

Patient preferences for endpoints in fecal incontinence treatment studies

S. Heymen¹ | O. Palsson¹ | M. Simren^{1,2} | W. E. Whitehead¹ 

¹Center for Functional Gastrointestinal and Motility Disorders, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

²Department of Internal Medicine & Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

Correspondence

William E. Whitehead, PhD, Center for Functional Gastrointestinal and Motility Disorders, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.
Email: William.Whitehead@med.unc.edu

Funding information

Funding for data collection was provided by an unrestricted grant from Salix Pharmaceuticals to WEW. Support for data analysis and writing was provided by an educational grant from Ferring Pharmaceuticals to MS.

Abstract

Background: Randomized controlled trials of treatments for fecal incontinence (FI) are difficult to compare because case definitions and study endpoints vary. Our aims were to assess patient perspectives on the case definition for FI and how treatment success should be measured.

Methods: In Phase 1, 28 FI patients participated in anonymous on-line focus groups, and in Phase 2, 186 people with FI, stratified by gender, race, and age completed an online survey.

Key Results: Focus group participants described frequency and urgency as the most important characteristics for defining FI. Most (80%) thought staining of underwear constitutes FI, but only 33% thought gas leakage was FI. When asked how the success of treatment should be defined, 77% said by a reduction in frequency or complete cure, but less than half thought a 50% reduction in frequency was enough. When asked how much reduction would be needed, responses averaged 80%. The Phase 2 survey confirmed that frequency, urgency, and intestinal discomfort are the most important characteristics for case definition, and that success should be defined by at least a 75% decrease in frequency. A 50% reduction was an acceptable endpoint for 58% overall but only 26% for those aged ≥ 65 . "Adequate relief" was acceptable to 78%.

Conclusions and Inferences: Inclusion criteria for trials should specify a minimum frequency of FI. Most patients would require a $\geq 75\%$ reduction in FI frequency to call a treatment successful but young adults and those with more severe FI would accept a $\geq 50\%$ reduction as meaningful.

KEYWORDS

accidental bowel leakage, adequate relief, focus group, inclusion criteria, outcome measure, patient reported outcome

1 | INTRODUCTION

Fecal incontinence (FI) is an important unmet healthcare need¹: It is highly prevalent affecting 8%-10% of non-institutionalized adults in the United States (U.S.)² and approximately half of nursing home residents,^{3,4} and it has a substantial impact on quality of life.^{5,6} However, only one-third of those affected receive medical evaluation and treatment.⁷

Surgical,⁸⁻¹⁰ medical,¹¹ and behavioral treatments¹¹⁻¹³ have been evaluated in randomized controlled trials and reported to be effective, but it has been difficult to compare these trials because the inclusion criteria (ie the case definition of FI) and study endpoints varied. Some studies required patients to have solid or liquid stool leakage to be enrolled^{8,10,11} while other studies included patients who only passed gas involuntarily.^{12,14,15} There is no consensus on whether the involuntary loss of flatus should be included in the case definition; it is included in

the definition of anal incontinence by some investigators¹⁶ because patients often seek treatment for it due to embarrassment, but it is excluded by others because 20.5% of women and 24% of men report involuntary loss of gas on a daily basis, suggesting that this is a normal occurrence.² There is also a lack of consensus on whether stains on underwear should be defined as FI, and it has been excluded in some studies.⁸

The primary outcome measures used in trials have also varied: A fecal incontinence severity score on a questionnaire was the primary outcome in several studies, but some of them used the Fecal Incontinence Severity Index¹¹ while others used the Cleveland Clinic scale¹⁷ or the Vaizey scale.^{13,18} (Involuntary loss of gas is scored as FI on all three of these scales.) Patient reports of “adequate relief” from fecal incontinence¹¹ or global ratings of symptom improvement¹² were the primary endpoints in other trials. Recently published studies, on the basis of which the U.S. Food and Drug Administration (FDA) approved new treatments for FI, defined a responder as someone who reported a reduction in FI frequency of at least 50% on a daily diary.⁸⁻¹⁰ Noelting and colleagues¹⁹ showed that patients reporting at least a 50% reduction in FI frequency were significantly more likely to exceed the Minimum Clinically Important Difference (MCID) on the Fecal Incontinence and Constipation Assessment (FICA) measure of FI symptom severity²⁰ compared to those with less than 50% reduction in frequency of FI, which provides evidence for the validity of this endpoint. However, the MCID in this study was defined by a well-accepted statistical standard (0.5 times the standard deviation and the standard error of measurement) without reference to whether patients found this amount of change satisfactory to them. This differs from the goal of our study, which was to determine whether patients with FI would consider a 50% reduction in FI frequency to be a successful outcome of treatment. Both patients⁶ and clinicians²¹ have questioned whether a 50% reduction in FI is sufficient for them to regard treatment as a success.

The FDA recommends that outcome measures for functional gastrointestinal and motility disorders should be patient-reported outcomes that are developed with input from focus groups.²² However, up to now none of the case definitions or primary outcome measures used in pivotal RCTs to evaluate the effectiveness of treatments for FI were developed with input from patients. The development and validation of patient-based FI severity measures²³ and trial endpoints²¹ have been identified as high priorities for future clinical research.

The purpose of this study was to incorporate patient perspectives into the choice of FI case definitions and the endpoint for clinical trials for FI. Specific aims were (i) to assess patients’ perspectives on what is important with respect to the case definition, (ii) to assess patients’ satisfaction with two current outcome measures—reduction of FI frequency by at least 50%, and adequate relief of FI; (iii) to determine with open-ended questions how people with FI believe treatment success should be measured; and (iv) to assess whether patient preferences for endpoints are influenced by (and potentially confounded with) baseline FI severity and demographic variables.

Key Points

- None of the current case definitions or primary outcome measures used to evaluate the effectiveness of treatments for fecal incontinence (FI) were developed with input from patients. Using focus groups and a national survey, we identified FI case definitions and trial endpoints that are acceptable to most people with FI.
- Staining of underwear is regarded as FI by most people, whereas loss of flatus is not. Inclusion criteria for clinical trials should include a minimum frequency of FI and a minimum level of urgency and intestinal discomfort.
- A 75% reduction in accidents is the endpoint preferred by most people with FI, and substitution of this more rigorous responder definition would not result in fewer treatments being approved.

