

Validation of a New Bowel Preparation Scale for Measuring Colon Cleansing for Colonoscopy: The Chicago Bowel Preparation Scale

David P. Gerard, MD¹, Diane B. Foster, BS, PA-C¹, Manfred W. Raiser, MD¹, John L. Holden, MD¹ and Theodore G. Karrison, PhD²

OBJECTIVES: Current bowel preparation scales (BPSs) have significant limitations including an inability to distinguish among bowel preparations that adequately cleanse a high percentage of colons. We assessed the reliability and validity of the new Chicago BPS and compared it with existing BPSs.

METHODS: We performed a prospective evaluation of the cleanliness of 150 colons. Inter-rater agreement was assessed using kappa and Pearson correlation coefficients. Each colon was rated by a gastroenterologist and physician's assistant using the Ottawa BPS, the Boston BPS, a dichotomous (adequate/inadequate) BPS (where adequate was defined as being able to visualize at least 95% of the mucosa), and the Chicago BPS.

RESULTS: Pearson correlation coefficients between the gastroenterologists and physician's assistant for total BPS scores were 0.79 (95% confidence interval (CI): 0.73, 0.85), 0.79 (95% CI: 0.72, 0.84), and 0.84 (95% CI: 0.79, 0.88) for the Ottawa, Boston, and Chicago BPSs, respectively. Kappa coefficients for right, middle, and distal colon segment ratings were 0.66, 0.53, and 0.49, respectively, for the Ottawa BPS; 0.64, 0.66, and 0.54, respectively, for the Boston BPS; and 0.70, 0.62, and 0.63, respectively, for the Chicago BPS. Differences between the Chicago BPS and the other BPSs were not statistically significant. The Chicago BPS exhibited the best correspondence between BPS total score and the adequate/inadequate BPS.

CONCLUSIONS: This study demonstrated the validity and reliability of the Chicago BPS. The better defined grading criteria, better designed numerical ratings scale, and better correspondence between Chicago BPS total score and the adequate/inadequate BPS make the Chicago BPS an attractive alternative to the Ottawa BPS and the Boston BPS.

Clinical and Translational Gastroenterology (2013) 4, e43; doi:10.1038/ctg.2013.16; published online 5 December 2013

Subject Category: Endoscopy

INTRODUCTION

The number of polyps and cancers detected during a colonoscopy is very dependent on the quality of the bowel preparation. During the past 40 years, trials comparing different bowel preparations have used a variety of bowel preparation scales (BPSs) to measure colon cleanliness for colonoscopy.¹ Most of these BPSs have not been shown to be valid (i.e., measure cleanliness accurately) or reliable (i.e., measure cleanliness in a reproducible manner).

Our review of the literature found only three BPSs (Table 1) that have been validated and had their reliability measured. The Aronchick BPS^{2,3} was created at the urging of the US Food and Drug Administration as a validated instrument to compare bowel preparation regimes. The Aronchick BPS rates the cleanliness of the whole colon with a single score and defines a good preparation as visualizing >90% of the mucosa. The Ottawa BPS⁴ was created to be a clearer, simpler and more objective rating scale and was claimed to be more reliable than the Aronchick BPS. The Ottawa BPS was the first validated BPS to rate three individual segments of the colon and the total volume of fluid in the whole colon. The Boston BPS^{5,6} was created to more accurately measure colon cleanliness; however, it was never compared with older BPSs. Each of these BPSs has significant limitations.

After using the Ottawa BPS, Boston BPS and an adequate/inadequate BPS to rate the cleanliness of 600 colons in a previous study,⁷ our observations led us to identify six features of an ideal BPS:

- (1) Produce a score that is reproducible from one endoscopist to another.
- (2) Be easy for the endoscopist to use.
- (3) Produce a score that is easily converted into the poor/fair/good/excellent subjective rating scale that is often used by gastroenterologists in their colonoscopy reports.
- (4) Produce a score that is easily converted into the adequate/inadequate subjective rating scale that is often used by gastroenterologists in their colonoscopy reports.
- (5) Recognize truly outstanding preparations where 100% of the mucosa is well visualized without any cleaning required.
- (6) Recognize adequate preparations (i.e., the ability to visualize the mucosa), but also measure the effort required to clean the colon of residual fecal material and liquid.

Recent studies^{7,8} have reported bowel preparation regimens that adequately cleanse the colon in 94.5–97.2% of

¹Gastroenterology Services, Downers Grove, Illinois, USA and ²Department of Health Studies, University of Chicago, Chicago, Illinois, USA Correspondence: David P. Gerard, MD, Gastroenterology Services, 3825 Highland Avenue, Suite 203, Downers Grove, Illinois 60515, USA. E-mail: davidpgerard@comcast.net

Received 19 June 2013; revised 11 September 2013; accepted 16 October 2013

Table 1 The Aronchick BPS,² the Ottawa BPS,⁴ the Boston BPS⁵ and the Chicago BPS*Aronchick BPS rating for the whole colon (individual segments not evaluated)*

- 5 = Inadequate (repeat preparation needed)
- 4 = Poor (semi-solid stool could not be suctioned and <90% of mucosa seen)
- 3 = Fair (semi-solid stool could not be suctioned, but >90% of mucosa seen)
- 2 = Good (clear liquid covering up to 25% of mucosa, but >90% of mucosa seen)
- 1 = Excellent (>95% of mucosa seen)

Ottawa BPS rating for each colon segment

- 4 = Inadequate (solid stool not cleared with washing and suctioning)
- 3 = Poor (necessary to wash and suction to obtain a reasonable view)
- 2 = Fair (necessary to suction liquid to adequately view segment)
- 1 = Good (minimal turbid fluid in segment)
- 0 = Excellent (mucosal detail clearly visible)

Ottawa BPS rating for the amount of fluid in the whole colon

- 2 = Large amount of fluid
- 1 = Moderate amount of fluid
- 0 = Small amount of fluid

Ottawa BPS total score is calculated by adding the scores of the right, transverse/descending, and sigmoid/rectum colon segments and the score for the fluid in the whole colon. The total Ottawa BPS score ranges from 14 (very poor) to 0 (excellent).

Boston BPS rating for each colon segment

- 0 = Unprepared colon segment with stool that cannot be cleared
- 1 = Portion of mucosa in segment seen after cleaning, but other areas not seen because of retained material
- 2 = Minor residual material after cleaning, but mucosa of segment generally well seen
- 3 = Entire mucosa of segment well seen after cleaning

Boston BPS total score is calculated by adding the scores of the right, transverse, and left colon segments. The total Boston BPS score ranges from 0 (very poor) to 9 (excellent).

Chicago BPS

- 0 = Unprepared colon segment with stool that cannot be cleared (>15% of the mucosa not seen)
- 5 = Portion of mucosa in segment seen after cleaning, but up to 15% of the mucosa not seen because of retained material
- 10 = Minor residual material after cleaning, but mucosa of segment generally well seen
- 11 = Entire mucosa of segment well seen after cleaning
- 12 = Entire mucosa of segment well seen without washing (suctioning of liquid allowed)

Chicago BPS rating for the amount of fluid in the whole colon

- 3 = Large amount of fluid (>300 cc)
- 2 = Moderate amount of fluid (151–300 cc)
- 1 = Minimal amount of fluid (51–150 cc)
- 0 = Little fluid (≤50 cc)

Chicago BPS total score is calculated by adding the scores of the right, transverse, and left colon segments. The total Chicago BPS score ranges from 0 (very poor) to 36 (outstanding). The score for the fluid in the whole colon is reported separately.

