Development and validation of the Bowel Cleansing Impact Review (BOCLIR)

Lynda Doward, Jeanette Wilburn, Stephen P McKenna, Roger Leicester, Owen Epstein, Vicki Hedley, Sanjeeva Korala, James Twiss, Deborah Jones, Mike Geraint

ABSTRACT

Objective Acceptability and tolerability of bowel cleansers influence whether patients are able to complete the prescribed dose and, consequently, the quality of the cleansing achieved. No standardised means of assessing patients’ experience of using bowel cleansing is currently available. The aim of the study was to develop the Bowel Cleansing Impact Review (BOCLIR) to assess patient response to bowel cleansing products.

Design Content was derived from qualitative interviews. Face and content validity were assessed via cognitive-debriefing interviews. Finally, patients completed the BOCLIR and a demographic questionnaire. Item response theory (Rasch analysis) was employed for item reduction and assessment of unidimensionality. Internal consistency and construct validity were also assessed.

Results Analysis of 40 interviews resulted in the production of three scales; patient satisfaction, symptomatic impact and activity limitations. Scales were designed to be used alone or together. 19 debriefing interviews demonstrated BOCLIR acceptability, relevance and ease of completion. The validation survey involved 166 patients (52% male, mean (SD) age 54.3 (15.2) years). After misfitting and redundant items were removed all scales fit the Rasch model confirming their unidimensionality. Cronbach’s α-coefficients were high (0.77–0.94) indicating good internal consistency. Scores on the BOCLIR were related to patients’ willingness to use the product in future and ease of drinking the full preparation (p<0.01 for each scale).

Conclusions The BOCLIR is a new measure consisting of three unidimensional scales (satisfaction, symptoms and activity limitations) with good psychometric and scaling properties. The BOCLIR will allow accurate assessment of patients’ response to bowel cleansing preparations.

INTRODUCTION

Procedures such as colonoscopy may be performed for the diagnosis, evaluation or therapeutic treatment of certain diseases of the gastrointestinal system. Adequate cleansing of the colon is a critical factor in achieving diagnostic accuracy and maximising the safety and efficacy of endoscopic or surgical resection. A number of options are available for colon cleansing. The most widely employed are the orthograde gut lavage preparations using sodium or balanced electrolyte solutions. While patients’ experiences of using bowel cleansing preparations have not been adequately researched, evidence suggests that this is dependent on the type of preparation used. The literature also suggests that the acceptability and tolerability of the specific product determines whether the patient is able to complete the prescribed dose and, hence, whether the bowel will be adequately cleansed. Thus, in addition to other factors such as dedicated time and information provided by the specialist screening practitioner, compliance has a direct impact on the accuracy and safety of the procedure. Serious consequences may follow from inadequate cleansing. These include missed underlying pathology, delayed treatment, complications and the cost of the failed procedures.

Studies assessing the relative effectiveness and tolerability of gut cleansing regimes have been hampered by the lack of a standardised measure. This is despite it becoming commonplace in many diagnostic areas to assess patient’s views. Patient-Reported Outcome (PRO) is an
umbrella term that covers a wide range of potential measurement endpoints but is used specifically to refer to outcomes reported directly by the patient, without interpretation by clinicians or others. PRO data are collected with standardised instruments or questions designed to measure an explicit concept (construct). Constructs of interest may include impairments (symptoms), activity limitations (measures of physical, social or psychological functioning), quality of life, patient satisfaction, compliance or treatment preferences.

A PRO measure is only capable of providing useful information if it meets certain developmental, psychometric and scaling standards. Specifically PROs should be:

- Based on an explicit description of the construct assessed.
- Relevant for, and well-targeted to, the patient group with which they are to be used. For a measure to be truly patient-based its content should be generated from relevant patients.
- Reliable, valid and responsive.
- Unidimensional, allowing scale items to be summed to provide a valid total score.

