Journal of the Canadian Association of Gastroenterology, 2019, 2(2), 57–62
doi: 10.1093/jcag/gwy027
Original Article
Advance Access publication 29 June 2018

OXFORD

Original Article

Comparison of the Boston Bowel Preparation Scale with an Auditable Application of the US Multi-Society Task Force Guidelines

Valérie Heron MD[1], Myriam Martel MS[1], Talat Bessissow MD CM, MS, FRCPC[1], Yen-I Chen MD[2], Etienne Désilets MD, FRCPC, MD, FRCPC[3], Catherine Dube MD, MS, FRCPC[4], Yidan Lu MD[1], Charles Menard MD[3], Julia McNabb-Baltar MD, MPH[5], Robin Parmar MD[1], Alaa Rostom MD, MS, FRCPC[4], Alan N Barkun MD, CM, FRCPC, MS[1,6]

[1]Division of Gastroenterology, The McGill University Health Center, Montreal General Hospital, McGill University, Montréal, Québec, Canada; [2]Division of Gastroenterology, Johns Hopkins University Hospital, Baltimore, Maryland, USA; [3]Division of Gastroenterology, University of Sherbrooke, Sherbrooke, Québec, Canada; [4]Division of Gastroenterology, University of Ottawa, Ottawa, Ontario, Canada; [5]Division of Gastroenterology, Hepatology and Endoscopy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA; [6]Division of Clinical Epidemiology, The McGill University Health Center, Montreal General Hospital, McGill University, Montréal, Canada

Correspondence: Dr. Alan Barkun, Montreal General Hospital, 1650 Cedar Avenue, #D7-148 Montreal, Quebec H3G 1A4, e-mail alan.barkun@muhc.mcgill.ca.

Abstract

Background: Existing bowel preparation scales (BPS) only modestly predict interval to next colonoscopy. The US Multi-Society Task Force (MSTF) recommends repeating colonoscopies within the year if the preparation does not allow detection of polyps over 5 mm.
Aim: This study aims to assess reliability and validity of an auditable application of the MSTF compared with the Boston BPS (BBPS).
Methods: We compared an auditable application of MSTF guidelines termed the Montreal BPS (MBPS) with the BBPS using a total cut-off score ≥6 with each segment score ≥2 (BBPS2-6). In sensitivity analyses, we applied the MBPS using a cut-off of 3 mm rather than 5 mm and also assessed the BBPS using an adequacy threshold of total score ≥5 (BBPS5). Videos of 83 colonoscopies (eight for intra-rater agreements) were independently evaluated by nine physicians. Weighted kappas quantified intra- and inter-rater agreements. Associations between scores and clinical outcomes were assessed.
Results: The BBPS2-6 and 5 mm MBPS showed moderate to substantial intra-rater agreements (κ=0.44 to 0.63 and κ=0.50 to 0.53, respectively); inter-rater agreements were only fair to moderate and slight to moderate (κ=0.25 to 0.48 and κ=0.19 to 0.50, respectively). Similar results were noted using alternate thresholds of BBPS5 and 3 mm MBPS. No significant associations were found between scores and clinical outcomes.
Conclusion: For all scales, intra-rater kappas were superior to inter-rater values, the latter reflecting at best moderate agreement. This modest performance may reflect the dichotomized interpretation of the scales (adequate versus inadequate), differing from previous publications assessing scores as continuous variables. Further studies are required to optimally interpret bowel preparation scales with regard to interval to next colonoscopy.

Keywords: *bowel preparation; validation; score.*

Colonoscopy is an integral part of colorectal cancer (CRC) screening programs. Although its potential for reducing morbidity and mortality associated to CRC is great (1), the quality and effectiveness of colonoscopy fluctuates according to several factors (2). Bowel preparation is one such factor. Indeed, bowel cleanliness is crucial to ensure optimal visualization of colonic mucosa and to allow detection and removal of polyps. Inadequate bowel preparation has been associated with missed lesions and longer procedural times (3, 4). Several scales have been developed to evaluate bowel preparations (5), and some societies have recommended that quality of bowel preparation be recorded as part of colonoscopy reports (6). However, no single bowel preparation scale has been adopted by current guidelines.

The impact of bowel preparation on colonoscopy has prompted the US Multi-Society Task Force (MSTF) to recommend that screening or surveillance colonoscopies be repeated within the year if the preparation does not allow for detection of lesions greater than 5 mm (7). Based on this recommendation, we have developed the Montreal Bowel Preparation Scale (MBPS). This simple scoring system aims to guide clinicians in determining appropriate follow-up based on bowel cleanliness and can easily be used as an auditing tool both for personal practice and in organized screening programs. In the present study, we assess reliability and validity of the MBPS and compare it to the Boston Bowel Preparation Scale (BBPS) (8–11).

METHODS

Study Design and Population

Video clip recordings of complete colonoscopies were prospectively collected between July and August 2014. Patients above the age of 18 presenting for a colonoscopy were consecutively recruited in a single-referral centre—at one of the sites of an academic hospital where over 12,000 colonoscopies are performed every year. Colonoscopies were performed by seven of the 15 endoscopists on staff. Patients with previous segmental colectomy were excluded from the study. Patients were also excluded if video recordings of all colonic segments were not obtained (right, transverse, and left that included descending, sigmoid and rectum).

Nine raters participated in the scoring of these video clips: seven staff gastroenterologists and two senior gastroenterology fellows.

Data Collection

Video recordings were collected during colonoscopies by two trained physicians using EndoWorks software (Olympus Corporation of the Americas, Center Valley facility, PA, USA). These two physicians did not participate in rating the preparations. Videos included clips of each of the three colonic segments (right, transverse, descending/sigmoid/rectum) during withdrawal after optimal washing had been achieved. An effort was made to ensure that clips were representative of the entire colonoscopy. Varying degrees of bowel cleanliness were deliberately included to ensure that the breadth of possible scores was represented.

At the time of colonoscopy, patient demographic information was collected including age, gender, indication for the procedure and type of preparation used. For each colonoscopy video, procedural information such as endoscopic findings, polyp detection and recommended interval to next colonoscopy by the performing endoscopist was also recorded. The number of adenomas detected was collected from pathology reports.

Montreal BPS

The US Multi-Society Task Force (MSTF) recommends early interval follow-up to the next colonoscopy if the preparation does not allow for the detection of polyps greater than 5 mm in size (7). A similar recommendation is made in the Canadian Association of Gastroenterology quality guidelines (12). We attempted to operationalize these recommendations by developing the Montreal Bowel Preparation Scale (MBPS): a simple instrument to be used in a clinical setting which may offer guidance to endoscopists in determining the appropriate follow-up based on bowel cleanliness. Furthermore, this scale can easily be transposed into an effective quality assurance audit tool. Like the BBPS and unlike other scales, the focus is not on evaluation of the effectiveness of preparation products but rather on clinical outcomes. Therefore, it is intended for use after optimal washing has been achieved.

The MBPS offers a global description of the preparation of the entire colon with only three outcomes: adequate, inadequate and unable to complete. A score of one (adequate preparation) is attributed if the preparation quality is judged adequate to detect lesions ≥5 mm in size after insertion and withdrawal. In this case, the patient may undergo regular scheduled follow-up colonoscopy according to current guidelines. If a preparation is considered inadequate to detect lesions ≥5 mm in size even after optimal washing, it receives a score of zero (inadequate preparation). The interval to next colonoscopy should in this case be shortened due to the inadequacy of the preparation. Finally, a score 'U' (unable to complete) describes the inability to complete colonoscopy to the proximal-most site of the existing colon due to factors unrelated to the preparation.

As a sensitivity analysis, we explored setting a more stringent definition of adequacy, applying the above score but rather using a cut-off of ≥3 mm in size. We have termed these two scales the 5 mm MBPS and the 3 mm MBPS.

