# Multicenter Validation of a Shortened Gastric-Emptying Protocol

Matthieu Pelletier-Galarneau[1], Oyebola O. Sogbein[1], Xuan Pham[1], Jason Hao[2], Jenna Le[3], Mark D. Strober[4], Michael L. Middleton[4], Janusz Kikut[2], Leonard M. Freeman[3], and Lionel S. Zuckier[1]

[1]The Ottawa Hospital, Division of Nuclear Medicine, Ottawa, Ontario, Canada; [2]University of Vermont Medical Center, Burlington, Vermont; [3]Montefiore Medical Center, Bronx, New York; and [4]Baylor Scott & White Health, Central Texas, Temple, Texas

Gastric emptying scintigraphy is considered to be the gold standard for detection of gastroparesis and other disorders of gastric motility; Society of Nuclear Medicine and Molecular Imaging guidelines are predicated on imaging over a period of 4 h, which is inconvenient for patients. Bonta et al. introduced 2-h criteria, which served to shorten the protocol in most patients, with negligible loss of accuracy. We have evaluated the Bonta criteria in a larger multicenter trial encompassing 4 academic institutions. **Methods:** Retrospective data from 4 academic medical centers were aggregated; 431 patients were included, 105 (24.4%) of whom demonstrated delayed gastric emptying defined by 4-h gastric retention of more than 10%. Bonta criteria (retention > 65% is considered abnormal and < 45% normal; otherwise, proceed to complete examination) were applied to the 2-h data. Sensitivity, specificity, accuracy, and resource use for the Bonta method were calculated. Results based on standard 4-h solid gastric emptying, performed according to current Society of Nuclear Medicine and Molecular Imaging guidelines, served as the gold standard. **Results:** Retention of 10% or less was achieved by 6, 77, 215, and 326 patients at 1, 2, 3, and 4 h, respectively. At 2 h, 261 of 431 patients (60.6%) had gastric retention of less than 45%, which according to Bonta would be classified as normal; 62 (14.4%) had gastric retention of more than 65%, which would be classified as delayed emptying; and 108 (25.1%) had intermediate values requiring further imaging through 4 h. The Bonta criteria yielded a sensitivity, specificity, and accuracy of 92.4%, 96.9%, and 95.8%, respectively, superior to any single cutoff point applied to the 2-h values. The criteria resulted in false-negative results in 8 (1.9%) patients, 6 of whom were borderline-positive at 4 h (gastric retention of 11%–14%). Using the Bonta criteria, 74.9% of studies would be terminated by 2 h, decreasing total camera use by 15.7%, from 1,768 to 1,490 images, and the average study duration would be reduced by 20.6%, from 3.1 to 2.5 h. **Conclusion:** In a multicenter cohort, use of the Bonta criteria shortened the duration of studies in most patients, resulting in an effective compromise between reduced resource use, improved patient convenience, and preserved accuracy.

**Key Words:** gastric emptying; Tougas protocol; Bonta criteria

Functional gastrointestinal disorders represent the most common complaint of patients referred to gastroenterologists and affect approximately 1 in 5 adults in the United States, especially young females (*1*). Most cases of gastroparesis are idiopathic, but delayed gastric emptying has been associated with diabetes, gastrointestinal tract surgery, Parkinson disease, collagen vascular disease, and intestinal pseudoobstruction (*2*). Gastric emptying scintigraphy (GES) is considered the gold standard for evaluation of patients with suspected delayed gastric emptying (*3*). In 2000, Tougas et al. published a 4-h protocol using a solid radioactive meal with a large normal database (*4*). Consensus recommendations by the American Neurogastroenterology and Motility Society proposed that the Tougas protocol become the standard gastric emptying protocol (*3*). This protocol is now incorporated into the Society of Nuclear Medicine and Molecular Imaging guidelines for GES (*5*). Because of its length, the standard protocol can be inconvenient for patients and may be difficult to integrate into the clinical workflow. To address this issue, Bonta et al. retrospectively analyzed 174 patients who underwent GES after ingestion of a standardized meal to determine whether 2-h gastric retention values could predict abnormal or normal results. They found that 2-h values of more than 65% and less than 45%, to indicate delayed and normal gastric retention, respectively, were predictive of the 4-h gold-standard results, with an overall accuracy of 96.6% (*6*). In a separate validation cohort of 158 patients, they obtained an accuracy of 94.9%. In both cohorts, use of the new criteria would have shortened the GES study to 2 h in approximately 80% of cases.

To further validate the Bonta criteria, we have applied the criteria to a larger group of patients aggregated from 4 North American medical centers, thereby evaluating whether the 4-h protocol can be shortened without a significant impact on the test accuracy.

## MATERIALS AND METHODS

### Patient Population

The population of this multicenter study included 431 patients, with approximately 25% derived from each of 4 sites (Table 1): The Ottawa Hospital, Ottawa, Ontario, Canada; Montefiore Medical Center, Bronx, New York; University of Vermont Hospital, Burlington, Vermont; and Baylor Scott & White Health, Central Texas, Temple, Texas. All patients were imaged according to the Society of Nuclear Medicine and Molecular Imaging protocol (*5*). Briefly, patients ingested a meal consisting of 120 mL (4 oz) of liquid egg whites admixed with 18.5–37 MBq (0.5–1 mCi) of $^{99m}$Tc-sulfur colloid, 2 slices of toasted white bread, 30 g of jam, and 120 mL of water. Anterior and posterior planar images centered on the upper abdomen

## TABLE 1
### Patient Characteristics for Each Center

| Center | n | Women (n) | Delayed 4-h emptying (n) |
|---|---|---|---|
| Montefiore Medical Center | 130 | 89 (68.5%) | 34 (26.2%) |
| Baylor Scott & White Health | 100 | 74 (74.0%) | 23 (23.0%) |
| Ottawa Hospital | 101 | 77 (76.2%) | 17 (16.8%) |
| University of Vermont Hospital | 100 | 66 (66.0%) | 31 (31.0%) |
| Total | 431 | 306 (71.0%) | 105 (24.4%) |

were acquired, and gastric activity was calculated using geometric mean (square root of the product of counts in the anterior and posterior regions of interest). Imaging was performed hourly over 4 h, unless gastric retention was equal to or less than 10%, whereupon imaging was terminated. Only patients who consumed the whole standardized meal were included in the study. This retrospective study was approved by The Ottawa Hospital Research Ethic Board as well as Institutional Review Boards at each medical center, and the requirement to obtain informed consent was waived.

### Data Analysis

Sensitivity, specificity, and accuracy of gastric emptying based on the Bonta 2-h criteria were computed using 4-h gastric retention values greater than 10% as the standard for delayed gastric emptying. Comparisons of mean were conducted using the Student $t$ test. Comparison of the 4 sites' population characteristics was performed using the $\chi^2$ test. $P$ values of less than 0.05 were considered statistically significant. Results are presented as mean $\pm$ SD. Statistical analyses were performed using Prism, version 6.01 (GraphPad Software), for Windows (Microsoft).

### RESULTS

A total of 431 patients were included. Patient characteristics for each center are presented in Table 1. The proportion of females ($P = 0.3335$) and the prevalence of delayed gastric

emptying ($P = 0.1279$) were not significantly different among the 4 sites.

The distribution of gastric retention values at 2 and 4 h is shown in Figure 1. Scatterplots of the 2-h versus 4-h gastric retention values are displayed in Figure 2. The receiver-operating-characteristic curve of 2-h gastric retention (using the 4-h gastric retention values as ground truth) yielded an area under the receiver-operating-characteristic curve of 0.926 ($P < 0.0001$, Fig. 3).

### Bonta Criteria

At 2 h, 261 of 431 patients (60.6%) had gastric retention of less than 45%, 62 (14.4%) had gastric retention of more than 65%, and 108 (25.1%) had intermediate values requiring further imaging. Therefore, 323 (74.9%) studies could be terminated at 2 h using the Bonta criteria. Overall sensitivity, specificity, and accuracy of the Bonta criteria, as applied to our multicenter cohort, were 92.4%, 96.9%, and 95.8%, respectively (Table 2; Fig. 3). Application of the Bonta criteria yielded 8 (1.9%) false-negative and 10 (2.3%) false-positive studies. The 4-h gastric retention values of
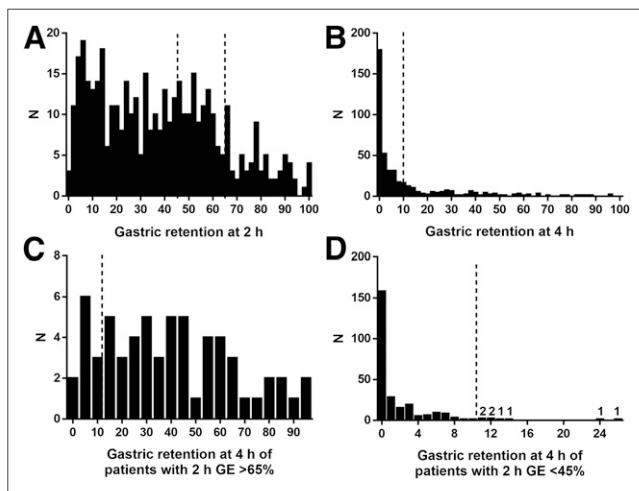