The methodology adopted recent advances in the conceptual and practical bases of measurement; combining strong theoretical underpinnings with the statistical and diagnostic power of the Rasch model. The application of Rasch analysis ensures that the fundamental scaling properties of an instrument (eg, unidimensionality and level of measurement) are assessed in addition to the traditional psychometric assessments of reliability and external construct validity.

**METHODOLOGY**

The development of the BOCLIR involved three key stages:

1. Item generation and production of draft scales.
2. Assessment of face and content validity via cognitive debriefing interviews.
3. Validation by means of a survey designed to assess the psychometric and scaling properties of the scales.

The methodology adopted recent advances in the conceptual and practical bases of measurement; combining strong theoretical underpinnings with the statistical and diagnostic power of the Rasch model. The application of Rasch analysis ensures that the fundamental scaling properties of an instrument (eg, unidimensionality and level of measurement) are assessed.

**Patient samples**

Approval for the study was granted by London-Surrey Borders Research Ethics Committee (LREC 06/Q0806/110). Patients for all stages of the study were recruited through two colonoscopy centres in London; St George’s Hospital Tooting and the Royal Free Hospital in Hampstead. All patients attending for colonoscopy were invited to participate in the study. Although different patients were involved at each stage of the study the reasons for the colonoscopies were comparable. Informed written consent was obtained from all participants.

Patients were eligible for the study if they were aged 18 or above, were able to provide written informed consent and could understand and complete questionnaires independently (as judged by the clinical team). Patients with major morbidity were excluded from the study.

**Conceptual basis of the BOCLIR scales**

The Symptom and Activity Limitations scales took as their conceptual basis the WHO classification of impairments (physiological and anatomical) and activity limitations (capacity and performance). WHO defines impairments as loss or abnormality of physiological, physiological or anatomical structure or function which represents disturbances at the level of the organ. Activity limitation (functioning) is defined as any restriction or lack of ability to perform an activity in the manner or within the range considered normal for a human being.

Patient satisfaction is a more difficult concept to define as there is no current consensus on what ‘satisfaction’ means or how best to measure it. However, this construct has become increasingly popular as an endpoint as patient satisfaction may influence a variety of other outcomes, including compliance, treatment success and continued use of health services.

**Stages in the development of the BOCLIR scales**

**Stage 1: Item generation and production of draft scales**

Qualitative interviews were conducted with patients in a private room following bowel cleansing and prior to their colonoscopy. The aim of the interviews was twofold. First, to identify the key impacts of the bowel cleansing process. Second, to provide statements that could form the basis of questionnaire items. Interviews took the form of an informal, focused conversation and participants were encouraged to talk freely about all aspects of their experience. Interviews were audio-recorded and verbatim transcripts produced. Thematic and interpretive phenomenological analyses were applied to the interview transcripts to identify key constructs for measurement and potential questionnaire items. Each interview transcript was analysed independently by two of five experienced researchers. Item pools were scrutinised to remove items that were idiosyncratic, ambiguous, badly phrased or duplicated.

**Stage 2: Assessment of face and content validity**

Semi-structured cognitive debriefing interviews were conducted with patients to assess the applicability, clarity, relevance, comprehensiveness and practicality of the draft measure. Respondents completed the BOCLIR at clinic, again after bowel cleansing and prior to their colonoscopy, in the presence of an
interviewer. Items found to be problematic were removed and a second draft version of the BOCLIR was prepared.

Stage 3: Validation survey
The purpose of the survey was to reduce the number of items in the BOCLIR scales and to assess the scaling and psychometric properties of the final scales. Participants completed the BOCLIR and a demographic questionnaire.

Scaling and item reduction analysis
Rasch analysis (one-parameter logistic item response theory) was conducted to determine the unidimensionality of the BOCLIR scales using the RUMM 2020 programme. Rasch analysis is a key tool in the development and improvement of questionnaires with advantages over traditional test theory and factor analysis. The Rasch model tests the assumption that a questionnaire scale has the basic property of unidimensionality and establishes whether all items capture a single underlying construct and can be combined to derive a valid total score.