Given that the MBPS provides a global assessment of the entire preparation, an adequate score can only be achieved if all segments meet the given definition.

Boston BPS

The BBPS was originally developed as a descriptive scale of three individual segments (right, transverse and left) with a total cleanliness score from zero to nine, nine being the cleanest. However, no clear cut-off for adequacy was predetermined. Validation studies of the BBPS have subsequently aimed to establish a threshold of cleanliness to better guide clinical management. Originally, a BBPS score ≥5 was proposed as an adequate preparation (9, 10). More recent studies have suggested that a total BBPS score ≥6 with each segment score ≥2 be retained as the definition of adequacy for the BBPS (13, 14).

Therefore, based on these previous validation studies, we assessed the BBPS using two possible thresholds of adequacy for the statistical analysis: total BBPS ≥5 or total BBPS ≥6 with each segment score ≥2. In this text, we refer to these two scores as the BBPS 5 and BBPS 2–6.

Calibration

In order to standardize agreement, participating raters underwent a calibration exercise using six colonoscopy videos of varying levels of cleanliness. Each participant rated these videos based on a detailed written description of the BBPS, the 3 mm MBPS and the 5 mm MBPS (descriptions available upon request). A calibration meeting was subsequently held during which a consensus was reached for the scoring of each of these videos. Seven of the nine raters attended this calibration meeting and were instructed to use this consensus as a point of reference during subsequent scoring. The other two raters were purposefully not informed of the outcome of the calibration meeting to provide an estimate of its impact.

A calibration image illustrating a standard 3 mm and 5 mm measurement was provided to all raters. Raters were instructed to interpret the videos as being recorded after optimal washing had been accomplished by the endoscopist.

Reliability Assessment

Inter-rater agreement was assessed based on nine raters' evaluations of 83 colonoscopy videos using the BBPS, 5 mm MBPS and the 3 mm MBPS. Eight colonoscopy videos were evaluated a second time to allow for assessment of intra-rater agreement one year apart. Video labels were changed so raters would be blinded to which videos were being used for intra-rater agreement and thus avoid possible bias. We further assessed intra-rater and inter-rater agreements among the subgroup of five staff gastroenterologists (senior group) as compared with agreements between two senior gastroenterology fellows who were grouped with two gastroenterologists in their first year on staff (junior group). We also assessed the possible impact of calibration on reliability testing by comparing the subgroup of seven raters who underwent calibration (calibration group) with the two raters who did not (noncalibration group).

Validity Assessment

Face validity of the MBPS was assessed using a standardized survey completed by five gastroenterologists. Following a modified Delphi process, a conference call was held to discuss the results of this questionnaire, and a final formulation of the scale was agreed upon. Construct validity was assessed by evaluating the association between MBPS or BBPS scores and the following clinical outcomes: polyp detection rate, adenoma detection rate, recommended interval to repeat colonoscopy, and withdrawal time. Raters were blinded to clinical outcomes.

Statistical Analysis

In primary analysis, both intra-rater and inter-rater agreements for the BBPS 5, BBPS 2–6, 5 mm MBPS and 3 mm MBPS were quantified using Kappa scores for nominal values with 95% CI following the Landis-Koch benchmarks (15). The strength of agreement of the Kappa values was characterized as follows: <0 poor; 0 to 0.20 slight; 0.21 to 0.40 fair; 0.41 to 0.60 moderate; 0.61 to 0.80 substantial; 0.81 to 1.00 almost perfect. In an exploratory analysis, interclass correlation coefficient (ICC) was calculated for the BBPS scores as continuous variables with a two-way random average measure and reported with 95% confidence intervals (CI) (16). The ICC coefficient was characterized as follows: values below 0.4 represent poor reliability; values above 0.75 represent excellent reliability; and values between 0.4 and 0.75 represent fair to good reliability (17).

If we expect kappa agreement of 60% between junior raters and senior, the estimated sample size would be of 69 patients for a relative error of 20% (18, 19). Associations between scores cut-off and clinical outcomes were assessed for independent samples with a Chi-square test (or Fisher exact test) for categorical variables. A two-sided P value threshold of 0.05 was adopted for statistical significance. All statistical analyses were performed using SAS version 9.4 (SAS Institute Inc., Cary, NC).

Ethical Considerations

This study was approved by the ethics committee of the McGill University Health Centre, and written informed consent was obtained from each patient prior to video recording.

RESULTS

Patient Demographic and Endoscopic Data

Eighty-three colonoscopy videos were included in the study, each containing individual clips of the three colonic segments. The average duration of the total video clip footage for each colonoscopy was 119.7 ± 4.4 seconds. The average age in the patient population was 64.4 ± 12.4 years. Men and women were evenly represented (49.4% women).

The majority of colonoscopies were performed for screening or surveillance (72.3%). Other indications included anemia (6.0%), rectal bleeding after age 40 (6.0%) and suspicion of active inflammatory bowel disease (3.6%). (A full list of indications is available upon request.) Sodium picosulfate was administered in 67.5% of patients before their colonoscopy, while the other 32.5% received polyethylene glycol electrolyte solution. Forty-eight percent of patients also received magnesium citrate as an adjuvant (in the sodium picosulfate group). A recommendation to repeat the colonoscopy because of inadequate preparation was made in 4.8% of cases. The average withdrawal time was 7 minutes and 38 seconds. Adenoma detection rate in this specific patient population was 22.9%. Polyp detection rate was 30.1%.

Reliability Testing

The proportion of colonoscopy videos receiving a BBPS score ≥5 (BBPS 5) by individual raters ranged between 65.1% and 89.2% in the senior group and between 25.3% and 94.0% in the junior group. Using a threshold of total BBPS ≥6 and each segment score ≥2 (BBPS 2–6), the proportion of colonoscopy videos receiving an adequate BBPS score ranged between 34.9% and 74.7% in the senior group and between 9.6% and 68.7% in the junior group. Adequate 5 mm MBPS scores were attributed to 36.1% to 91.6% of videos in the senior group and to 10.8% to 62.7% in the junior group. Adequate 3 mm MBPS scores were given to 28.9% to 55.4% and 9.6% to 38.6% in the senior and junior groups, respectively.

Total inter-rater and inter-rater agreements among subgroups of calibrated, senior and junior raters were assessed for the BBPS 5, BBPS 2–6, 5 mm MBPS and 3 mm MBPS (Table 1). Inter-rater agreements ranged between <0 and 0.35 for the BBPS 5, 0.25 and 0.48 for the BBPS2-6, 0.19 and 0.50 for the 5 mm MBPS, and 0.39 and 0.52 for the 3 mm MBPS. In a sensitivity analysis, inter-rater agreement was assessed excluding one rater who had consistently rated preparations lower than other raters (Appendix 1).

Similarly, intra-rater agreement was assessed for these same subgroups (Table 2). Kappa scores ranged between 0.58 and 0.74 for the BBPS 5 and between 0.44 and 0.63 for the BBPS 2-6. Intra-rater agreement ranged between 0.50 and 0.53 and between 0.73 and 0.78 for the 5 mm and 3 mm MBPS, respectively.

Secondary outcomes of inter- and intra-rater agreements for the BBPS employed as a continuous scale are shown in Table 3.

Validity Testing

Face validity

A survey revealed that 60% of participants believe the most important aim of a bowel preparation scale should be to provide information on recommended follow-up, while 40% expect a scale to principally evaluate bowel cleanliness. All participants (100%) agreed that "inadequacy to detect lesions" should result in a shortened colonoscopy follow-up interval due to preparation. All participants agreed that a cut-off of 5 mm for detection of lesions was appropriate for the MBPS, while 60% felt that a cut-off of 3 mm should also be explored. One hundred percent of participants agreed the scale should be applied after optimal washing had been achieved, and 60% felt the preparation should be rated upon combination of insertion and withdrawal, while the other 40% felt it should be rated upon withdrawal alone. The final formulation of the MBPS was approved by all survey participants.

Construct validity

No statistically significant correlations were identified between any of the predetermined BBPS or MBPS cut-offs and the following clinical outcomes: polyp detection rates (Appendix 2), adenoma detection rate (Appendix 3), recommendation for repeat colonoscopy (Appendix 4) and withdrawal time (Appendix 5).

DISCUSSION

We assessed the reliability and validity of a new bowel preparation scale—the MBPS—as well as that of the existing BBPS (8–11), comparing two previously proposed thresholds of adequacy: BBPS score ≥5 (BBPS 5) (9, 10) and total BBPS score ≥6 with each segment score ≥2 (BBPS 2–6) (13, 14).Though previous publications assessed the validity of the BBPS using these dichotomized adequacy thresholds to determine association with clinical outcomes, reliability of the BBPS has always been assessed as a continuous scale using ICCs. Our study is the first to assess its performance as a dichotomized scale based on proposed adequacy thresholds using weighted kappas in order to better reflect expected interpretation in a clinical setting.

Table 1. Inter-rater agreement

	BBPS 5 (κ, 95% CI)	BBPS 2-6 (κ, 95% CI)	5 mm MBPS (κ, 95% CI)	3 mm MBPS (κ, 95% CI)
All raters	0.22 (0.18; 0.25)	0.39 (0.35; 0.43)	0.31 (0.28; 0.35)	0.44 (0.40; 0.47)
Calibration group	0.22 (0.18; 0.27)	0.38 (0.34; 0.43)	0.33 (0.29; 0.38)	0.42 (0.37; 0.47)
Senior group	0.35 (0.28; 0.42)	0.48 (0.41; 0.55)	0.19 (0.12; 0.26)	0.39 (0.32; 0.46)
Junior group	<0	0.25 (0.16; 0.34)	0.50 (0.41; 0.58)	0.52 (0.43; 0.61)

Table 2. Intra-rater agreement

	BBPS 5 (κ, 95% CI)	BBPS 2-6 (κ, 95% CI)	5 mm MBPS (κ, 95% CI)	3 mm MBPS (κ, 95% CI)
All raters	0.68 (0.55; 0.88)	0.55 (0.53; 0.90)	0.52 (0.59; 0.83)	0.76 (0.45; 0.74)
Calibration group	0.58 (0.45; 0.71)	0.44 (0.30; 0.59)	0.50 (0.35; 0.65)	0.73 (0.55; 0.91)
Senior group	0.65 (0.34; 0.99)	0.50 (0.21; 0.72)	0.52 (0.22; 0.74)	0.78 (0.50; 1.00)
Junior group	0.74 (0.40; 1.00)	0.63 (0.15; 1.00)	0.53 (0.15; 0.92)	0.73 (0.57; 0.89)

Table 3. Inter- and intra-rater agreement for BBPS analyzed as a continuous score using intra-class correlation coefficient

	Inter-rater agreement (ICC, 95% CI)	Intra-rater agreement (ICC, 95% CI)
All raters	0.94 (0.91; 0.95)	0.79 (0.41; 0.66)
Calibration group	0.79 (0.68; 0.86)	0.73 (0.59; 0.87)
Senior group	0.91 (0.88; 0.94)	0.76 (0.59; 0.87)
Junior group	0.89 (0.85;0.92)	0.82 (0.46; 1.00)

The BBPS 5 and BBPS 2–6 both showed moderate to substantial intra-rater agreement. However, inter-rater agreement was only poor to fair and fair to moderate for BBPS 5 and BBPS 2–6, respectively. Similarly, intra-rater agreements were moderate for the 5 mm MBPS and substantial for the 3 mm MBPS. Inter-rater agreement was slight to moderate for the 5 mm MBPS and fair to moderate for the 3 mm MBPS. Overall, with regard to reliability, the 3 mm MBPS performed slightly better than its counterparts, with better intra-rater than inter-rater agreement.

A potential limitation to agreement is the subjective nature of the bowel preparation scales studied. Indeed, appreciation of a 5 mm lesion may vary among raters. Likewise, the BBPS is dependent on raters' interpretation of subjective visual descriptions such as "*minor* amount of residual staining" and "*portion* of mucosa of the colon segment seen" (9).

Another underlying factor which may have limited agreement is the adoption of a dichotomized definition of adequacy geared toward clinical outcomes. Indeed, when inter- and intra-rater agreement among these nine raters was assessed for the BBPS using ICCs for continuous variables, the BBPS performed much better in keeping with findings from previous studies. This finding may be explained by the fact that when an arbitrary threshold is chosen, though the absolute difference between scores may be small, these may fall on either side of the cut-off into opposing categories—adequate and inadequate—leading to a narrower variance of scoring for continuous scales compared with binomial scales. Furthermore, the superior agreement observed with the 3 mm MBPS may be related to the sternness of the scale, causing more unified agreement with a much larger proportion of inadequate scores because most preparations do not meet this rigid standard.

In comparing senior raters with junior raters, one group was not consistently associated with better agreement than the other, suggesting that both scales may be used by physicians with varying levels of experience. Interestingly, calibration did not seem to improve agreement among raters. Indeed, both intra-rater and inter-rater agreements of the calibrated group were comparable to or slightly lower than overall agreement among all raters. Given the small size of the noncalibrated group, further studies may be warranted to confirm this finding, especially with regard to its generalizability.

Though our data suggest possible trends toward positive clinical outcomes with adequate BBPS and MBPS scores, associations between scores and clinical outcomes lacked statistical significance.

These findings may be related in part to the low adenoma detection rate among colonoscopies included in this study. We believe the low adenoma detection rate is due to patient factors, as a recently published quality improvement project including the same group of endoscopists reported polyp detection rates of 43.8% overall and 45.6% for screening (20). Patient-related factors potentially contributing to this include the indications for colonoscopies because only 72.3% were performed for screening or surveillance, with some having already undergone recent colonoscopy. Furthermore, certain risk factors such as smoking, age and male gender have recently been identified as risk factors for both adenomas and suboptimal bowel preparation (21). Therefore, the presence of these patient characteristics may decrease adenoma detection.

Strengths of our study include validation of bowel preparation scales with a clear adequacy threshold adapted to guide clinical decisions. Furthermore, we investigated the effect of both calibration and clinical experience on reliability of the MBPS and BBPS. We benefited from a large number of colonoscopies of varying degrees of cleanliness and a group of raters with different levels of expertise, thus maximizing generalizability.

The use of representative video clips, though not as complete as using full colonoscopies, is certainly superior to the use of pictures alone. Furthermore, this allowed raters to be blinded to

colonoscopy findings, such as polyp detection, which may have introduced a bias when rating preparations.

The MBPS poses an advantage compared with other scales because it provides a global assessment of the entire preparation after washing rather than individual segments. This allows for easy interpretation of the score even in patients having undergone segmental colonic resections. Due to its dichotomous nature, it also prevents preparations that are inadequate in only one segment from receiving an overall adequate score, which is the case with other scales, as this may lead to missed polyps, particularly in the right colon (22). The BBPS has also addressed this issue with the addition of a secondary criterion for adequacy of each segment score ≥2 (13, 14). However, the MBPS may be easier to use for auditing purposes because it aims to directly reflect a single score rather than a total score conditional on minimal segmental scores. It is important to note that this may not correlate with clinical outcomes as suggested in recent meta-analysis, showing that intermediate preparations may allow for similar clinical outcomes to clean preparations (23).

CONCLUSION

Overall, the MBPS and BBPS performed similarly in reliability assessment, with intra-rater agreement being superior to inter-rater agreement. The dichotomized definitions adopted for each scale may account for the lower agreements than expected. Further validation studies of existing ordinal bowel preparation scales should aim to assess reliability in a categorical manner based on proposed adequacy thresholds that aim to determine optimal interval to next colonoscopy.

Supplementary data

Supplementary data are available at *Journal of the Canadian Association of Gastroenterology* Online.

References

1. Winawer SJ, Zauber AG, Ho MN, et al. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. N Engl J Med 1993;329:1977–81.
2. Hewett DG, Rex DK. The big picture: Does colonoscopy work? Gastrointest Endosc Clin N Am 2015;25:403–13.
3. Harewood GC, Sharma VK, de Garmo P. Impact of colonoscopy preparation quality on detection of suspected colonic neoplasia. Gastrointest Endosc 2003;58:76–9.
4. Froehlich F, Wietlisbach V, Gonvers JJ, et al. Impact of colonic cleansing on quality and diagnostic yield of colonoscopy: The European Panel of Appropriateness of Gastrointestinal Endoscopy European multicenter study. Gastrointest Endosc 2005;61:378–84.
5. Parmar R, Martel M, Rostom A, et al. Validated scales for colon cleansing: A systematic review. Am J Gastroenterol 2016;111:197–204; quiz 5.
6. Hassan C, Bretthauer M, Kaminski MF, et al. Bowel preparation for colonoscopy: European Society of Gastrointestinal Endoscopy (ESGE) guideline. Endoscopy 2013;45:142–50.
7. Johnson DA, Barkun AN, Cohen LB, et al. Optimizing adequacy of bowel cleansing for colonoscopy: Recommendations from the US multi-society task force on colorectal cancer. Gastroenterology 2014;147:903–24.
8. Calderwood AH, Jacobson BC. Comprehensive validation of the Boston Bowel Preparation Scale. Gastrointest Endosc 2010;72:686–92.
9. 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;69:620–5.
10. Gao Y, Lin JS, Zhang HD, et al. Pilot validation of the Boston Bowel Preparation Scale in China. Dig Endosc 2013;25:167–73.
11. Schindler AE, Chan WW, Laborde CJ, et al. Reliability of the Boston Bowel Preparation Scale in the endoscopy nurse population. Clin Gastroenterol Hepatol 2016;14:775–6.
12. Leddin D, Enns R, Hilsden R, et al. Colorectal cancer surveillance after index colonoscopy: Guidance from the Canadian Association of Gastroenterology. Can J Gastroenterol 2013;27:224–8.
13. Calderwood AH, Schroy PC, 3rd, Lieberman DA, et al. Boston Bowel Preparation Scale scores provide a standardized definition of adequate for describing bowel cleanliness. Gastrointest Endosc 2014;80:269–76.
14. Clark BT, Protiva P, Nagar A, et al. Quantification of adequate bowel preparation for screening or surveillance colonoscopy in men. Gastroenterology 2016;150:396–405; quiz e14–5.
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
16. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 1979;86:420–8.
17. Fleiss J. Design and Analysis of Clinical Experiments, chapter 1. New York, NY, USA: Wiley Classic Library, 1999.
18. Gwet KL. Handbook of Inter-Rater Reliability, 2nd ed. 2010.
19. Cantor AB. Sample size calculations for the log rank test: A Gompertz model approach. J Clin Epidemiol 1992;45:1131–6.
20. Kherad O, Restellini S, Martel M, et al. Polyethylene glycol versus sodium picosulfalte bowel preparation in the setting of a colorectal cancer screening program. Can J Gastroenterol Hepatol 2015;29:384–90.
21. Wong MC, Ching JY, Chan VC, et al. Determinants of bowel preparation quality and its association with adenoma detection: A prospective colonoscopy study. Medicine 2016;95:e2251.
22. East JE, Vieth M, Rex DK. Serrated lesions in colorectal cancer screening: Detection, resection, pathology and surveillance. Gut 2015;64:991–1000.
23. Clark BT, Rustagi T, Laine L. What level of bowel prep quality requires early repeat colonoscopy: Systematic review and meta-analysis of the impact of preparation quality on adenoma detection rate. Am J Gastroenterol 2014;109:1714–23; quiz 24.