BPS, bowel preparation scale.

subjects. As bowel preparation regimens are developed that adequately cleansed close to 100% of colons, it becomes important to distinguish between preparations that result in colons that contain residual fecal material (requiring irrigation to remove) and liquid so that mucosal visualization is optimized and procedure time reduced.

The choice of a BPS (Table 1) can markedly affect the outcome of a bowel preparation study. For example, if a split-dose bowel preparation (resulting in a wetter, but slightly cleaner colon after suctioning the liquid) is compared with a day-prior bowel preparation (resulting in a dryer, but slightly less clean colon), the outcome would depend on the BPS chosen as the primary end point. The Ottawa BPS would favor the day-prior bowel preparation because total scores as high as 8 (low is good, range 0–14) can result from large volumes of clear liquid throughout the colon. The Boston BPS would favor the split-dose bowel preparation because there is no penalty on that BPS for residual liquid in the colon.

None of the existing validated BPSs are ideal primary end points for bowel preparations studies where close to 100% of colons are adequately cleansed. The Ottawa BPS is overly

sensitive to the effort required to remove residual fecal material and liquid from the colon resulting in a final score that does not always reflect the visualization of the mucosa. The Boston BPS is not sensitive to the effort required to cleanse the colonic mucosa. Neither the Ottawa BPS nor the Boston BPS produces final scores that are easily converted into the subjective ratings often used by gastroenterologists in their colonoscopy reports.⁷

A new BPS was created (the Chicago BPS) to address the limitations of existing BPSs. The purpose of this investigation was to validate the new Chicago BPS, assess its reproducibility, and compare it with existing BPSs.

METHODS

Study design and oversight. The protocol was approved by the Adventist Midwest Region Institutional Review Board (AMH 2012–08 1220120880, ClinicalTrials.gov NCT01635725). The trial was designed by the authors with no outside input. Adventist Hinsdale Hospital waived the fee for the Institutional Review Board. All other expenses associated with this study

were paid for by the authors. Study participants received no financial reward. Each subject signed an informed consent document.

The study was a prospective study of four BPSs: (1) the Ottawa BPS; (2) the Boston BPS; (3) the Chicago BPS; and (4) a dichotomous adequate/inadequate BPS in which adequate was defined as being able to see at least 95% of the colon mucosa after the mucosa was cleaned. We chose not to study the Aronchick BPS because its definition of a good preparation (visualizing >90% of the mucosa) was considered as setting too low of a standard for a modern bowel preparation.

Subjects were allowed to use any bowel preparation method. Subjects were on a clear liquid diet the day before the colonoscopy and told to consume nothing by mouth after midnight except for medications taken in the morning. They were instructed to take Fleet Enemas (CB Fleet, Lynchburg, VA) if their stool was not clear on the morning of the colonoscopy.

Subjects had their colonoscopy performed in our office endoscopy suite between 9 July 2012 and 15 October 2012 by one of three gastroenterologists (DPG, JLH and MWR) with 19 to 36 years of experience. A physician's assistant (DBF) with 29 years of gastrointestinal endoscopy experience and who has personally performed many flexible sigmoidoscopies was also present. The only major responsibility of the physician's assistant during the colonoscopy was to watch the procedure and grade the quality of the bowel preparation.

Subjects who were at least 18 years of age and planning to undergo an elective colonoscopy in our office endoscopy suite were eligible for inclusion in this study. Individuals were excluded if they were pregnant, had undergone a colon resection, or if the physician's assistant was not working on the day of their colonoscopy.

Before the study began, it was calculated that with 150 subjects the study would be adequately powered (80%) to test whether the Pearson correlation between raters (gastroenterologists and physician's assistant) exceeded 0.85 and whether the kappa coefficient exceeded 0.75 for the Chicago BPS, assuming true values of 0.90 and 0.87, respectively.

Our staff was trained to screen and recruit all patients who met the inclusion criteria. A total of 176 consecutive patients who underwent an elective colonoscopy in our office endoscopy suite were evaluated for inclusion in the study and 151 of these patients met the inclusion criteria and were invited to participate in the study. All but one patient agreed to participate and no subjects withdrew from the study. Of the 25 excluded patients, 22 had their colonoscopies on a day the physician's assistant was not working and 3 had undergone a colon resection.

One author (DBF) was responsible for monitoring the data for any adverse outcomes.

Creation of the Chicago BPS. The Chicago BPS (Table 1 and Supplementary Page 1 online) was developed using data from 600 colonoscopies reported in a previous study⁷ that rated each colon's cleanliness using the Ottawa BPS, Boston BPS, and the same adequate/inadequate BPS used in this study. The Chicago BPS was designed to incorporate most of the features of an ideal BPS and overcome the limitations of existing BPSs. The database did not include all

the information needed to calculate an exact Chicago BPS score, but allowed the Chicago BPS scores to be calculated for each subject within a small range. For example, a Boston BPS segment score of 3 corresponded to a Chicago BPS segment score of 11 or 12.

The Chicago BPS rates the right, transverse, and left colon segments with a score of 0 to 12 for each segment and adds the three segment scores together resulting in a total score that can range from 0 (very poor) to 36 (outstanding). The segments are rated as follows:

0 = Unprepared colon segment with stool that cannot be cleared (> 15% of the mucosa not seen).

5 = Portion of mucosa in segment seen after cleaning, but up to 15% of the mucosa not seen because of retained material.

10 = Minor residual material after cleaning, but mucosa of segment generally well seen.

11 = Entire mucosa of segment well seen after washing.

12 = Entire mucosa of segment well seen before washing (suctioning of liquid allowed).

The Chicago BPS rates the total volume of fluid in the whole colon (excluding any fluid injected into the colon) with the fluid score is reported separately from the total score. The fluid scores are as follows:

3 = Large amount of fluid (> 300 cc).

2 = Moderate amount of fluid (151–300 cc).

1 = Minimal amount of fluid (51–150 cc).

0 = Little fluid (≤ 50 cc).

The Chicago BPS features several major innovations:

- (1) The grading scale used to measure each colon segment's cleanliness is more clearly defined to reduce inter-observer variability. The Chicago BPS uses precise terms ("> 15%" and "up to 15%") to define how much of the mucosa is not visualized in dirty segments while the Boston BPS uses vague terms ("stool that cannot be cleared" and "areas not seen because of retained material") to define the cleanliness of dirty segments. The Chicago BPS scores for the fluid in the whole colon are precisely defined (from "≤ 50 cc" to "> 300 cc") while the Ottawa BPS uses vague terms (from "small" to "large") to describe the amount of fluid in the whole colon.
- (2) The scoring system for each segment's cleanliness grade is scaled so that small differences in segment cleanliness result in small differences in segment scores, whereas large differences in segment cleanliness result in larger differences in segment scores. On the Chicago BPS, colon segments with 100% of the mucosa visualized receive scores of 10, 11, or 12, whereas segments where the mucosa is not 100% visualized after cleaning maneuvers receive scores of 0 or 5. Segment scores for consecutive grades on the Ottawa BPS and Boston BPS are separated by single point no matter if the difference in cleanliness between consecutive grades is small or large.
- (3) The anatomic borders of each colon segment are defined by easily recognized endoscopic landmarks (similar to the Boston BPS). On the Chicago BPS, the right colon extends from the proximal cecum to the mid-hepatic

flexure, the transverse colon segment extends from the mid-hepatic flexure to the mid-splenic flexure, and the left colon segment extends from the mid-splenic flexure to the distal rectum. The Ottawa BPS defines the mid-colon segment as extending to the junction of the descending and sigmoid colon, which is difficult to precisely locate during a colonoscopy.