Fit to the model is represented by a non-significant chi-square (χ²) fit statistic. A significance level of 0.01 was used to account for multiple testing. Items were assessed to see if they fit the Rasch model. Those that showed overall misfit to the model or had excessive fit residuals (≥ 2.5 or < −2.5) were considered for removal. Highly negative residuals indicate overfit to the model (suggesting that the item is duplicating other items in the scale) while highly positive residuals indicate that the item violates unidimensionality. The person separation index (PSI) or the extent to which items distinguish between distinct levels of the underlying construct was also assessed. For PSI, 0.7 is considered the minimal required value. Finally, differential item functioning (DIF), which represents the stability of the instrument irrespective of the group being evaluated, was assessed. Assessment of DIF ensures that responses to individual items are not affected by factors that are external to the measurement tool, such as age and gender. Such DIF was assessed using analysis of variance models. Items failing to fit the Rasch model and/or that showed DIF by gender or age (above or below the median) were removed. Throughout the item reduction process the content and meaning of items was taken into account before their removal, to maintain content validity.

Traditional psychometric analysis
Internal consistency was assessed using Cronbach’s α. A correlation of 0.70 is accepted as indicating adequate consistency. Construct validity was determined by examining BOCLIR scores of respondents who differed according to certain known factors. For all three scales, respondent scores were related to: the amount of cleansing product the patient was able to consume (‘full’ or ‘partial’), ease of drinking the whole volume (‘very easy’ to ‘very difficult’) and how willing they would be to use the same preparation in the future (‘very happy’ to ‘very unhappy’). In addition, scores on the symptoms and activity limitations scales were related to how well the patient felt after taking the product, (‘very well’ to ‘very unwell’), satisfaction with the overall experience (‘very pleasant’ to ‘very unpleasant’) and the convenience of the preparation (‘very convenient’ to ‘very inconvenient’). Individuals reporting a worse experience and those who had greatest difficulty consuming the product were expected to have significantly higher BOCLIR scores. Non-parametric tests for independent samples (Mann-Whitney U test for two groups and Kruskal-Wallis one-way analysis of variance for three or more groups) were employed. Psychometric testing was performed using the SPSS V16.0 statistical package.

RESULTS
Participant details for all study stages of the study are shown in table 1.

Item generation and production of draft scales
Interviews were conducted with 20 patients from each centre and lasted between 20 and 45 min. Three key areas of impact were identified; satisfaction with preparation, symptomatic impact and impact on the patient’s ability to carry out their usual daily activities. Common themes reported within these key areas are presented in table 2. Using these areas as a guide, interview transcripts were searched to identify statements that could be used as the basis for items. This allowed a draft questionnaire to be produced consisting of three scales; patient satisfaction (12 items), symptoms (29 items) and activity limitations (13 items). In order to ensure that individual symptoms were expressed in the most appropriate way for patients, some alternative expressions for the same concept were included in the draft questionnaire. It was intended that the less appropriate items would be removed from the scale during the later testing stages. Polytomous response formats were selected for the scales; five options for the patient satisfaction items and four for the symptoms and activity limitations items.

Assessment of face and content validity
Cognitive debriefing interviews were conducted with 19 patients. Interviews lasted approximately 20 min. Patients reported that they found the BOCLIR clear, acceptable and relevant. Minor changes were made to item wording to improve the flow of the scale. A new draft BOCLIR was produced for use in the validation survey.

Validation survey
The BOCLIR and a demographic questionnaire were completed by 166 patients following bowel cleansing and prior to colonoscopy.
Scaling properties and item reduction
The initial 12-item satisfaction and 29-item symptom scales showed misfit to the Rasch model. Four items were removed from the satisfaction scale due to item misfit, extreme fit residuals or item redundancy. After item reduction the scale fit the Rasch model (table 3). Low levels of uniform DIF by age were observed for three items and DIF by gender for one item. However, these factors were judged unlikely to compromise the scale’s overall fit. Fifteen items were

Table 2 Key areas of impact

<table>
<thead>
<tr>
<th>Patient satisfaction</th>
<th>Symptoms</th>
<th>Activity limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>General satisfaction</td>
<td>Bloating</td>
<td>General ability to do usual activities</td>
</tr>
<tr>
<td>Volume</td>
<td>Cramps/gurgling/stomach pains</td>
<td>Household chores for example, cooking</td>
</tr>
<tr>
<td>Taste</td>
<td>Nausea/dizziness</td>
<td>Family impact</td>
</tr>
<tr>
<td>Drinking full volume/keeping it down</td>
<td>Energy/fatigue/concentration problems</td>
<td>Social impact</td>
</tr>
<tr>
<td>Preparation/instructions</td>
<td>Sleep disturbance</td>
<td>Work impact</td>
</tr>
<tr>
<td>Adhering to diet/fast</td>
<td>Headache</td>
<td></td>
</tr>
<tr>
<td>Certainty of thoroughness of cleansing</td>
<td>Need to urinate</td>
<td></td>
</tr>
<tr>
<td>Speed of effect</td>
<td>Bowel evacuation</td>
<td></td>
</tr>
<tr>
<td>Comparison to other preparations</td>
<td>► Urgency/control</td>
<td></td>
</tr>
<tr>
<td></td>
<td>► Frequency</td>
<td></td>
</tr>
<tr>
<td></td>
<td>► Leakage</td>
<td></td>
</tr>
<tr>
<td></td>
<td>► Pain/soreness</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hunger</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dry mouth/thirst</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Irritability</td>
<td></td>
</tr>
</tbody>
</table>
removed from the symptom scale due to item redundancy, misfit, extreme fit residuals, DIF, poor item total correlations or low endorsement rate. The final scale fit the Rasch model. After item reduction only one item exhibited low levels of DIF by age. The draft 13-item activity limitations scale fit the Rasch model. However, one item displayed a high fit residual and was removed from the scale. Low levels of DIF by age were observed for three items and by gender for two items. Again, this was judged not to compromise overall scale fit. The PSI was good for all scales indicating appropriate inter-relatedness of items. The Rasch analyses demonstrated that the final versions of all scales were unidimensional (p<0.05) indicating that they each measure a single construct (table 3). Consequently, scale responses can be summed to provide valid scale scores.

### Traditional psychometric analysis

The final satisfaction scale contains eight items scored 0 (highly satisfied) to 32 (highly dissatisfied). The symptoms scale includes 14 items scored 0 (no symptoms) to 42 (severe symptoms). The activity limitations scale is made up of 12 items with possible scores ranging from 0 (no effect on activities) to 36 (activities greatly affected). Table 4 shows descriptive statistics for the final scales. Limited ceiling or floor effects were found for all scales. Internal consistency was good for all three scales. Cronbach’s α values were Satisfaction 0.84, Symptoms 0.77 and Activities 0.94. Table 5 shows symptoms and activity limitation scale scores by satisfaction factors. Both scales were able to distinguish between groups of patients categorised by how well or unwell they felt after taking the bowel cleansing preparation, how pleasant or unpleasant they rated the experience of using the product and self-rated convenience of product use.

Figures 1–3 show BOCLIR scale scores by product consumption factors; specifically, volume of cleanser

### Table 3

Rasch fit statistics for final Bowel Cleansing Impact Review scales

<table>
<thead>
<tr>
<th>Scale</th>
<th>Item-person interaction</th>
<th>Item-trait interaction χ²</th>
<th>PSI</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfaction</td>
<td>p&lt;0.13</td>
<td>0.86</td>
<td>0.22</td>
<td>1.12</td>
<td>−10.29</td>
<td>1.09</td>
<td></td>
</tr>
<tr>
<td>Symptoms</td>
<td>p&lt;0.38</td>
<td>0.71</td>
<td>−0.07</td>
<td>0.96</td>
<td>−0.17</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td>p&lt;0.43</td>
<td>0.94</td>
<td>0.05</td>
<td>1.4</td>
<td>−0.28</td>
<td>1.30</td>
<td></td>
</tr>
</tbody>
</table>

PSI, person separation index.

