Journal of Nuclear Medicine, published on October 19, 2017 as doi:10.2967/jnumed.117.197053

August 20, 2017

Intragastric Meal Distribution during Gastric Emptying Scintigraphy for Assessment of Fundic Accommodation: Correlation with Symptoms of Gastroparesis

Perry Orthey (Section of Gastroenterology, Temple University, Philadelphia, PA), Daohai Yu (Department of Clinical Sciences, Temple Clinical Research Institute, Temple University School of Medicine, Philadelphia, PA), Mark L. Van Natta (Johns Hopkins University Data Coordinating Center, Baltimore, MD), Frederick V. Ramsey (Department of Clinical Sciences, Temple Clinical Research Institute, Temple University School of Medicine, Philadelphia, PA), Jesus R. Diaz (Nuclear Medicine Section, Texas Tech University, El Paso, TX), Paige Bennett (Nuclear Medicine Section, Wake Forest University, Winston Salem, NC), Andrei Iagaru (Nuclear Medicine Section, Stanford University, Palo Alto, CA), Roberto Salas Fragomeni (Nuclear Medicine Section, Johns Hopkins University, Baltimore, MD), Richard W. McCallum (Section of Gastroenterology, Texas Tech University, El Paso, TX), Irene Sarosiek (Section of Gastroenterology, Texas Tech University, El Paso, TX); William L. Hasler (Division of Gastroenterology, University of Michigan, Ann Arbor, MI), Gianrico Farrugia (Section of Gastroenterology, Mayo Clinic, Rochester, MN); Madhusudan Grover (Section of Gastroenterology, Mayo Clinic, Rochester, MN), Kenneth L. Koch (Section of Gastroenterology, Wake Forest University, Winston Salem, NC); Linda Nguyen (Division of Gastroenterology, Stanford University, Palo Alto, CA); William J. Snape (Division of Gastroenterology, California Pacific Medical Center, San Francisco, CA); Thomas L. Abell (Division of Gastroenterology, University of Louisville, Louisville, KY); Pankaj J. Pasricha (Section of Gastroenterology, Johns Hopkins University, Baltimore, MD), James Tonascia (Johns Hopkins University Data Coordinating Center, Baltimore, MD); Frank Hamilton (National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD); Henry P. Parkman (Section of Gastroenterology, Temple University, Philadelphia, PA); Alan H. Maurer (Nuclear Medicine Section, Temple University, Philadelphia, PA), for the NIH Gastroparesis Consortium.

Word counts: Abstract: 299. Manuscript (Introduction through Discussion): 3,584.

Key words: fundic accommodation; gastric emptying; gastroparesis; diabetic gastroparesis; idiopathic gastroparesis; Gastroparesis Cardinal Symptom Index (GCSI); Patient Assessment of Upper GI Symptoms (PAGI-SYM).

No conflicts of interest exist.

The Gastroparesis Clinical Research Consortium is supported by the National Institute of Diabetes and Digestive and Kidney Diseases (grants U01DK073975, U01DK073983, U01DK073985, U01DK074007, U01DK073974, U01DK074008).

ClinicalTrials.gov Identifier: NCT01696747

Correspondence to: Henry P. Parkman, M.D.
Gastroenterology Section
Temple University School of Medicine, Philadelphia, PA 19140
Telephone: 215-707-7579
Fax: 215-707-2684
Email: henry.parkman@temple.edu

ABSTRACT

Impaired fundic accommodation (FA) limits fundic relaxation and ability to act as a reservoir for food. Assessing intragastric meal distribution (IMD) during gastric emptying scintigraphy (GES) allows for a simple measure of FA. **Goals**: 1) Evaluate nuclear medicine and radiology physician trained readers' visual assessment of FA from solid-meal GES; 2) Develop software to quantify GES intragastric meal distribution (IMD); 3) Correlate symptoms of gastroparesis with IMD and GE. **Methods:** After training to achieve consensus interpretation of GES FA, 4 readers interpreted FA in 148 GES studies from normal volunteers and patients. Mixture distribution and Kappa agreement analysis assessed reader consistency and agreement of scoring of FA. Semi-automated software quantified IMD (ratio of gastric counts in the proximal stomach to total stomach) at 0, 1, 2, 4 hrs postprandially. ROC analysis was performed to optimize diagnosis of abnormal IMD^0 (IMD at 0 min) with impaired FA. IMD^0, GES, water loading test and symptoms were then compared in 177 patients with symptoms of gastroparesis. **Results:** Reader pairwise weighted Kappas visual assessment of FA averaged 0.43 (moderate agreement) for normal vs impaired FA. Readers achieved 84.0 % consensus and 85.8% reproducibility in assessing impaired FA. IMD^0 based on division of the stomach into proximal and distal halves averaged 0.809±0.083 (SD) for normal FA compared to 0.447±0.132 ($p<0.01$) for impaired FA. Optimal cutoff for IMD^0 based on ROC analysis was 0.568 to discriminate normal versus impaired FA (sensitivity 86.7%, specificity 91.7%). Of 177 patients with symptoms of gastroparesis, 129 (72.9%) had delayed GE; 25 (14.1%) had abnormal IMD^0. Low IMD^0 (impaired FA) was associated with increased early satiety ($p=0.02$). **Conclusions:** FA can be assessed visually during routine GES with moderate agreement and high reader consistency. Visual and quantitative assessment of FA during GES can yield additional information on gastric motility to help explain patient symptoms.

INTRODUCTION

Gastric emptying scintigraphy (GES) is routinely used to measure overall gastric emptying (GE) in patients with dyspeptic symptoms of gastroparesis which include nausea, vomiting, early satiety, postprandial fullness, and in some patients, upper abdominal pain. Gastroparesis symptoms correlate, albeit weakly, with delayed global GE (1,2). This weak association may reflect separate mechanisms that together contribute to gastric emptying.

With normal meal ingestion, the proximal stomach relaxes and increases in volume, to accommodate the meal (3). During GE, the solid meal progresses from the proximal stomach into the distal stomach. Impaired fundic accommodation (FA) compromises the ability of the upper stomach to act as a reservoir for ingested food and can result in enhanced transit from the proximal to the distal stomach (4). Abnormal FA may potentially explain some dyspeptic symptoms (3). Studies using the gastric barostat suggested impaired accommodation is associated with early satiety and weight loss (5). While the barostat study is considered the gold standard for assessing FA, it is invasive and not widely available. In addition, the barostat balloon itself can alter gastric physiology (6). Several alternate methods have been developed to measure FA, such as single-photon emission computed tomography (SPECT), and magnetic resonance imaging (MRI) but these are not in widespread use or use technology not widely available (7).

Quantitative measurement of intragastric meal distribution (IMD) comparing the proportion of the meal in the proximal to distal stomach can be used as an indirect measure of FA (8). Assessment of IMD may yield additional information from a standard GES study permitting a better assessment of symptoms of gastroparesis to abnormal gastric motility. This may then lead to potential therapy directed to improve gastric accommodation (9,10,11).

The aims of this study were: 1) Evaluate the agreement and consistency among trained readers visually assessing FA during routine solid-meal GES; 2) Develop computer software for quantitative analysis of IMD during solid-meal GES and establish quantitative criteria for determining normal vs abnormal FA; 3) Correlate symptoms of gastroparesis with different measures of gastric function, including IMD, GES, and water load satiety testing.

METHODS

Overview of study

In this study, we developed visual and quantitative assessment tools to measure FA and IMD utilizing standard solid-meal GES. We developed a numerical cutoff value for abnormal IMD based on an expert panel of nuclear medicine and radiology physicians' visual assessments of FA. We applied these cutoffs to solid-meal GES tests of patients from the NIH Gastroparesis Consortium Centers to relate abnormal IMD to defined symptoms of gastroparesis. We also looked at the relationship of other measures of gastric function (total gastric emptying and water loading) to symptoms.

These studies were approved by the IRB at Temple University School of Medicine and each of the NIH Gastroparesis Consortium Centers contributing patient studies for analysis.

Gastric emptying scintigraphy (GES)

Solid-meal GES was measured using the 4-hour protocol of Tougas et al (12) and recommended by the consensus report of the Society of Nuclear Medicine and Molecular Imaging and the American Neurogastroenterology and Motility Society (8). Patients stopped prokinetic agents, narcotic analgesics, and anticholinergic agents for 2 days prior to the GES test. Subjects were studied in the morning after fasting overnight. The meal consisted of 120 grams liquid egg white radiolabeled with 18.5-37.0 mBq (0.5 to 1 mCi) Tc-99m sulfur colloid served with two pieces of white bread and 30 grams jelly with 120 mL of water as recommended in the Society of Nuclear Medicine and Molecular Imaging procedure guideline (13).. Patients were asked to finish ingesting the meal within 10 minutes. Imaging was performed at 0, 1, 2, and 4 hours with the patient upright. Some centers also performed imaging at 30 min and 3 hours. Anterior and posterior images were acquired for 60 seconds. Total GE was analyzed as percent of radioactivity retained in the whole stomach over time using the geometric mean of the decay-corrected anterior and posterior counts for each time point. Gastric retention of the Tc-99m labeled solid meal >60 % at 2 hours and/or >10% at 4 hours was considered delayed gastric (8,12).

Reader Validation of Visual Assessment of Fundic Accommodation during GES

We assessed if normal and abnormal FA can be consistently assessed visually during GES. An instruction guide on how to interpret FA was first developed and sent to four nuclear medicine and radiology physicians of the NIH Gastroparesis Clinical Research Consortium (GpCRC) (Supplemental Appendix 1). Examples of normal and impaired FA are shown in Figure 1. After completing the instruction guide, all readers had to successfully complete a training set of 24 studies to identify 16 normal volunteers, 8 patients with abnormal FA.

Following initial training, the readers then interpreted 148 GES studies performed at Temple University Hospital. This study set included 19 normal volunteers and 99 patients (18 impaired FA, 81 normal FA). In addition, 30 patient studies were repeated (4 impaired, 26 normal) in the study set to test reader reproducibility. The readers used a FA interpretive score of 1-5: 1=definitely abnormal accommodation; 2=probably abnormal accommodation; 3=possibly abnormal accommodation; 4=probably normal accommodation; and 5=definitely normal accommodation. These scores were based on Kundell et al, for the mixture distribution analysis (MDA) (14).

Computer Assessment of Intragastric Meal Distribution

Semi automated software development and gastric segmentation for IMD analysis

There are no prior well-established criteria for how to divide the stomach into proximal and distal segments to assess FA. We developed semi-automated MatLab® (MathWorks®, Natick, Massachusetts) software to automatically derive threshold based gastric borders to outline the gastric region of interest. Supplemental Appendix 2 has the details of the software approach.

To quantitate whether there is normal vs abnormal FA in the GES images obtained immediately post meal ingestion (t=0 minutes), we defined IMD^0 as the ratio of gastric counts in the proximal stomach to the entire stomach. For this study, we evaluated three different quantitative approaches to measure IMD^0. The first approach divided the stomach into two equal parts (proximal and distal halves) using the

midpoint of the longitudinal axis of the stomach. This "halves" approach was previously described by Piessevaux et al who evaluated gastric images at different time intervals post meal ingestion and then summed the images (15). The longitudinal axis of the stomach was defined as an the curve which runs from the most proximal point of the stomach wall to the most distal point along which the stomach was divided in two parts based on the mid-length of this axis (15). The second approach divided the stomach into three parts using three equal segments of the longitudinal axis (16). A third approach divided the stomach into proximal and distal parts visually using the anatomic incisura (17). These methods for dividing the stomach are illustrated in Figure 2.

Correlation with Symptom Scores from Gastroparesis Registry Patients

For this second study group, a set of 177 patients with recorded symptoms of gastroparesis undergoing GES enrolled at 8 centers in the NIH Gastroparesis Registry from September 2012 to March 2016 (18) was used Patients met specific entry criteria: being 18 years or older with symptoms suggestive of gastroparesis (nausea, vomiting, early satiety, and postprandial fullness) of at least 12 weeks in duration, having had GES using the 4-hour liquid egg white protocol, and having no structural abnormality on upper endoscopy within one year of enrollment. Each patient completed the 20-item Patient Assessment of Upper Gastrointestinal Symptoms (PAGI-SYM) questionnaire to assess symptoms of gastroparesis, dyspepsia, and gastroesophageal reflux disease (19). It includes the nine symptoms of the Gastroparesis Cardinal Symptom Index (GCSI): nausea, retching, vomiting, stomach fullness, inability to finish meal, excessive fullness, loss of appetite, bloating, and abdominal distension (20). Patients are asked to assess the severity of their symptoms during the previous two weeks using a 0 to 5 scale where no symptoms=0, very mild=1, mild=2, moderate=3, severe=4, and very severe=5.

GES studies were performed as described above at each NIH Gastroparesis Registry center (8,12). Patients also underwent a non-caloric liquid water satiety test (21). The water load test is a standardized test to induce gastric distension. Patients reported after fasting overnight and were

instructed to drink maximal volumes of water using an opaque 150 mL cup over 5 minutes until they felt completely full (21). The volume of water consumed was recorded.

Statistical Analysis Methods

As there is currently no routinely available, gold standard test that is clinically employed to assess if normal vs abnormal FA is present in patients undergoing GES for suspected gastroparesis, we used a statistical method described by Kundel and Polansky called mixture distribution analysis (MDA) to assess the consistency of scoring for normal vs abnormal FA observed during solid-meal GES. MDA was originally applied to interpretation of plain film x-rays and is used to evaluate the consistency of assessments among different readers, where a gold standard does not exist to allow for a determination of whether an accurate diagnosis can be made. It assumes that high consistency among readers can be used to establish a correct diagnosis (14).

To perform MDA, an initial set of GES studies (n=148) was independently evaluated by the four trained readers and given a score from 1 to 5 as per Kundel and Polansky (14): 1=definitely abnormal accommodation, 2=probably abnormal accommodation, 3=possibly abnormal accommodation, 4=probably normal accommodation, and 5=definitely normal accommodation. This scoring system was then condensed into a two-level scoring as per Kundel and Polansky (14): abnormal/impaired accommodation (scores 1, 2, 3) and normal/non-impaired accommodation (scores 4, 5). These scores were compiled and analyzed using an iterative method (expected-maximization (EM) algorithm) (22), to calculate the maximum likelihood estimates of the model parameters and derive the proportion "correctly" or consistently diagnosed by the way of consensus of the four readers. In other words, how much the readers agree with one another as a panel, indicating a reproducible diagnosis. A more complete description of MDA statistical analysis is included in Supplemental Appendix 3 (23-28).

We then correlated the characteristics and symptoms of a second set of patients from the NIH Gastroparesis Registry with their IMD values. Descriptive statistics (means, standard deviations, frequencies, and percentages) were used to characterize subgroups of gastroparesis patients.

Demographics, medical history, gastroparesis history, symptom severity, and quality of life characteristics at the enrollment visit were compared across subgroups of IMD^0 as determined in ROC analyses: normal (>0.643), borderline (0.568-0.643), and abnormal (<0.568) (See Supplemental appendix 3). P-values were derived from a Cochran-Armitage trend test for binary variables, or linear regression of continuous variables on the 3 categories of IMD modelling IMD as an ordinal variable (29,30). Multiple logistic models regressing abnormal or abnormal/borderline IMD on enrollment characteristics were selected based on Akaike Information criteria (AIC) from a candidate set of all characteristics (31,32).

We also analyzed each of the other assessments of gastric motility, specifically water load and gastric emptying, to explore their relationships with PAGI-SYM items. Symptom severity was determined for different degrees of abnormalities for each test.

All p-values are two-sided and nominal with p-values <0.05 considered statistically significant. Analyses were performed using methods described in SAS version 9.3 (SAS Institute Inc., Cary, NC) or Stata version 13.1 (StataCorp).

RESULTS

Visual Assessment of Fundic Accommodation/Intragastric Meal Distribution

The results of the readers' assessment of FA in the first set of 99 patient studies are shown. Fifteen patients had impaired accommodation based on FA scores (majority 3 or 4 readers rating impaired accommodation, i.e., a score of 1-3), 73 patients had normal accommodation (majority 3 or 4 readers rating normal accommodation, i.e., a score of 4-5), and 11 patients had indeterminate accommodation by the readers (2 rating impaired and 2 rating normal). Pairwise weighted Kappas among the 4 readers' assessments of FA using the 5 levels of normal/impaired accommodation averaged 0.38 (fair agreement; 95% CI: 0.25-0.51) (Table 2). Kappa agreement measures among the 4 readers for normal vs. impaired accommodation classifications (i.e., 4-5 vs. 1-3) averaged 0.43 (moderate agreement; 95% CI: 0.25-0.62).

Using MDA, 84% of the time a panel of could reach a "correct" consensus about the impairment of FA (Table 1 of Appendix 3). MDA of the 99 patients' data further revealed that approximately 11% of

cases were easy positives and 62% were easy negatives with remaining 27% a difficult assessment. In addition, approximately 87.2% of easy positives and 5.3% of easy negatives were called positives while 41.0% of difficult cases were called positives.

To analyze the reproducibility of the readers' assessments of FA, a set of 30 patient studies were repeated (4 impaired, 26 normal) by the 4 readers giving a total of 120 repeated readings. Of 90 readings initially read as normal FA (score 4 or 5), 77 (85.6%) were read as normal on the second reading. Of the 30 readings initially read as impaired FA (score 1-3), 26 (86.7%) were read as impaired FA on the second reading. This resulted in an overall reproducible of 85.8% .