- (4) The fluid score is reported as a secondary measure of colon cleanliness rather than being incorporated into the primary BPS score (as with the Ottawa BPS) or ignored (as with the Boston BPS).
- (5) The effort required to wash the colon is incorporated into the segment scores (as with the Ottawa BPS). If only a measurement of mucosal visualization is desired (as with the Boston BPS), a modified Chicago BPS total score can be calculated by collapsing the best segment scores of 12 and 11 into a score of 11. The modified Chicago BPS and Boston BPS would then have identical definitions for the best segment grade.

The Chicago BPS scoring system was designed so that a total score of ≥ 25 (Table 2) could only be obtained if at least 95% of the colon mucosa is visualized (the same definition we used for an adequate grade on the adequate/inadequate BPS) and a total score of ≤ 24 is always obtained if $< 95\%$ of the colon mucosa is visualized. It is possible to obtain a total score of 15–22 and still have an adequate preparation. For example: (1) three segments could receive a score of 5 with $< 5\%$ of the mucosa in each segment not visualized; or (2) two segments could receive a score of 5 with $< 7.5\%$ of the mucosa in each segment not visualized and one segment could receive a score of 10–12.

Measurements. Immediately after the colonoscopy, separate questionnaires were filled out by the gastroenterologist and physician's assistant (see Supplementary Pages 2 and 3 online), which included the subject's age, sex, race, the bowel preparation used, the time the colonoscopy was scheduled to begin, the indications, the findings, and whether the cecum was reached. Both questionnaires rated the cleanliness of the colon on the Ottawa BPS, Boston BPS, Chicago BPS, and adequate/inadequate BPS. Any segment not seen because of technical difficulties advancing the colonoscope around the colon was not rated. Any segment not seen because the more distal colon was too dirty to pass received the worst rating on each scale.

The gastroenterologist and physician's assistant were not allowed to discuss the cleanliness of the colon before the questionnaires were completed, nor were they allowed to discuss the segment of colon through which the colonoscope was being passed. They were allowed to discuss the locations of the appendiceal orifice and ileocecal valve and the possible locations of the hepatic flexure and splenic flexure. They were allowed to observe the number of centimeters the colonoscope was inserted into the colon.

The total volume of fluid suctioned through the colonoscope was measured in a 600 cc Gomco suction bottle (Allied Health Products, St Louis, MO). A graduated cylinder was used to confirm that the calibration markings on the Gomco bottle were accurate to $\pm 2\%$. The suction bottle was covered so

that neither the gastroenterologist nor the physician's assistant could see the volume of fluid being suctioned into the bottle.

The fluid used to irrigate the colon during the colonoscopy was injected into the colonoscope using a 60-cc syringe and the total volume of irrigation fluid used was recorded by a nurse. The gastroenterologist and physician's assistant were permitted to observe the volume of fluid injected. The net volume of fluid suctioned from the colon was calculated by taking the total volume of fluid suctioned into the Gomco bottle and subtracting the volume of fluid used to irrigate the colon.

A few colonoscopies required a small volume of irrigation fluid to remove blood from a polypectomy site. The total volume of fluid suctioned through the colonoscope and the volume of irrigation fluid reported in this study were reduced by the volume of fluid used to remove blood from a polypectomy site.

Before the study began, the gastroenterologists and physician's assistant watched a video (media.bmc.org/Media/BostonBowelPreparationScale.wmv) as a calibration exercise for the BPSs and the results were discussed among the raters.

Unreported data from a previously published study⁷ of 600 subjects who had their bowel preparations graded on the Ottawa BPS, Boston BPS, and the same adequate/inadequate BPS used in this study are summarized in this article.

Data analysis. Mean scores between the gastroenterologists and the physician's assistant were compared using a paired two-tailed *t*-test. These comparisons addressed whether there was a *systematic* difference between the two types of raters. To assess inter-rater agreement in segmental colon and fluid scores, which were rated using three, four, and five-point scales, kappa coefficients were computed together with 95% confidence intervals (CIs). Values of 0.41–0.60 generally reflect moderate agreement, values of 0.61–0.80 substantial agreement, and values of 0.81–1.0 near perfect agreement.⁹ Inter-rater agreement between the total bowel preparation scores was evaluated using Pearson's correlation coefficient and two-way scatter plots. Fisher's *z*-transformation was applied to obtain 95% CIs. The association between fluid scores and net fluid suction was assessed similarly.

To determine whether the magnitude of the kappa coefficients and Pearson correlation coefficients were significantly different among the Chicago, Ottawa, and Boston BPSs, a bootstrap procedure was performed.¹⁰ This method accounts for the fact that ratings were obtained using each BPS on each patient, resulting in a "within-patient correlation." Bootstrap re-sampling was conducted stratified by gastroenterologist; 1,000 bootstrap replications were performed.

RESULTS

Subjects and endoscopic findings. The baseline characteristics of the 150 subjects, indications for the colonoscopies and findings are shown in Table 3.

There were no significant complications related to the colonoscopies.

Table 2 The relationship between Chicago BPS segment scores, Chicago BPS total scores, and a description of the overall colon cleanliness

| Colon segment scores | Permutations | Total scores | Description of cleanliness |
|--|--------------|--------------|---|
| Any segment 0 | 61 | 0–24 | Poor (inadequate) |
| Two or three segments 5 and no segment 0 | 10 | 15–22 | Usually Poor (inadequate), rarely fair (adequate) |
| One segment 5, other two 10–12 | 27 | 25–29 | Fair (adequate) to good |
| All segments 10–12 and total score 30–33 | 17 | 30–33 | Good to excellent |
| All segments 10–12 and total score 34–36 | 10 | 34–36 | Excellent to outstanding |

BPS, bowel preparation scale.

The Chicago BPS has five possible grades for the right, transverse, and left colon segments for total of 125 permutations (5^3) of the three segment scores, which add up to a total score of 0–36 for each colon.

Colon polyps were found in 22 of 54 (40.7%) average risk subjects undergoing a screening colonoscopy. Colon polyps were found in 22 of 96 subjects (22.9%, $P=0.035$ by Yates continuity-corrected χ^2 test vs. average risk subjects) undergoing colonoscopy because of increased risk or active symptoms.

The cecum was reached in 149 of 150 (99.3%) subjects. One colonoscopy was technically difficult and reached the hepatic flexure. In no subjects was the cecum not reached because of an inadequate preparation.

Preparation quality. The cleanliness data for each BPS are shown in Table 4 for both the gastroenterologists and physician's assistant. The gastroenterologists and physician's assistant's mean fluid scores on both the Ottawa BPS and Chicago BPS were significantly different. There were also statistically significant differences, but of a very small magnitude, between the gastroenterologists and physician's assistant's mean right colon scores using the Boston BPS, and between the Ottawa BPS and Boston BPS total scores.