### Table 4

Bowel Cleansing Impact Review scores

<table>
<thead>
<tr>
<th></th>
<th>Satisfaction</th>
<th>Symptoms</th>
<th>Activity limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>10.7 (5.9)</td>
<td>11.9 (5.7)</td>
<td>17.0 (10.0)</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>10 (6–15)</td>
<td>11.5 (8–15)</td>
<td>17 (9–25)</td>
</tr>
<tr>
<td>Range</td>
<td>0–24</td>
<td>2–28</td>
<td>0–36</td>
</tr>
<tr>
<td>% Scoring min</td>
<td>2.6</td>
<td>0.0</td>
<td>4.6</td>
</tr>
<tr>
<td>% Scoring max</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

### Table 5

Known group validity for Bowel Cleansing Impact Review symptoms and activity limitations scales

<table>
<thead>
<tr>
<th>How well or unwell did you feel as a result of taking the bowel preparation?</th>
<th>n</th>
<th>Symptoms Mean (%)</th>
<th>Activity limitations Mean (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very well</td>
<td>41</td>
<td>7.9 (3.3)</td>
<td>12.0 (9.0)</td>
</tr>
<tr>
<td>Quite well</td>
<td>48</td>
<td>11.1 (4.8)</td>
<td>15.8 (9.6)</td>
</tr>
<tr>
<td>Neither well nor unwell</td>
<td>43</td>
<td>13.1 (5.5)</td>
<td>19.6 (9.5)</td>
</tr>
<tr>
<td>Quite / very unwell</td>
<td>21</td>
<td>18.3 (4.4)</td>
<td>24.8 (8.5)</td>
</tr>
<tr>
<td>p</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall, how would you rate your experience of using this bowel preparation?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very pleasant</td>
<td>14</td>
<td>6.7 (4.1)</td>
<td>7.4 (9.1)</td>
</tr>
<tr>
<td>Quite pleasant</td>
<td>35</td>
<td>9.3 (4.1)</td>
<td>14.3 (9.2)</td>
</tr>
<tr>
<td>Neither pleasant nor unpleasant</td>
<td>62</td>
<td>11.2 (4.4)</td>
<td>15.5 (8.5)</td>
</tr>
<tr>
<td>Quite / very unpleasant</td>
<td>42</td>
<td>16.5 (5.6)</td>
<td>24.5 (8.6)</td>
</tr>
<tr>
<td>p</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How convenient was it to fit taking the bowel preparation into your daily routine?</th>
<th>n</th>
<th>Symptoms Mean (%)</th>
<th>Activity limitations Mean (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very convenient</td>
<td>33</td>
<td>8.8 (3.6)</td>
<td>12.8 (9.3)</td>
</tr>
<tr>
<td>Quite convenient</td>
<td>54</td>
<td>10.5 (4.8)</td>
<td>15.3 (10.6)</td>
</tr>
<tr>
<td>Neither convenient nor inconvenient</td>
<td>27</td>
<td>13.0 (6.0)</td>
<td>17.9 (9.4)</td>
</tr>
<tr>
<td>Quite inconvenient</td>
<td>28</td>
<td>15.4 (6.7)</td>
<td>20.9 (8.3)</td>
</tr>
<tr>
<td>Very inconvenient</td>
<td>10</td>
<td>13.9 (3.7)</td>
<td>26.2 (5.8)</td>
</tr>
<tr>
<td>p</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ENDOSCOPY**

---

consumed, ease of use and willingness to use in the future. Patients who reported a worse experience in terms of these factors, scored higher on the BOCLIR scales.

DISCUSSION

Adequate colon cleansing is a prerequisite to effective endoscopic examination and surgical procedures associated with the large colon and distal portion of the small bowel. The accuracy with which colorectal polyps or lesions can be detected is largely associated with the cleanliness of the luminal environment. Poor cleansing is also associated with an increased risk of peritoneal contamination should perforation occur. Similarly, poor cleansing may increase the risk of colonic gas explosion during colonoscopy with electrocautery. Indeed, up to a third of failed procedures have been attributed to inadequate bowel cleansing.

