Pancreatology 19 (2019) 182–190

Contents lists available at ScienceDirect

Pancreatology

journal homepage: www.elsevier.com/locate/pan

Psychometric evaluation of a patient-reported outcome measure in pancreatic exocrine insufficiency (PEI)

Colin D. Johnson [a], Nicola Williamson [b, *], Gwendolyn Janssen-van Solingen [c], Rob Arbuckle [b], Chloe Johnson [b], Sarah Simpson [b], Doris Staab [d], Enrique Dominguez-Munoz [e], Phillippe Levy [f], Gary Connett [g], Markus M. Lerch [h]

[a] University Hospital Southampton NHS Foundation Trust, Southampton, United Kingdom
[b] Adelphi Values, Cheshire, England, United Kingdom
[c] Abbott Product Operations AG, EPD HQ, Allschwil, Switzerland
[d] Charité University Hospital, Berlin, Germany
[e] University Hospital of Santiago de Compostela, Santiago de Compostela, Spain
[f] Université Denis Diderot-Paris VII Hôpital Beaujon, Clichy Cedex, France
[g] Southampton Children's Hospital, Hampshire, England, United Kingdom
[h] University Medicine Greifswald, Greifswald, Germany

ARTICLE INFO

Article history:
Received 30 August 2018
Received in revised form
20 November 2018
Accepted 27 November 2018
Available online 27 November 2018

Keywords:
Pancreatic exocrine insufficiency
Psychometric validation
Symptoms
Health-related quality of life

ABSTRACT

Background/objectives: Pancreatic exocrine insufficiency (PEI) is commonly caused by chronic pancreatitis (CP) or cystic fibrosis (CF). There are no PEI-specific patient-reported assessments of symptoms and impacts. The PEI Questionnaire (PEI-Q) was developed through qualitative research with PEI patients and expert clinical input. This study evaluated the psychometric properties of the PEI-Q.
Methods: 162 PEI patients (CF = 71 and CP = 91), 62 diarrhoea-specific irritable bowel syndrome (IBS-D) patients and 60 healthy controls completed the 26-item PEI-Q and the Gastrointestinal Quality of Life Index (GIQLI) at baseline. PEI patients completed the measures again two weeks later to assess the test-retest reliability of the PEI-Q. Analyses supported item reduction and scoring algorithm development, followed by psychometric evaluation.
Results: Over 90% of PEI patients completed at least 23 of the 26 items at baseline. Item responses and clinical relevance supported retention of 18 items. Factor analysis supported a three-factor solution (abdominal symptoms, bowel movements, impacts) with adequate model fit. PEI-Q scores had good internal consistency (Cronbach's alpha: 0.77–0.82) and test-retest reliability (ICC: 0.73–0.87). Correlations between PEI-Q and GIQLI supported convergent validity. Known-groups and receiver operating characteristic analyses demonstrated that PEI-Q scores discriminated ($p < 0.001$) between differing PEI severities, and PEI patients and controls.
Conclusions: The PEI-Q has good validity and reliability. Results indicate that the PEI-Q could be used to aid identification and diagnosis of PEI, assist in the management of patients already diagnosed with PEI, ensuring correct and optimum treatment as well as enhance patient-clinician communication.

Introduction

Pancreatic exocrine insufficiency (PEI) is the reduction in the synthesis and secretion of pancreatic digestive enzymes into the duodenum, resulting in an inadequate enzymatic response to a meal. The most common causes of PEI are chronic pancreatitis (CP) or cystic fibrosis (CF). PEI also occurs in patients with pancreatic cancer (PC), acute pancreatitis (AP) and after pancreatic surgery. If untreated, PEI results in only 50–60% of dietary fats and proteins being absorbed, leading to health problems such as malnutrition, impaired growth, increased infection rates (CF) and cardiovascular events [1,2]. PEI patients can experience steatorrhoea, diarrhoea, weight loss and abdominal distension/discomfort.

Patients in Europe and the USA are treated with pancreatic

---

* Corresponding author. Adelphi Values, Adelphi Mill, Grimshaw Lane, Bollington, Cheshire, SK10 5JB, United Kingdom.
E-mail address: nicola.williamson@adelphivalues.com (N. Williamson).

https://doi.org/10.1016/j.pan.2018.11.013

enzyme replacement therapy (PERT), an enzyme formulation that helps patients digest fat [3]. Current evidence indicates that PERT provides relief of maldigestion-related symptoms (e.g. steatorrhoea, weight loss, flatulence) and normalization of the nutritional status of the patients [4]. Many patients are however undiagnosed or under-treated [5,6]. which potentially leads to intensified symptoms, reduced health-related quality of life (HRQoL) and potentially increased malnutrition-related morbidity and mortality particularly in CP patients.

Tests, such as fecal elastase, pancreatic function test, FE-1 or c-mixed triglyceride breath test, can be used to aid diagnosis [4]. However, there is no gold-standard test and, in clinical practice, often patients are diagnosed based on symptom experience. Assessment of response to treatment often depends on symptom severity; there is no routinely available test of treatment effectiveness. As PEI symptoms are subjective, they are best measured by patient self-report. Thus, a patient-reported outcome (PRO) measure (questionnaire) of PEI symptoms and impacts could be valuable in routine clinical practice as a standardized assessment to guide treatment decisions, monitor patients' symptoms and improve patient-clinician communication.

Although numerous PROs assess gastrointestinal diseases, there are no existing PEI-specific PRO symptom or HRQoL assessments. The United European Gastroenterology evidence-based guidelines for the diagnosis and therapy of CP highlight the importance of using validated HRQoL assessments in patients with CP and PEI, for both in- and out-patients and during their follow-up, to allow multiple dimensions of their HRQoL to be assessed. The guidelines also highlight the importance of assessing PEI in patients with mild, moderate and severe symptoms, and including HRQoL assessments as endpoints in clinical studies.

To address the need for a measure to assess PEI symptoms and impacts, a 45-item PEI-specific PRO (PEI-Q) was developed through a literature review and qualitative concept elicitation interviews with clinicians ($n = 10$) and patients ($n = 61$ PEI patients with CF/CP) in France, Germany and the UK. Content validity was then assessed through cognitive debriefing interviews with patients in the same countries, and 18 items were deleted due to conceptual redundancy or lack of relevance, resulting in a 26-item measure.

This study evaluated the psychometric properties of the 26-item PEI-Q using data from an observational study. Analyses aimed to support possible item reduction and finalise scoring, and to evaluate the psychometric validity of the resulting scores.

Methods

Study design

This was a multi-centre observational study in France, Germany, Spain and the UK which aimed to collect quantitative data from approximately 325 participants (205 PEI adult and adolescent patients, 60 patients with diarrhoea-specific irritable bowel syndrome (IBS-D) and 60 healthy controls). Analysis was conducted in two phases: stage 1 used dimensionality analyses to assess item performance and support item reduction and scoring development; stage 2 assessed the measurement properties of the resulting PEI-Q scores.