Of the 146 colons that were rated adequate on the adequate/inadequate BPS and had all three segments graded, both the gastroenterologists and the physician's assistant rated 145 of 146 colons in the 25–36 range on the Chicago BPS. One colon rated as adequate by the gastroenterologists had a Chicago BPS total score of 20 (5 right, 5 transverse, and 10 left) and a different colon rated as adequate by the physician's assistant had a Chicago BPS total score of 20 (5 right, 10 transverse, and 5 left). In a previously published bowel preparation study⁷ of 600 subjects (591 with all three segments graded), 1 of 591 subjects had an adequate preparation and a Chicago BPS total score of 15–22. The three colons from this study rated as inadequate on the adequate/inadequate BPS had Chicago BPS total scores of 20 and 15, 20 and 20, and 10 and 15, respectively, by the gastroenterologists and physician's assistant.

With regard to inter-rater agreement, kappa coefficients for the Ottawa BPS segment scores ranged from 0.493 to 0.655 and those from the Boston BPS ranged from 0.545 to 0.661. Corresponding kappa statistics for the Chicago BPS were generally higher, 0.624 to 0.702, but the differences were not statistically significant. Kappa coefficients for the Ottawa BPS and Chicago BPS fluid scores were very similar. Pearson correlation coefficients were highest for the Chicago BPS but not significantly greater than those obtained for the other BPSs. Kappa coefficients are not reported for the modified

Chicago BPS because its segment scores are calculated from the Chicago BPS segment scores.

Inter-rater agreement for the BPS total scores, as assessed by the Pearson correlation coefficient, were 0.793, 0.795, and 0.843 for the Ottawa BPS, Boston BPS, and Chicago BPS, respectively. The higher degree of reliability exhibited by the Chicago BPS compared with the Ottawa BPS and Boston BPS was not statistically significant. Scatter plots are shown in Figure 1.

Correlations between fluid scores and net fluid suctioned are shown in Table 4 for all subjects and for the subset with no fluid injected. For the Chicago BPS, correlations were significantly higher ($P=0.024$) or marginally so ($P=0.085$) for the physician's assistant compared with the gastroenterologists. The Chicago BPS outperformed the Ottawa BPS under the physician's assistant ($P=0.004$), but not for the gastroenterologists. The correlation between fluid scores and net fluid suctioned was much stronger when no fluid was injected.

Additional cleanliness data comparing the results on the Ottawa BPS, Boston BPS, Chicago BPS, and adequate/inadequate BPS are shown in Table 5 from this study and the previously published study.⁷ From this study, in colons with all three segments graded, overlap between adequate and inadequate grades on the adequate/inadequate BPS and total scores was observed: for the Ottawa BPS from 8 to 10, 3/15th of the scale (6/65 colons with total scores ≥ 8 were graded as inadequate); for the Boston BPS at 4, 1/10th of the scale (6/8 colons with total scores ≤ 4); and for the Chicago BPS at 20, 1/37th of the scale (6/8 colons with total scores ≤ 24). In the previous study of 600 subjects (591 with all three segments graded) overlap was observed: for the Ottawa BPS from 3 to 10, 8/15th of the scale (27/249 colons with total scores ≥ 3 were graded as inadequate); for the Boston BPS from 4 to 7, 4/10th of the scale (27/166 colons with total scores ≤ 7); and for the Chicago BPS in 1/28 colons with total scores ≤ 24 , 1/37th of the scale.

Cleanliness data calculated from the Chicago BPS data and reported on the modified Chicago BPS are shown in Table 5. The modified Chicago BPS scores (mean \pm s.d., $n=300$) were 10.25 ± 1.70 , 10.69 ± 1.01 , 10.76 ± 1.01 , and 31.69 ± 3.05 , respectively, for the right segment, transverse segment, left segment, and total.

DISCUSSION

The Chicago BPS was created because there was a need for a primary end point for bowel preparations studies where

Table 3 Data from the questionnaires filled out by the gastroenterologists and physician's assistant immediately after the colonoscopies

| | |
|--|-------------------------------------|
| <i>N</i> | 150 |
| Sex ratio (M: F) | 90: 60 (60%: 40%) |
| MD (Gerard: Raiser: Holden) | 57: 44: 49 (38.0%: 29.3%: 32.7%) |
| Age in years (mean \pm s.d.) | 58.5 \pm 11.3 |
| <i>Race</i> | |
| Caucasian | 128 (85.3%) |
| African American | 4 (2.7%) |
| Hispanic | 5 (3.3%) |
| Other | 13 (8.7%) |
| Procedure start time (mean hour \pm s.d.) | 10.1 \pm 0.9 |
| <i>Preparation consumed</i> | |
| Gatorade + PEG-3350 | 113 (75.3%) |
| Other clear liquid + PEG-3350 | 3 (2.0%) |
| NuLYTELY or TriLyte | 34 (22.7%) |
| <i>Indications (multiple indications possible)</i> | |
| Screening | 54 (36.0%) |
| History of polyps | 45 (30.0%) |
| History of colon cancer | 1 (0.7%) |
| Family history of colon neoplasia | 24 (16.0%) |
| Rectal bleeding | 17 (11.3%) |
| Anemia | 3 (2.0%) |
| History of inflammatory bowel disease | 9 (6.0%) |
| Diarrhea | 3 (2.0%) |
| Recent diverticulitis | 4 (2.7%) |
| Abnormal gastrointestinal X-ray | 1 (0.7%) |
| Change in bowel habits | 1 (0.7%) |
| <i>Findings (multiple findings possible)</i> | |
| Normal | 44 (29.3%) |
| Colon polyps | 61 (40.7%) |
| Colon cancer | 0 |
| Diverticulosis | 44 (29.3%) |
| Crohn's disease | 4 (2.7%) |
| Ulcerative colitis | 8 (5.3%) |
| Other colitis | 1 (0.7%) |
| Incomplete colonoscopy | 1 (0.7%) |
| <i>All colonoscopies—fluid (ml)</i> | |
| Total suctioned (mean \pm s.d.) | 198 \pm 126 |
| Irrigation fluid injected (mean \pm s.d.) | 111 \pm 122 |
| Net fluid suctioned (mean \pm s.d.) | 87 \pm 84 |
| Chicago BPS fluid score (mean \pm s.d.) | 0.83 \pm 0.81 |
| <i>Colonoscopies with no fluid injected (n = 50, ml)</i> | |
| Total suctioned (mean \pm s.d.) | 121 \pm 83 |
| Chicago BPS fluid score (mean \pm s.d.) | 1.12 \pm 0.80 |

BPS, bowel preparation scale.

The Chicago BPS fluid score is calculated based on the net fluid suctioned.

close to 100% of colons are expected to be adequately cleansed. The Chicago BPS also was designed to overcome the limitations of existing BPSs and incorporate as many features of an ideal BPS as practical. Suggested instructions for using the Chicago BPS in future bowel preparation studies are included in Supplementary Page 1 online.

Older bowel preparation studies often evaluated preparations that adequately cleansed <75% of colons for colonoscopy.¹¹ For these studies, BPSs needed to be sensitive to the gross contamination of the colonic mucosa with residual stool. Subtle differences such as the effort required to irrigate the colon or suction residual liquid were much less important.

The colon preparations reported in this study adequately cleansed 98.5% of colons and other recent studies^{7,8} have reported adequate cleansing in 94.5–97.2% of colons. An ideal bowel preparation would produce a dry, clean colon where 100% of the mucosa could be visualized without any cleaning maneuvers. As preparations are developed that cleanse close to 100% of colons adequately, it becomes important to distinguish between preparations that result in colons that contain residual fecal material requiring irrigation to remove because this can prolong the procedure time and interfere with mucosal visualization. Even in colons that contain only large amounts of turbid liquid, small puddles of liquid left behind after cleaning can decrease mucosal visualization.

In this study, there was a perfect correspondence between the gastroenterologists and physician's assistant's grades on the adequate/inadequate BPS, but the adequate/inadequate BPS is a very crude measure of bowel cleanliness.

The Chicago BPS was designed so that any colon that received a total score of 25–36 had at least 95% of its mucosa visualized, which by definition is an adequate grade on the adequate/inadequate BPS. As expected, all Chicago BPS total scores of 25–36 were observed in colons that were graded adequate on the adequate/inadequate BPS. The Chicago BPS was designed so that any colon that received an inadequate grade on the adequate/inadequate BPS had a total score of 0–24 and had <95% of its mucosa visualized. As expected, all colons graded as inadequate on the adequate/inadequate BPS had Chicago BPS total scores of 0–24.

Although any cutoff between an adequate and inadequate is somewhat arbitrary, this study selected 95% mucosal visualization for a number of reasons: (1) our previous study⁷ used this cutoff and it seemed to correspond well colons cleansed adequately of solid stool; (2) other bowel preparation studies¹¹ have used cutoffs up to 95%, but never higher; (3) multiple bowel preparations in use today¹¹ are able to achieve this degree of cleansing in a high percentage of colons; and (4) a number less effective bowel preparations¹¹ do not achieve this degree of cleansing on a regular basis.

BPSs with imprecisely defined grading criteria have two important contributors to the errors of the scores: (1) the ability of the rater to make the estimate; and (2) the rater's interpretation of the BPS grading scale. The precisely defined grading criteria of the Chicago BPS removes the second source of error from the Chicago BPS scores.

A limitation in the design of the Chicago BPS is that adequate preparations can have total scores of 15–22. In this study, two of 298 total scores were in the 15–22 range on the Chicago BPS and were graded adequate on the adequate/inadequate BPS. Retrospective data from a previous study⁷ showed that one of 591 colons had a Chicago BPS total score of 15–22 and was graded adequate on the adequate/inadequate BPS. The Chicago BPS total scores come close to the perfect correspondence with the adequate/inadequate BPS grades that would be expected of an ideal BPS.

In this study, the Ottawa BPS total scores did not correlate as well with the grades on the adequate/inadequate BPS, with

Table 4 Colon cleanliness data from the questionnaire filled out by the gastroenterologists and physician's assistant immediately after the colonoscopies

| (a) Descriptive statistics | | | |
|---|----------------------------|-------------------------------------|----------------|
| | Gastroenterologists | Physician's assistant | P value |
| <i>Ottawa BPS</i> | | | |
| Right colon (mean ± s.d., n = 149) | 2.18 ± 0.96 | 2.15 ± 1.02 | 0.52 |
| TV/descending colon (mean ± s.d.) | 1.49 ± 1.10 | 1.48 ± 1.12 | 0.92 |
| Sigmoid/rectum (mean ± s.d.) | 1.34 ± 1.10 | 1.27 ± 1.20 | 0.29 |
| Fluid total colon (mean ± s.d.) | 0.51 ± 0.55 | 0.35 ± 0.52 | < 0.001 |
| Total OBPS (mean ± s.d.) | 5.52 ± 2.52 | 5.24 ± 2.43 | 0.034 |
| <i>Boston BPS</i> | | | |
| Right colon (mean ± s.d., n = 149) | 2.53 ± 0.66 | 2.64 ± 0.63 | 0.004 |
| TV colon (mean ± s.d.) | 2.79 ± 0.49 | 2.81 ± 0.46 | 0.44 |
| Left colon (mean ± s.d.) | 2.80 ± 0.42 | 2.85 ± 0.43 | 0.11 |
| Total BBPS (mean ± s.d.) | 8.12 ± 1.37 | 8.29 ± 1.36 | 0.016 |
| <i>Chicago BPS</i> | | | |
| Right colon (mean ± s.d., n = 149) | 10.59 ± 2.01 | 10.74 ± 1.87 | 0.13 |
| TV colon (mean ± s.d.) | 11.37 ± 1.30 | 11.43 ± 1.28 | 0.33 |
| Left colon (mean ± s.d.) | 11.40 ± 1.22 | 11.45 ± 1.29 | 0.43 |
| Fluid total colon (mean ± s.d.) | 0.99 ± 0.79 | 0.83 ± 0.74 | 0.008 |
| Total CBPS (mean ± s.d.) | 33.36 ± 3.84 | 33.63 ± 3.88 | 0.13 |
| Inadequate preparation (number) | 3 | 3 | 1.0 |
| (b) Inter-rater agreement | | | |
| | Kappa (95% CI) | Pearson correlation (95% CI) | |
| <i>Ottawa BPS</i> | | | |
| Right colon (n = 149) | 0.655 (0.553, 0.757) | 0.792 (0.726, 0.846) | |
| TV/descending colon | 0.531 (0.437, 0.625) | 0.756 (0.678, 0.818) | |
| Sigmoid/rectum | 0.493 (0.402, 0.583) | 0.729 (0.730, 0.797) | |
| Fluid total colon | 0.337 (0.195, 0.479) | 0.382 (0.236, 0.511) | |
| Total score | — | 0.793 (0.726, 0.846) | |
| <i>Boston BPS</i> | | | |
| Right colon (n = 149) | 0.644 (0.518, 0.769) | 0.756 (0.677, 0.818) | |
| TV colon | 0.661 (0.522, 0.800) | 0.777 (0.703, 0.834) | |
| Left colon | 0.545 (0.398, 0.692) | 0.651 (0.551, 0.735) | |
| Total score | — | 0.795 (0.721, 0.843) | |
| <i>Chicago BPS</i> | | | |
| <i>Right colon (n = 149)</i> | 0.702 (0.603, 0.801) | 0.796 (0.731, 0.848) | |
| Chicago vs. Ottawa | P = 0.43 | P = 0.96 | |
| Chicago vs. Boston | P = 0.063 | P = 0.35 | |
| <i>TV colon</i> | 0.624 (0.510, 0.737) | 0.792 (0.726, 0.848) | |
| Chicago vs. Ottawa | P = 0.23 | P = 0.71 | |
| Chicago vs. Boston | P = 0.56 | P = 0.84 | |
| <i>Left colon</i> | 0.630 (0.514, 0.745) | 0.834 (0.778, 0.876) | |
| Chicago vs. Boston | P = 0.22 | P = 0.11 | |
| <i>Fluid total colon</i> | 0.335 (0.226, 0.445) | 0.515 (0.387, 0.624) | |
| Chicago vs. Ottawa | P = 0.99 | P = 0.079 | |
| <i>Total score</i> | — | 0.843 (0.786, 0.881) | |
| Chicago vs. Ottawa | — | P = 0.40 | |
| Chicago vs. Boston | — | P = 0.15 | |
| (c) Correlation between fluid scores and net fluid suctioned | | | |
| | Gastroenterologists | Physician's assistant | P value |
| <i>1. All colonoscopies</i> | | | |
| Ottawa score vs. net fluid | 0.511 (0.380, 0.620) | 0.565 (0.442, 0.664) | 0.59 |
| Chicago score vs. net fluid | 0.507 (0.375, 0.618) | 0.700 (0.604, 0.774) | 0.024 |
| Chicago score vs. Ottawa score | P = 0.94 | P = 0.004 | |
| Chicago score vs. net fluid (collapsed) | 0.546 (0.422, 0.647) | 0.649 (0.544, 0.733) | 0.18 |
| <i>2. No fluid injected (n = 50)</i> | | | |
| Ottawa score vs. net fluid | 0.637 (0.336, 0.726) | 0.704 (0.529, 0.821) | 0.50 |
| Chicago score vs. net fluid | 0.684 (0.500, 0.808) | 0.793 (0.658, 0.876) | 0.085 |
| Chicago score vs. Ottawa score | P = 0.60 | P = 0.22 | |
| Chicago score vs. net fluid (collapsed) | 0.748 (0.592, 0.848) | 0.720 (0.551, 0.831) | 0.70 |