Bowel cleansing may be achieved by several different methods that vary by volume of liquid required for ingestion, the length of time over which the preparation is consumed and taste. Preparations are taken at prescribed intervals either all on the day before the colonic procedure or on a combination of the day before and the morning of the procedure. In addition, all preparations require the patient to maintain a liquid diet for at least 12 h prior to the procedure. It has been reported that the requirement to use a bowel preparation may be a major factor in deterring patients from attending for screening colonoscopies. Reports on adverse events and on ability to complete prescribed doses suggest that many patients find them unpleasant to use or even distressing. Symptoms such as nausea, abdominal cramps and vomiting are commonly reported. Problems with sleep disturbance have also been reported.

For clinicians and commissioners to judge the best cleansing agent it is necessary to employ Patient-Reported Outcome Measures with good psychometric properties. The BOCLIR was developed to address this need. It consists of three stand-alone scales (satisfaction, symptoms and activity limitations) which together form a suite of instruments. The content of the BOCLIR was generated from qualitative interviews with relevant patients and the scales represent their key areas of concern. Respondents find the instrument acceptable, comprehensive and relevant. The BOCLIR is completed by the patient independently and takes approximately 5 min. It can be completed at any stage following bowel cleansing and prior to the procedure. As it is quick and easy to complete it is suitable for use in clinical settings.

Application of the Rasch model showed each BOCLIR scale to be unidimensional and to have minimal DIF. Confirmation of unidimensionality...
means that each scale provides an index or valid single total score. Construct validity was demonstrated by evaluating the scales’ ability to distinguish between groups of patients categorised by known factors. Participants who reported a worse experience (as indicated by several variables) had worse scores on the BOCLIR scales.

It is concluded that the BOCLIR will provide a valuable tool for measuring patients’ response to bowel cleansing preparations prior to colonoscopy.

The study has limitations. As it was designed to develop and validate the new measure it did not allow the comparison of BOCLIR scores with clinical judgement or markers of the degree of bowel cleansing achieved. It would be useful for future studies to compare BOCLIR scores with clinical assessments of bowel cleansing. It is anticipated that further use of the BOCLIR will provide such data. The work was conducted entirely in London. However, there is no reason to suggest that individuals from other locations would respond differently to bowel cleansing.

What is already known on this topic

- Satisfactory colonic cleansing is considered a crucial factor for the efficacy and safety of colonic procedures.
- Acceptability and tolerability of bowel cleansers influence the ability of the patient to complete the prescribed dose and, consequently, the quality of the cleansing achieved.
- The available literature suggests that there are clear differences in patient tolerability of different bowel cleansing preparations.
- No standardised means of assessing patients’ experience of bowel cleansers is available.

What this study adds

- Three key factors affecting patient response to, and compliance with, bowel cleansing preparations were identified: satisfaction with the preparation, symptomatic impact and effect on patients’ ability to carry out their usual daily activities.
- A new suite of Patient-reported outcome scales, the Bowel Cleansing Impact Review (BOCLIR), has been developed covering these key areas.
- The BOCLIR scales have been shown to be unidimensional and to have excellent content and construct validity.

Acknowledgements The authors would like to acknowledge all those patients who helped with the study.

Contributors JW was involved with the design of the study, acquisition, analysis and interpretation of data and drafting of the manuscript. LD and SPM were involved in the conception and design of the study, acquisition and interpretation of data and contributed to the manuscript. RL and OE were involved with the design of the study, acquisition of data and reviewed and contributed to the manuscript. VH and SK were involved in the acquisition of data and reviewed and contributed to the manuscript. DJ and MG were involved in the design of the study and contributed to the manuscript. All authors read and approved the final manuscript.

Funding Norgine Pharmaceuticals Ltd.

Competing interests None.

Ethics approval London-Surrey Borders Research Ethics Committee.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES


