A validation study of 3 grading systems to evaluate small-bowel cleansing for wireless capsule endoscopy: a quantitative index, a qualitative evaluation, and an overall adequacy assessment

Corey Brotz, MD, Neilanjan Nandi, MD, Mitchell Conn, MD, Constantine Daskalakis, ScD, Michael DiMarino, MD, Anthony Infantolino, MD, Leo C. Katz, MD, Theodore Schroeder, BS, David Kastenberg, MD
Philadelphia, Pennsylvania, USA

Background: Capsule endoscopy (CE) is a powerful tool for evaluating the small bowel. Assessment of small-bowel cleansing for CE is an essential quality measure.

Objective: Our purpose was to validate 3 new scales that grade small-bowel cleansing for CE.

Design: Prospective, randomized, single-center study.

Setting: Tertiary university hospital.

Intervention: Five experienced capsule endoscopists read 40 CEs twice, separated by 1 month, to grade small-bowel cleansing on 3 scales—quantitative index (QI; 0-10), qualitative evaluation (QE; poor, fair, good, excellent), and overall adequacy assessment (OAA; inadequate, adequate). The QI and QE evaluated both the entire and distal small bowel. Investigators received no prior training in these scales.

Main Outcome Measurements: Intraclass correlation coefficients to assess intraobserver (test-retest) and interobserver reliability.

Patients: Forty patients who underwent 1 CE between June 2005 and May 2006 and who satisfied entry criteria.

Results: Intraobserver reliability was moderate to substantial for the QI (0.60-0.66), moderate for the OAA (0.56), and fair to moderate for the QE (0.37-0.47). Interobserver scores were lower: QI and OAA moderate (0.47-0.52, 0.41, respectively) and slight to fair for the QE (0.20-0.24). QI scores for the entire and distal small bowel were highly correlated for each reader (0.57-0.87), and distal small-bowel scores were lower by 1.3 points, indicating poorer cleansing ($P < .001$). A dichotomized QE of excellent/good versus fair/poor had moderate to substantial intraobserver and interobserver reliability (0.58-0.66, 0.41-0.49, respectively). There was a strong and highly significant association among all 3 scales ($P < .001$ between QI and both QE and OAA).

Conclusion: We have described and validated 3 scales for grading small-bowel cleansing for CE. An evaluation of small-bowel cleansing should be routinely incorporated into the CE report. (Gastrointest Endosc 2009;69:262-70.)

Abbreviations: CE, capsule endoscopy; DQE, dichotomized qualitative evaluation; ICC, intraclass correlation coefficient; OAA, overall adequacy assessment; OR, odds ratio; QE, qualitative evaluation; QI, quantitative index; TIMI, thrombolysis in myocardial infarction; TJUH, Thomas Jefferson University Hospital.

DISCLOSURE: The following authors disclosed financial relationships relevant to this publication: M. Conn, M. DiMarino: Speaker and consultant for Given Imaging, Ltd. A. Infantolino: Speakers bureau for Given Imaging, Ltd. All other authors disclosed no financial relationships relevant to this publication.

Copyright © 2009 by the American Society for Gastrointestinal Endoscopy

0016-5107/$36.00

Wireless capsule endoscopy (CE) has revolutionized imaging of the small bowel, specifically with regard to occult GI bleeding, Crohn’s disease, and neoplasms. More than 500,000 small-bowel CEs have been performed to date, and utilization of this technology grew 22% worldwide in the last year alone. Even greater expansion of this technology has been limited by issues of cost and reimbursement, physician training, time requirements for study interpretation, and the absence of therapeutic capability to date. With respect to cost, the need for repeat studies because of poor visualization of the small bowel may occur in as many as one third of studies. Thus,
optimal small-bowel cleansing is critical for the evolution of CE.

The colon purgative literature best demonstrates that inadequate cleansing affects quality and cost. For instance, poor preparation interferes with the diagnosis of colon neoplasia. Furthermore, the need for repeat colonoscopy because of inadequate visualization increases cost. Cleansing is of even greater importance for CE because there is no possibility of flushing or suctioning. The current standard preparation for CE is a liquid diet after lunch until 10 PM on the day before CE with fasting thereafter. However, there is evidence that this protocol is inadequate. A variety of factors have been evaluated thereafter. However, there is evidence that this protocol is inadequate. A variety of factors have been evaluated in an attempt to improve small-bowel visualization, including patient position, prokinetic agents, simethicone, and use of the established colon purgatives polyethylene glycol or sodium phosphate. To date, these investigations have relied on unvalidated grading scales.

As with colonoscopy, a CE report should include an assessment of small-bowel cleansing. This quality measure is necessary to judge the reliability of the findings. Reproducible assessment of small-bowel cleansing is also important for the future development of small-bowel purgatives. Toward these ends, we have developed objective (quantitative index) and subjective (qualitative evaluation, overall adequacy assessment) grading systems to assess small-bowel cleansing for CE. The purpose of this study was to validate these scoring systems.

METHODS

A chart review identified 40 outpatients who had undergone CE at Thomas Jefferson University Hospital (TJUH) between June 2005 and May 2006 and who met entry criteria. These criteria included (1) age ≥ 18 years, (2) capsule reached the cecum, (3) no prior esophageal, gastric or small-bowel surgery, and (4) follow-up at TJUH available. There were 541 studies retrospectively screened to find 40 CEs that met entry criteria. Most screening failures were due to an absence of clinical follow-up at our institution because many of the patients had been referred exclusively for CE. The preparation before CE at TJUH during the time these 40 CEs were performed consisted of a clear liquid diet after lunch and fasting after 10 PM on the day before the study. This study was approved by the TJUH Institutional Review Board.

Of the 40 CEs included in the study, 39 were originally read in clinical practice by one of the study’s 5 participating capsule endoscopists. The 5 study clinicians (M. C., A. I., D. K., M. D., L. K.) contributed 13, 12, 7, 4, and 3 readings, respectively. The remaining study was originally read by a nonparticipating capsule endoscopist. A little more than 4 months elapsed between the last clinical reading of a CE and the first investigational reading performed for the purpose of validating small-bowel grading scales.

In this validation study, these 40 CE videos were deidentified and findings were removed except for landmarks. The videos were then randomized (T1) and prospectively read by 5 experienced capsule endoscopists who evaluated the quality of small-bowel cleansing by use of the 3 scoring systems described below. One month later, the 40 videos were rerandomized (T2) and again evaluated for small-bowel cleansing quality by the same 5 endoscopists using the same 3 scoring systems. Readers were blinded to patient history, demographics, and findings of the originally read CE. The average length of time for each capsule endoscopist to complete the readings during T1 and T2 was 4.3 days, with a median length of time of 4 days and a range from 2 to 8 days. Only one reader was given access to a CE study at any given time. The Given PillCam M2A capsule was used and images were viewed using Given Rapid 4 software, version 4.0 in the QuadView at 30 frames/second (Given Imaging Ltd, Yoqneam, Israel).

With readers blinded to the results of the T1 and T2 readings for small-bowel cleansing, an assessment of the relationship between the indication for CE and CE findings was performed. Findings were categorized as “none,” unrelated to the clinical indication (“unrelated”), or related to the clinical indication (“related”).

Grading scales

Each CE was evaluated for small-bowel cleansing by 3 scoring systems: a quantitative index (QI), a qualitative evaluation (QE), and an overall adequacy assessment (OAA). On a 3-point scale (0 = severe impairment, 1 = moderate impairment, 2 = minimal impairment), the QI assessed 5 elements: (1) mucosal visualization, (2)
 fluid and debris, (3) bubbles, (4) bile/chyme staining, and (5) brightness (Table 1, Appendix 1, available online at www.giejournal.org). The QI of the small-bowel preparation was obtained by summing the scores of these 5 elements (total score = 0 to 10, with higher scores indicating better cleansing). The QE of the small-bowel preparation was scored as poor, fair, good, or excellent. Each of these 4 grades was accompanied by a description to which the evaluator could refer (Table 1, Appendix 2, available online at www.giejournal.org). For both the QI and the QE, the reader provided separate scores for the entire and distal half of the small bowel. The distal small bowel was defined as any point beyond the reader’s estimate of the midpoint of the small bowel. The final scoring system, the OAA, graded small-bowel cleansing as “adequate” or “inadequate” for CE (Table 1, Appendix 2).

### Statistical analysis

The intraclass correlation coefficient (ICC) was used to assess intrareader (test-retest) and interreader reliability of the QI. The unweighted $k$ statistic was used to assess the reliability of the QE and the OAA. For interpretation of intrareader and interreader reliability values, we used previously established $k$ interpretations by Landis and Koch. To account for the correlated nature of the multiple readings of the same CE (multiple readers and time-points), CIs and $P$ values for the reliability measures were computed by the bootstrap. For the same reason, analyses of the association between the different scoring systems were based on the generalized estimating equations approach with the robust (“sandwich”) variance estimator. Statistical analyses were carried out in SAS 9.1 (SAS Institute, Cary, NC) and Stata 8.2 (StataCorp, College Station, Tex).
RESULTS

The characteristics of the 40 study patients are summarized in Table 2. Patients could have more than one indication for CE, and the most common indications were GI bleeding, abdominal pain, diarrhea, and anemia. Most patients (75%) were taking at least 1 medication at the time of the CE, with acid suppression and antihypertensive medications most common. All but 1 patient had at least one prior endoscopic (colonoscopy, EGD) or radiologic study (small bowel radiograph, US, CT) before CE.

The 40 CEs were read independently by 5 clinicians on 2 separate occasions. Two ratings (both on the second reading, T2) were inadvertently not recorded, thus leading to a total of 398 ratings for these analyses. Across all 5 readers and their 2 readings, the QI had a mean of 8.3 with an SD of 2.1. Distal small-bowel scores were consistently lower than were entire small-bowel scores by an average of 1.3 points ($P < .001$, Fig. 1). Comparing QI scores between the 2 readings, scores from the first reading (T1) were lower than those from the second reading (T2) by about 0.3 points for the entire small bowel and by about 0.6 points for the distal small bowel ($P = .013$ and $P = .001$, respectively). QI scores for the entire and distal small bowel were highly correlated with Spearman correlation coefficients ranging from 0.57 to 0.87 across the 2 time points for each of the 5 physician readers. Regarding the QE of the entire small bowel, 26% of the ratings were classified as excellent, 44% as good, 20% as fair, and 10% as poor. Ratings were poorer for QE of the distal small bowel, with a corresponding breakdown of 15%, 37%, 23%, and 25%. Finally, 72% of the OAA (entire small bowel) ratings were adequate.

Reliability of scoring systems

The intrareader and interreader reliability of the scoring systems is summarized in Table 3. Intrareader reliability appeared higher than interreader reliability. Also, for scales scoring both the entire and distal small bowel (QI and QE), reliability was consistently better for the distal small bowel.

**Intrareader (test-retest) reliability.** The QI reliability was in the moderate-to-substantial range, with an estimated overall intrareader ICC of 0.60 for the entire small bowel and 0.66 for the distal small bowel. The reliability of the QE was in the fair-to-moderate range, with an overall intrareader $\kappa$ of 0.37 for the entire small bowel and 0.47 for the distal small bowel. Reliability improved to the moderate-to-substantial range when the QE was dichotomized (DQE) as excellent/good versus fair/poor. For the DQE, the overall intrareader $\kappa$ for the entire small bowel was 0.58 and for the distal small bowel 0.66. Finally, the overall intrareader $\kappa$ for the OAA was in the moderate range at 0.56.

**Interreader reliability.** Across all 5 readers, the interreader reliability was lower than the intrareader reliability for all 3 scoring systems. For the QI, interreader reliability was moderate. The overall ICC was estimated as 0.47 for the entire small bowel and 0.52 for the distal small bowel. The 4-category QE (excellent, good, fair, poor) had slight-to-fair inter-reader reliability with an overall estimated $\kappa$ of 0.20 for the entire small bowel and 0.24 for the distal small bowel. Reliability improved to the moderate range for the DQE (excellent/good versus fair/poor), with an overall $\kappa$ of 0.41 for the entire small bowel and 0.49 for the distal small bowel. Finally, the OAA was found to have moderate interreader reliability with an overall estimated $\kappa$ of 0.41.

![Figure 1. Mean and 95% CI of the quantitative indices of the quality of the wireless CE prep for the entire and the distal small bowel, by reader and time point.](image-url)
Relationship between scoring systems

There was a strong and highly significant association among all 3 scoring systems (QI, QE, OAA).

**QI and QE.** The QI was significantly associated with the QE for both the entire and distal small bowel (for both, \( P = .001 \)). This association remained significant in multivariable analysis that controlled for age, sex, weight, and the most common indications and medications as listed in Table 2. The distribution of QI scores for each QE category is presented in Fig. 2. For the entire small bowel, the mean QI score was 4.1 for poor, 6.6 for fair, 9.1 for good, and 9.9 for excellent. Looking at the relationship between the QI and DQE, each 1-point increase in the QI was associated with an almost 10-fold increase in the odds of an excellent/good score (odds ratio [OR] = 9.52, \( P = .001 \)). For the distal small bowel, there was also a significant association between the QI and the DQE (OR = 4.44, \( P = .001 \)).

