of novel treatments. Representative of this approach is the IBS Symptom Severity Scale (IBS-SSS), a global measure of IBS symptoms that aggregates patient ratings of different, well-defined domains of IBS into a single overall score. The IBS-SSS [6] has been recommended by the Rome Foundation [7] as the global end point for measuring IBS symptom severity in clinical trials. This scale asks individuals to rate four symptom dimensions, each measured on a 0 to 100 rating scale: 1) the severity of abdominal pain, 2) the severity of abdominal distention/tightness (bloating), 3) satisfaction with bowel habits, and 4) life interference due to IBS symptoms. A final item asks the number of days out of previous 10 when the patient experiences abdominal pain, with the answer multiplied by 10 to create a 0 to 100 metric for it.

Like other composite measures developed under the PRO initiative [8], the IBS-SSS generates a total aggregate by summing its items to derive an index of overall symptom severity. Combining items across multiple symptom domains for the purpose of generating a global score is a common practice in PRO development [9]. Theoretically, this approach maps onto the requisite conceptual framework (see Fig. 1) the FDA recommends for constructing a PRO instrument whose overarching measurement focus (concept) is a product of multiple domains (e.g., signs and symptoms). Methodologically, however, aggregating scores from composite items potentiates at least two problems that can make scales misleading. The first problem is aggregation bias. The underlying mechanisms that impact one set of symptoms may differ from the underlying mechanisms that impact another set of symptoms. Aggregating across items obscures such dynamics. One is left with a total score whereby component scores may behave differently in response to a treatment or where the overall score masks a relationship between a component symptom and some other variable of clinical import. For example, a pharmacological treatment may impact defecatory symptoms but have limited, if any, effect on abdominal pain or discomfort. An overall score would then reflect a mixture of change on the item associated with bowel habits and random or systematic “noise” due to the pain items. The result is an index that can mask the true effects of the treatment or make it harder to detect those effects. By the same token, a novel agent may provide some pain relief but little, if any, benefit for defecatory symptoms. An aggregate index will capture both the genuine therapeutic change and the “noise” due to the other components measured by the scale. Aggregation bias represents a serious threat to accurately characterizing patients’ experience of their disease and associated treatment, which is the penultimate goal of FDA’s PRO guidance document [10].

A second problem is that variation in the total score, all else being equal, is dominated by whichever subcomponent has the most variability across people. If patients exhibit considerable variation on the pain subscale but only modest variation on a subscale about defecatory symptoms such as stool frequency, then variations on the total score will primarily reflect variations in pain, not bowel symptoms. Such a dynamic would advantage agents with strong analgesic properties over those that are primarily designed to relieve defecatory symptoms when using the overall composite score to evaluate a therapeutic agent. This is particularly germane to the IBS-SSS because three of its five items tap abdominal pain or discomfort (pain and bloating severity, number of pain days) and thus likely contribute more variability to the total score. In effect, the IBS-SSS “triple weights” items assessing abdominal pain/discomfort over those assessing nonpain aspects of the IBS experience.

These issues are not necessarily problematic if individual items comprising a multi-item scale are highly correlated and unidimensional in nature. Unidimensional scales measure a single dimension or group of dimensions that cluster with one another. To the extent that a scale is unidimensional and items are highly correlated, the behavior of one item parallels the behavior of other items. The extent to which composite measures of IBS (or for that matter other diseases) are unidimensional is thus an important and largely overlooked psychometric matter critical to the development of sound, meaningful, and sensitive PROs. To our knowledge, IBS end points have been developed without regard to documenting their unidimensional versus multidimensional properties. One purpose of the present study was to examine the unidimensional versus multidimensional structure of the IBS-SSS.

To the extent that individual IBS symptoms are characterized by both nontrivial unique and common variance, it can be useful to identify symptom clusters that characterize significant numbers of patients with IBS. Different treatment regimens might then be implemented depending on the observed symptom patterns, with some therapeutic agents being more appropriate for some types of patients but less appropriate for other types. The present study applied cluster analytic methods to the core symptoms measured by the IBS-SSS with the objective of identifying distinct patient profiles that may require different approaches to symptom resolution. In contrast to traditional approaches that treat each symptom as a separate construct, person-centered cluster analysis uses an idiographic approach that represents the presentation of multiple symptoms as an organized whole [11].