BPS, bowel preparation scale; BBPS, Boston BPS; CBPS, Chicago BPS; CI, confidence interval; OBPS, Ottawa BPS; TV, transverse. Net fluid (collapsed) is net fluid collapsed into four intervals: ≤ 50 ml, 51–150 ml, 151–300 ml, and > 300 ml. N = 150 unless otherwise noted. P values in right-hand column compare gastroenterologists with physician's assistant. P values in body of table compare Chicago BPS with the Ottawa BPS and Boston BPS. Bold values indicate statistical significance.

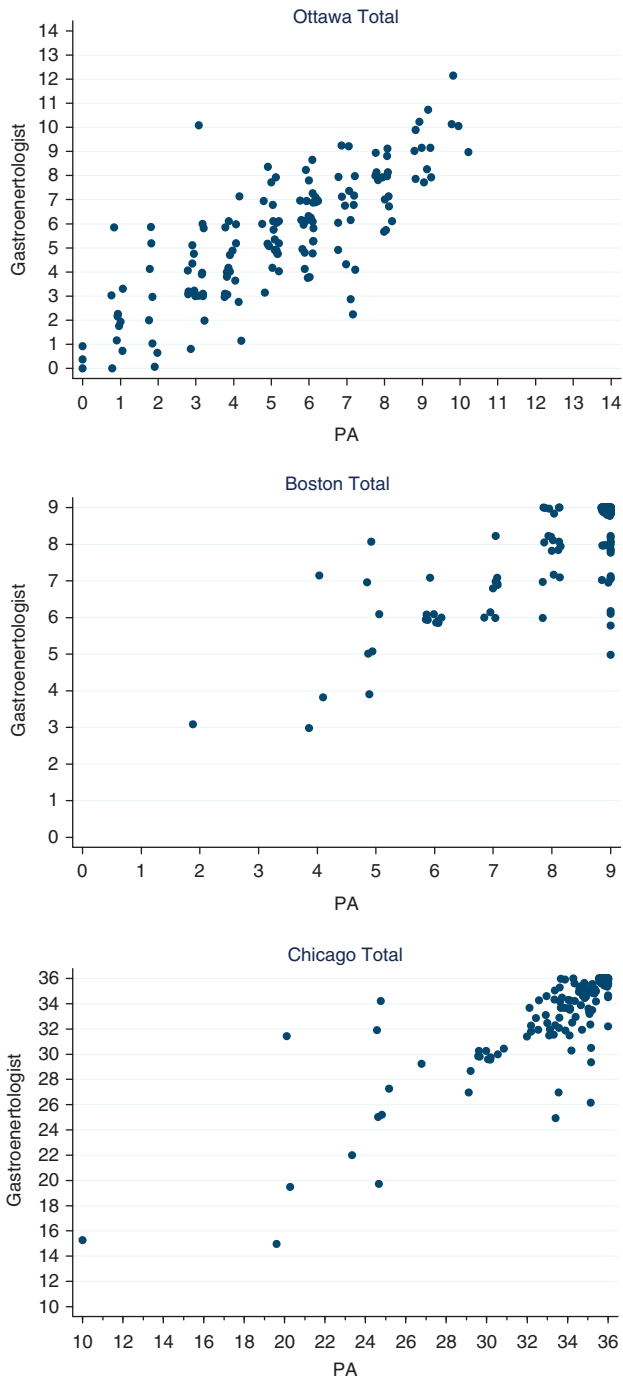


Figure 1 Scatter plots of gastroenterologist vs. physician's assistant (PA) total scores ($n=150$) for the Ottawa bowel preparation scale (BPS; top), Boston BPS (middle), and Chicago BPS (bottom). All scores are recorded as integers. Points have been "jittered" to better display multiple observations. For example, for the Boston BPS, 84 patients had scores of 9 recorded by both raters. For the Chicago BPS, 24 patients had scores of 35 recorded by both raters and 45 patients had scores of 36 by both raters.

a substantial overlap in total scores between colons that were graded adequate and inadequate. A limitation of this study was that only 3 of 150 colons were graded inadequate on the adequate/inadequate BPS. Prospective data from a previous study⁷ showed that there was substantial overlap in total

scores on both the Ottawa and Boston BPSs between colons graded as adequate and inadequate.

It is possible to have an Ottawa BPS total score as low as 3 (low is good) and have an inadequate preparation or a score as high as 11 and have adequate preparation. It is possible to have a Boston BPS total score as high as 7 (high is good) and have an inadequate preparation or a score as low as 3 and have an adequate preparation. This result is a consequence of poorly designed scoring systems in which the difference between numerical segment scores is not proportional to the difference in cleanliness. For example, on the Ottawa BPS, there is a one point difference between the numerical scores assigned to segments that are not visualized because of retained stool and segments well visualized after washing (a major difference in cleanliness) and a one point difference between the numerical scores assigned to segments with minimal turbid fluid and segments that are well visualized (a minor difference in cleanliness).

There was a moderate inverse correlation between the total score for the Ottawa BPS and the Boston BPS and a moderate inverse correlation between the total score for the Ottawa BPS and the Chicago BPS. We found a strong correlation between the total score for the Boston BPS and the Chicago BPS. These findings are not surprising as both the Boston BPS and Chicago BPS are largely measures of cleanliness after cleaning maneuvers are performed, whereas the Ottawa BPS is a measure of cleanliness before cleaning maneuvers are performed. This finding is further evidence that the Chicago BPS is at least as valid a measure of colon cleanliness as the Boston BPS. The ultimate goal of a colonoscopy preparation is to allow the colon mucosa to be visualized, which appears to be better reflected in the grading approach taken by the Chicago BPS and Boston BPS.

Another way to assess the validity of the Chicago BPS is to look at all 125 permutations of the three possible colon segment scores (5 possible scores for 3 segments or 5^3 permutations). Almost all the permutations of the three segment scores yield a total score that correctly corresponds to the poor/fair/good/excellent/outstanding description of the cleanliness, as shown in Table 2. It is possible to have a Chicago BPS of 15–22 and fall in range associated with poor (inadequate) preparations when, in fact, 95% or more of the mucosa was visualized and the preparation should have been described as fair (adequate). The only way the Chicago BPS can overcome this inconsistency would be to have the endoscopist quantify the percentage of mucosa not well visualized; however, this quantification would create a BPS that would be much more complex to use and increase the inter-observer variation.

The Chicago BPS had higher kappa coefficients of inter-rater agreement for the right and left colon segments than the Ottawa and Boston BPSs and a higher Pearson correlation coefficient for the total score, but the differences did not reach statistical significance. Of note, the observed kappa coefficients and Pearson correlation for the Chicago BPS were not as high as projected when the study was designed. The Ottawa BPS has less clearly defined grading criteria, which accounted for a small portion of the variation observed and both the Ottawa and Boston BPSs have poorly designed scoring systems in which the difference between numerical segment

Table 5 Data from 298 colon total scores from this study (149 by the gastroenterologists and 149 by the physician's assistant) and 591 colon total scores from a previous study⁷ where all colon segments were visualized

| | Ottawa BPS total scores | | | | | | | | | | | | | | |
|---|-------------------------|-----|-----|----|----|----|----|----|----|----|----|----|----|----|----|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| Number Ottawa BPS total scores (<i>n</i> = 298) | 7 | 17 | 15 | 32 | 34 | 43 | 50 | 35 | 34 | 20 | 9 | 1 | 1 | | |
| Number of inadequate scores (<i>n</i> = 6 of 298) | | | | | | | | | 2 | 1 | 1 | 1 | 1 | | |
| Number Ottawa BPS total scores (<i>n</i> = 591) | 122 | 114 | 106 | 67 | 47 | 41 | 41 | 22 | 14 | 10 | 4 | 1 | 1 | | 1 |
| Number of inadequate scores (<i>n</i> = 27 of 591) | | | | 1 | | 2 | 3 | 4 | 5 | 6 | 3 | 1 | 1 | | 1 |
| Boston BPS total score 9 (<i>n</i> = 194 of 591) | 7 | 17 | 15 | 29 | 26 | 29 | 30 | 21 | 11 | 5 | 4 | | | | |
| Boston BPS total score 8 (<i>n</i> = 39 of 591) | | | | 2 | 6 | 9 | 6 | 4 | 9 | 2 | 1 | | | | |
| Boston BPS total score 7 (<i>n</i> = 25 of 591) | | | | | 2 | 2 | 7 | 4 | 5 | 4 | 1 | | | | |

| | Boston BPS total scores | | | | | | | | | |
|---|-------------------------|-----|----|----|----|---|---|---|---|---|
| | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 |
| Number Boston BPS total scores (<i>n</i> = 298) | 194 | 39 | 25 | 23 | 9 | 5 | 2 | 1 | | |
| Number of inadequate scores (<i>n</i> = 6 of 298) | | | | | | 3 | 2 | 1 | | |
| Number Boston BPS total scores (<i>N</i> = 591) | 305 | 120 | 70 | 63 | 17 | 8 | 4 | 2 | 1 | 1 |
| Number of inadequate scores (<i>n</i> = 27 of 591) | | | 1 | 1 | 10 | 7 | 4 | 2 | 1 | 1 |

| | Chicago BPS total scores | | | | | | | | | | | | | |
|--|--------------------------|----|----|-----|----|----|----|----|----|----|----|----|----|----|
| | 36 | 35 | 34 | 33 | 32 | 31 | 30 | 29 | 27 | 26 | 25 | 20 | 15 | 10 |
| Number Chicago BPS total scores (<i>n</i> = 298) | 100 | 66 | 43 | 20 | 20 | 4 | 19 | 4 | 4 | 1 | 9 | 5 | 2 | 1 |
| Number modified Chicago BPS total scores (<i>n</i> = 298) | | | | 194 | 40 | 19 | 19 | 0 | 4 | 5 | 9 | 5 | 2 | 1 |
| Number of inadequate scores (<i>n</i> = 6 of 298) | | | | | | | | | | | 3 | 2 | 1 | |
| Boston BPS total score 9 (<i>n</i> = 194 of 298) | 100 | 66 | 24 | 4 | | | | | | | | | | |
| Boston BPS total score 8 (<i>n</i> = 39 of 298) | | | 19 | 15 | 5 | | | | | | | | | |
| Boston BPS total score 7 (<i>n</i> = 25 of 298) | | | | 1 | 15 | 4 | | 4 | | | 1 | | | |

BPS, bowel preparation scale.
The Chicago BPS vs. Boston BPS data are from this study only.

scores is not proportional to the difference in cleanliness, which accounted for much of the remaining variation observed.

The volume of residual fluid in the colon is a grading criterion used by some BPSs including the Ottawa BPS and Chicago BPS, but the accuracy of these fluid measurements has never been evaluated. This study measured the total volume of fluid suctioned from the colon, but this measurement overestimates the residual fluid in the colon when irrigation fluid is added to the colon. The net volume of fluid suctioned from the colon was calculated by subtracting the volume of fluid used to irrigate the colon from the total fluid suctioned from the colon, but this measurement underestimates the residual fluid in the colon when all the fluid used to irrigate the colon is not suctioned out during the colonoscopy.

There was a moderate-to-high correlation between the calculated Chicago BPS fluid score (from the net fluid suctioned for the colon) and the estimate of the residual fluid in the colon by both the gastroenterologists and the physician's assistant. When only the colons in which no irrigation fluid was injected are evaluated, the correlation between the net fluid and both the gastroenterologists and physician's assistant's scores is much stronger and better reflects the accuracy of the observers' fluid estimates.

The mean physician's assistant's Ottawa BPS and Chicago BPS fluid scores were significantly lower than those of the

gastroenterologists, but the magnitude of the differences was relatively small.

In this study, 65.1% of colons received the highest total score of 9 on the Boston BPS, which makes it difficult to distinguish between excellent and outstanding preparations. Only 33.6% of colons received the highest total score of 36 on the Chicago BPS (31.5% of colons received a total score of 9 on the Boston BPS and a total score of 33–35 on the Chicago BPS) allowing a meaningful distinction to be made between excellent and outstanding preparations. The modified Chicago BPS (Table 5) collapses the Chicago BPS highest segment grades of 11 or 12 into a grade of 11 which results in both the modified Chicago BPS and Boston BPS having identical definitions for the best segment grade. Not surprisingly, 65.1% of colons received the highest total score of 33 on the modified Chicago BPS.

An ideal validation study would have many gastroenterologists view the colonoscopy as it was being performed, but no previous study has ever attempted such a comprehensive validation. The validation study³ for the Aronchick BPS was only published in abstract form and used 80 videotaped colonoscopies to test four different BPS using four observers. The Ottawa BPS was validated in 99 subjects by a gastroenterologist and a fellow both of whom viewed the colonoscopies "live" as they were being performed.⁴ This

study had more subjects (150 vs. 99) than the Ottawa BPS validation study⁴ and more observers (4 vs. 2).