The study was conducted in accordance with the Declaration of Helsinki. Ethical oversight was provided by the National Research Ethics Service in the UK and the Freiburger Ethics commission review board in Germany. In Spain, site-specific approval was obtained at the University Hospital of Santiago de Compostela. In France, the study was declared to the National Board of Physicians and an ethical waiver granted.

Eligibility criteria

Patients were eligible to participate if they had an expert clinician-confirmed diagnosis of PEI and either CF or CP. CP patients had to have been on PEI treatment for six years or less. Both CP and CF patients had to have experienced PEI symptoms within the past three months. All patients were adults (18 years or more) apart from a subset of CF adolescents with PEI (aged 12–17 years) who were also included. Adult IBS-D patients and healthy controls were included to provide comparison groups for evaluating the discriminant properties of the measure. Patients with IBS-D were included if they met the ROME III criteria (at the time of data collection the Rome IV had not yet been published) for IBS, and experienced PEI-like bowel symptoms at least three days per month over the past three months. PEI-like bowel symptoms were defined as diarrhoea, increased frequency of bowel movements and/or bowel urgency based on the recruiting clinician's clinical opinion. IBS-D patients were also required to have experienced abdominal symptoms (i.e. abdominal pain, bloating, distension, abdominal noises or flatulence) within the past 3 months. PEI patients were excluded if they had history of other gastrointestinal conditions or had undergone gastrointestinal surgery in the past five years. Healthy controls were excluded if they had significant signs of abdominal/bloating symptoms.

Recruitment

PEI patients (including both CF and CP patients) were recruited via their clinical site; potential participants were invited to participate in the study by their clinician during a normal routine appointment. All potential participants were provided with an information letter and Informed Consent Form (ICF) to read providing information about the study, and to sign if they wished to participate. CF adolescents with PEI were asked to read and sign an informed assent form and their parent/guardian was asked to provide informed consent for them to participate in the study. Written informed consent was obtained prior to any study-related activities. For each eligible participant, their physician completed a Case Report Form (CRF) confirming eligibility and diagnosis. Participants were also asked to complete a demographic form providing background and demographic information to characterize the study sample. Participants were remunerated for participation. Purposive sampling was used to ensure good representation across demographic characteristics including age, gender, ethnicity and disease condition.

Assessments

The version of the PEI-Q evaluated consisted of 26 items across 10 domains assessing the symptoms and impacts of patients' PEI in the past seven days (Fig. 1). The measure had two sections; section 1 consisted of 17 symptoms item (i.e. pain, bloating symptoms, bowel movement symptoms, nausea/vomiting and eating) and section 2 consisted of 9 HRQoL impact items (i.e. impact on daily activities, emotional wellbeing, diet, social functioning and sleep). Most items use a five-point Likert scales with verbal descriptors. Other response options include from ‘less than one per day’ to ‘more than four per day’ (1 item), and a nine-segment abdominal diagram (1 item). Higher scores indicate worse symptoms and greater impact on HRQoL.

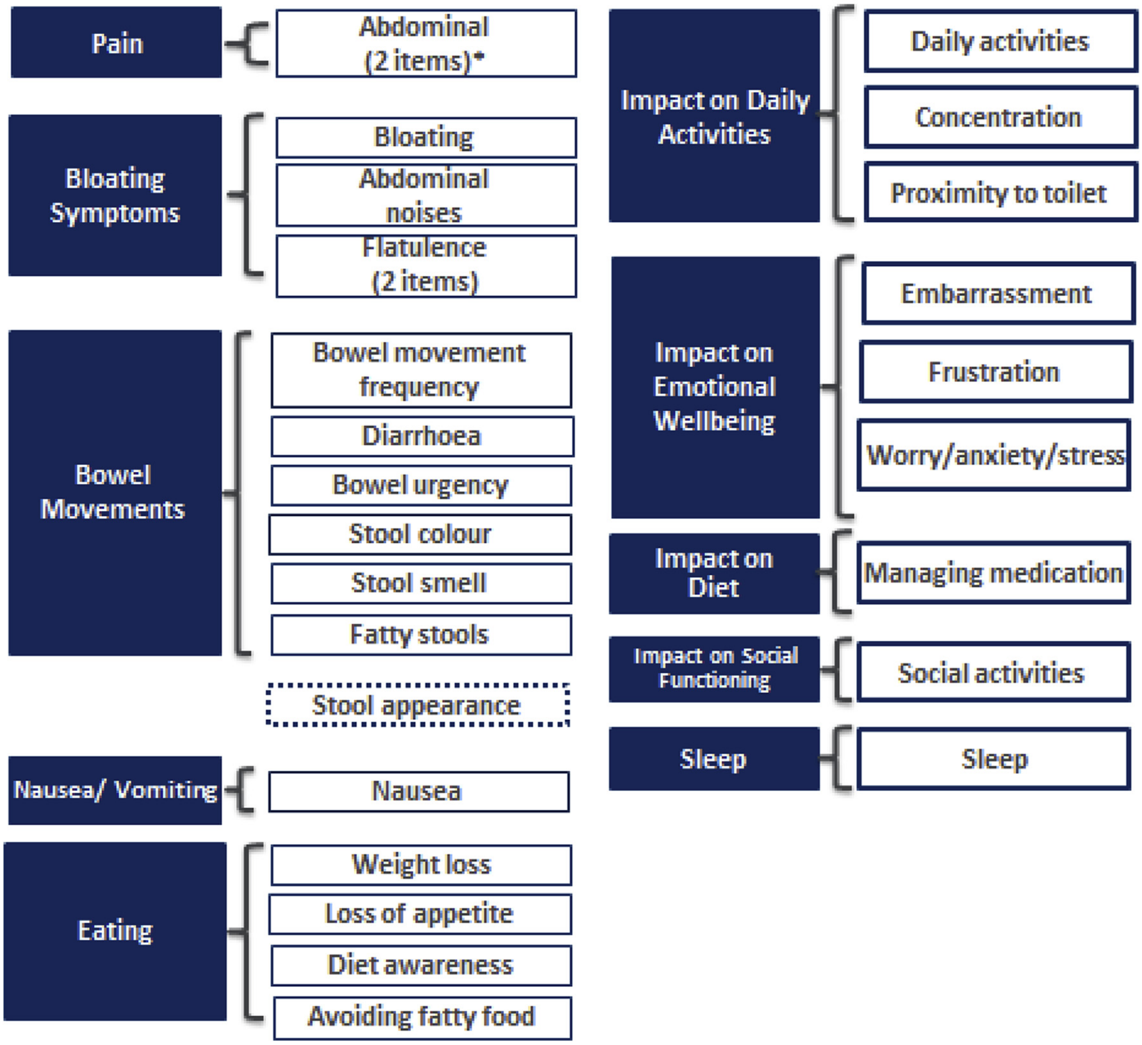


Fig. 1. Conceptual framework.

In addition, participants were administered the Bristol Stool Form Scale alongside the PEI-Q. The Bristol Stool Form Scale is a one item assessment designed to characterize type and consistency of stool [7]. Participants were asked to circle the type of stool they had most of the time over the past seven days.

Participants also completed a global impression of severity (PGI-S) ('please could you rate the severity of your abdominal and bowel symptoms today') with a 4-point response scale and a global impression of change (PGI-C; 'compared to approximately 14 days ago, would you describe your abdominal and bowel symptoms today as') with a seven-point response scale. PGI-S and PGI-C measures are widely accepted measures [8] of severity and change and are recognized by regulatory agencies including the Food and Drug Administration to define groups for responder definitions and anchors [9].

The GIQLI is a validated HRQoL measure for patients with a gastrointestinal disease in a clinical setting. The GIQLI consists of 36 items assessing four domains of HRQoL including physical function (7 items), emotional (5 items), social (4 items), gastrointestinal symptoms (19 items) and one item on subject treatment assessment. All items use a two-week recall period and a five-point response scale ranging from 0 (lease desirable option) to 4 (most desirable option), with higher scores indicating better HRQoL.

Clinicians completed a Clinical Global Impression of Severity (CGI-S) item rating their patient's PEI symptom severity on the same day from 1 'no symptoms' to 4 'severe'.

Schedule of assessments

All participants completed all measures in the native language of their country (i.e. English, German, Spanish or French). All participants completed section 1 of the PEI-Q, the Bristol Stool Form Scale and GIQLI at baseline. Additionally, PEI patients also completed section 2 of the PEI-Q and the PGI-S. Patients completed the measures (PEI-Q, Bristol Stool Form Scale, PGI-S and PGI-C) a second time at home approximately two weeks (i.e. 12–17 days) following their first visit, with the paper copies of the measures returned to the clinical site via post.

Analysis

Table 1 details the item-level and scale-level psychometric analyses conducted. Following item-level analyses, item reduction and scoring algorithm decisions were made and scale-level analyses were performed.

Table 1
Overview of analyses.

Analysis	Description
Item-level and dimensionality analyses	
Quality of completion	• Quality of completion was assessed by examining the proportion of missing data at the item-level for each subsample.
Item response distribution and floor and ceiling effects	• Item response distributions were examined to assess the frequency and proportion of participants who selected each response, to identify any items or response options that were overly favoured. • Floor and ceiling effects refer to evidence of high proportions of PEI patients endorsing the lowest (floor: > 60%) and highest (ceiling: > 60%) possible response options, respectively for each item on the PEI-Q. The threshold was chosen as large proportions of PEI patients were expected to select the lowest response option, given that their PEI was treated.
Inter-item correlations	• Inter-item correlations were examined for the PEI-Q to ensure each item measured a distinct concept without any redundancy. Items which correlated with one another at >0.80 were flagged for review.
Item discrimination indices [17]	• Item discrimination indices were calculated to assess whether items could differentiate among groups of participants based on pre-specified groupings. Analyses were conducted separately for each PEI subgroup (all PEI patients, CF patients and CP patients). • Participants were divided into two symptom severity groups based on their PGI-S response at baseline, with PGI-S responses 'no symptoms' and 'mild' grouped into the 'least severe' group and responses 'moderate' and 'severe' in the 'most severe' group. • An item discrimination index was calculated for each item. The following thresholds can be used to assess discrimination: <0.19 indicates poor discriminant ability, 0.20–0.39 indicates reasonable discriminant ability and >0.40 good discriminant ability [17].
Exploratory factor analysis (EFA)	•Exploratory factor analysis was conducted to explore potential factor structures for the PEI-Q. • Model fit was assessed with Bentler's Comparative Fit Index (CFI; > 0.90) [18] root mean square error approximation (RMSEA; < 0.08) [19], and standardized root-mean square residuals (between 0 and 0.80) [20,21]. • Factor loadings of >0.40 and similar in magnitude were among items belonging to the same domain. Factor loadings of <0.40 for any item were examined further. • Modification indices were also examined to assess the extent to which any items and/or domains had co-dependence. Any items/domains with a modification index of >15 were examined further [22].
Scale-level analyses	
Internal consistency reliability	• Internal consistency, concerned with the homogeneity of items belonging to the same domain, was evaluated using Cronbach's alpha coefficient for all multi-item domains (>0.70 for good internal consistency) [23]. • The alpha-if-item-deleted method was also conducted to assess whether the internal consistency of each domain would improve with the removal of each item in turn.
Test-retest reliability (TRT)	• Test retest reliability (TRT) was evaluated by examining the stability of scores between baseline and two-week follow-up for stable PEI patients [24]. Two approaches were used to define patients as stable. First, PEI patients were considered 'stable' if they completed the PGI-C item and selected the 'no change' response option. Second, patients were considered stable if they showed no change between baseline and two-week follow up in their PGI-S score (i.e. selected the same response at baseline and two-week follow-up). • Intra-class correlation coefficients (ICCs) were calculated and an ICC of 0.70 or greater for the stable sample was considered evidence of good TRT.
Concurrent validity	• Concurrent or convergent validity of the PEI-Q was evaluated by examining the correlations between the scores of the PEI-Q and the domain scores of a pre-existing, validated measure, the GIQLI [25]. Correlations were assessed between the domain and total scores of the PEI-Q and the GIQLI. Pearson correlation coefficients were calculated between baseline scores of the PEI-Q and GIQLI. • Domains assessing similar or related concepts were expected to correlate at 0.40 or higher, and domains assessing unrelated concepts were expected to show small (<0.40) or negligible correlations. • Concurrent validity was assessed for all PEI patients and separately for CF and CP patients.
Known-groups validity	• Construct validity was also assessed using the known-groups method [26], to evaluate differences in PEI-Q scores among participants who differ on health/disease related variables (PEI symptom severity, presence of symptoms, length of time since diagnosis, and participant subgroup). t-tests (for comparison of pairs of groups) and ANOVAs (comparison of more than two groups) were used to evaluate if differences were statistically significantly ($p < 0.05$). PEI-Q scores were obtained between PEI participant subgroups [26].
Domain-level Receiver Operating Characteristic (ROC) analysis	• Univariate logistic regression was conducted with diagnosis as the outcome variable and each item in section 1 of the PEI-Q as a covariate, to assess if any items can be used to screen for PEI. • Statistical significance of the odds ratios was assessed using the upper and lower 95% confidence intervals (CIs), where a significant odds ratio is indicated if a value of 1.00 is not included in the CI range. Statistically significant odds ratios for an item indicated that the item can discriminate between participant groups. Odds ratios and 95% confidence intervals were calculated comparing each participant group (PEI patients with IBS-D patients, and PEI patients with healthy controls). Comparisons between CF/CP patients and IBS-D patients and healthy controls, respectively, were also performed. • Receiver operating characteristic (ROC) analysis was conducted to assess the sensitivity and specificity of the PEI-Q domain and total scores at baseline [27], to confirm whether the PEI-Q domains were able to discriminate between PEI and non-PEI patients. Specifically, comparisons were made between PEI (all, CF and CP) and IBS-D patients and between PEI patients (all, CF and CP) and healthy controls. • An area under the curve (AUC) statistic was calculated for each item. An AUC value close to 1.0 maximizes sensitivity and specificity, indicating greater ability of the measure to differentiate between participant subgroups. An AUC value of >0.80 was used to represent good discriminant ability, 0.65–0.80 moderate discriminant ability and a value between 0.50 and 0.65 poor discriminant ability [28]. • For each domain, sensitivity and specificity statistics were also calculated to establish the optimal cut-point on the domain and total scores, to discriminate between PEI and other participants. A good cut-point was the response option that achieved a sensitivity and specificity of 80% or more, fair between 65 and 80% and poor <65% [28].

Results

The sample enrolled consisted of 284 participants (162 PEI patients, 62 patients with IBS-D and 60 healthy controls). Just under half of the PEI patients were reported to have been diagnosed by a faecal elastase test ($n = 72$, 44.4%) and breath test ($n = 7$, 4.3%), whereas the use of any diagnostic test was not reported for 83 PEI patients (51.2%). Age was higher among CP (57.1 years) than CF patients (25.7 years) in part due to the inclusion of adolescents in the CF sample (Supplementary Table 1). There were higher proportions of females in the CF (female: 64.8%, male: 35.2%) and IBS-D (female: 71.0%, male: 29.0%) samples but more males than females in the CP sample (male: 69.2%, female: 30.8%). Most participants in the UK, Germany and Spain were Caucasian/White (range: 64.8–88.3%); ethnicity data were not collected in France as it is considered inappropriate.

Item-level and dimensionality results

Completion of the PEI-Q was high among PEI and IBS-D patients; over 90% of PEI patients completed at least 23 of the 26 items as baseline, while for IBS-D patients all patients (100%) completed at least 15 items out of the 18 they were asked to complete. This was weaker among healthy controls (3.3% missing no items), however the average number of missing items was still low: 1.6 (SD = 1.1). Three items had relatively poor quality of completion by health controls: items 6 (gas odour: 28.3%, n = 17), 11 (stool colour: 23.3%, n = 14) and 15 (lack of appetite: 26.7%, n = 16). This suggests that possibly some healthy individuals are offended by the ‘odour’ and ‘stool colour’ items.

Item response distributions were heavily skewed towards the lower end of the scale for PEI patients, indicating that many participants experienced minimal PEI symptoms and impacts. Seven items had substantial floor effects (range: 60.0%–67.9%). Item response distributions were more evenly spread among IBS-D patients, indicating that IBS-D patients reported more severe symptoms than those with PEI. For the healthy controls, as would be expected, substantial floor effects (range: 60.4–95.0%) were present for all but four items.

For the symptom items (section 1 of the PEI-Q), most items correlated relatively weakly, with very few strong correlations >0.70, and the majority were in the range of r = 0.20–0.40. Inter-item correlations were weaker among symptom items compared to among the impact items (Table 2). Items 19 (daily activities) and 26 (social activities) correlated most strongly (r = 0.81) indicating possible item redundancy.

Item discrimination indices grouped participants as ‘severe’ or ‘mild’ based on their PGI-S score; the ability of the items to discriminate between those two groups was examined. Eight items discriminated poorly between PEI severity levels: items assessing *number of bowel movements* (item discrimination index = 0.08), *diarrhoea* (0.19), *stool colour* (0.11), *stool odour* (0.17), *fat or oil on toilet paper* (0.09), *weight loss* (0.07), *take the right amount of enzymes* (0.11) and *sleep* (0.17).

Item reduction and scoring decisions

A second-order confirmatory factor analysis (CFA) was performed to evaluate whether the PEI-Q items fit the 10-domain hypothesized conceptual framework. However, due to the large number of hypothesized domains, the model did not converge. Exploratory factor analysis (EFA) was therefore conducted to identify a suitable factor structure.

EFA was conducted using the 26-item PEI-Q. Eight PEI-Q items were removed following the first factor analysis, due to either lack of clinical relevance or poor item performance, resulting in an 18-item measure. The item assessing *number of bowel movements* was not discriminative and was conceptually related to the other bowel movement symptoms and so was deleted. The seven-day recall period for the item assessing *weight loss* was not considered appropriate to assess weight loss and the item performed poorly and so was deleted. The item asking whether the patient was *careful about the food they ate* correlated strongly (r = 0.69) with the item asking about *avoiding fatty food* indicating redundancy. The item related to *avoiding fatty food* was retained as it is more specific to PEI. The items assessing *daily activities* and *taking the right amount of enzymes* were deleted due to substantial floor effects and lack of clinical importance. The *daily activities* item also correlated highly with the *social impact* item, which was considered more relevant. Finally, the item which captured *stomach pain location* lacked clinical importance and the items assessing *frustration* and *sleep* were not considered sufficiently specific and so were deleted.

For scoring, each item is assigned a score from 0 to 4 on the response selected by the patient. Domain scores were calculated by taking a mean of all items within the domain. A total summary score was calculated by taking a mean of each domain in the PEI-Q, while a total symptom score consisted of averaged domain scores of the bowel movement and abdominal symptoms domains. The total symptom score (mean) can be calculated for all patients to provide information to help doctors to determine a diagnosis of PEI, while the total summary score (mean) should only be calculated for patients who have been diagnosed with PEI. Domain scores can only be calculated if more than 50% of items are completed within each domain (i.e. four or more items in the abdominal symptoms domain, three or more items in the bowel movement symptoms domain, and three or more items in the impact domain). Total

Table 2
Inter-item correlations of the PEI-Q among PEI patients (n = 162) at baseline.

Factor loading (SE) / Items	Factor 1 – Abdominal symptoms	Factor 2 – Bowel movement symptoms	Factor 3 - Impacts
1. Stomach pain	0.499*	0.102	0.411*
3. Feel bloated	0.460*	0.172	**0.309**
4. Stomach noises	0.517*	0.128	0.199
5. Pass gas	0.739*	0.094	-0.034
6. Gas odor	0.934*	0.078	-0.046
12. Fat/oil on toilet paper	**0.691***	0.217	0.008
13. Feel sick	0.429*	0.272	0.127
15. Lack of appetite	**0.471***	-0.077	0.413*
8. Diarrhea	-0.026	0.678*	-0.046
9. Rush to the toilet	0.145	0.815*	0.119
10. Stool color	0.220	0.675*	0.150
11. Stool odor	0.336	**0.355**	0.142
21. Close to toilet	-0.038	0.778*	-0.009
18. Avoid fatty food	0.293	-0.147	**0.343**
20. Concentration	0.456*	-0.326	**0.473***
22. Embarrassment	-0.098	0.161	0.840*
24. Worried/anxious/stressed	-0.147	0.136	0.870*
26. Social activities	0.418*	-0.074	**0.574***

scores can only be calculated if two or more domains are present for each patient. If more items/domains are missing, the domain/total score cannot be calculated and will be considered missing.

Scale-level results

A final EFA model was conducted on the revised 18-item PEI-Q (Table 3). Considering both model fit and the conceptual grouping of items, a three-factor model was selected. Most items loaded most strongly onto a factor with conceptually similar items, with the three factors (domains) assessing: abdominal symptoms (7 items; factor loadings: 0.43–0.93), bowel movement symptoms (6 items; 0.36–0.82) and impacts (5 items: 0.34–0.87). Although item 12 (fat/oil on toilet paper) loaded highest on the abdominal symptoms domain (0.69) it was considered more conceptually related to the bowel movement symptoms domain (0.22) and was included in that domain.

Internal consistency was examined using Cronbach's alpha to assess the homogeneity of items belonging to the same domain. All alpha coefficients surpassed 0.70 indicating good internal consistency (abdominal symptoms: 0.77, bowel movement symptoms: 0.82, impacts: 0.78, total symptom: 0.86, total summary score: 0.90). The alpha-if-item-deleted method was also conducted to assess whether the internal consistency of each domain would improve with the removal of each item in turn. Cronbach's alpha coefficients calculated with each item deleted in turn were lower than the overall domain/total score alphas, indicating that all items contribute to the overall domain/total score. The overall internal consistency improved slightly with the removal of items 18 (avoid fatty foods) and 15 (lack of appetite); however, given the marginal difference in the Cronbach's alpha coefficient, this result was not considered problematic.

Test-retest reliability was evaluated to examine the stability of scores at baseline and two-week follow-up, for stable PEI patients. When stability was defined using either the PGI-C or PGI-S, all Intraclass Correlation Coefficients (ICC) surpassed 0.70 supporting test-retest reliability (Table 4). In addition, changes in mean scores for all domains and both summary total scores between baseline and two-week follow-up were also evaluated using a *t*-test and found to be non-significant ($p > 0.05$), providing further evidence of the reproducibility of PEI-Q scores in stable patients.

Concurrent validity was evaluated by examining the pattern of correlations between the PEI-Q and GIQLI scores, to assess whether domains in each measure assessing conceptually similar constructs correlate. A logical pattern of correlations was observed with stronger correlations for scores expected to be related than for unrelated scores (Table 5). For example, the PEI-Q symptom domains and total score correlated moderately with the GIQLI symptom domain (range: 0.64-0.77), providing evidence of convergent validity. Similarly, the PEI-Q impact score correlated moderately with the GIQLI impact domain (range: 0.61-0.67). Evidence of discriminant validity was provided by weak correlations (range: 0.21–0.27) between the GIQLI medical treatment domain and each PEI-Q domain/total score.

Known-groups analyses were conducted to assess the construct validity of the PEI-Q by examining the relationship between PEI-Q scores and pre-specified groups, to evaluate differences in PEI-Q scores among participants who differ on health or disease-related variables. Known-groups analyses compared PEI-Q domain and total scores according to groups defined by PEI symptom severity, time since diagnosis and comparing PEI patients with healthy controls (Table 6). For each PEI-Q domain and total score, mean scores differed significantly ($p < 0.05$) among groups defined by PEI symptom severity/presence and between PEI patients and healthy controls, with mean estimates monotonically increasing in a logical pattern as expected. Comparison of groups by time since diagnosis provided mixed, non-significant results with no logical pattern.

In addition, exploratory known-groups analyses were

Table 3
Exploratory factor analysis of the revised PEI-Q at baseline.2

Domain	Items	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Pain	1. Pain	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2. Pain area	0.54	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Bloating symptoms	3. Bloated	0.42	0.38	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4. Noises	0.35	0.24	0.43	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5. Gas	0.30	0.21	0.23	0.47	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	6. Smelly gas	0.37	0.22	0.27	0.35	0.67	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Bowel movement symptoms	7. Bowel movements	0.09	-0.01	0.05	-0.16	0.02	0.20	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	8. Diarrhoea	0.32	0.08	0.20	0.19	0.13	0.28	0.43	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	9. Urgency	0.32	0.14	0.29	0.18	0.25	0.36	0.40	0.66	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	10. Colour	0.17	-0.01	0.28	0.13	0.17	0.30	0.30	0.35	0.41	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	11. Smelly poo	0.36	0.32	0.40	0.29	0.41	0.65	0.25	0.41	0.46	0.43	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	12. Fat/oil	0.30	0.15	0.25	0.14	0.25	0.40	0.32	0.40	0.26	0.45	0.45	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Nausea/Vomiting	13. Nausea	0.46	0.18	0.30	0.30	0.23	0.31	-0.07	0.30	0.29	0.15	0.32	0.21	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Eating	14. Weight loss	0.20	0.05	0.15	0.10	0.07	0.08	-0.03	0.20	0.21	0.05	0.12	0.22	0.33	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-
	15. Appetite loss	0.38	0.31	0.26	0.24	0.07	0.17	-0.07	0.17	0.17	0.11	0.21	0.05	0.39	0.29	1.00	-	-	-	-	-	-	-	-	-	-	-	-
Bowel movement Symptoms	16. Bristol stool	0.19	-0.04	0.12	0.08	0.13	0.16	0.40	0.57	0.45	0.30	0.30	0.31	0.09	-0.07	-0.11	1.00	-	-	-	-	-	-	-	-	-	-	-
Eating	17. Diet awareness	0.35	0.28	0.32	0.23	0.21	0.16	0.05	0.20	0.29	0.27	0.23	0.29	0.19	0.19	0.23	0.13	1.00	-	-	-	-	-	-	-	-	-	-
	18. Avoid fatty food	0.43	0.25	0.26	0.14	0.17	0.18	0.01	0.16	0.20	0.20	0.20	0.26	0.31	0.17	0.19	0.08	0.69	1.00	-	-	-	-	-	-	-	-	-
Daily activities	19. Daily activity	0.52	0.45	0.40	0.34	0.33	0.35	0.12	0.35	0.35	0.40	0.39	0.32	0.37	0.15	0.44	0.18	0.48	0.45	1.00	-	-	-	-	-	-	-	-
	20. Concentration	0.50	0.40	0.40	0.36	0.33	0.37	0.07	0.33	0.39	0.45	0.40	0.25	0.41	0.15	0.48	0.18	0.43	0.41	0.77	1.00	-	-	-	-	-	-	-
	21. Proximity	0.40	0.25	0.40	0.33	0.35	0.39	0.32	0.54	0.61	0.44	0.38	0.37	0.38	0.19	0.31	0.34	0.41	0.32	0.60	0.58	1.00	-	-	-	-	-	-
Emotional wellbeing	22. Embarrassment	0.34	0.24	0.42	0.27	0.33	0.37	0.35	0.54	0.45	0.41	0.45	0.55	0.31	0.25	0.26	0.29	0.31	0.24	0.46	0.31	0.62	1.00	-	-	-	-	-
	23. Frustration	0.50	0.42	0.37	0.31	0.42	0.40	0.21	0.39	0.40	0.42	0.44	0.40	0.33	0.12	0.46	0.23	0.41	0.24	0.71	0.64	0.54	0.43	1.00	-	-	-	-
	24. Worry/anxiety	0.54	0.43	0.37	0.32	0.32	0.35	0.17	0.35	0.33	0.41	0.42	0.40	0.40	0.09	0.48	0.20	0.44	0.34	0.74	0.63	0.49	0.42	0.77	1.00	-	-	-
Diet	25. Medication	0.24	0.17	0.24	0.18	0.32	0.26	0.11	0.23	0.20	0.34	0.26	0.38	0.20	0.07	0.12	0.11	0.26	0.17	0.32	0.27	0.37	0.31	0.44	0.29	1.00	-	-
Social functioning	26. Social	0.49	0.34	0.38	0.30	0.33	0.35	0.10	0.28	0.27	0.43	0.39	0.34	0.35	0.12	0.40	0.11	0.42	0.34	0.81	0.66	0.49	0.44	0.70	0.75	0.44	1.00	-
Sleep	27. Sleep	0.47	0.38	0.38	0.38	0.27	0.27	0.10	0.26	0.25	0.33	0.40	0.29	0.36	0.19	0.47	0.07	0.48	0.41	0.72	0.67	0.37	0.31	0.53	0.63	0.17	0.62	1.00

Key: <0.20 ░ 0.20-0.40 ▒ 0.40-0.70 ▓ >0.70 █

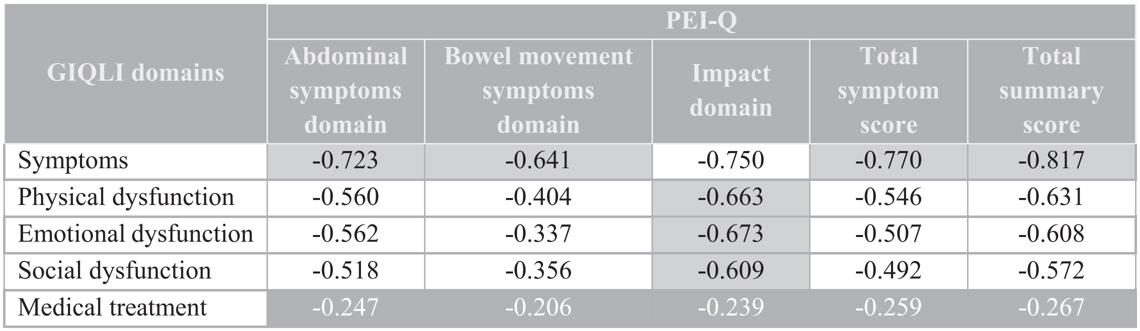
*Factor loading above 0.40 threshold. The cells highlighted in grey the factor each item loads most strongly onto. Bold factor loadings indicate the stronger factor loading among items that crossloaded.

Table 4
Test-retest reliability for the PEI-Q between baseline and follow-up with the stable sample defined using the PGI-C ($n = 64$) and PGI-S ($n = 90$).

Domain	N	Reliability (ICC)	95% Confidence Interval		Mean change (SD)	P-value*
			Lower	Upper		
PGI-S						
Abdominal symptoms	64	0.837	0.745	0.897	−0.0 (0.9)	0.962
Bowel movements	64	0.838	0.748	0.899	0.1 (1.0)	0.765
Impact score	62	0.731	0.590	0.828	0.2 (0.9)	0.175
Total symptom score	64	0.864	0.785	0.915	0.0 (0.9)	0.883
Total summary score	62	0.872	0.796	0.921	0.1 (0.8)	0.618
PGI-C						
Abdominal symptoms	80	0.722	0.598	0.812	−0.1 (0.9)	0.533
Bowel movements	80	0.783	0.682	0.855	−0.0 (0.9)	0.907
Impact score	80	0.864	0.796	0.911	0.1 (0.9)	0.516
Total symptom score	80	0.787	0.686	0.858	−0.1 (0.8)	0.695
Total summary score	80	0.846	0.770	0.898	−0.0 (0.8)	0.988

*P-values are derived from a paired *t*-test testing the within group difference in mean scores to the two time-points.

Table 5
Concurrent validity correlations of the PEI-Q and GIQLI at baseline ($n = 162$).3



GIQLI domains	PEI-Q				
	Abdominal symptoms domain	Bowel movement symptoms domain	Impact domain	Total symptom score	Total summary score
Symptoms	-0.723	-0.641	-0.750	-0.770	-0.817
Physical dysfunction	-0.560	-0.404	-0.663	-0.546	-0.631
Emotional dysfunction	-0.562	-0.337	-0.673	-0.507	-0.608
Social dysfunction	-0.518	-0.356	-0.609	-0.492	-0.572
Medical treatment	-0.247	-0.206	-0.239	-0.259	-0.267

*Light shaded cells indicate hypothesized correlations >0.40 indicating concurrent validity; Dark shaded cells indicate correlations expected to be weak (<0.40) indicating discriminant validity.

Table 6
Known-groups analysis comparisons of total scores at baseline.

Group	Category	Total symptom score		Total summary score	
		N	Mean (SD)	N	Mean (SD)
Severity of PEI symptoms (PGI-S)	No symptoms	36	0.7 (0.6)*	36	0.6 (0.4)*
	Mild	58	1.2 (0.6)*	57	0.9 (0.6)*
	Moderate	50	1.6 (0.7)*	50	1.5 (0.7)*
	Severe	16	2.0 (0.7)*	16	1.9 (0.7)*
Severity of PEI symptoms (CGI-S)	No symptoms	20	0.6 (0.5)*	20	0.5 (0.3)*
	Mild	66	1.1 (0.7)*	66	0.9 (0.6)*
	Moderate	54	1.5 (0.7)*	54	1.4 (0.7)*
	Severe	17	1.9 (0.6)*	16	1.8 (0.7)*
PEI symptoms	PEI symptoms	137	1.4 (0.7)*	136	1.2 (0.7)*
	No symptoms	20	0.6 (0.5)*	20	0.5 (0.3)*
Length of time since diagnosis	Less than one year	11	1.2 (0.6)	11	0.9 (0.5)
	One to two years	16	1.4 (0.6)	15	1.2 (0.5)
	Two to five years	27	1.5 (0.9)	27	1.4 (0.9)
	More than five years	91	1.2 (0.7)	91	1.0 (0.7)
Participant group	PEI patients	160	1.3 (0.8)*	159	1.1 (0.7)*
	Healthy controls	60	0.4 (0.5*)	24	0.5 (0.5)*

*Statistically significant differences among mean scores ($p < 0.05$).

conducted to provide further insight into whether there were differences in PEI-Q scores between CF and CP patients, and CF adults and adolescents. Mean scores differed significantly ($p < 0.05$) among CP and CF patients, with higher scores among CP patients on the ‘abdominal symptoms’ and ‘impact’ domains, and on the total symptom and summary scores. In contrast, the differences in the mean scores between CF adults and CF adolescents were largely non-significant, suggesting that the PEI experience is similar for CF adults and CF adolescents. There were, however, significant differences in mean scores ($p < 0.05$) on the ‘bowel movements’ domain, suggesting that CF adults experience slightly more severe bowel movement symptoms than CF adolescents.

Logistic regression was conducted to establish how well the PEI-Q measure could discriminate between PEI patients and healthy controls to further assess the construct validity of the measure. Logistic regression evaluated the relative ability of PEI-Q scores to discriminate between PEI patients and healthy controls. The symptom domains and total symptom score significantly discriminated ($p < 0.001$) between PEI patients and healthy controls. Each PEI-Q domain and total score discriminated between PEI patients and IBS-D patients. However, scores were higher among IBS-D patients; for the comparison between PEI patients and healthy controls, higher scores were observed among PEI patients.

The ROC analysis assessed the sensitivity and specificity of each PEI-Q domain/total score to confirm whether PEI-Q scores could discriminate PEI patients and healthy controls. For discrimination between PEI patients and healthy controls, the abdominal symptoms and bowel movements domains and total symptom score surpassed 0.80 (AUC range: 0.81–0.85), indicating good discriminant ability (Supplementary Table 2).

With regard to interpretation of scores, evidence suggests that patients with a total symptom score (mean) of greater than or equal to 0.60 is consistent with a diagnosis of PEI, if the patient does not have a diagnosis of another gastrointestinal condition such as irritable bowel syndrome with diarrhoea. Evidence also suggests that a total symptom score (mean) of greater than or equal to 1.8 is indicative of severe or poorly controlled PEI, scores of 1.4–1.8 are indicative of moderate PEI symptoms and scores of 0.60–1.4 are indicative of minimal or mild PEI symptoms.

Discussion

The PEI-Q is the first-known PEI-specific PRO developed specifically for use in PEI patients. The analyses described in this study involved evaluating the psychometric properties of the PEI-Q using data collected from an observational, non-interventional study with treated PEI patients (CF and CP patients), IBS-D patients and healthy controls in the UK, France, Germany and Spain. These results support reliability, validity and ability to discriminate based on severity or diagnosis of PEI, in a sample of treated patients from four European countries.

There are concerns that PEI is under-diagnosed and potentially under-treated [10], leading to a more severe symptom experience, greater impacts on HRQoL [11,12] and potentially less treatment satisfaction. The PEI-Q could be of value in clinical practice to help monitor symptoms and impacts in a standardized way and improve patient-physician communication. It is suggested that patients could complete the PEI-Q prior to a clinical appointment and specific responses could potentially be discussed with their clinician. Furthermore, there is also the possibility of the PEI-Q being used as an endpoint in future PEI clinical trials.

Quality of completion was good among all participant groups, indicating that the PEI-Q was not burdensome to complete. Item deletion decisions were based on prior qualitative research, item-level performance, preliminary testing of the 26-item PEI-Q using EFA and clinical expertise. For example, where substantial floor effects (>60%) were observed, indicating that a large proportion of PEI patients did not experience the symptoms/impacts assessed within the past seven days, the items were considered less relevant and were removed. The final PEI-Q with 18-items is clearly focused on three domains: abdominal symptoms, bowel movements and impacts of these symptoms.

All PEI-Q domain and total scores produced evidence of good internal consistency and test-retest reliability. The pattern of correlations with the GIQLI and evidence of significant differences among groups expected to differ, supported the construct validity of the finalized PEI scores. Notably, the known-groups analysis produced similar results using the total symptom and summary scores. A total symptom score could be used to monitor symptoms, while the total summary score could be used to aid patient-clinician communication. The PEI-Q could discriminate between participants according to PEI symptom severity/presence, length of diagnosis and PEI diagnosis among treated patients versus healthy controls. Logistic regression and ROC analyses both provided further evidence that the PEI-Q scores could discriminate between PEI patients and healthy controls.

These findings suggest potential use of the PEI-Q to monitor treatment response and identify inadequately treated PEI. Because treatment of PEI is aimed at symptom control, it will aid clinical decision-making, and comparison of different treatment regimens, to have a standardized measure of symptoms, such as the PEI-Q. Until now that has been lacking. It is also recommended that the Bristol Stool Form Scale is administered alongside the PEI-Q in clinical practice to aid monitoring of type and consistency of bowel movement symptoms.

The study is not without limitations. Patients were considered to have PEI based on diagnosis by clinicians with expertise in PEI. However, data about whether a diagnostic test was used was not collected for half of the sample. Use of a specific diagnostic test was not required, to maximize the feasibility of conducting the study across multiple sites and countries and of recruiting an adequate sample, given that different methods of diagnosis are used in clinical practice. However, patients were identified by clinicians at specialist clinical sites with particular expertise in identifying PEI; therefore, while possible it is unlikely that many of the PEI patients had been incorrectly diagnosed. Additionally, IBS-D patients in the sample were diagnosed based on the Rome III criteria [13]. Since the completion of this study, the Rome IV criteria have been introduced, but were not available at the time of data collection [13]. The key distinction is that the frequency of abdominal pain required has increased from ‘three days per month’ to ‘at least one day per week’. This may mean that patients with a Rome III diagnosis may not meet the Rome IV criteria for IBS-D [14]. There is the possibility that some patients included in the IBS-D sample also had PEI, due to the overlap in symptom presentation. However, as the IBS-D patients were recruited by specialist clinicians at clinical sites with extensive experience identifying both PEI and IBS-D, it is considered unlikely that these clinicians would have misdiagnosed PEI patients as having IBS-D.

The two-week follow-up produced good evidence of test-retest reliability, however data were not collected before and after an intervention, to confirm the ability of the PEI-Q to detect changes over time. Analysis of longitudinal data from an intervention sample will be needed to establish clinically meaningful change thresholds in PEI-Q scores and generate evidence to support the ability of PEI-Q scores to detect changes over time.

The study included PEI patients with CF or CP, as these are the most common causes of PEI. Additionally, patients in the sample were receiving treatment for their condition. However, it is recommended that the PEI-Q is further evaluated in patients with other causes of PEI such as pancreatectomy, pancreatic cancer or acute pancreatitis and untreated and newly diagnosed PEI patients to confirm reliability and validity in those populations [15,16]. Future studies should also explore the cross-cultural validity of the PEI-Q in additional countries. Although the psychometric evaluation performed here was limited to the UK, Germany, France and Spain, recruited patients had diverse demographic characteristics, so findings are likely representative of the PEI experience among CF/CP patients across a range of countries.

Conclusion

Findings provide evidence of strong psychometric properties for the PEI-Q in a sample of PEI patients in the UK, France, Germany and Spain. Results indicate that the PEI-Q could monitor PEI symptoms and impacts and enhance patient-clinician communication.

Declaration of conflicting interests

This study was funded by Abbott. Mr. Arbuckle, Ms Williamson, Ms Johnson and Ms Simpson were contracted by Abbott as consultant to perform the research. They had access to all data for the purposes of analyzing and interpreting the results. Professors Johnson, Staab, Dominguez-Munoz, Levy and Lerch and Dr. Connett were engaged as scientific advisors by Adelphi Values on behalf of Abbott and received honoraria for their participation. The authors had access to the data in summary tables and text. The authors declare that there are no other conflicting interests. Ms Janssen-van Solingen is an employee of Abbott.

Acknowledgements

Professors Johnson, Staab, Dominguez-Munoz, Levy and Lerch and Dr. Connett provided input into the study design, reviewed and provided input into the psychometric analyses and results and provided input into the clinical relevance of symptoms and impacts to PEI. Mr. Arbuckle, Ms Williamson, Ms Johnson and Ms Simpson were employees of Adelphi Values during the conduct of the study, a healthcare outcomes research company which was paid by Abbott to conduct this research. They were involved in all research activities, leading the study design, development of study documents, data collection and analysis and write up of the data. All authors contributed to the writing of the manuscript and reviewed and approved the final manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pan.2018.11.013.

References

[1] Dominguez-Munoz JE. Pancreatic exocrine insufficiency: diagnosis and treatment. J Gastroenterol Hepatol 2011;26(Suppl 2):12–6.

[2] Borowitz D, Lin R, Baker SS. Comparison of monoclonal and polyclonal ELISAs for fecal elastase in patients with cystic fibrosis and pancreatic insufficiency. J Pediatr Gastroenterol Nutr 2007;44(2):219–23.

[3] Shafiq N, Rana S, Bhasin D, et al. Pancreatic enzymes for chronic pancreatitis. The Cochrane Library; 2009.

[4] Löhr JM, Dominguez-Munoz E, Rosendahl J, et al. United European Gastroenterology evidence-based guidelines for the diagnosis and therapy of chronic pancreatitis (HaPanEU). United European Gastroenterology Journal 2017;5(2):153–99.

[5] Dominguez-Munoz JE, Iglesias-Garcia J. Oral pancreatic enzyme substitution therapy in chronic pancreatitis: is clinical response an appropriate marker for evaluation of therapeutic efficacy? JOP : J Pancreas 2010;11(2):158–62.

[6] Sikkens EC, Cahen DL, van Eijck C, Kuipers EJ, Bruno MJ. Patients with exocrine insufficiency due to chronic pancreatitis are undertreated: a Dutch national survey. Pancreatology : official journal of the International Association of Pancreatology (IAP) [et al] 2012;12(1):71–3.

[7] Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit time. Scand J Gastroenterol 1997;32(9):920–4.

[8] Guy W. ECDEU assessment manual for psychopharmacology. Rockville, MD: US Department of Heath,Education, and Welfare Public Health Service Alcohol, Drug Abuse, and Mental Health Administration; 1976.

[9] Nixon A, Doll H, Kerr C, Burge R, Naegeli AN. Interpreting change from patient reported outcome (PRO) endpoints: patient global ratings of concept versus patient global ratings of change, a case study among osteoporosis patients. Health Qual Life Outcome 2016;14:25.

[10] Sikkens ECM, Cahen DL, van Eijck C, Kuipers EJ, Bruno MJ. Patients with exocrine insufficiency due to chronic pancreatitis are undertreated: a Dutch national survey. Pancreatology : official journal of the International Association of Pancreatology (IAP) [et al] 2012;12(1):71–3.

[11] Gooden HM, White KJ. Pancreatic cancer and supportive care—pancreatic exocrine insufficiency negatively impacts on quality of life. Support Care Canc 2013;21(7):1835–41.

[12] Johnson CD, Arbuckle R, Bonner N, et al. QQualitative assessment of the symptoms and impact of pancreatic exocrine insufficiency to inform development of a patient-reported outcome. The Patient - Patient Centered Outcomes Research 2017;10(7):615–28.

[13] Lacy BE, Mearin F, Chang L, et al. Bowel disorders. Gastroenterology 2016;150(6):1393–407. e1395.

[14] Aziz I, Törnblom H, Palsson OS, Whitehead WE, Simrén M. How the change in IBS criteria from Rome III to Rome IV impacts on clinical characteristics and key pathophysiological factors. Am J Gastroenterol 2018:1.

[15] Mössner J, Keim V. Pancreatic enzyme therapy. Deutsches Arzteblatt international 2011;108(34–35):578.

[16] Kahl S, Schutte K, Glasbrenner B, et al. The effect of oral pancreatic enzyme supplementation on the course and outcome of acute pancreatitis: a randomized, double-blind parallel-group study. JOP : J Pancreas 2014;15(2):165–74.

[17] Matlock-Hetzel S. Basic concepts in item and test analysis. 1997.

[18] Hatcher L. A step-by-step approach to using the SAS system for factor Analysis and structural equation modeling. SAS Publishing; 1994.

[19] Hancock GR, Freeman MJ. Power and sample size for the root mean square error of approximation test of not close fit in structural equation modeling. Educ Psychol Meas 2001;61(5):741–58.

[20] Yu CBM. Evaluation of model fit indices for latent variable models with categorical and continuous outcomes. 2001.

[21] Hooper D, Coughlan J, Mullen M. Structural equation modelling: guidelines for determining model fit. Electron J Bus Res Methods 2008;6(1):53–60.

[22] Awang Z. A handbook on SEM. 2012.

[23] Cohen J. Statistical power analysis for the behavioral sciences. revised ed. New York: Academic Press; 1977.

[24] Nunally J, Bernstein IH. Psychometric theory. New York. 1994. p. 248–92.

[25] Krause MS. The implications of convergent and discriminant validity data for instrument validation. Psychometrika 1972;37(2):179–86.

[26] Hedges L, Olkin I. Statistical methods for meta-analysis. San Diego, CA: Academic Press; 1985.

[27] Swets JA. Measuring the accuracy of diagnostic systems. Science 1988;240(4857):1285–93.

[28] Cantor SB, Kattan MW. Determining the area under the ROC curve for a binary diagnostic test. Med Decis Making 2000;20(4):468–70.