**Methods**

**Participants**

Participants included 98 consecutively evaluated patients with IBS recruited primarily through local media coverage and community advertising and referral by local physicians to a tertiary care center at an academic medical center. To qualify, participants must have met Rome II IBS diagnostic criteria [12] without organic GI disease (e.g., IBD and colon cancer) as determined by a board-certified study gastroenterologist. Because this study was conducted as part of a clinical trial for patients more severely affected with IBS, participants must have also reported IBS symptoms of at least moderate intensity (i.e., symptom occurring at least twice weekly for 6 months and causing life interference). Institutional review board approval and written, signed consent
were obtained before the study began. This study was completed in full compliance with the Declaration of Helsinki.

Procedure

After a brief telephone interview to determine whether participants were likely to meet basic inclusion criteria, participants were scheduled for a medical examination to confirm IBS diagnosis [12,13] and psychometric testing, which for the purposes of this study included the following test battery. Detailed information about study procedure can be found elsewhere [14].

Measures

The IBS-SSS [6] is a five-item instrument used to measure severity of abdominal pain, frequency of abdominal pain, severity of abdominal distension, dissatisfaction with bowel habits, and interference with quality of life on a 100-point scale. For four of the items, the scales are represented as continuous lines with end points 0% and 100%, with different descriptors at the end points and adverb qualifiers (e.g., “not very,” “quite”) strategically placed along the line. Respondents mark a point on the line reflecting the extremity of their judgment between the two end points, and the proportional distance from zero is the score assigned for that scale (hence scores range from 0 to 100). The end points for the severity items are “no pain” and “very severe,” for satisfaction, the end points are “not at all satisfied” and “very satisfied,” and for interference, they are “not at all interferes” to “completely interferes.” A final item asks the number of days out of 10 the patient experiences abdominal pain, and the answer is multiplied by 10 to create a 0 to 100 metric. The items were summed, and thus the total score could range from 0 to 500. The IBS-SSS was used as an end point of the clinical study from which secondary analyses were derived for this study.

We also obtained a single-item, global measure of symptom severity by using the University of California at Los Angeles Global Severity of Gastrointestinal Symptoms Scale. It [15] is a 21-point rating scale where participants rate the overall severity of their symptoms (where the specific symptoms are unspecified) on a 0- (no symptoms) to 20- (most intense symptoms imaginable) point scale.

Data Analytic Plan

To place variables on a common metric and to ease interpretation of some analyses, all measures were rescaled to a 0 to 10 metric. This was accomplished by subtracting the lowest possible score attainable for each variable from an individual’s observed score and then dividing this by the highest possible score on that variable. This result was then multiplied by 10. Thus, scores for every variable range from 0 to 10, with a midpoint of 5. This strategy represents a simple linear transformation that preserves differential variability, differential central tendencies, and distributional forms for the variables, all while establishing a common 0 to 10 metric. We first tested whether the five component measures of the IBS-SSS behaved in accord with a single-factor model (i.e., unidimensionality) by using confirmatory factor analysis based on maximum likelihood methods. Once unidimensionality was rejected, we used exploratory factor analysis to gain insights into the underlying dimensional structure of the scale, based on principal components extraction with an oblimin rotation. The oblimin rotation permits the underlying factors to be correlated, which seemed more theoretically compelling than a varimax rotation. Finally, as described later, cluster analysis of the measures was applied to the data using K-means clustering algorithms, which are described below. This method of analysis identifies clusters of individuals who show similar profiles across the component symptoms but who have a distinct profile relative to individuals in other clusters. The goal is to classify individuals into categories, with each category containing individuals who are similar to each other and different from individuals in other categories. Such an approach is useful in IBS research where there is known heterogeneity in the complexion of symptoms that characterize patients with a common syndrome. The K-means algorithm specifies a fixed number of clusters, k, and assigns cases to clusters so that the means for all variables are as different from each other as possible for the different clusters. The value of k is determined through successive solutions that increment the value of k by 1, with the chosen solution being one that yields theoretically meaningful cluster profiles that meaningfully increment estimates of explained variance relative to solutions with smaller k and whose cluster sizes are not trivial. Although it would have been preferable to use more modern methods of classification analysis, such as latent profile analysis, the sample size was judged to be too small for such methods. As such, the results reported here should be viewed as illustrative and tentative.

Results

Preliminary Analyses

There were minimal missing data, amounting to no more than a few cases on most variables and representing less than 2% of the cases. Missing values were imputed by using the Markov Chain Monte Carlo method based on Schafer [16]. Nonnormality of the variables was evaluated by computing values of skewness and kurtosis. These values as well as the means and SDs of the variables are presented in Table 1. We evaluated the data for multivariate outliers by using a leverage statistic. An individual was identified as an outlier if he or she had a leverage score four times larger than the mean leverage index. No outliers were identified.

The Unidimensionality of the IBS-SSS Measure

To determine whether the five component items of the IBS-SSS are unidimensional in structure, a single-factor confirmatory factor analysis model was fit to the 5 × 5 covariance matrix for the items. The single-factor model yielded a poor fit to the data (χ² with 5 degrees of freedom = 19.99; P < 0.01; Comparative Fit Index = 0.80; Root Mean Square Error of Approximation = 0.18; P value for close fit < 0.01), suggesting a multidimensional structure. An exploratory factor analysis on the component measures was performed by using principal components extraction and an oblimin rotation. A two-factor solution was suggested by the analysis, with the two pain and bloating items loading on one abdominal pain/discomfort factor and the life interference and satisfaction with bowel items loading on the second symptom burden factor. Columns 2 through 6 of Table 2 present the 5 × 5 correlation matrix between the component measures as well as the standardized factor loadings and standardized unique variances for the two-factor solution. Even though a two-factor model reasonably accounts for the correlations between the measures, several of the individual measures have considerable unique variance associated with them, independent of the two common factors, ranging from about 25% to 50%. These data are presented in Table 2. The presence of substantial unique variance was even true of pain items, one of which taps the number (frequency) of pain days and the other taps pain intensity. This suggests that even forming two “total scores,” one for each of the factors observed in the factor analysis, may obscure important dynamics for some of the component measures in their own
right when mapping out the mechanisms underlying IBS symptoms. The estimated correlation between the factors was 0.20.

**Exploratory Cluster Analysis of Items of the IBS-SSS**

Given the failure of unidimensionality for the IBS-SSS and the pattern of correlations in Table 3, we conducted exploratory cluster analyses on the items of the IBS-SS by using a K-means clustering algorithm. A four-group solution was settled upon. The cluster sizes and profile means on the five items are presented in Table 3. All means are on a 0 to 10 metric, with 5 representing the scale midpoint.

The first cluster of patients, representing about 25% of the cases, is characterized by elevated scores on pain frequency and bowel dissatisfaction, with somewhat less elevated but still high mean scores on life interference. However, the pain severity ratings tend to be low. The second cluster of patients, representing about 19% of the cases, is characterized by intermediate scores on the pain dimensions, with somewhat more elevated scores on bowel dissatisfaction. The third cluster of patients, representing about 18% of the cases, is characterized by elevated scores on all subcomponents. Finally, the fourth cluster of patients, representing about 37% of the cases, has relatively low scores on all dimensions except bowel dissatisfaction, with a somewhat elevated mean for life interference due to IBS symptoms.

It is interesting to compare the four clusters on the overall global ratings of symptom severity on the independently assessed University of California at Los Angeles Symptom Severity Scale. Table 3 shows, not surprisingly, that the mean global characterization of IBS symptom severity is highest for cluster 3 patients who scored high on all IBS-SSS subscales and somewhat elevated for cluster 1 and 2 patients.

**Variance Contributions of Individual Items to the Total Score**

The first row of Table 2 presents the correlations between the individual items of the IBS-SS and the total IBS-SS score. The correlations are strongest for the two pain items. The squared semipartial correlation between the two pain items and the total score was 0.19 (holding the other items constant), suggesting that the pain items uniquely account for 19% of the variation in the total scores. The squared semipartial correlation for bloating and the total score was 0.07, suggesting that it uniquely accounts for 7% of the variation in IBS-SSS scores. For bowel dissatisfaction, the squared semipartial correlation was 0.05, suggesting that it uniquely accounts for 5% of the variation in the IBS-SSS total score, and for life interference, it was 0.04, suggesting that it

---

**Table 1 – Descriptive statistics.**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>88</td>
</tr>
<tr>
<td>Married</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>29</td>
</tr>
<tr>
<td>Employed full time</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>46</td>
</tr>
<tr>
<td>Graduating college</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>52</td>
</tr>
<tr>
<td>Age (y)</td>
<td>46.7</td>
<td>16.9</td>
<td>.03</td>
<td>−99</td>
<td></td>
</tr>
<tr>
<td>Months with IBS</td>
<td>197.8</td>
<td>184.0</td>
<td>1.05</td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td>UCLA-SSS</td>
<td>5.49</td>
<td>2.03</td>
<td>−.36</td>
<td>−.35</td>
<td></td>
</tr>
<tr>
<td>IBS-SSS</td>
<td>5.97</td>
<td>1.54</td>
<td>−.11</td>
<td>−.29</td>
<td></td>
</tr>
<tr>
<td>Bowel subtypes</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>IBS-Diarrhea</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>41</td>
</tr>
<tr>
<td>IBS-Constipation</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>39</td>
</tr>
<tr>
<td>IBS-Mixed</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>18</td>
</tr>
<tr>
<td>Missing</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>2</td>
</tr>
</tbody>
</table>

Note. Untransformed mean for UCLA-SSS is 11.52 and for IBS-SSS it is 298.5.

IBS, irritable bowel syndrome; IBS-SSS, IBS Symptom Severity Scale; UCLA-SSS, University of California at Los Angeles Symptom Severity Scale.

**Table 2 – Correlations between IBS-SSS component items, the total IBS-SSS score, and results of a factor analysis on the items.**

<table>
<thead>
<tr>
<th></th>
<th>Pain-Severity</th>
<th>Pain-Frequency</th>
<th>Bloating</th>
<th>Bowel dissatisfaction</th>
<th>Interference</th>
<th>F1 Loading</th>
<th>F2 Loading</th>
<th>Unique variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>.76*</td>
<td>.76*</td>
<td>.66*</td>
<td>.42*</td>
<td>.57*</td>
<td>.87</td>
<td>.17</td>
<td>.24</td>
</tr>
<tr>
<td>Pain-Severity</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain-Frequency</td>
<td>.45*</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td>.65</td>
<td>.45</td>
<td>.48</td>
</tr>
<tr>
<td>Bloating</td>
<td>.53*</td>
<td>.29*</td>
<td>1.00</td>
<td></td>
<td></td>
<td>.80</td>
<td>.02</td>
<td>.34</td>
</tr>
<tr>
<td>Bowel dissatisfaction</td>
<td>.00</td>
<td>.23*</td>
<td>.02</td>
<td>1.00</td>
<td></td>
<td>.00</td>
<td>.85</td>
<td>.25</td>
</tr>
<tr>
<td>Interference</td>
<td>.30*</td>
<td>.25*</td>
<td>−.14</td>
<td>.95*</td>
<td>1.00</td>
<td>.33</td>
<td>.75</td>
<td>.41</td>
</tr>
</tbody>
</table>

Notes. "Total" refers to the total score on the IBS-SSS scale; factor loadings are standardized; unique variances are standardized and reflect the proportion of variance in a measure unaccounted for by the two factors: F1 = Factor 1, F2 = Factor 2.

IBS-SSS, irritable bowel syndrome Symptom Severity Scale.

* A statistically significant correlation, $P < 0.05$. 

---
across all patients runs the risk of imposing through end point clinical significance of IBS symptoms is necessarily uniform by their bowel symptoms [18,19]. The assumption that the significant IBS symptom, while for others, the illness is defined symptoms and pain as equally problematic. But for some stressing, impactful). To be sure, some patients experience bowel the patient's experience (i.e., equally intense, bothersome, dis-
dynamics among component symptoms. One option is to adopt coprimary end points reflecting the core symptoms that define IBS (i.e., abdominal pain and bowel symptoms). Beyond issues of sample size, trial costs, and drug development timelines, this approach presumes that the two symptoms accurately capture the patient's experience (i.e., equally intense, bothersome, distressing, impactful). To be sure, some patients experience bowel symptoms and pain as equally problematic. But for some patients, abdominal pain or discomfort is the most clinically significant IBS symptom, while for others, the illness is defined by their bowel symptoms [18,19]. The assumption that the clinical significance of IBS symptoms is necessarily uniform across all patients runs the risk of imposing through end point selection a value judgment on patients' symptom experience. To us, this would undermine a core principle of the PRO movement, namely, incorporating into therapy development the patient's perspective on the impact of a disease and its treatment on his or her functioning and well-being. An alternative approach based on the recently released FDA guidance [20] is to calibrate the selection of end points to the target symptom and its putative mechanism of action; for example, drugs that target abdominal pain would use a single end point of pain intensity. Of course, if one's goal is merely to document a consistently low score across all symptoms without concern for the different mechanisms that can produce moderate to high scores, then aggregating across the diverse items can be justified. However, aggregating items that are multidimensional in structure carries a realistic risk of introducing a degree of noise that makes it difficult to understand the patients' experience and detect treat-
ment effects. This is particularly problematic for agents that differentially impact different aspects of a given disease "where the intention is not necessarily to cure but to ameliorate symp-
toms, facilitate functioning, or improve quality of life" [8]. For these reasons, we believe that such practices are not conducive to building an informative and useful knowledge base surrounding IBS [21]. Nor will it lend itself to characterizing precisely the therapeutic benefit of novel agents for specific symptoms. In our opinion, it is preferable to distinguish between the different components defining IBS symptom severity and to model these components separately or multivariately, such as in a person-
centered approach.

To this end, we conducted an exploratory cluster analysis of the five symptom items of the IBS-SSS to identify meaningful, distinct patient profiles on the basis of their IBS-SSS responses. We found evidence for four discrete patient subtypes or profiles. One cluster of patients (25%) was characterized by elevated scores on pain frequency and bowel dissatisfaction, with somewhat less elevated but still high mean scores on life interference. The pain severity ratings, however, tended to be relatively low. Although these patients do not report severe pain, what pain they experience is persistent and a source of moderate life interfer-
ence. A second cluster of patients (19%) was characterized by intermediate scores on the pain dimensions, but with more elevated scores on bowel dissatisfaction. These patients seem to endure IBS without experiencing the levels of pain of other groups of patients, but their symptoms take a toll on their well-
being. A third cluster of patients (18%) had elevated scores on all subcomponents of the IBS-SSS. Finally, the fourth and the most common cluster of patients (37%) had relatively low scores on all dimensions except dissatisfaction with bowels and, to a lesser extent, IBS interfering with one's life. Interestingly, the mean

<table>
<thead>
<tr>
<th>Table 3 – Profile means for K-means cluster analysis.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cluster 1</strong></td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Pain-severity</td>
</tr>
<tr>
<td>Pain-frequency</td>
</tr>
<tr>
<td>Bloating</td>
</tr>
<tr>
<td>Bowel dissatisfaction</td>
</tr>
<tr>
<td>Life interference</td>
</tr>
<tr>
<td>Percentage of cases</td>
</tr>
<tr>
<td>UCLA severity mean</td>
</tr>
</tbody>
</table>

Notes. Within a row, cluster means are statistically significantly different at P < 0.05 if they do not share a common superscript. The UCLA severity scale was not a formal part of the cluster analysis. The mean values as a function of cluster are reported here for their interest value. UCLA, University of California at Los Angeles.
global characterization of IBS symptom severity was highest for cluster 3 patients, lowest for cluster 4 patients, and somewhat elevated for cluster 1 and 2 patients. This analysis, of course, was based on a relatively small sample and a restricted population from a single investigative team, and so it must be interpreted judiciously. However, future research might benefit from adapting a person-centered approach to data analysis to characterize common combinations of IBS symptoms. Both the factor analyses and the cluster analyses underscore the utility of differentiating rather than aggregating symptom dimensions.

Clinically, these data have important implications. Classifying patients into subgroups on the basis of features they share may expand our understanding of a disorder marked by dramatic heterogeneity. To date, the Rome criteria are the most sophisticated mechanism for making sense of the clinical diversity of IBS [12]. The goals of Rome criteria, like any classification system, are identifying common features within groups of individuals (within-group homogeneity) and identifying differences with groups of individuals (between-group heterogeneity). The nomothetic focus of any diagnostic system means that it is concerned with the grouping of individuals not heterogeneity within these groups (within-group heterogeneity). As such, it is often assumed that once subjects have been assigned a diagnostic label (e.g., met Rome criteria for IBS) they are sufficiently similar to be randomly assigned to treatment arms to determine the efficacy of a novel agent. The tacit assumption that all patients with a given disorder are one homogeneous group—what is referred to as the patient uniformity myth [22]—is intrinsic to most clinical research that features randomized clinical trials. But just because patients enrolled into a clinical trial satisfy diagnostic criteria does not necessarily mean that they suffer from the same set of problems any more than family members who share the same surname have identical attributes. Not all patients diagnosed with IBS are the same. The severity and complexity of their symptoms, underlying pathophysiology, and treatment goals (e.g., pain reduction, bowel symptom relief) differ often dramatically between patients. Assigning a diagnostic label links individuals of a heterogeneous population on the basis of shared clinical attributes, but it does not “homogenize” pathogenic processes, diverse symptom dimensions, or patients’ treatment goals. We are not confident that global end points (e.g., adequate relief, IBS-SSS, and composite symptom severity scores) are psychometrically sound enough to correct for the within-group heterogeneity that characterizes patients with IBS of so many different stripes. Indeed, there are some reasons to believe that they may actually introduce error variance that makes it more difficult to detect the beneficial effect of therapy [23]. Disregarding individual differences that make up the symptom topography of patients who share a diagnostic label may very well explain why so many studies testing the therapeutic benefit of novel agents yield disappointing results. Peripherally acting therapies for IBS have demonstrated partial efficacy at best with numbers needed to treat in the double digits [24]. Central nervous system pharmacotherapies that have proven efficacious for non-GI pain disorders [25,26] have a disappointing track record with IBS in spite of a growing body of evidence highlighting the abnormal pain processing pathways in IBS [27].

Our data raise the question of whether meaningful conclusions regarding the therapeutic benefit of a given therapy can be confidently drawn until patient heterogeneity is recognized and factored into the design and interpretation of randomized controlled trials. If not, then researchers would be wise to move beyond the “horserace” questions of does treatment X work better than Y for patients with a shared diagnosis and confront the question of what treatments produce what changes for which kind of symptoms [21]. Establishing subtypes of IBS may help develop more targeted, effective interventions that are calibrated to the symptoms for which patients seek relief not the diagnosis they have been assigned. For an illness as heterogeneous as IBS, a claim based on diagnosis confers little information about the specific symptoms for which a treatment is beneficial. A more tailored approach may optimize treatment benefits, minimize harms, and conserve the use of limited health resources. The process of subgrouping patients may yield indirect benefits to physicians and their patients. Because unrecognized heterogeneity among patients can undermine the power of a randomized trial to detect a truly beneficial treatment [28,29], efforts to systematically minimize heterogeneity may help accelerate the development of novel agents and index more accurately their genuine treatment effects that differences among study participants may obscure. This may in the long term improve the quality of care and reduce the disease burden of IBS.

The results of this study are by no means the last or first word on subtyping patients with IBS. Previous efforts to delineate IBS subgroups have focused on patterns of symptoms (e.g., predominant bowel type), their severity, visceral hyperalgesia, intestinal transit, psychological distress, and illness behaviors [30–37]. We chose to use the IBS-SSS as a basis for subtyping because the Rome Foundation recommends it as an end point in clinical trials [7]. That said, we understand that its scope is limited to subjective complaints. Other factors that make up the multifactorial puzzle of IBS include gut-related dysfunctions (motility, secretion, immune function, permeability) as well as brain-related alterations, including increased stress reactivity [38], impaired coping styles [23], and selective attention to threat and symptom-related fears [39]. Further research that integrates the biobehavioral variables and instruments (PROs, biological assessments) that define a subgroup scheme with the greatest clinical and research utility is urgently needed.

The results of the present research, of course, must be interpreted in light of limitations. The sample is relatively small and representative of only a single clinic in an academic medical center located in the northeast. Because the measures of all constructs were based on self-report data, they are subject to some bias and measurement error, which can bias parameter estimates. K-means classification algorithms are less than optimal, but we could not use more advanced latent class analyses because of sample size restrictions. Because this is a cross-sectional study, data were collected at a single point in time. In light of research showing that subtyping of bowel habits among a significant proportion of patients with IBS changes over time [40], it will be important to determine the temporal stability of subtyping derived from the IBS-SSS using longitudinal research designs. Despite such limitations, the results of the research are interesting and set the stage for more refined empirical work on the important issues identified. The research suggests the need to distinguish between different GI symptoms defining IBS severity, without obscuring matters through aggregation into total scores unless unidimensionality can be first demonstrated. This is an empirical question that demands more than leaps of faith to heed the FDA’s call for quality PROs. Even if there is evidence of unidimensionality, if one or more of the individual items has large amounts of unique variance, then, averaging items relegates the unique features of a single symptom to error variance and diverts attention from what might be clinically meaningful facets of patients’ experience of IBS. These are weighty psychometric issues. In the end, attention to these important matters will go a long way in creating an empirically sound, meaningful, and clinically useful PRO that is sensitive to the treatment effects of agents for IBS and other disease states whose symptoms lack a reliable biomarker and are best understood from the patient perspective.

Source of financial support: This research was supported in part by research grants 67878 and 77738 from the National Institutes of Health/NIDDK.
Supplemental Materials
Supplemental material accompanying this article can be found online as a hyperlink at http://dx.doi.org/10.1016/j.jval.2012.08.2207 or, if a hard copy of article, at http://www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES