Multicenter Validation of a Shortened Gastric-Emptying Protocol

Matthieu Pelletier-Galarneau¹, Oyebola O. Sogbein¹, Xuan Pham¹, Jason Hao², Jenna Le³, Mark D Strober⁴, Michael L Middleton⁴, Janusz Kikut², Leonard M. Freeman³, Lionel S. Zuckier¹

(1) The Ottawa Hospital, Division of Nuclear Medicine, Ottawa Ontario
(2) University of Vermont Medical Center, Burlington, VT
(3) Montefiore Medical Center, Bronx, NY
(4) Baylor Scott & White Health, Central Texas, Temple, TX

Correspondence: Matthieu Pelletier-Galarneau ( Resident), Division of Nuclear Medicine, The Ottawa Hospital, General Campus, 501 Smyth Road, University of Ottawa, Ottawa Ontario, Canada K1H 8L6. Phone: 613-737-8528 x13350, Fax: 613-737-8705, Email: mattgalarneau@gmail.com

No disclosure. No financial support.

Word Count: 2126 words.

Running Title: Shortened Gastric-Emptying Protocol
Abstract

Gastric emptying scintigraphy is considered to be the gold standard for detection of gastroparesis and other disorders of gastric motility; SNMMI guidelines are predicated upon imaging over a period of 4 hours, which is inconvenient for patients. Bonta et al. introduced 2-hour criteria which served to shorten the protocol in the majority of 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-hour gastric retention > 10%. Bonta criteria (retention >65% abnormal, <45% normal, otherwise proceed to complete exam) were applied to the 2-hour data. Sensitivity, specificity, accuracy and resource utilization for the Bonta method were calculated. Results based upon standard 4-hour solid gastric emptying, performed according to current SNMMI guidelines, served as the gold standard.

Results: 10% retention or less was achieved by 6, 77, 215 and 326 patients at 1, 2, 3 and 4 hours, respectively. At 2 hours, 261 of 431 patients (60.6%) had gastric retention <45% and according to Bonta would be classified as normal, 62 (14.4%) had gastric retention>65% and would be classified as delayed emptying, while 108 (25.1%) had intermediate values requiring further imaging through 4 hours. The Bonta criteria yielded a sensitivity, specificity, and accuracy of 92.4%, 96.9%, and 95.8% respectively, superior to any single cut-off point applied to the 2 hour values. They resulted in false negative results in 8 (1.9%) patients, 6 of whom were borderline positive at 4-hours (gastric retention of 11-14%). Using the Bonta criteria, 74.9% of
studies would be terminated by 2 hours, decreasing total camera usage 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 hours.

**Conclusions:** In a multicenter cohort, use of the Bonta criteria shortened the duration of studies in a majority of patients, resulting in an effective compromise between reduced resource utilization, improved patient convenience, and preserved accuracy.

**Keywords:** 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). The majority of cases of gastroparesis are idiopathic but delayed gastric emptying has been associated with diabetes, gastrointestinal tract surgery, Parkinson’s disease, collagen vascular disease, and intestinal pseudoobstruction (2). Gastric emptying scintigraphy (GES) is considered the gold standard for evaluation of patient with suspected delayed gastric emptying (3). In 2000, Tougas et al. published a 4-hour 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 becomes the standard gastric emptying protocol (3). This protocol is now incorporated into the Society of Nuclear Medicine and Molecular Imaging (SNMMI) guidelines for GES (5). Because of its length, the standard protocol can be inconvenient for the 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 following ingestion of a standardized meal to determine if 2-hour gastric retention values could predict abnormal or normal results. They found that 2-hour values of >65% and <45%, to indicate delayed and normal gastric retention respectively, were predictive of the 4-hour “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 hours in approximately 80% of cases.

In order 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-hour protocol can be shortened without significant impact on the test accuracy.
MATERIALS AND METHODS

Patient Population

The patient population of this multicenter study included 431 patients, with approximately 25% of the patients derived from each of 4 sites (Table 1); The Ottawa Hospital, Ottawa (TOH), ON, Canada, Montefiore Medical Center (MMC), Bronx, NY, University of Vermont Hospital (UVM), Burlington, VT, and Baylor Scott & White Health (S&W), Central Texas, Temple, Texas. All patients were imaged according to the SNMMI 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, 30g of jam, and 120 mL of water. Anterior and posterior planar images centered on the upper abdomen 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 hours, unless gastric retention was equal 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-hour criteria were computed using 4-hour 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-square test. P-values less than 0.05 were considered statistically significant. Results are presented as mean ±
RESULTS

A total of 431 patients were included. Patient characteristics for each center are presented in Table 1. Proportion of females \( (p=0.3335) \) and prevalence of delayed gastric emptying \( (p=0.1279) \) were not significantly different between the 4 sites.

Distribution of gastric retention values at 2- and 4-hours is shown in Figure 1. Scatter plot of the 2 versus 4 hour gastric retention values are displayed in Figure 2. ROC curve of 2-hour gastric retention (using the 4-hour gastric retention values as ground truth) yielded an area under the ROC curve of 0.926 \( (p<0.0001, \text{Figure 3}) \).

Bonta Criteria

At 2 hours, 261 of 431 patients \( (60.6\%) \) had gastric retention <45%, 62 \( (14.4\%) \) had gastric retention >65%, and 108 \( (25.1\%) \) had intermediate values requiring further imaging. Therefore, 323 \( (74.9\%) \) studies could be terminated at 2 hours 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 and Figure 3). Application of the Bonta criteria yielded 8 \( (1.9\%) \) false negative and 10 \( (2.3\%) \) false positive studies. The 4-hour gastric retention values of the 8 false negative studies were 11%, 11%, 12%, 12%, 13%, 14%, 24% and 26% (Figure 1D).

Number of Images

The cumulative number of patients with gastric retention of 10% or less at 1, 2, 3, and 4 hours 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 images, representing 84% of baseline. Use of the Bonta criteria resulted in a significantly shorter average per-patient imaging time of 2.46 ± 0.85 vs. 3.10 ± 0.73 hours (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 cut-off 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.

The population studied was similar at all four sites and with the original Bonta et al. cohorts with comparable proportion of females and disease prevalence (Table 3). In our multicenter cohort of patients, accuracy (95.6%) using the shortened criteria was similar to those 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-hour cut-off value that could be applied to the multicenter cohort (Figure 2). On first analysis, sensitivity using the Bonta criteria (92.4%) appears low as application of these criteria resulted in 8 false negative studies. In reality, 6 of these 8 patients had 4-hour 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-hour 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-hour criteria could be applied to the 2 hour gastric retention values in order to improve sensitivity and specificity; however, this would also reduce the number of patients in whom the GES would be shortened (Figure 2). We have therefore chosen
to prospectively evaluate the criteria as set by Bonta rather than retrospectively choose another set of
criteria which 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 Bonta et al. original findings (6).

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

Shortening the gastric emptying protocol to 2 hours in a majority of patients, with minimal
impact on test accuracy, can be convenient for the patients and reduce camera time utilisation.
Nevertheless, it may be prudent to weigh the advantage of the shortened test with information
lost. Finally, patients classified as normal based on 2 hour values may actually present
abnormally rapid gastric emptying, or “dumping”, defined as retention <30% at 1 hour (4,5).
Therefore, imaging at 1 hour 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 Bonta criteria shortened studies in a majority of patients, and resulted in an effective
compromise between preserved accuracy (95.8%), reduced resource utilization and improved
patient convenience. Our results support raise the possibility of using these shortened criteria, or
a variation thereof, in clinical practice.
REFERENCES


Acknowledgment

Ms. Carla Bailey, Baylor Scott & White Health, Central Texas, Temple, TX
Figure 1: Distribution of gastric retention values at 2-hours (A) and 4-hours (B) for all 431 participants. Distribution of gastric retention values at 4-hours for patients with gastric retention >65% at 2-hours (C) and <45% at 2-hours (D).
Figure 2: Scatter plot of the 2 hour versus 4 hour gastric retention values. The 4 hour cut off of 10% for delayed gastric emptying is illustrated by the vertical dashed line. Bonta 2 hour criteria for delayed (>65%) and normal (<45%) gastric retention are shown by the horizontal dashed lines. False positive results (2 hour retention >65%; 4 hour retention ≤10%) corresponds to the shaded area in the upper left of the image whereas false negative results (2 hour retention <45%; 4 hour retention >10%) correspond to the lightly shaded area in the lower right. GE, gastric emptying. Baylor Scott & White Health, Central Texas; TOH, The Ottawa Hospital; UVM, University of Vermont Hospital; MMC, Montefiore Medical Center
Figure 3: Receiver operating characteristic (ROC) curve for 2-hour gastric retention for the diagnosis of delayed 4-hour gastric emptying. Sensitivity and specificity of the Bonta criteria (●) are more accurate than any single 2 hour cut-off. AUC, Area under the curve.
Table 1: Patient characteristics for each center.

<table>
<thead>
<tr>
<th>Centers</th>
<th>N</th>
<th>Females N (%)</th>
<th>Delayed 4 hour emptying N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMC</td>
<td>130</td>
<td>89 (68.5)</td>
<td>34 (26.2)</td>
</tr>
<tr>
<td>S&amp;W</td>
<td>100</td>
<td>74 (74.0)</td>
<td>23 (23.0)</td>
</tr>
<tr>
<td>TOH</td>
<td>101</td>
<td>77 (76.2)</td>
<td>17 (16.8)</td>
</tr>
<tr>
<td>UVM</td>
<td>100</td>
<td>66 (66.0)</td>
<td>31 (31.0)</td>
</tr>
<tr>
<td>Total</td>
<td>431</td>
<td>306 (71.0)</td>
<td>105 (24.4)</td>
</tr>
</tbody>
</table>
Table 2: Contingency table of shortened Bonta protocol compared to 4-hour retention.

<table>
<thead>
<tr>
<th></th>
<th>4 hour &gt;10%</th>
<th>4 hour ≤10%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shortened protocol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>positive</td>
<td>97</td>
<td>10</td>
<td>107</td>
</tr>
<tr>
<td>negative</td>
<td>8</td>
<td>316</td>
<td>324</td>
</tr>
<tr>
<td></td>
<td>105</td>
<td>326</td>
<td>431</td>
</tr>
</tbody>
</table>
Table 3: Comparison of current and Bonta-based results (6).

<table>
<thead>
<tr>
<th></th>
<th>Multicenter</th>
<th>Bonta Original Dataset</th>
<th>Bonta Validation Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>431</td>
<td>174</td>
<td>158</td>
</tr>
<tr>
<td>Prevalence of delayed gastric emptying (%)</td>
<td>24.4</td>
<td>26</td>
<td>N/A</td>
</tr>
<tr>
<td>Proportion of patient stopped at 2 hr (%)</td>
<td>74.9</td>
<td>79</td>
<td>78</td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>95.8</td>
<td>96.6</td>
<td>94.9</td>
</tr>
</tbody>
</table>