Comprehensive validation of the Boston Bowel Preparation Scale

Audrey H. Calderwood, MD, Brian C. Jacobson, MD, MPH, FASGE
Boston, Massachusetts, USA

Background: The Boston Bowel Preparation Scale (BBPS) is a novel bowel cleanliness rating scale that has undergone partial validation previously.

Objective: To fully validate the BBPS and assess the ease of its dissemination.

Design: Observational study.

Setting: Various endoscopy units worldwide.

Subjects: Endoscopists.

Methods: Video recordings of colonoscopies with varying degrees of cleanliness were viewed twice by gastroenterologists at 1 medical center. For each video, participants assigned segment and total BBPS scores. Endoscopists worldwide were also surveyed about their experience with the BBPS after viewing an instructional video.

Main Outcome Measurements: Intraclass correlation coefficients and weighted κ values assessed inter- and intrarater reliability, respectively. The BBPS was used among 983 patients undergoing screening colonoscopy.

Results: The BBPS demonstrated near-perfect interrater reliability (intraclass correlation coefficient = 0.91) and substantial intrarater reliability (weighted κ = 0.78; 95% CI, 0.73-0.84). Among 983 colonoscopies, right and left colon segment scores of 2 or 3 had a multivariate odds ratio of 1.60 (95% CI, 1.01-2.55) and 2.58 (95% CI, 1.34-4.98), respectively, for polyp detection compared with segment scores of 0 or 1. Endoscopists from a variety of settings worldwide found the BBPS easy to implement and applicable to their patient population.

Limitations: Single-center reliability testing.

Conclusions: The BBPS is a valid and reliable instrument for assessing bowel cleanliness during colonoscopy. Segment scores may represent a standardized way to determine bowel preparation adequacy. The BBPS can be easily disseminated through the use of a brief instructional video. (Gastrointest Endosc 2010;72:686-92.)
by colonoscopy for right- and left-sided advanced neoplasia, a standardized bowel preparation rating scale that has the ability to capture colon segment differences may be advantageous. Furthermore, a standardized bowel preparation rating scale should be readily disseminated among endoscopists worldwide.

The Boston Bowel Preparation Scale (BBPS) is a bowel cleanliness rating scale originally designed and validated for use during colonoscopy-oriented research. It relies on the summation of 3 individual colonic segment scores (for the right, transverse, and left colons) to indicate the degree of bowel visualization. Total BBPS scores have been associated with clinical outcomes such as polyp detection rates, recommendations for repeated procedures, and colonoscope insertion and withdrawal times. Reliability testing, however, was based only on 3 truncated colonoscopy video clips and was limited to the intermediate range of bowel cleanliness. Individual segment scores have not been validated and total BBPS scores have not yet been shown to be reliable across the full spectrum of bowel cleanliness. The current study was designed to assess inter- and intrarater reliability of total and segment BBPS scores across the full spectrum of possible degrees of bowel cleanliness, to use the BBPS among a diverse patient population, to examine the relationship between individual segment scores and polyp detection rates, and to assess whether the BBPS can be easily disseminated to other medical centers for both clinical and research purposes.

METHODS

The study was approved by the Institutional Review Board of the Boston University Medical Center (BUMC) with waiver of informed consent.

The BBPS

The BBPS was described in detail previously. Briefly, the BBPS is applied during the withdrawal phase of colonoscopy, after all washing, suctioning, and other cleaning maneuvers have been performed by the endoscopist. Good clinical practice dictates that the endoscopist should always clean as much as possible to obtain the highest possible score and ensure the best possible bowel cleanliness. Each of the 3 segments of the colon (right, including the cecum and ascending colon; transverse, including the hepatic and splenic flexures; and left, including the descending colon, sigmoid, and rectum) is given a score from 0 to 3 defined as follows: 0 = unprepared colon segment with mucosa not seen because of solid stool that cannot be cleared; 1 = portion of mucosa of the colon segment seen, but other areas of the colon segment not well seen because of staining, residual stool, and/or opaque liquid; 2 = minor amount of residual staining, small fragments of stool and/or opaque liquid, but mucosa of colon segment seen well; 3 = entire mucosa of colon segment seen well with no residual staining, small fragments of stool, or opaque liquid.

Each of the 3 segment scores is then summed for a total score of 0 to 9, in which 0 is unprepared and 9 is entirely clean. If an endoscopist aborts a procedure because of inadequate preparation, then any nonvisualized proximal segments are assigned a score of 0. An instructional video demonstrating how to use the BBPS is available online at http://www.bmc.org/gastroenterology/research.htm.

Assessment of reliability

We recorded 119 colonoscopies at BUMC by using DVD Recorders (Multi-Function DVD Recorder VRD-MC3; Sony, Tokyo, Japan) connected to our endoscopy unit’s video processors (EVIS Exera II CLV-180; Olympus Medical Systems, Tokyo, Japan). Recordings contained only endoscopic video footage and were free of patient identifiers. From this video library, one author selected 10 video clips demonstrating the full range of bowel cleanliness encountered in clinical practice, from excellent to poor, in which no polyps were detected. Each video clip demonstrated the entire withdrawal portion of a colonoscopy, including 2 cases when the preparation was so poor that only the left-sided colon was seen. Using video editing software (Adobe Premiere Elements 3.0; Adobe Systems Inc, San Jose, Calif), we created DVDs, termed bowel preparation DVDs (BP-DVDs), that contained these 10 colonoscopy video clips but in different random orders. The standard bowel preparation in our endoscopy unit at the time the video clips were recorded was a 4-L polyethylene glycol lavage solution (GoLyte, NuLyte, Braintree Laboratories Inc, Braintree, Mass; Colyte, Alaven Pharmaceutical LLC, Marietta, Georg; or TriLyte; Schwarz Pharma Inc, Milwaukee, Wisc) taken the night before the scheduled colonoscopy. Some patients may have also taken bisacodyl 20 mg before beginning their lavage solution. Split dosing was not used during the time of this study.

Attending gastroenterologists and fellows-in-training at BUMC, a tertiary care academic hospital, viewed all 10 colonoscopies on the DVDs on 2 occasions at least 4 weeks apart. Different BP-DVDs that contained videos in different orders were used for the first and second view-

Take-home Message

- A standardized bowel preparation rating scale for colonoscopy is important for clinical practice, quality assurance, and outcomes research. In this comprehensive validation study, the Boston Bowel Preparation Scale (BBPS) was found to be a valid and reliable instrument for assessing bowel cleanliness during colonoscopy. The BBPS is easily disseminated through use of a brief instructional video.
ings to minimize the physician’s ability to remember his or her previous responses (ie, to limit recall bias). For each video, participants assigned segment and total BBPS scores, categorized the bowel preparation as adequate or inadequate to exclude polyps larger than 5 mm, and recommended the timing of the next colonoscopy, assuming the colonoscopy was performed on a 60-year-old healthy man with no family history of colonic neoplasia undergoing routine screening.

Use of the BBPS during screening colonoscopy and relationship to polyp detection

After all of the endoscopists had viewed the instructional video describing use of the BBPS, they applied the BBPS prospectively during 983 screening colonoscopies at our institution (633 described in our preliminary analysis and 350 additional procedures). The 983 cases were those in which a BBPS score was recorded during an unrelated study of screening colonoscopy (NCT00643682). We excluded patients who had an incomplete procedure for reasons other than bowel preparation (n = 7) or had a history of colectomy (n = 2). After each screening colonoscopy, the endoscopist was asked to record the quality of the bowel preparation by using the BBPS. The endoscopists also recorded the location and size of all polyps found during the examinations. We defined diminutive polyps as those estimated by the endoscopist to be 5 mm or smaller.