The Boston BPS was initially validated using three digital video disc (DVD) recordings of “truncated” colonoscopies with moderate degrees of cleanliness viewed twice by 13 gastroenterologists, 8 fellows and 1 physician’s assistant.⁵ The Boston BPS was subsequently validated using 10 DVD recordings of “truncated” colonoscopies with varying degrees of cleanliness viewed twice by nine gastroenterologists and three fellows.⁶ This study observed 150 colonoscopies as they were being performed rather than 3 to 10 DVD recordings as with the Boston BPS validation studies^{5,6} and is the first study to compare the Boston BPS to other BPSs. The Boston BPS validation studies did have more observers than this study (12 to 22 vs. 4). This study is the first to validate the Boston BPS using “live” colonoscopies, which provides a more realistic validation of the Boston BPS. Viewing the colonoscopy “live” allows the raters to assess the length of scope inserted into the patient, better assess the fluid being suctioned through the scope, and provides the opportunity to revisit areas of the colon, which may not have been well visualized.

A limitation of this study was that only four observers were used to validate the BPSs, one was a physician’s assistant who scored all patients and three were gastroenterologists who scored about a third of the patients each. Comparing the cleanliness grades of a physician’s assistant to those of gastroenterologists may be less than ideal, but there was little evidence that the physician’s assistant ability to grade colon cleanliness was different from the gastroenterologists. Inter-rater agreement for the BPS total scores, as assessed by the Pearson correlation coefficient, correlated well that that of the gastroenterologists. In addition, the mean segment scores and mean scores for total colon cleanliness were very similar between the physician’s assistant and gastroenterologists. In fact, the physician’s assistant estimates of the fluid suctioned were closer to the measured values than that of the gastroenterologists.

This study used a physician’s assistant who has witnessed thousands of colonoscopies in her 29 years of practice, far more than almost any fellow. The physician’s assistant personally had performed flexible sigmoidoscopies and dealt with the challenges of working with poorly prepared colons. As a co-author of our previous bowel preparation study,⁷ the physician’s assistant had experience dealing with BPSs. Owing to the high cost of having multiple gastroenterologists view colonoscopies over a 2- to 3-month period, it is doubtful a validation study using only gastroenterologists to view “live” colonoscopies will ever be performed. Other limitations of this study include: (1) the observers were very experienced in using BPSs making the applicability of these results to less experienced observers less certain; (2) poorly cleansed colons were rarely seen making the applicability of these data to preparations that do not cleanse the colon well less certain; and (3) the level of intra-rater agreement was not evaluated, although this would require the use of recorded colonoscopies rather than “live” viewings, which would not be equivalent.

In conclusion, this prospective study demonstrated the validity and reliability of the Chicago BPS, although it did not yield significantly higher indices of inter-rater agreement than the Ottawa BPS and the Boston BPS. The Chicago BPS is easy to use, produces a score that is easily converted into

poor/fair/good/excellent and adequate/inadequate subjective rating scales that are often used by gastroenterologists in their colonoscopy reports, is able to recognize truly outstanding preparations and measures the effort required to clean the colon of residual liquid.

The Chicago BPS has the flexibility to rate mucosal visualization (modified Chicago BPS), the effort required to wash the colon (Chicago BPS total score) and the residual fluid in the colon (Chicago BPS fluid score). With this information, it is possible to distinguish the pros and cons of various bowel preparations that adequately cleanse a high percentage of colons.

CONFLICT OF INTEREST

Guarantor of the article: David P. Gerard, MD.

Specific author contributions: David P. Gerard: all. Diane B. Foster: acquisition of data, analysis, drafting, critical revisions, and administrative. Manfred W. Raiser: concept, acquisition of data, analysis, drafting, and critical revisions. John L. Holden: concept, acquisition of data, analysis, drafting, and critical revisions. Theodore G. Karrison: power calculations, data analysis, drafting, and critical revisions.

Financial Support: None.

Potential Competing Interests: None.

Acknowledgments. We thank our office personnel for their help with this study: Kathy, Karina, Arianna, Kasia, Melissa, Tammy and Carla.

Study Highlights

WHAT IS CURRENT KNOWLEDGE

- Most bowel preparation scales (BPS) have not been shown to be valid (i.e., measure cleanliness accurately) or reliable (i.e., measure cleanliness in a reproducible manner).
- The few BPSs that have been validated have significant limitations including an inability to distinguish among bowel preparations that adequately cleanse a high percentage of colons.
- The scores produced by the validated BPSs are not easily converted into the subjective rating scale that is often used by gastroenterologists in their colonoscopy reports.

WHAT IS NEW HERE

- This study demonstrated the validity and reliability of the Chicago BPS.
- The Chicago BPS comes closer to being an ideal BPS than existing scales such as the Ottawa BPS and Boston BPS and produces a score that is easily converted into poor/fair/good/excellent and adequate/inadequate subjective rating scales that are often used by gastroenterologists in their colonoscopy reports.
- The Chicago BPS has the flexibility to rate mucosal visualization (modified Chicago BPS), the effort required to wash the colon (Chicago BPS total score) and the residual fluid in the colon (Chicago BPS fluid score). With this information, a bowel preparation study can distinguish the pros and cons of various bowel preparations that adequately cleanse a high percentage of colons.

1. ASGE. A consensus document on bowel preparation before colonoscopy: prepared by a Task Force from the ASCRS, the ASGE, and the SAGES. *Gastrointest Endosc* 2006; **63**: 894–909.
2. Aronchick CA, Lipshutz WH, Wright SH *et al.* A novel tableted purgative for colonoscopic preparation: efficacy and safety comparisons with Colyte and Fleet Phospho-Soda. *Gastrointest Endosc* 2000; **52**: 346–352.
3. Aronchick CA, Lipshutz WH, Wright SH *et al.* Validation of an instrument to assess colon cleansing (abstract). *Am J Gastroenterol* 1999; **94**: 2667.
4. Rostom A, Jolicoeur E. Validation of a new scale for the assessment of bowel preparation quality. *Gastrointest Endosc* 2004; **59**: 482–486.
5. Lai EJ, Calderwood AH, Doros G *et al.* The Boston Bowel Preparation scale: a valid and reliable instrument for colonoscopy-oriented research. *Gastrointest Endosc* 2009; **69**: 620–625.
6. Calcerwood A, Jacobson BC. Comprehensive validation of the Boston bowel preparation scale. *Gastrointest Endosc* 2010; **72**: 686–692.
7. Gerard DP, Holden JL, Foster DB *et al.* Randomized trial of Gatorade/polyethylene glycol with or without bisacodyl and NuLYTELY for colonoscopy preparation. *Clin Transl Gastroenterol* 2012; **3**: 1–9.
8. DiPalma JA, Rodriguez DO, McGowan J *et al.* A randomized clinical study evaluating the safety and efficacy of a new, reduced-volume, oral sulfate colon-cleansing preparation for colonoscopy. *Am J Gastroenterol* 2009; **104**: 2275–2284.
9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–174.
10. Efron B, Gong G. A leisurely look at the bootstrap, the jackknife, and cross-validation. *Am Stat* 1983; **37**: 36–48.
11. Belsey J, Epstein O, Heresbach D. Systematic review: oral bowel preparation for colonoscopy. *Aliment Pharmacol Ther* 2007; **25**: 373–384.



Clinical and Translational Gastroenterology is an open-access journal published by Nature Publishing Group. This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>

Supplementary Information accompanies this paper on the Clinical and Translational Gastroenterology website (<http://www.nature.com/ctg>)