**QI and OAA.** For the entire small bowel, inadequate cleansing had a mean QI of 5.7 and adequate cleansing had a mean QI of 9.3. For the distal small bowel, the mean QIs for inadequate and adequate cleansing were 3.5 and 8.4, respectively. Thus, the QI for both the entire and distal small bowel was strongly associated with the odds of adequate cleansing (OR = 5.59 and 4.49, respectively, \( P = .001 \) for both). When the QI measures for both the entire and distal small bowel were considered together in the same model, they were both significantly associated with the OAA (OR = 3.27 and 3.26, respectively, \( P = 0.001 \) for both). These results remained significant in multivariable analyses. The QI’s predictive ability with respect to the OAA seemed best with a cut point of 8 or above for the entire small bowel (sensitivity = 95%, specificity = 84%) and 6 or more for the distal small bowel (sensitivity = 94%, specificity = 89%).

**DQE and OAA.** Defining good/excellent as adequate, and fair/poor as inadequate, there was very close agreement between the DQE of the entire small bowel and the OAA, with a sensitivity of 94% and a specificity of 93%. Of the 110 instances when the OAA was adequate, the QE was fair or poor in 102 and good in the remaining 8. Of the 288 instances when the OAA was adequate, the QE was fair in 18 and good or excellent in the remaining 270. For the distal small bowel, the DQE had a sensitivity of 71% and specificity of 100% with respect to the OAA assessment.

Relationship between findings and preparation quality. With respect to the primary indication for CE, we found 30% (12/40) of CEs had no finding, 50% (20/40) had unrelated findings, and 20% (8/40) had related findings. With use of all available data (all physicians and readings), the QI of the entire small bowel was not significantly associated with the presence or absence of a CE finding. The absence of an association was demonstrated by adjusted analyses that controlled for the patient’s age, sex, weight, use of antihypertensive medication, and prior diagnostic studies (adjusted OR = 0.95, \( P = .650 \)).

**DISCUSSION**

The reliability of data is dependent on the instrument used for measurement. To date, there has been no validated tool to assess small-bowel cleansing for CE. To meet this need, we developed 3 different grading scales because the demands of such a grading scale vary depending on whether it is applied to clinical practice or research.

The QI is a quantitative system that evaluates 5 different aspects of small-bowel cleansing. This permits finer differentiation between various purgatives and regimens used for small-bowel cleansing. The QE is based on the QI but is a simpler classification of excellent, good, fair, and poor, similar to a widely used convention for colon cleansing. Because excellent and good have come to be a surrogate for adequate colon cleansing and fair and poor, we also evaluated the validity of a DQE, excellent and good versus fair and poor.

The designation of “adequate” and “inadequate” have also been widely adopted for describing colonoscopy preparation. The OAA is a global assessment that uses these descriptors in evaluating the small-bowel preparation for CE.

Our study assessed both intrareader and interreader reliability of these 3 scoring systems. For all readers, the second reading of the CE yielded a slightly higher score. There was a surprisingly high number of poor small-bowel preparations as measured by each of the 3 grading scales, and we suspect that over time readers became more tolerant of preparation deficiencies. This self-calibration may have raised scores during the second reading. All analyses except for the OAA were performed for both the entire and distal half of the small bowel. The importance of providing this additional information on cleansing quality was based on the observation that visualization of the distal small bowel is typically inferior to that of the proximal small bowel. Our findings confirmed significantly inferior cleansing of the distal small bowel. Reliability \( k \) scores for the less-well-cleansed distal small bowel were superior to those of the entire small bowel. This reflects the fact that differentiation between good and bad preparations is more reliable when there are bigger variations in quality (such as those that exist for the distal small bowel).

For each scoring system, intrareader (test-retest) reliability exceeded interreader reliability. The QI had the greatest reliability, while the QE performed more poorly although its reliability improved when dichotomized as excellent and good versus fair and poor. Finally, reliability for the OAA was in the moderate range.

Validating a measuring tool designed for clinical use is important and often overlooked. Even within the
literature for an established procedure such as colonoscopy, only a small fraction of studies evaluating purgatives use validated instruments for assessing colon cleansing. In fact, we are aware of only 2 such systems for measuring colon cleansing, the Aronchick and Ottawa scales.21,23

A brief look at the Aronchick and Ottawa scales is relevant to our study. The Aronchick scale uses the clinical descriptors of excellent, good, fair, poor, and inadequate. Prospective validation of this system yielded interreader \( \kappa \) values of 0.76 (substantial) for the “cecum,” 0.41 (moderate) for the “right colon,” and 0.31 (fair) for “distal to hepatic flexure.” Similar to our findings, agreement was strongest in the part of the large intestine where cleansing is typically inferior. The validation trial of the Ottawa scale began by training readers through the use of calibration exercises and group discussions between investigators “to ensure a high level of agreement for the scale items.”23 Once this thorough training was completed, formal reliability testing of this scale resulted in \( \kappa \) values in the almost perfect range for the right colon (0.92), the mid colon (0.88), and the rectosigmoid (0.89), with an overall reliability of 0.94. These investigators then applied the Aronchick grading system to the colon segments pertinent to the Ottawa scale and found a substantial overall reliability for the Aronchick scale (0.77).23

Although prevalidation training can substantially improve reliability of measurement, we chose a more pragmatic approach and did not train readers for our study. Because most physicians in clinical practice would likely not receive formal training with such instruments, we felt our protocol would produce a truer representation of each grading system’s performance. However, investigators were provided with descriptions of each grading system (Appendix 1 and 2) with each CE reviewed. Similar to the superior performance demonstrated with prestudy investigator training in the Rostom study,23 we suspect our scoring systems would have had even higher reliability if readers had been subjected to calibration exercises before the study. In support of this hypothesis, among the 5 readers in our study, the investigator who developed the scoring systems indeed had the highest intrareader reliability. Furthermore, we compared the QI interreader reliability for the first 20 CEs versus that for the final 20 CEs and found that reliability for both the entire and distal small bowel was significantly higher and within the substantial range for the later cases (for both entire and distal small bowel, \( P < .001 \)).

There was strong and highly significant agreement among the scoring systems. A numerically higher QI score predicted superior cleansing as judged by the QE, with each 1-point increase in the QI associated with an almost 10-fold increase in the odds of having excellent or good cleansing by the QE. Similarly, each 1-point increase in the QI (both entire and distal small bowel) was associated with an approximately 5-fold increase in the odds of having adequate small-bowel cleansing as assessed by the OAA. Finally, a QE of excellent or good was also strongly associated with an adequate evaluation with the OAA.

We considered the preparation of clear liquid diet after lunch and fasting after 10 PM on the day before CE “standard” for several reasons. To begin, the majority of studies published in peer-review journals that evaluate agents aimed at improving small-bowel visualization for CE use this (or a similar protocol) as the comparator.11,12,24,25 In addition, there is no mention of using a purgative or any other means of improving small-bowel cleansing before CE besides fasting at the capsule manufacturer’s (Given Imaging and Olympus Corporation) Web sites.26,27 Third, the American Society for Gastrointestinal Endoscopy status evaluation report on wireless CE only makes mention of fasting and does not comment on use of a purgative.28 Finally, our informal discussions with other gastroenterologists at national and international meetings focused on
CE lead us to believe that many, if not most, clinicians currently follow a regimen similar to the “standard” preparation as described and are not using a small-bowel purgative. Subsequent to this study and other reports that demonstrate suboptimal bowel preparation with the “standard” regimen, the primary investigator of this study (D. K.) has begun using a polyethylene glycol–based purgative before CE.

The study has certain limitations that have to be addressed. First, as a practical matter, scoring for all systems was conducted concomitantly after each CE was read (QI first, QE second, OAA third). It is possible that the scoring of one system affected another. In turn, this could affect the degree of agreement among the 3 systems. This procedure was adopted for practical purposes to limit the number of times (80 for our study vs 240 if each scale were evaluated separately) a reader would have to read each CE.

Second, we note that our use of a simplified dichotomized QE (DQE, excellent and good versus fair and poor) can be questioned. We feel that this simplification of the original 4-level QE has clinical implications and usefulness analogous to colonoscopy, where fair or poor cleansing often requires an immediate or early repeat procedure because of the risk of a missed lesion. Nevertheless, the DQE did not appear to perform substantially better than the global OAA. Thus, when a simple “adequate” versus “inadequate” distinction is desired, the very simple qualitative OAA might be as good as the use of more specific information that leads to the QE.

Additionally, one may debate the process by which readers determined what constituted the distal half of the small bowel as well as issues related to the variable rate of capsule transit. Although timers and other tools have been used in an attempt to bring precision to this issue, we again chose a practical and easily applicable method that depended on the experienced reader’s estimate of the midpoint of the small bowel. We recognize that variable rates of capsule transit could affect the interpretation of small-bowel cleansing quality. For instance, rapid transit through clean areas and delayed transit through poorly cleansed areas could potentially result in low scores that underestimate the true overall cleansing adequacy. Readers did take such transit issues into consideration, but this is one potential explanation for lower reliability across observers compared with with-in-observer reliability. In an attempt to minimize this effect, all CEs were read at the same speed and using the same view.

Furthermore, it is important to note that this study does not establish, nor was it powered to establish, a relationship between the quality of small-bowel cleansing and the diagnostic yield of CE. In fact, when we looked at the relationship between preparation and CE findings, we found no significant association. The validation portion of our study used a total of 400 readings (40 studies read by 5 investigators on 2 separate occasions) for preparation quality alone, but the capsule findings were based on a retrospective chart review of 40 readings. The relationship between preparation quality and CE findings will need to be answered by future and much larger studies.

Our reliability estimates are similar to those of other scales used within, and beyond, the field of gastroenterology. Corazza et al assessed reproducibility of the Marsh-Oberhuber (“Marsh”) classification system for duodenjejunal mucosal lesions for celiac disease compared with a new simplified histologic classification system. The Marsh system yielded a mean κ value of 0.35, indicating fair reproducibility, whereas the newly proposed classification system yielded a mean κ value of 0.55, indicating moderate reproducibility. Another study validated the Los Angeles endoscopic classification system for esophagitis by using 46 endoscopists from around the world to assess 22 patients with variable ranges of esophagitis. With a mean κ of 0.40 indicating fair interobserver agreement, this system has become accepted as the classification of choice among gastroenterologists. Similarly, within the field of interventional cardiology, assessment of reperfusion after thrombolysis in myocardial infarction (TIMI) relies on the TIMI flow grade scale. This scale depends on subjective interpretation of angiographic images that possess interreader variability. The corrected TIMI flow grading system has an overall κ value of 0.59 for all TIMI grade flow scores. Our proposed systems (QI, QE, and OAA) have comparable, if not higher, reliability than these methods, and their moderate to substantial intrarater and interreader agreement supports their incorporation into practice. Furthermore, evidence supports improved reliability of at least the QI with increasing clinical experience with it.

How should these scales be used in clinical practice? We have essentially developed 2 qualitative scales (OAA, QE) and 1 quantitative scale (QI). One problem with qualitative scales is that those who use them may not adequately familiarize themselves with their original descriptions. Even if these descriptions are well studied, there is still subjective interpretation that varies from clinician to clinician. This is reflected in the lower interreader κ scores for these scales. The advantage of such scales is rooted in their ease of use and familiarity related to their common application for colonoscopy—most clinicians feel comfortable differentiating adequate from inadequate or classifying a preparation as excellent, good, fair, or poor. Because there is very strong agreement between the dichotomized QE (excellent and good vs fair and poor) and the OAA, the use of either of these qualitative scales is acceptable and the choice may be guided by physician preference. Unlike colonoscopy, we now offer a quantitative scale (QI) to support the clinician’s qualitative impression of the small-bowel preparation. One approach to evaluating the small-bowel preparation might include a subjective evaluation (either OAA or QE) in the report and accompanying...
this with supportive data provided by a QI score. For ease of use, the QI could easily be incorporated into software reporting programs for the currently available capsule manufacturers. In addition, such a program could automatically provide the clinician with statistical information regarding the probability of an adequate preparation on the basis of the QI score of the entire, distal, or both entire and distal small bowel.

Similar to colonoscopy, adequacy of small-bowel cleansing affects both quality and cost. The quality of the preparation is relevant to capsule interpretation and the necessity for repeat CE. Ultimately, in clinical practice, one aim of a scale is to aid in a go/no-go decision. In the case of CE, one such important decision is whether the preparation should be redone. When an obvious solution to a clinical problem is solved by CE, then preparation quality becomes a lesser issue. However, when CE fails to reveal a definitive diagnosis, then preparation quality becomes an important consideration in determining the next step. Because the 3 scales (OAA, DQE, QI) presented strongly correlate with one another, we recommend the clinician choose a single standard for deciding whether a CE should be repeated because of suboptimal small-bowel preparation. This may be defined as a preparation that is inadequate (OAA), fair or poor (DQE), or has a QI <8 for the entire small bowel.

Although we have not tracked the frequency within our practice with which CE has been repeated because of inadequate small-bowel cleansing since concluding this study, we strongly believe that development and validation of these scales have affected our clinical practice. Although commentary regarding preparation quality was inconsistent in the past, it is now standard practice to include this in the capsule report. Furthermore, although repeating a capsule study due to preparation quality was very infrequent in the past (1 of 40 studies in the group that we studied), we now repeat the CE when the findings are inconclusive and the preparation was inadequate.

In conclusion, we have developed and validated 3 scoring systems that evaluate small-bowel cleansing for CE. Potential applications include clinical research to optimize small-bowel cleansing and incorporation into the standard CE report as an important quality measure.

REFERENCES


Received October 23, 2007. Accepted April 12, 2008.
Current affiliations: Department of Gastroenterology (C. B., M. C., M. D., A. I., L. K.), Department of Internal Medicine (N. N.), and Department of Biostatistics (C. D.), Thomas Jefferson University Hospital, and Thomas Jefferson Medical College (T. S.), Philadelphia, Pennsylvania, USA.
Reprint requests: David Kastenberg, MD, Main Building, Suite 480, 132 S 10th St, Philadelphia, PA 19107.
If you want to chat with an author of this article, you may contact him at david.kastenberg@jefferson.edu.
### APPENDIX 1. Quantitative small-bowel scoring sheet

<table>
<thead>
<tr>
<th>Investigator Name: __________________</th>
<th>Randomization Number: ______________</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date: ______________</td>
<td>Round:  □ First □ Second</td>
</tr>
</tbody>
</table>

1. Percent of mucosa visualized

**Entire Small Bowel**
- □ ≥ 90%
- □ 80%-89%
- □ < 80%

**Distal Half of Small Bowel**
- □ ≥ 90%
- □ 80%-89%
- □ < 80%

2. Fluid and debris

**Entire small bowel**
- □ Minimal to mild
- □ Moderate
- □ Severe

**Distal half of small bowel**
- □ Minimal to mild
- □ Moderate
- □ Severe

3. Bubbles

**Entire small bowel**
- □ Minimal to mild
- □ Moderate
- □ Severe

**Distal half of small bowel**
- □ Minimal to mild
- □ Moderate
- □ Severe

4. Bile/chyme staining

**Entire small bowel**
- □ Minimal to mild
- □ Moderate
- □ Severe

**Distal half of small bowel**
- □ Minimal to mild
- □ Moderate
- □ Severe

5. Brightness

**Entire small bowel**
- □ Minimal to mild reduction
- □ Moderately reduced
- □ Severely reduced

**Distal half of small bowel**
- □ Minimal to mild reduction
- □ Moderately reduced
- □ Severely reduced

---

### APPENDIX 2. Qualitative small-bowel scoring sheet

<table>
<thead>
<tr>
<th>Investigator Name: __________________</th>
<th>Randomization Number: ______________</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date: ______________</td>
<td>Round:  □ First □ Second</td>
</tr>
</tbody>
</table>

**Definition of terms**

Excellent: Visualization of ≥ 90% of mucosa; no, or minimal, fluid and debris, bubbles, and bile/chyme staining; no, or minimal, reduction of brightness.

Good: Visualization of ≥ 90% of mucosa; mild fluid and debris, bubbles, and bile/chyme staining; mildly reduced brightness.

Fair: Visualization of < 90% of mucosa; moderate fluid and debris, bubbles, and bile/chyme staining; moderately reduced brightness.

Poor: Visualization of < 80% of mucosa; excessive fluid and debris, bubbles, and bile/chyme staining; severely reduced brightness.

6. Choose the category that best describes the small bowel preparation:

**Entire small bowel**
- □ Excellent
- □ Good
- □ Fair
- □ Poor

**Distal half of small bowel**
- □ Excellent
- □ Good
- □ Fair
- □ Poor

7. The overall quality of the small bowel prep was:

- □ Adequate
- □ Not adequate