Assessment of the ability to disseminate the BBPS to other medical centers

Since the publication of our initial BBPS validation study, 35 endoscopists in numerous countries and practice settings have contacted us requesting access to our BBPS instructional video. This instructional video was made available as a free download from our gastroenterology division’s Web site (as listed previously), and the link was sent to the requesting endoscopists. We later invited these endoscopists to complete an institutional review board–approved, 26-question electronic survey to learn about their experience with the instructional video describing use of the BBPS, they applied the BBPS to other medical centers. Since the publication of our initial BBPS validation study, 35 endoscopists in numerous countries and practice settings have contacted us requesting access to our BBPS instructional video. This instructional video was made available as a free download from our gastroenterology division’s Web site (as listed previously), and the link was sent to the requesting endoscopists. We later invited these endoscopists to complete an institutional review board–approved, 26-question electronic survey to learn about their experience with the instructional video describing use of the BBPS, they applied the BBPS to other medical centers. Since the publication of our initial BBPS validation study, 35 endoscopists in numerous countries and practice settings have contacted us requesting access to our BBPS instructional video. This instructional video was made available as a free download from our gastroenterology division’s Web site (as listed previously), and the link was sent to the requesting endoscopists. We later invited these endoscopists to complete an institutional review board–approved, 26-question electronic survey to learn about their experience with the instructional video describing use of the BBPS, they applied the BBPS to other medical centers.

Statistical analysis

To assess interrater reliability, we calculated the intraclass correlation coefficient (ICC) (2,1) among total and segment BBPS scores applied after viewing the BP-DVDs, using the methods of Shrout and Fleiss. To assess intrarater reliability, we calculated weighted κ values for total and segment BBPS scores, according to Fleiss and Cohen. We calculated the percentage of physicians deeming a bowel preparation adequate to exclude polyps larger than 5 mm for each BBPS score (0-9) assigned during the BP-DVD viewings. The distribution of continuous variables was tested for normality. We then calculated the mean (standard deviation) interval of follow-up to next colonoscopy recommended by participants. Among 983 screening colonoscopies, we compared mean BBPS scores among various demographic categories including sex (male, female), age (younger than 65, 65 and older), and self-identified race/ethnicity (white, black, Hispanic, Asian). In this last comparison, white patients were the reference population. Chi-square tests and analysis of variance were used to compare BBPS scores across groups. Among these 983 colonoscopies, we also determined the polyp detection rate for each colon segment based on BBPS segment scores. Unconditional logistic regression was used to determine the odds ratio for polyp detection controlling for segment scores, age, race, and sex. Descriptive statistics were used to analyze survey results. All calculations were performed by using SAS version 9.1 (SAS Institute, Cary, NC), and 2-sided P values <.05 were considered significant.

Sample size estimation

The primary outcomes of this study were the inter- and intrarater reliability for total BBPS scores. Because the intrarater reliability is usually higher than or equal to the interrater reliability, we performed a power calculation based on expected ICCs to estimate appropriate sample sizes. Based on our preliminary BBPS validation study, we assumed that at least 10 gastroenterologists (raters) would participate and that the expected ICC would be 0.74 or greater. We then estimated that approximately 6 to 11 videos (subjects) would be needed to have a 95% confidence interval with a width of ±0.20 for an ICC between 0.7 and 0.8. If the ICC approached 0.9, then approximately 8 videos would be needed to have a 95% confidence interval with a width of ±0.10. Given these sample size considerations and the time commitment required by our volunteer raters, we selected 10 colonoscopy videos for this study. A post hoc analysis showed that we had 75% power to detect meaningful differences in polyp detection for left segment scores of 0 or 1 compared with 2 or 3. We only had 45% power in the right side of the colon to detect such a difference. A total of 5453 right segment scores (or colonoscopies) would have been needed to achieve 80% power to detect such differences.

RESULTS

Reliability testing

The BP-DVDs were viewed by 9 full-time faculty and 3 fellows at BUMC. Previous work showed that the BBPS can be used reliably by clinicians with various levels of experience with similar inter- and intrarater reliability, and therefore results for all 12 participants were analyzed and reported together. Some of the participants had been introduced to the BBPS during a previous research study, but the scale was not being used routinely for...
rating bowel preparation quality. Individuals viewed a different BP-DVD on 2 occasions, with a mean (± standard deviation) of 61 ± 35 days between viewings. Table 1 shows the spectrum of bowel cleanliness demonstrated by the 10 colonoscopy videos clips in the BP-DVDs.

All physicians deemed the bowel preparation adequate to exclude polyps larger than 5 mm when the BBPS score was 8 or greater compared with 88% when the score was 7, 82% when the score was 6, 33% when the score was 5, and 0% when the score was 4 or less.

When the BBPS score was less than 5, all physicians recommended repeat colonoscopy within 1 year. When the BBPS score was between 5 and 6, the mean recommended interval to next colonoscopy was 4.3 (± 3.9) years. When BBPS score was 7 or greater, all physicians recommended the next colonoscopy occur in 10 years (Table 1).

The BBPS demonstrated near-perfect interrater reliability with an ICC of 0.91 over the full range of possible total BBPS scores. Intrarater agreement was substantial with a weighted κ value of 0.78 (95% CI, 0.73-0.84) over the full range of possible total BBPS scores.14 Inter- and intrarater reliability for BBPS segment scores by location (right, transverse, left) were also similar (Table 2).

<table>
<thead>
<tr>
<th>Video</th>
<th>Total BBPS score, mean (SD)*</th>
<th>% stating preparation was adequate to exclude polyps &gt;5 mm</th>
<th>Recommended screening interval, y, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0 (0)</td>
<td>0</td>
<td>0 (0)</td>
</tr>
<tr>
<td>B</td>
<td>1.7 (0.9)</td>
<td>0</td>
<td>(0)</td>
</tr>
<tr>
<td>C</td>
<td>1.9 (0.7)</td>
<td>0</td>
<td>(0)</td>
</tr>
<tr>
<td>D</td>
<td>3.1 (0.8)</td>
<td>0</td>
<td>(0)</td>
</tr>
<tr>
<td>E</td>
<td>5.0 (1.2)</td>
<td>33</td>
<td>75</td>
</tr>
<tr>
<td>F</td>
<td>5.2 (0.5)</td>
<td>25</td>
<td>100</td>
</tr>
<tr>
<td>G</td>
<td>6.3 (0.9)</td>
<td>75</td>
<td>100</td>
</tr>
<tr>
<td>H</td>
<td>7.2 (0.8)</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>I</td>
<td>7.3 (1.2)</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>J</td>
<td>8.0 (1.1)</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

BBPS, Boston Bowel Preparation Scale; SD, standard deviation. *Mean of 12 raters assigning a Boston Bowel Preparation Scale score to the same video clips.

TABLE 2. Intraclass correlation coefficients and weighted kappa values for total BBPS scores and segment scores by location demonstrated consistent inter- and intrarater reliability, respectively

<table>
<thead>
<tr>
<th></th>
<th>Total score</th>
<th>Right side of the colon score</th>
<th>Transverse colon score</th>
<th>Left side of the colon score</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC</td>
<td>0.91</td>
<td>0.88</td>
<td>0.83</td>
<td>0.79</td>
</tr>
<tr>
<td>Weighted κ value (95% CI)</td>
<td>0.78 (0.73-0.84)</td>
<td>0.78 (0.71-0.85)</td>
<td>0.73 (0.64-0.81)</td>
<td>0.75 (0.67-0.83)</td>
</tr>
</tbody>
</table>

BBPS, Boston Bowel Preparation Scale; CI, confidence interval; ICC, intraclass correlation coefficient.

Use during screening colonoscopy and segment score validation

The BBPS was used prospectively by 12 attending gastroenterologists during 983 screening colonoscopies. The mean (± SD) BBPS score was 6.2 ± 1.6, and the median score was 6.0 (range 0.0-9.0; interquartile range 6.0-7.0). Women had slightly higher BBPS scores compared with men (6.4 vs 6.0, \( P < .001 \)), and Asian patients had slightly higher scores compared with whites (6.6 vs 6.2, \( P = .05 \)), but there was no significant difference in mean BBPS scores among other races using whites as the reference population (black 6.2, \( P = .76 \); Hispanic 6.1, \( P = .94 \)) or by age (6.2 for age younger than 65 vs 6.3 for age 65 and older, \( P = .42 \)).