2 | MATERIALS AND METHODS

The study aims were addressed by a two-step process: In the first phase 28 patients with FI were recruited by advertisement to participate in anonymous focus groups, which were conducted as internet chat rooms with groups of 2-5 patients. Phase 2 was a survey of a nationally representative sample of patients with FI who were identified from the registry of a market-research company (CINT USA Inc., Lawrenceville, NJ, USA). The national survey was used to confirm the findings from the focus groups and to address whether patient preferences for outcome measures are influenced by FI severity and demographic characteristics in a large sample.

2.1 | Focus groups

Participants were recruited via online advertisements, mass e-mails to the students and staff of the University of North Carolina at Chapel Hill (UNC), and a research participant registry maintained by the UNC Center for Functional Gastrointestinal and Motility Disorders. Prospective participants first logged onto a website to give informed consent and answered standardized questionnaires on FI severity (Fecal Incontinence Severity Index,²⁴ Fecal Incontinence and Constipation Assessment questionnaire²⁵) and quality of life (Fecal Incontinence Quality of Life scale²⁶). The on-line consent informed them that they would be asked to participate in chat-room discussions with the investigators and with other patients with FI, and that their participation was anonymous; only arbitrarily assigned numbers (IDs) were used to identify patients in the focus groups. They were told that the topics to be covered in these two-hour sessions would include (i) how FI should be defined, and (ii) what is the best way to measure success when treating bowel leakage. Open-ended questions were combined with more structured probes as shown in Table 1. The initial set of prompts were selected based on a review of outcome

TABLE 1 Prompts used to guide the online focus group discussions

1. How should fecal incontinence be defined, and what should be included in it?
2. Should the volume of stool loss be included? Should there be a minimum amount of leakage?
3. Should the frequency of stool loss be included in the definition? Should it include a minimum number of times?
4. Should how bothersome or embarrassing the bowel leakage is be part of the definition?
5. Should stool consistency (for example, only solid or liquid stool) be included?
6. Should passing gas be part of the definition?
7. If you only stain your underwear, should this be called fecal incontinence?
8. Rate the importance of each of the following terms for the definition of fecal incontinence on a 0-10 scale: Volume, Frequency, Bothersomeness, Consistency (ie only solid or liquid stool loss), ^aUrgency, and ^aIntestinal discomfort.
9. What is the best way to measure success when treating bowel leakage?
10. Is a 50% reduction in fecal incontinence episodes enough improvement?
11. How much reduction in frequency of bowel leakage would be enough for you to consider treatment successful (0-100%)?
12. Would complete continence—a cure—be necessary?
13. Is a question on adequate relief of FI a good way to measure treatment results?

^aRatings of the importance of Urgency, and Intestinal discomfort were added to the focus group probes after other participants mentioned them in response to open ended questions.

measures used in FI treatment trials and FI severity scales, but additional prompts were added based on the responses of patients to open-ended questions. The prompts were provided one at a time, followed by the participants' responses. Once all input was completed, the next prompt was provided.

Focus group members participated from different locations on their own computers and interacted in a chat-room. They typed their responses and comments, which were viewed by all participants in the focus group and also retained as a transcript of the session. The investigators discussed these transcripts among themselves and grouped responses into common themes by consensus. Recruitment and scheduling of focus groups continued until "saturation" was achieved, ie the investigators concluded that successive focus groups were identifying no new content. This is a standard procedure in qualitative research.

TABLE 2 Questions in the national survey on the best endpoint to use in treatment trials

1. In research on treatments for fecal incontinence, a 50% reduction in frequency of incontinence episodes is often considered a success. In your opinion, is that enough improvement to consider treatment successful? [Possible responses: "Yes" or "No"]
2. How much improvement in frequency of fecal incontinence episodes do you think is necessary for treatment to be considered successful? [Possible responses ranged from "10% or less" up to 100% ("total cure") in 10% increments].
3. In research on treatments for FI, a report of "adequate relief" is often used as a measure of positive outcome. In your opinion, is that a good measure of success for this type of problem? [Possible responses: "Yes" or "No."]

2.2 | National survey

Focus group participants rated the importance of six factors for defining FI (Table 1). To confirm the findings from the focus groups, these six items were reassessed in a nationally representative sample where a new sample of subjects with FI were asked to rate the importance of these factors to the definition of FI. Responses of the focus group participants to the questions on the best endpoints for treatment trials (items 9-13 in Table 1) enabled us to formulate three more precise questions about trial endpoints for the national survey; they are listed in Table 2.

The sample for the National Survey was recruited by CINT USA, Inc., which is a market research company (<http://www.cint.com>) with a large registry of survey respondents. The health and continence status of people in this registry was unknown to CINT USA. To minimize possible bias through self-selection for participation, the invitation to potential subjects only described the study as a health survey, and a screening question was embedded among questions on other physical symptoms to identify subjects with FI. Quota sampling was employed to recruit equal numbers of subjects with and without FI; equal numbers of males and females; balanced age strata (ie 40% aged 20-40 years, 40% aged 41-60 years, and 20% aged 61+); and approximately 60% Caucasians, 20% African Americans and 20% Hispanics.

The survey included 43-53 questions (dependent on branching) and required approximately 20-25 minutes to complete. (Some questions addressed other research aims such as the terminology subjects would prefer their doctors to use when they discuss FI with them, and a continent control group was also included in the survey; these data will be reported in a separate publication.) Two questions from an early part of the questionnaire were repeated unexpectedly towards the end of the survey as a check on data quality, and subjects whose responses to these multiple choice questions differed by more than one step on a 5-point ordinal scale were judged unreliable and were excluded from the analysis.

2.3 | Statistical analysis

In Phase 1 descriptive statistics, primarily frequency counts, were used to describe the responses of focus group participants. Participants' ordinal scale ratings of the importance of different factors to the definition of a treatment responder were compared by t-tests for related samples in both Phase 1 and Phase 2. Phase 2 was an observational study of a single cohort of people with FI, and descriptive statistics (means and 95% confidence intervals) were used to address aims 1 and 2. Analysis of how FI severity and demographic characteristics of participants influenced the acceptability of key responder definitions

was assessed by Chi square statistics, followed by binary logistic regression analysis to identify which variables made independent contributions to the participant's views on the acceptability of these study endpoints. The independent variables included in these regression models were age, gender, race, education, household income, relationship (living alone vs cohabiting or married), and severity of fecal incontinence (FISI total score). All independent variables were entered as a single block. Correlations between ordinal scaled variables were assessed by nonparametric Spearman correlation coefficients.

Both phases of this study were reviewed and approved by the Institutional Review Committee for the Protection of Human Research Participants at the University of North Carolina. Phase 2 of the study was approved as IRB exempt because all data collection was done anonymously.

3 | RESULTS

3.1 | Phase 1—Focus groups

Participants in the focus groups included 27 women and one man ranging in age from 21 to 87 years. To reassure participants about their anonymity in these small focus groups, no other demographic information was collected. The average Fecal Incontinence Severity Index (FISI, range 0-61) for these subjects was 23.4 (95% confidence interval=18.7, 28.2). Nineteen of these 27 subjects (70%) reported at least weekly solid or liquid FI, which included 7/27 (26%) who reported daily solid or liquid FI. The reported volume of solid or liquid stool accidents was at least a small amount (ie more than staining) for 23/27 (85%) and was at least a moderate amount (3-5 teaspoons) in 11/27 (41%). Thus, most focus group participants had moderately severe or very severe FI.

3.1.1 | Definition of FI

When focus group participants were asked whether staining of underwear and/or passing gas should be included in the definition of FI, 78.6% felt staining should be counted as FI, but only 33.3% felt that unintentionally passing gas should be included. When asked whether the following factors should be included in the definition of FI, most answered "yes" to consistency of stool loss (82.1%) and frequency of FI (57.1%); however, only 42.9% indicated that embarrassment or bother should be included in the definition and 37% said volume of stool lost should be included. When these focus group participants were asked to rate the importance of these four characteristics plus two others that were volunteered by patients in response to open ended question, namely "urgency" and "intestinal discomfort", to the definition of FI (Table 3), the highest ratings were assigned to urgency and frequency. Frequency was rated significantly more important to the case definition of FI compared to volume ($P<.001$) and consistency ($P=.002$) but not compared to urgency, intestinal discomfort, or bothersomeness ($P>.01$ for all). Urgency was rated significantly more important to the case definition compared to volume ($P=.002$), consistency ($P=.004$), and bothersomeness ($P=.047$) but not compared to intestinal discomfort ($P>.05$ for both). None of the other ratings

TABLE 3 Patient ratings of the importance of six factors to the definition of fecal incontinence

| | Rating of Importance to Definition of FI on 0-10 Scale (Mean±95% CI) | |
|-------------------------------|--|--------------------------|
| | Phase 1: Focus groups | Phase 2: National survey |
| Volume of stool lost | 5.00 (3.44, 6.56) | 6.89 (6.49, 7.29) |
| Frequency | 8.44 (7.60, 9.29) | 7.63 (7.28, 7.98) |
| Bothersome/ embarrassing | 7.26 (5.85, 8.67) | 6.86 (6.47, 7.25) |
| Consistency (liquid or solid) | 5.74 (4.25, 7.24) | 6.88 (6.51, 7.25) |
| Urgency | 8.70 (7.63, 9.77) | 7.56 (7.22, 7.91) |
| Intestinal discomfort | 7.94 (6.61, 9.28) | 7.42 (7.08, 7.77) |

made by focus group participants were significantly different from each other.

3.1.2 | Endpoints for clinical trials

When focus group participants were asked an open-ended question, "In your opinion, what is the best way to measure success when treating bowel leakage?" 13 of the 22 who responded to this question said the frequency of FI episodes would have to decrease, and an additional 4 of 22 said only a complete cure would be considered a success; thus 17 of 22 (77.3%) endorsed reductions in the frequency of FI as an important measure of successful treatment. Other responses mentioned by at least one participant were decreased use of continence pads, smaller amount of bowel leakage, and having less messy episodes of FI. Two responses were not interpretable as answers to this question (ie "I have had no success due to the IBS" and "How can there be success if it still occurs?").

When asked whether a 50% reduction in the frequency of bowel accidents would be enough to consider a treatment successful, 11/27 (41% of focus group participants) said "yes" but several of them qualified their answer: a typical comment of a participant who answered "yes" was "it's a whole lot better but I wouldn't call it enough..." When asked what percent reduction in FI frequency would be necessary to call the treatment a success, the median response was 80%.

When they were asked whether "adequate relief" of FI was a good way to measure the success of treatment, 11/26 responders (42%) said "yes". However, 10/26 responders (39%) said that "adequate relief" is too subjective or needs to be defined, and one thought that, except for medical professionals, people would not understand the concept.

3.2 | Phase 2—National survey

Participants completing the national survey included 234 with FI. However, 48 (20.5%) were excluded from analysis because they gave inconsistent responses to two questions that were repeated as a quality control, leaving 186 for analysis. One hundred three (55%) reported that they had consulted a physician for management of FI. The characteristics of the final sample of 186 respondents with FI are given in Table 4.

3.2.1 | Definition of FI

In the Phase 2 survey a nationally representative sample of 184 people with FI was asked to rate the importance of the same six characteristics for defining FI that the focus group participants rated (Table 3). Frequency and urgency were rated the most important to the definition of FI, followed by intestinal discomfort. Volume and consistency of stool lost and bothersomeness were significantly less important to these participants than frequency, urgency, and intestinal pain ($P<.005$ for all comparisons).

TABLE 4 Characteristics of National Survey Participants

| Participant characteristic | Sample distribution |
|--|---------------------|
| Sex ^a | |
| Females, n (%) | 97 (52) |
| Males, n (%) | 89 (48) |
| Race ^a | |
| White, n (%) | 154 (83) |
| Black, n (%) | 15 (8) |
| Hispanic, n (%) | 17 (9) |
| Age (mean (SD) in years) for all subjects with FI ^b | |
| Aged 18-34.9, n (%) | 50 (27) |
| Aged 35-64.9, n (%) | 94 (50) |
| Aged 65+, n (%) | 42 (23) |
| Marital status | |
| Single, widowed, or divorced, n (%) | 68 (37) |
| Married or cohabiting, n (%) | 118 (63) |
| Education | |
| High school or less, n (%) | 27 (15) |
| Some college, n (%) | 86 (46) |
| Post-baccalaureate, n (%) | 73 (39) |
| Household income | |
| Less than \$35 000, n (%) | 62 (33) |
| \$35 000 to \$75 000, n (%) | 45 (24) |
| Over \$75 000, n (%) | 72 (39) |
| Missing data, n (%) | 7 (4) |
| Fecal Incontinence Severity Index (mean, CI) | |
| Mildest tertile (mean, CI) | 12.1 (11.03, 13.22) |
| Middle tertile (mean, CI) | 26.0 (24.5, 27.5) |
| Most severe tertile (mean, CI) | 51.8 (50.7, 53.0) |
| Fecal Incontinence Quality of Life total score for all patients with FI (mean, CI) | |
| Most affected tertile on FIQOL (mean, CI) | 1.6 (1.6, 1.7) |
| Middle tertile on FIQOL (mean, CI) | 2.5 (2.4, 2.6) |
| Least affected tertile on FIQOL (mean, CI) | 3.5 (3.4, 3.6) |

CI, 95% confidence interval.

^aQuota sampling was used for the first three demographic characteristics to ensure sample sizes were adequate for all demographic segments.

3.2.2 | Outcome measures for clinical trials

When Phase 2 participants were asked whether a 50% reduction in FI frequency is enough to consider a treatment successful, 107/186 (58%) believed this endpoint was acceptable. However, as shown in Table 5, acceptability of this endpoint varied greatly as a function of FI severity, impact of FI on quality of life, and demographic characteristics. When FISI total severity scores were divided into tertiles, the third of participants with the most severe FI were 2.4 times more likely than those with the mildest FI severity to be satisfied with this responder definition ($P<.001$); and similarly, participants with the greatest impact of FI on their quality of life were 2.2 times more likely than those with the least quality of life impact to find this responder definition acceptable ($P<.001$). Participants who had consulted a physician for treatment of FI were 1.9 times as likely to find this responder definition acceptable compared to non-consulters ($P=.001$).

Demographic characteristics also had a major impact on the acceptability of defining a responder by a reduction in FI frequency of at least 50% (Table 5). The participants least likely to find a 50% reduction in FI frequency acceptable to them were female subjects aged 65 or older. Lower socioeconomic status, defined either by income or education, was also associated with a reduced likelihood of finding this endpoint acceptable. Hispanic subjects were more likely to find this endpoint acceptable, but the number of Hispanics in the sample was small ($n=9$).

Logistic regression was used to determine whether the subjects' satisfaction with a reduction in FI frequency of 50% or more is a satisfactory measure of treatment success; independent variables tested were age, sex, education, household income, relationship (ie single, divorced, married or cohabiting), race, and fecal incontinence severity. A test of the full model against a constant-only model was statistically significant, indicating that the predictors as a set reliably distinguished between those who found a 50% reduction to be a measure of treatment success vs those who were not satisfied with this endpoint ($\chi^2=83.96$, $P<.001$, $df=12$). Nagelkerke's R^2 of .504 indicated a modest relationship between the predictors and the dependent measure. Overall prediction success was 78.2%. The Wald criterion demonstrated that the significant independent predictors were of younger age (Wald=14.172, $P<.001$), Hispanic race (Wald=4.39, $P=.036$), and higher household income (Wald=9.603, $P=.008$).

The effects of age on satisfaction with a 50% reduction in FI and with "Adequate Relief" are shown in Figure 1.

Responses to the first two questions in Table 2 are key outcomes for this study, so it is important to know whether subjects answered these questions in a consistent manner. To address this question we compared the distribution of responses to Question 2, (how much improvement is necessary for treatment to be considered successful) to the responses to Question 1 (is a 50% reduction in frequency of FI enough to consider treatment successful, yes or no). Responses to these questions were highly correlated: the median response of subjects who said "yes" to Question 1 was 90% on Question 2 compared to 70% for those who said "no" to Question 1 ($\chi^2=41.6$, $P<.001$).

TABLE 5 Patient characteristics that moderate the acceptability of different outcome measures

| Possible moderator | Responder defined by ≥50% less FI (%) | Responder defined by adequate relief of FI (%) | How much decrease is required for treatment success? (%) |
|----------------------------|--|---|---|
| Sex | | | |
| Female | 45 | 68 | 77 |
| Male | 71 ^a | 77 ^a | 77 |
| Age ranges | | | |
| Age 18-34 years | 80 | 92 | 70 |
| Age 35-64 years | 60 ^{a,c} | 83 ^c | 80 ^a |
| Age ≥65 years | 26 ^{a,b} | 50 ^{a,b} | 79 ^a |
| Race | | | |
| White | 55 | 77 | 77 |
| Black | 40 | 67 ^c | 80 |
| Hispanic (any race) | 94 ^{a,b} | 100 ^{a,b} | 72 |
| Education | | | |
| High School or less | 44 | 59 | 79 |
| College | 47 ^c | 78 ^a | 76 |
| Post-graduate | 75 ^{a,b} | 85 ^a | 78 |
| Family income | | | |
| <\$35 000 | 42 | 65 | 77 |
| \$35 000-\$75 000 | 38 ^c | 76 ^c | 78 |
| >\$75 000 | 86 ^{a,b} | 93 ^{a,b} | 76 |
| Marital status | | | |
| Married or cohabiting | 63 | 84 | 78 |
| Single, widowed, divorced | 49 | 68 ^a | 75 |
| Consulter for FI | | | |
| Consulted MD for FI | 39 | 67 | 77 |
| Did not consult for FI | 73 ^a | 86 ^a | 77 |
| FI severity (FISI score) | | | |
| FISI <18 (mildest tertile) | 38 | 70 | 78 |
| FISI 19-38 (middle) | 43 ^c | 69 ^c | 76 |
| FISI >39 (most severe) | 92 ^{a,b} | 95 ^{a,b} | 78 |
| FIQOL Total Score | | | |
| Most affected tertile | 87 | 95 | 78 |
| Intermediate | 46 ^a | 72 ^a | 76 |
| Least affected | 40 ^a | 67 ^a | 78 |
| Volume of stool loss | | | |
| Small amount (staining) | 42 | 72 | 76 |
| Moderate amount | 45 ^c | 67 ^c | 76 |
| Large amount (full BM) | 88 ^{a,b} | 95 ^{a,b} | 80 |

Superscripts: ^aSignificantly different from row A; ^bSignificantly different from row B; ^cSignificantly different from row C.

3.2.3 | Defining a treatment responder by a report of Adequate Relief of FI

When participants were asked whether “adequate relief” of FI would be acceptable for identifying treatment responders, 145/186 (78%) said “yes”. Demographic variables that were found in univariate statistical tests to be significantly associated with

finding adequate relief an acceptable measure of treatment response included age, sex, race, family income, and education. This pattern of results is similar to that seen for the responder definition based on at least a 50% decrease in FI episodes. As shown in Table 5, the acceptability of “adequate relief” as a trial endpoint was related to the severity of FI: The third of participants with the most severe FISI scores were 1.4 times more likely to find adequate

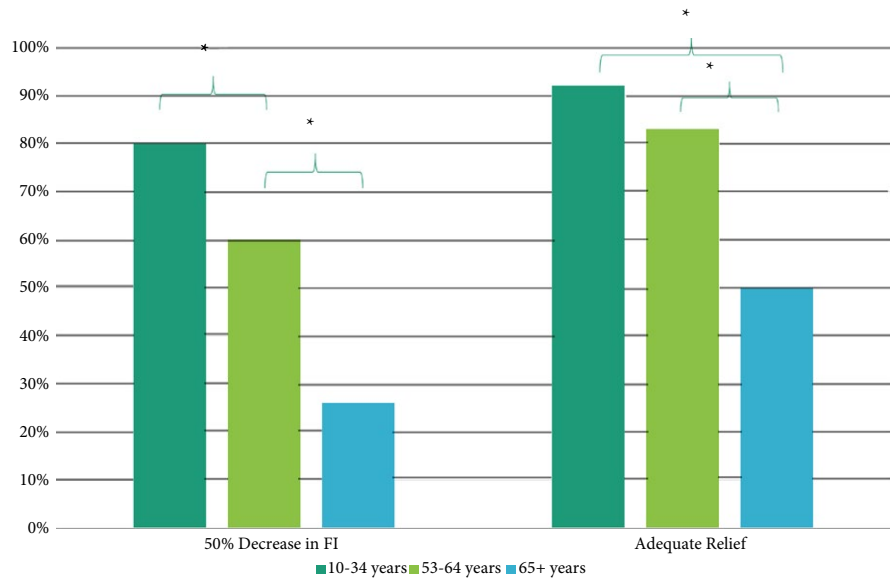


FIGURE 1 Proportion of each age group reporting satisfaction with two endpoints: at least a 50% reduction in FI episodes and “adequate relief” of FI. *Significantly different at $P < .05$

relief acceptable compared to participants with the least severe FISI scores ($P < .001$). Consulters were also significantly more likely than non-consulters to endorse adequate relief as a trial endpoint ($P = .002$).

Logistic regression was also used to determine whether the subjects’ satisfaction with “adequate relief” as a measure of successful treatment can be predicted by the same independent variables. A test of the full model against a constant-only model was statistically significant, indicating that the predictors as a set reliably distinguished between those who found “adequate relief” to be an acceptable measure of treatment success vs those who did not ($\chi^2 = 43.119$, $P < .001$, $df = 12$). Nagelkerke’s R^2 of .332 indicated a weak but statistically significant relationship between the predictors and the dependent measure. Overall prediction success was 79.9%. The Wald criterion demonstrated that the only significant independent predictor was younger age (Wald = 9.931, $P = .002$). Figure 1 shows the association of age with patient views on whether “adequate relief” of FI is an acceptable endpoint for treatment trials.

3.2.4 | How much reduction in FI frequency is enough?

When survey participants were asked to rate how much reduction in FI would be required before they considered a treatment successful, the average amount of reduction required was 77%, with a 95% confidence interval of 75%-80%. The response scale was in 10% increments. Figure 2 shows the distribution of responses to this question.

Linear regression was used to investigate whether the amount of improvement required to consider treatment successful could be predicted by the demographic and symptom severity measures that were found to be significantly associated with subjects reporting satisfaction with a 50% or greater reduction in FI or with “adequate relief”. However, no variable tested was a significant independent predictor. The adjusted R^2 was .003.

4 | DISCUSSION

The aims of this study were to ascertain patient perspectives on how fecal incontinence should be defined and what outcome measures should be used to measure the success of FI treatment. We approached this by beginning with focus groups of identified FI patients and then surveying a population-based sample of 184 adults with FI (including some non-patients with FI) to confirm and elaborate the focus group findings. While one other study²⁷ has used focus groups of patients with FI to address the definition of FI and trial endpoints, the methods used differed from ours, and so did many of the findings: Sung and colleagues began with a systematic review of the literature to develop a conceptual framework describing how patients think about FI. They used this framework to guide their interviews with patients, and the comments of patients confirmed the a priori conceptual framework without significant elaboration of it. By contrast, we emphasized open-ended questions such as “how would you define FI?” and “what do you think is the best way to measure success in treatment trials for FI?” We supplemented these open-ended questions with probes to address specific topics, and we also asked focus group participants to rate the importance of the factors they identified as part of the definition of FI.

4.1 | Definition of fecal incontinence

In both the Sung study²⁷ and our study, frequency of FI and urgency were felt to be part of the definition. However, in our study intestinal discomfort was listed as being a defining symptom of FI while it was mentioned only incidentally in the Sung study. For our subjects, the most important symptoms defining FI were frequency of FI episodes, urgency sensations, and intestinal discomfort while volume of stool lost, consistency of leakage (eg solid, liquid, or gas), and bothersomeness were rated as significantly less important.

“Intestinal discomfort” was queried in both Phase 1 and Phase 2 surveys because it was mentioned by 3 focus group participants.

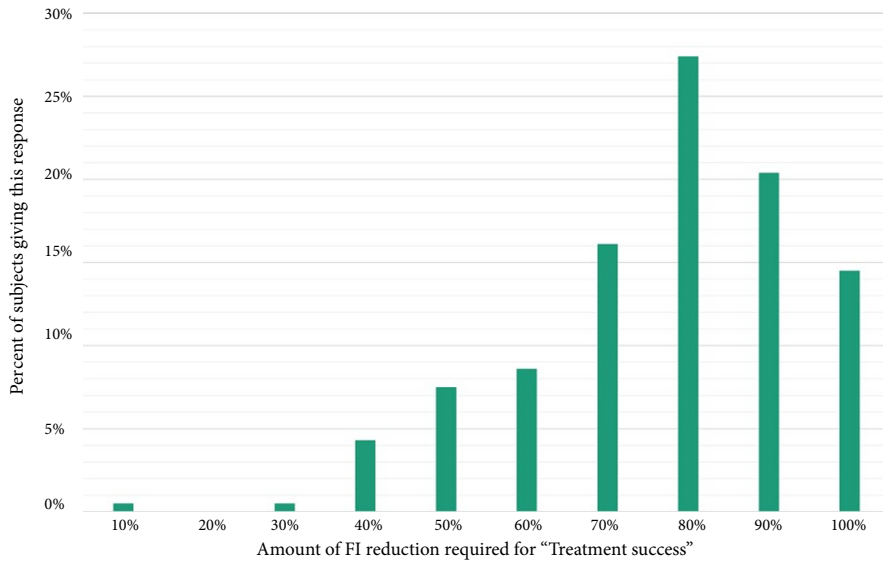


FIGURE 2 Frequency of responses to the question, “How much improvement in frequency of fecal incontinence episodes do you think is necessary for treatment to be considered successful?”

However, this is an ambiguous term. We did not include questions to clarify its meaning because its mention by patients in Phase 1 was unanticipated, and because we wanted the two surveys to reflect as much as possible the patients’ perspectives on FI. However, in post hoc analyses we found that ratings of the importance of intestinal discomfort were significantly correlated with ratings of urgency (Spearman $\rho = .559, P < .001$), and that intestinal discomfort was not correlated with a screening question on whether patients had abdominal pain at least once a month ($\rho = -.016, P = .829$). These post hoc analyses suggest that the sensation of urge preceding FI may be the primary determinant of patient reports of intestinal discomfort, but further research is needed to better understand what patients mean by this symptom.

The conceptual model of FI developed by Sung et al. included two factors not detected in our study, namely the predictability of accidental bowel leakage and the association of leakage with constipation. These differences may be a consequence of sample variability as the focus group samples in both studies were relatively small. Further research is needed to confirm the importance of predictability of FI as a dimension for defining FI and evaluating its severity, and the role of constipation as a cause of FI should be acknowledged although it may be limited to a subset of patients.

A novel aspect of our study was the inclusion of specific questions about whether passing gas or staining underwear should be considered FI. We found that 78.6% felt that staining of underclothes should be called FI but only 33.3% thought passing gas should be defined as FI.

4.2 | Endpoints in clinical trials

Anecdotal reports suggest that many patients are not satisfied with defining a 50% reduction in frequency of FI as a treatment success^{6,21} even though this has been the primary endpoint in recent clinical trials and was the basis for the FDA’s approval of sacral neuromodulation⁸ and dextranomer¹⁰ injections for treatment of FI. One aim of our study was to investigate this claim. Overall 58% of adults with FI

in our community population survey found this to be an acceptable definition of treatment success. However, the people in the age group at greatest risk of FI were least satisfied with it: only 26% of people with FI aged 65 and older found this endpoint acceptable.²

When we compared the responses of subjects in the national survey to a yes/no question about whether a 50% reduction in FI frequency was an acceptable measure of success in a clinical trial to their responses to a more open-ended question in which they were asked how much improvement in FI frequency is necessary for them to consider treatment successful, we found that the responses to the two questions were significantly correlated, supporting the validity of the study design. However, many subjects who answered that they were willing to accept a 50% decrease as a meaningful endpoint in a clinical trial would nevertheless prefer a larger decrease in FI frequency to consider treatment successful. For example, one subject who answered yes to this question said “it’s a whole lot better but I would not call it enough”. The regression analyses also reinforce this interpretation: responses to the question on whether a 50% reduction in FI would be enough to qualify as successful treatment in a clinical trial were significantly influenced by demographic factors such as age, sex, and race/ethnicity and measures of FI severity and quality of life impact, but responses to the question on how much improvement in FI frequency would be needed to regard a treatment as successful were not significantly influenced by these demographic variables or measures of FI severity and quality of life impact.

Several patient characteristics were identified that influence whether people with FI agree that a 50% reduction in frequency of episodes is an acceptable outcome measure (Table 5). Individuals with the most severe FI and those with the greatest impact of FI on their quality of life were the most likely to be satisfied with a 50% reduction in FI in a treatment trial. Moreover, the demographic variables in Table 5 that are significant moderators of satisfaction with this trial endpoint are even more important than FI severity and quality of life impact because these are potential confounders in clinical trials: In addition to age (see Figure 2), we found that female sex and lower family

income were significantly associated with a reduced likelihood of finding a 50% reduction in FI frequency an acceptable measure of success.

A recent reanalysis²⁸ of the data from the dextranomer trial¹⁰ showed that a 50% reduction in FI episodes was significantly correlated with improvements in the Cleveland Clinic Fecal Incontinence Score,²⁹ the Fecal Incontinence Quality of Life scale,²⁶ and several diary-based measures of fecal incontinence. The authors concluded that the responder definition based on a 50% decrease in FI episodes is valid and responsive from a psychometric perspective. Moreover, Noelting and colleagues¹⁹ showed that patients who achieved at least a 50% reduction in FI frequency were significantly more likely to exceed the MCID for two different measures of symptom improvement. We accept the conclusions of these two studies that a 50% reduction in FI frequency identifies patients who show a statistically significant improvement, but we remind the reader that our study addressed a different question, namely whether people with FI believe that a 50% reduction in FI episodes is enough for them to regard the treatment as successful.

We also assessed satisfaction with a second endpoint that has been used in biofeedback trials,¹¹ namely “adequate relief of FI”. This measure is akin to global rating scales of improvement which have been used in some studies.^{12,13} The proportion of people with FI who found this endpoint satisfactory was higher than for the 50% reduction in FI frequency criterion (78% vs 58%), but only half (50%) of people aged 65 or older were satisfied with this endpoint (Figure 1). When focus group participants were asked about “adequate relief”, 39% of them said they believed this phrase was very subjective and they would need to have it defined for them, and one said this is the way doctors talk and not the way patients talk. On this basis we expected to find that acceptance of this endpoint was related to the participant’s education, and this was confirmed in the national survey: those with a college or postgraduate education were more likely to find “adequate relief” an acceptable way of identifying treatment responders than those with only a high school education.

The second aim of our study was to use qualitative research techniques to learn how people with FI think treatment success should be defined. When asked “What is the best way to measure treatment success in the treatment of accidental bowel leakage?” an overwhelming majority (77%) of focus group participants said there should be a reduction in FI frequency or a complete cure. The national survey confirmed that patients with FI assign greater importance to a reduction in frequency of episodes than they do to the volume or consistency of stool lost or how bothersome the bowel accident is to them. When we asked people with FI how much reduction in the frequency of FI episodes would be needed to consider treatment successful, the average response was 77% and the 95% confidence interval was 75% to 80%. Moreover, responses to this question were not significantly influenced by FI severity or demographic characteristics other than age, and the impact of age was modest compared to the other two outcome measures we tested. Based on these data we suggest that success in clinical trials of FI should be defined as at least a 75% reduction in FI episodes (or days with FI) measured by a daily symptom log. However, to insure that studies using this outcome measure can be compared to

already published studies, we recommend that a $\geq 50\%$ reduction be included as a secondary responder definition.

4.3 | Should decisions about the success of treatment be based on a multidimensional scale?

Our data and those of Sung et al.²⁷ suggest that patients define FI by several symptoms in addition to the frequency of FI events; they rate urgency to defecate as equally important to frequency, and the symptom of intestinal discomfort is nearly equivalent in importance. Based on the study by Sung et al., unpredictability should be added to this list. However, with the exception of the frequency of accidental bowel leakage, these symptoms are neither necessary nor sufficient to define the occurrence of FI, and it would be more logical to treat them as component measures of the severity of FI. These data suggest that the amount of urgency, intestinal discomfort, and predictability associated with bowel leakage should be considered for inclusion criteria for clinical trials, especially when these characteristics are relevant to the mechanism of action of the investigational intervention. These symptoms should also be incorporated into multicomponent measures of FI severity.

Several questionnaires^{5,24,29-33} have been developed and validated for the measurement of FI severity which attempt to capture the multidimensional aspects of accidental bowel leakage. These were recently reviewed by Bharucha.²³ All of these scales ask separate questions about the frequency of different categories of leakage (gas, staining or mucus, liquid stool, or solid stool), and some incorporate the volume of leakage, the presence of a warning sensation before stool loss, the sensation of urgency, and the use of pads or antidiarrheal medications. Responses to these detailed questions can provide valuable guidance to the clinician in planning a management strategy, and these scales can be recommended as secondary endpoints in clinical trials and as clinical assessment tools. However, there is no consensus so far on the optimal FI severity questionnaire: the summary scores on these questionnaires are not comparable and they are not easily translated into clinically meaningful differences between patients.^{23,34} Our suggestion that the primary outcome in RCTs for treatment of FI should be based on a $\geq 75\%$ reduction in FI frequency is simple, straightforward in its interpretation, reflects patient preferences, and is consistent with the outcome measure of $\geq 50\%$ reduction in FI frequency used in several recent pivotal trials.^{8,10}

If our suggestion to define treatment as successful only if FI frequency decreases by 75% instead of 50% were adopted, how would the interpretation of recently published randomized controlled trials be affected? Wexner⁸ reported that 89% of patients treated with sacral nerve stimulation met the 50% reduction criterion, and this rate would drop to 65% if success required a 75% reduction in frequency. Our group¹¹ reported that 76% of the patients treated with biofeedback reported adequate relief at 3 months follow up, and when we reanalyzed our data we found that 93% met the 50% responder definition and 78% met the 75% responder definition [unpublished observations]. Graf¹⁰ reported that 52% met the 50% responder definition,

and this decreased to approximately 28% for a 75% responder threshold when reanalyzed by Sanchez.²⁸

A concern has been raised that adopting this more restrictive responder definition could result in treatments being disapproved which are potentially useful for patients with more severe FI. In all three of the reference trials described, only patients with severe FI (defined as at least 1¹¹ or 2^{8,10} solid or liquid FI episodes per week) were included, and the average baseline frequency of FI was greater than 3 days per week in all three studies. When the more restrictive 75% decrease in FI frequency was compared to a 50% decrease, the responder rate decreased by 15%¹¹ to 24%^{8,10}, but the responder rates in the control conditions also decreased, and in neither of the two RCTs^{10,11,28} would the adoption of the 75% criterion have resulted in a non-significant difference between active treatment and control groups.

Our study has limitations: The questions used to assess how much improvement in FI frequency subjects would require to call a treatment successful were developed for this study and have not been independently validated. The focus group participants were a convenience sample and may not be representative of all patients with FI. Only one male was included in the focus groups so the qualitative data may not be generalizable to males. This is a concern because the etiologies for FI often differ in males and females, and future research should address sex differences in the perception of what constitutes a successful therapeutic outcome. The survey sample, while nationally representative and unselected with respect to prior knowledge of what the survey was about, nevertheless contained more subjects with post-baccalaureate education than expected. Institutionalized and home-bound patients with FI were not included in our study, and they are known to have a higher prevalence of FI than the non-institutionalized subjects we studied; however, they are rarely included in clinical trials. Strengths of the national survey are that we stratified the sample by age, sex, and race to ensure adequate representation of these subgroups, and we excluded subjects who gave inconsistent responses to quality control questions.

Our findings have implications for clinical practice and the conduct of clinical trials. With regard to the definition of FI, staining of underwear is regarded as FI by ¾ of patients with FI, but only 1/3 regard the loss of flatus as FI. Our data also suggest that inclusion criteria for clinical trials should include a minimum frequency of FI and a minimum level of urgency and intestinal discomfort, but the volume of accidents, consistency of stools lost, and bothersomeness are of less importance. With regard to the outcome of treatment, clinicians and investigators should be aware that many of their patients with FI will not be satisfied with treatments that are FDA approved but that produce relatively modest decreases in stool frequency of 50%; many patients expect a reduction of at least 75% before they consider a treatment as successful. Clinical researchers should be aware that age, sex, and socioeconomic status influence response rates for some outcome measures and should insure that their groups are balanced on these dimensions.

ACKNOWLEDGMENTS

Sung Min Kim and Stefanie Twist provided administrative support to the study.

AUTHOR CONTRIBUTION

WEW, SH, and OP designed the study and wrote the paper; SH, OP, and WEW conducted focus groups and analyzed the focus group data; OP developed the online survey; OP, WEW, and MS analyzed the survey data; MS participated in writing the paper.

DISCLOSURE

Salix Pharmaceuticals provided support for data collection and Ferring Pharmaceuticals supported the role of MS in data collection.

REFERENCES

1. Landefeld CS, Bowers BJ, Feld AD, et al. National Institutes of Health State-of-the-Science Statement: prevention of fecal and urinary incontinence in adults. *Ann Intern Med.* 2008;148:449–458.
2. Whitehead WE, Borrud L, Goode PS, et al. Fecal incontinence in US adults: epidemiology and risk factors. *Gastroenterology.* 2009;137:512–517, 7 e1-2.
3. Dey AN. *Characteristics of Elderly Nursing Home Residents: Data From the 1995 National Nursing Home Survey.* Advance data from vital and health statistics; no. 289. Hyattsville, MD: National Center for Health Statistics; 1997.
4. Borrie MJ, Davidson HA. Incontinence in institutions: costs and contributing factors. *CMAJ.* 1992;147:322–328.
5. Bharucha AE, Zinsmeister AR, Locke GR, et al. Prevalence and burden of fecal incontinence: a population-based study in women. *Gastroenterology.* 2005;129:42–49.
6. Norton NJ. The perspective of the patient. *Gastroenterology.* 2004;126(1 Suppl. 1):S175–S179.
7. Dunivan GC, Heymen S, Palsson OS, et al. Fecal incontinence in primary care: prevalence, diagnosis, and health care utilization. *Am J Obstet Gynecol.* 2010;202:493.e1–e6.
8. Wexner SD, Collier JA, Devroede G, et al. Sacral nerve stimulation for fecal incontinence: results of a 120-patient prospective multicenter study. *Ann Surg.* 2010;251:441–449.
9. Tjandra JJ, Chan MK, Yeh CH, Murray-Green C. Sacral nerve stimulation is more effective than optimal medical therapy for severe fecal incontinence: a randomized, controlled study. *Dis Colon Rectum.* 2008;51:494–502.
10. Graf W, Mellgren A, Matzel KE, Hull T, Johansson C, Bernstein M. Efficacy of dextranomer in stabilised hyaluronic acid for treatment of faecal incontinence: a randomised, sham-controlled trial. *Lancet.* 2011;377:997–1003.
11. Heymen S, Scarlett YV, Jones KR, Ringel Y, Drossman DA, Whitehead WE. Randomized controlled trial shows biofeedback to be superior to pelvic floor exercises for fecal incontinence. *Dis Colon Rectum.* 2009;52:1730–1737.
12. Norton C, Chelvanayagam S, Wilson-Barnett J, Redfern S, Kamm MA. Randomized controlled trial of biofeedback for fecal incontinence. *Gastroenterology.* 2003;125:1320–1329.
13. Solomon MJ, Pager CK, Rex J, Roberts R, Manning J. Randomized, controlled trial of biofeedback with anal manometry, transanal ultrasound, or pelvic floor retraining with digital guidance alone in the treatment of mild to moderate fecal incontinence 7. *Dis Colon Rectum.* 2003;46:703–710.
14. Schwandner T, König IR, Heimerl T, et al. Triple target treatment (3T) is more effective than biofeedback alone for anal incontinence: the 3T-AI study. *Dis Colon Rectum.* 2010;53:1007–1016.
15. Damon H, Siproudhis L, Faucheron JL, et al. Perineal retraining improves conservative treatment for faecal incontinence: a multicentre randomized study. *Digest Liver Dis.* 2014;46:237–242.
16. Norton C, Christiansen J, Butler U, et al. Anal Incontinence. In: Abrams P, Cardozo L, Khoury S, Wein A, eds. *Incontinence.*

- International Consultation on Incontinence*. 2002. Plymouth: Health Books; 2002.
17. Peirce C, Murphy C, Fitzpatrick M, et al. Randomised controlled trial comparing early home biofeedback physiotherapy with pelvic floor exercises for the treatment of third-degree tears (EBAPT Trial). *BJOG*. 2013;120:1240-1247; discussion 6.
 18. Dehli T, Stordahl A, Vatten LJ, et al. Sphincter training or anal injections of dextranomer for treatment of anal incontinence: a randomized trial. *Scand J Gastroenterol*. 2013;48:302-310.
 19. Noelting J, Zinsmeister AR, Bharucha AE. Validating endpoints for therapeutic trials in fecal incontinence. *Neurogastroenterol Motility*. 2016;28:1148-1156.
 20. Bharucha AE, Locke GR III, Seide BM, Zinsmeister AR. A new questionnaire for constipation and faecal incontinence. *Aliment Pharmacol Ther*. 2004;20:355-364.
 21. Wald A. New treatments for fecal incontinence: update for the gastroenterologist. *Clin Gastroenterol Hepatol*. 2014;12:1783-1788.
 22. Guidance for industry. patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Quality Life Out*. 2006;4:79.
 23. Bharucha AE, Dunivan G, Goode PS, et al. Epidemiology, pathophysiology, and classification of fecal incontinence: state of the science summary for the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) workshop. *Am J Gastroenterol*. 2015;110:127-136.
 24. Rockwood TH, Church JM, Fleshman JW, et al. Patient and surgeon ranking of the severity of symptoms associated with fecal incontinence: the fecal incontinence severity index. *Dis Colon Rectum*. 1999;42:1525-1532.
 25. Bharucha AE, Zinsmeister AR, Locke GR, Schleck C, McKeon K, Melton LJ. Symptoms and quality of life in community women with fecal incontinence. *Clin Gastroenterol Hepatol*. 2006;4:1004-1009.
 26. Rockwood TH, Church JM, Fleshman JW, et al. Fecal Incontinence Quality of Life Scale: quality of life instrument for patients with fecal incontinence. *Dis Colon Rectum*. 2000;43:9-16.
 27. Sung VW, Rogers RG, Bann CM, et al. Symptom outcomes important to women with anal incontinence: a conceptual framework. *Obstet Gynecol*. 2014;123:1023-1030.
 28. Sanchez JE, Brenner DM, Franklin H, Yu J, Barrett AC, Paterson C. Validity of the $\geq 50\%$ response threshold in treatment with NASHA/Dx injection therapy for fecal incontinence. *Clin Transl Gastroenterol*. 2015;6:e70.
 29. Jorge JM, Wexner SD. Etiology and management of fecal incontinence. *Dis Colon Rectum*. 1993;36:77-97.
 30. Vaizey CJ, Carapeti E, Cahill JA, Kamm MA. Prospective comparison of faecal incontinence grading systems. *Gut*. 1999;44:77-80.
 31. Cotterill N, Norton C, Avery KN, Abrams P, Donovan JL. Psychometric evaluation of a new patient-completed questionnaire for evaluating anal incontinence symptoms and impact on quality of life: the ICIQ-B. *Dis Colon Rectum*. 2011;54:1235-1250.
 32. Cotterill N, Norton C, Avery KN, Abrams P, Donovan JL. A patient-centered approach to developing a comprehensive symptom and quality of life assessment of anal incontinence. *Dis Colon Rectum*. 2008;51:82-87.
 33. Sansoni J, Hawthorne G, Fleming G, Marosszeky N. The revised faecal incontinence scale: a clinical validation of a new, short measure for assessment and outcomes evaluation. *Dis Colon Rectum*. 2013;56:652-659.
 34. Bols EMJH, Hendricks HJ, Berghmans LCM, Baeten CGMI, de Bie RA. Responsiveness and interpretability of incontinence severity scores and FIQL in patients with fecal incontinence: a secondary analysis from a randomized controlled trial. *Int Urogynecol J*. 2013;24:469-478.