**FIGURE 1.** (A and B) Distribution of gastric retention values at 2 h (A) and 4 h (B) for all 431 participants. (C and D) Distribution of gastric retention values at 4 h for patients with gastric retention > 65% at 2 h (C) and < 45% at 2 h (D).
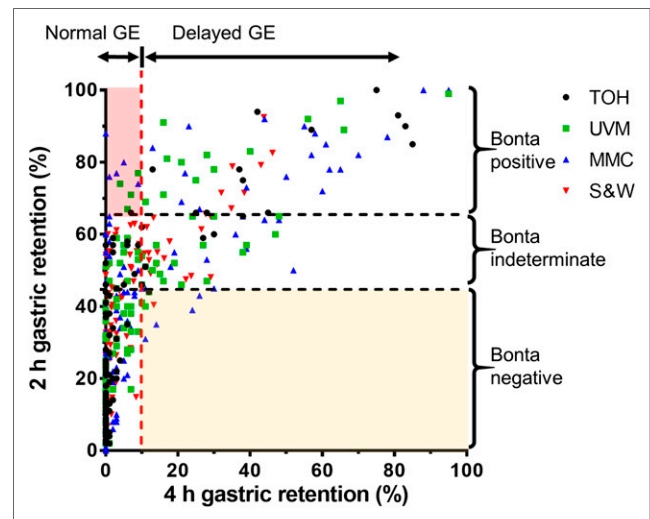


**FIGURE 2.** Scatterplot of 2-h vs. 4-h gastric retention values. Four-hour cutoff of 10% for delayed gastric emptying is illustrated by vertical dashed line. Bonta 2-h criteria for delayed (>65%) and normal (<45%) gastric retention are shown by horizontal dashed lines. False-positive results (2-h retention > 65%; 4-h retention ≤ 10%) correspond to shaded area at upper left, whereas false-negative results (2-h retention < 45%; 4-h retention > 10%) correspond to shaded area at lower right. GE = gastric emptying; MMC = Montefiore Medical Center; S&W =Baylor Scott & White Health; TOH = The Ottawa Hospital; UVM = University of Vermont Hospital.
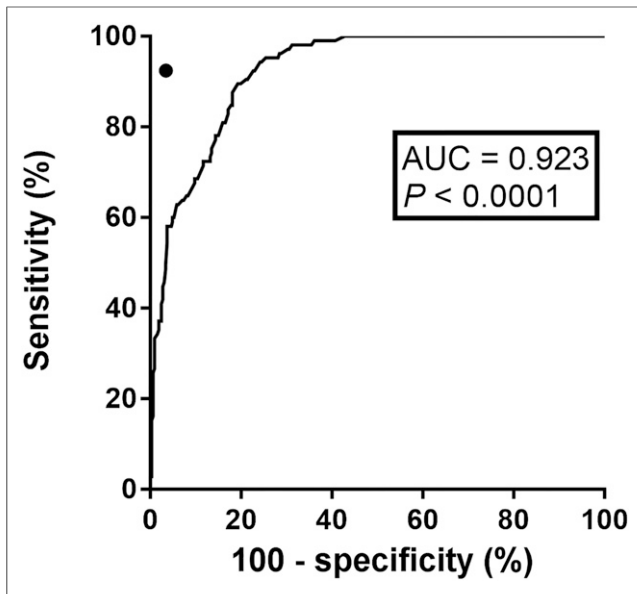
**FIGURE 3.** Receiver-operating-characteristic (ROC) curve for 2-h gastric retention for diagnosis of delayed 4-h gastric emptying. Sensitivity and specificity of Bonta criteria (●) are more accurate than any single 2-h cutoff. AUC = Area under curve.

the 8 false-negative studies were 11%, 11%, 12%, 12%, 13%, 14%, 24%, and 26% (Fig. 1D).

### Number of Images

The cumulative number of patients with gastric retention of 10% or less at 1, 2, 3, and 4 h were 6, 77, 215, and 326, respectively. The total number of images needed to complete the study using standard criteria was 1,768. The total number of images acquired when applying the Bonta criteria was 1,490, representing 84% of baseline. Use of the Bonta criteria resulted in a significantly shorter average per-patient imaging time of 2.46 ± 0.85 versus 3.10 ± 0.73 h ($P < 0.0001$) for the standard method, a reduction of 20.6%.

### DISCUSSION

Bonta et al. proposed an algorithm to shorten gastric emptying imaging based on the standardized GES protocol (6). They retrospectively optimized cutoff criteria in a cohort of 174 patients and validated these criteria in a second independent cohort of 154 patients, in whom an accuracy of 94.9% was demonstrated. The purpose of the current study was to validate the Bonta criteria applied to a broader population in a multicenter study.

**TABLE 2**
Contingency Table of Shortened Bonta Protocol Compared with 4-Hour Retention

| Finding | 4 h > 10% | 4 h ≤ 10% | Total |
|---|---|---|---|
| Shortened protocol positive | 97 | 10 | 107 |
| Shortened protocol negative | 8 | 316 | 324 |
| Total | 105 | 326 | 431 |

The population studied was similar at all 4 sites, as well as similar to the original cohorts of Bonta et al., with comparable proportions of female patients and disease prevalence (Table 3). In our multicenter cohort of patients, accuracy (95.6%) using the shortened criteria was similar to that obtained by Bonta (94.9%) in the validation group. The specificity (96.9%) and accuracy (95.8%) obtained using the shortened protocol were excellent, and superior to any single 2-h cutoff value that could be applied to the multicenter cohort (Fig. 2). On first analysis, sensitivity using the Bonta criteria (92.4%) appears low, because application of these criteria resulted in 8 false-negative studies. In reality, 6 of these 8 patients had 4-h gastric retention of 11%–14% and 4 had gastric retention of 11%–12%, which, depending on the precision of the examination, may well be indistinguishable from a normal value of 10% and unlikely to have any clinical significance. Indeed, small changes in regions of interests can lead to significant changes in gastric retention values, especially when adjacent bowel activity is present. Were we to reclassify the 4-h values of 11%–12% to be normal gastric retention, then the sensitivity of the examination would have been 96.0% and the corresponding accuracy 96.8%. Of course, revised 2-h criteria could be applied to the 2-h gastric retention values to improve sensitivity and specificity; however, this would also reduce the number of patients in whom the GES would be shortened (Fig. 2). We have therefore chosen to prospectively evaluate the criteria as set by Bonta rather than retrospectively choose another set of criteria that would require further independent validation, though this strategy could be pursued by individual centers. Indeed, application of these criteria over the multicenter cohort resulted in excellent results, consistent with the original findings of Bonta et al. (6).

When applying the Bonta criteria to our multicenter cohort, the study length was reduced in 55.7% of patients, whereas the total number of images required decreased by 15.7%. Per-patient imaging time declined to 2.46 h as compared with 3.10 h, a reduction of 20.6%. These parameters translate into decreased resource use, more rapid patient discharge, and presumably improved patient satisfaction.

Shortening the gastric emptying protocol to 2 h in most patients, with minimal impact on test accuracy, can be convenient for patients and reduce camera time. Nevertheless, it may be prudent to weigh the advantage of the shortened test against the information lost. Finally, patients classified as normal on the basis of

**TABLE 3**
Comparison of Current and Bonta-Based Results (6)

| Parameter | Multicenter | Bonta original dataset | Bonta validation dataset |
|---|---|---|---|
| *n* | 431 | 174 | 158 |
| Prevalence of delayed gastric emptying (%) | 24.4 | 26 | Not applicable |
| Proportion of patients stopped at 2 h (%) | 74.9 | 79 | 78 |
| Accuracy (%) | 95.8 | 96.6 | 94.9 |

2-h values may actually present abnormally rapid gastric empty-ing, or dumping, defined as retention of less than 30% at 1 h (*4,5*). Therefore, imaging at 1 h remains important since it can provide additional diagnostic information.

## CONCLUSION

In our application of the Bonta criteria to a multicenter cohort of patients, we have found that the criteria shortened studies in most patients and resulted in an effective compromise between preserved accuracy (95.8%), reduced resource use, and im-proved patient convenience. Our results support the possibility of using these shortened criteria, or a variation thereof, in clinical practice.

## DISCLOSURE

The costs of publication of this article were defrayed in part by the payment of page charges. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734. No potential conflict of interest rele-vant to this article was reported.

## REFERENCES

1. Parkman HP, Hasler WL, Fisher RS; American Gastroenterological Association. American Gastroenterological Association medical position statement: diagnosis and treatment of gastroparesis. *Gastroenterology.* 2004;127:1589–1591.
2. Soykan I, Sivri B, Sarosiek I, Kiernan B, McCallum RW. Demography, clinical characteristics, psychological and abuse profiles, treatment, and long-term follow-up of patients with gastroparesis. *Dig Dis Sci.* 1998;43:2398–2404.
3. Abell TL, Camilleri M, Donohoe K, et al. Consensus Recommendations for Gastric Emptying Scintigraphy: A Joint Report of the American Neurogastroen-terology and Motility Society and the Society of Nuclear Medicine. *Am J Gastro-enterol.* 2008;103:753–763.
4. Tougas G, Eaker EY, Abell TL, et al. Assessment of gastric emptying using a low fat meal: establishment of international control values. *Am J Gastroenterol.* 2000;95:1456–1462.
5. Donohoe KJ, Maurer AH, Ziessman HA, et al. Procedure guideline for adult solid-meal gastric-emptying study 3.0. *J Nucl Med Technol.* 2009;37:196–200.
6. Bonta DV, Lee HY, Ziessman HA. Shortening the 4-hour gastric-emptying pro-tocol. *Clin Nucl Med.* 2011;36:283–285.