Individual BBPS segment scores in the right side of the colon showed a positive trend with polyp detection rates (Table 3). We found that higher BBPS segment scores (2 and 3 vs 0 and 1) were associated with improved polyp detection in the left side of the colon (odds ratio 2.58; 95% CI, 1.34-4.98) and right side of the colon (OR 1.60, 95% CI 1.01-2.55) when controlling for age, sex, and race. There was no association found between scores and polyp detection in the transverse colon (odds ratio 0.7; 95% CI, 0.48-1.96).

Ability to disseminate the BBPS

Twenty-four of 35 invited participants (69%) completed our survey. Of these, 16 (67%) were adult gastroenterologists, 7 (29%) were gastroenterology fellows, and 1 (4%) was a pediatric gastroenterology nurse. The number of years in practice of the adult gastroenterologists ranged from less than 5 years (37%) to 11 to 20 years (25%) to more than 20 years (37%). Participants were from the United States (California, Illinois, Minnesota, North Caro-
pared with 0 and 1, were associated with an increased population.

found the BBPS to be generally applicable to their patient population. Almost all respondents experience all felt confident in using the BBPS after viewing throughout several countries and with varying degrees of rates. Gastroenterologists and gastroenterology fellows between individual segment scores and polyp detection demonstrated significant associations. Prospective use of the BBPS during more than 900 screen reliabilities over the full range of possible segment scores. Gastroenterologists and gastroenterology fellows for rating colonoscopies, with 16 (67%) using it in more than 50 cases. Eighteen reported using the BBPS for research purposes, 14 for routine patient care, 3 for teaching and training, and 1 for quality improvement initiatives. Ninety-two percent of respondents found the BBPS to be generally applicable to their patient population.

**DISCUSSION**

In this study, we further established the validity and reliability of the BBPS as an instrument for assessing bowel cleanliness during colonoscopy for both research and clinical purposes. We also demonstrated that the BBPS can be easily disseminated to a wide range of practice locations worldwide by using a simple online video. Total BBPS scores demonstrated strong inter- and intrarater reliability over the full range of possible scores and bowel cleanliness. Total BBPS scores were associated with perception of bowel preparation adequacy for excluding polyps larger than 5 mm and recommended timing of repeat colonoscopy, both of which provide additional evidence of the construct validity of the BBPS. Individual segment BBPS scores also demonstrated strong inter- and intrarater reliability over the full range of possible segment scores. Prospective use of the BBPS during more than 900 screening colonoscopies demonstrated significant associations between individual segment scores and polyp detection rates. Gastroenterologists and gastroenterology fellows throughout several countries and with varying degrees of experience all felt confident in using the BBPS after viewing a short instructional video. Almost all respondents found the BBPS to be generally applicable to their patient population.

Dichotomized BBPS segment scores of 2 and 3, compared with 0 and 1, were associated with an increased likelihood of finding polyps in the right and left sides of the colon. Surprisingly, no association was found between transverse colon segment scores and polyp detection. Fewer absolute polyps were detected in the transverse colon compared with the right and left colons, perhaps limiting the power to detect meaningful differences.

The lack of association between transverse colon segment scores and polyp detection raises the possibility of a simplified BBPS that only includes assessments of the right and left sides of the colon. Although future research may indeed justify exclusion of the transverse colon segment score, we believe that it is likely premature to draw this conclusion because there may be important information captured by inclusion of all 3 segment scores for research or clinical purposes.

Prospective use of the BBPS in more than 900 colonoscopies showed small variations in quality of bowel cleanliness by sex and race. Women had slightly cleaner preparations compared with men, a finding consistent with other studies. In our study, we also found that Asians had cleaner preparations compared with whites, a trend seen in 1 previous study. This information may be relevant for those planning to use the BBPS for research purposes in special populations, although, in general, the differences observed in mean scores were small and likely of limited clinical significance.

Expanding further on our previous validation of total BBPS scores, we showed that the BBPS is sensitive to differences in cleanliness within the colonic segments, and therefore it may become a useful tool in helping to understand segment-specific risks for missed pathology based on the degree of bowel cleanliness. This feature may prove clinically relevant given reported discrepancies in the effectiveness of colonoscopy for preventing colorectal cancer of the left versus the right side of the colon. Such discrepancies may reflect location-specific differences in bowel preparation adequacy, and BBPS segment scores may highlight where endoscopists should focus more time and effort in cleaning. Through its ability to reflect the clinical effectiveness of colonoscopy and to preserve variation in cleanliness by segment, the BBPS may be a useful tool in quality assurance for colonoscopy. The BBPS can be presented as a total score (eg, 6), individual segment scores (eg, 1-3-2), or both (eg, 1-3-2 = 6) to fit the user’s
needs, recognizing that only individual segment scores preserve segmental differences in bowel preparation quality.

Our study has several strengths including assessment of both inter- and intrarater reliability of total and segment scores, correlation of BBPS score with the perception of bowel preparation adequacy in excluding polyps larger than 5 mm, and prospective validation among a large number of ethnically diverse patients. Nevertheless, we acknowledge certain limitations. The BBPS requires assessment of bowel preparation quality in the right, transverse, and left colon segments, and landmarks distinguishing these segments may be poorly defined or difficult to recognize, raising the possibility of the difficulty of identifying specific segments. Although this has the potential to result in assignment of segment scores that do not reflect the intended location, this limitation is not unique to the BBPS, but rather inherent to the practice of colonoscopy and affects other factors, such as the reporting of polyp location. However, we also hypothesize that the excellent interrater reliabilities observed for various segments indicates at least reasonable agreement among endoscopists about where the colonoscopes were during score assignments.

Reliability testing was performed by using only the withdrawal portion of 10 colonoscopies and did not include visualization during colonoscope insertion. However, because the BBPS is applied only during the withdrawal portion of a colonoscopy, our study methods represent real-world use of the scale.

Our study was limited to a single, urban, academic medical center that serves a diverse, underserved, predominantly black population (39% black, 23% white, and 20% Hispanic), and therefore our findings may not be generalizable. However, the BBPS performed very similarly among a heterogeneous group of patients, suggesting that it should be useful in various clinical settings. Furthermore, a broad range of endoscopists from multiple national and international settings found the BBPS applicable to their patient populations. The BBPS was also not evaluated for use during other bowel-imaging procedures, such as CT colonography or capsule endoscopy, or in colonoscopies after colorectal surgery, and, as such, we are unable to comment on its utility in these settings. Until the BBPS is validated in patients who have undergone colorectal surgery or whose procedures are aborted for reasons other than bowel preparation quality, we would suggest that segment scores, rather than total BBPS scores, should be used. Although the BBPS has not been validated specifically for use during colonoscopy with the water infusion technique, it should be applicable in this setting because clear fluid that does not obscure visualization of the underlying mucosa does not affect the BBPS scoring system. Similarly, if an area of mucosa is initially obscured by overlying residue that is shifted away by a change in patient position such that the underlying mucosa is now well visualized, then the BBPS score should reflect this improvement in visualization.

In summary, the BBPS is a valid and reliable instrument for assessing bowel preparation adequacy during colonoscopy regardless of the degree of cleanliness. It may be useful during colonoscopy-related research when needing to control for bowel preparation quality. Furthermore, segment scores may represent a new way to determine bowel preparation adequacy in clinical practice. The BBPS can also be disseminated easily through use of a brief, instructional, Web-based video, making it a strong candidate for a standardized tool for rating bowel preparations.

ACKNOWLEDGMENT

We acknowledge the efforts of our colleagues at BUMC who reviewed the BP-DVDs as well as the national and international clinicians who responded to our survey.

REFERENCES