Software analysis

Three different approaches for dividing the stomach (equal halves, equal thirds, and division at the incisura) were evaluated. Using division of the stomach into proximal and distal halves resulted in a mean IMD^0 of 0.768±0.107 for normal FA and 0.488±0.132 for impaired FA ($p<0.0001$). Using division of the stomach into proximal, middle, and distal thirds resulted in a mean IMD^0 of 0.506±0.120 for normal FA and 0.271±0.091 for impaired FA ($p<0.0001$). The incisura method for segmentation resulted in a mean IMD^0 of 0.850±0.083 for normal FA and 0.584±0.122 for impaired FA ($p<0.0001$). While all three methods were able to differentiate normal from abnormal FA, we found that the incisura was difficult to be reliably identified in some of the images (Figure 2), so this method (15) was not further utilized for correlation to symptom scores. The equal thirds division approach often left the antrum with very low counts for analysis and this also was therefore not utilized for further analysis. Thus, gastric division into proximal and distal halves along the long axis approach was used for analyzing the remaining results in this study.

The time course for IMD from the 19 normal subjects and patients with abnormal FA are shown in Figure 3. IMD^0 (using image immediately post meal ingestion) averaged 0.672±0.092. For these 19 normal subjects, their mean FA scores from the four readers averaged 4.2±0.6, ranging from 3.25 to 5.00.

IMD^0 correlated with the mean FA score by the readers for fundic accommodation (r=0.660; p<0.01). The 99 patients had an average score from the visual reading of 4.1±0.9, ranging from 1.0 to 5.0. The correlation between the mean visual FA score and IMD^0 for the 99 patients was 0.832 (p<0.01). The correlation between the mean visual FA score and IMD^0 for the combined 99 patients and 19 normal subjects was 0.812 (p<0.01).

Patients with normal FA had an average IMD^0 of 0.809±0.083 compared to 0.447±0.132 (p<0.01) for those with impaired FA, defined as ≥3 of the 4 readers reading the FA as impaired. Based on ROC analysis (Figure 4), the AUC was 0.93 and the optimal cutoff was 0.568 for the IMD^0 ratio of proximal to total gastric counts to discriminate normal versus impaired FA (Table 2 of Appendix 3). This resulted in a sensitivity of 86.7% and specificity of 91.7%.

Correlation of GES, IMD, and water loading in patients with symptoms of gastroparesis

A total of 129 (72.9%) of 177 NIH Gastroparesis Registry patients had delayed GES (gastric retention >60% at 2 hours and/or >10% at 4 hours). Twenty five (14.1%) of 177 patients had abnormal IMD^0 < 0.568 with an additional 20 (11.3%) patients having borderline IMD^0 (0.568-0.642). (Figure 1 of Appendix 4 and Table 1 of appendix 4). Impaired FA status as defined by low IMD^0 was associated with low BMI (p=0.006), loss of weight since diagnosis of gastroparesis (p=0.06), nondiabetic patients (p=0.01), and prior pyloric botulinum toxin treatment (p=0.04). Low IMD^0 was also associated with less gastric retention of solid meal at 1 hr (p=0.001), 2 hr (p=0.002), and 4 hr (p=0.05).

Associations with measured clinical parameters did not vary significantly by whether or not patient had delayed GES. We found that 77% of the 132 patients with normal IMD^0 had delayed GE, compared to 64% of the 25 patients with impaired IMD^0 (p=0.13). Of the 47 patients with normal GE, nine (19%) had low IMD^0. Of the 130 patients with delayed GE, 16 (12%) had impaired IMD^0.

Low IMD^0 was associated with more severe early satiety (p=0.02), but not nausea (p=0.39), vomiting (p=0.49), postprandial fullness (p=0.34), or upper abdominal pain (p=0.68) (Table 1 of

Appendix 4). Of the 130 patients with delayed GE, the 16 patients with impaired IMD^0 had greater severity of early satiety than those with normal IMD^0 (4.2±1.3 vs. 3.4±1.6, p=0.01). Of the 47 patients with normal GE, the 9 patients with impaired IMD^0 tended to have greater nausea than those with normal IMD^0 (3.9±0.09 vs. 3.1±1.3; p=0.06).

Increased solid gastric retention at 4 hours was associated with increased severity of vomiting (p=0.03). (Table 1 of Appendix 5). Abnormal water load test was associated with severity of nausea (p=0.05), lower abdominal pain (p=0.0001), and diarrhea (p=0.01, (Table 2 of Appendix 5).

DISCUSSION

Several different tests have been proposed to assess the FA response to an ingested meal, but no one test has gained widespread clinical use (6). In this study, we evaluated IMD^0 which is readily obtained from routine GES, as a marker of the FA response. The current "gold standard" to assess for FA is the gastric barostat. The barostat test, however, is invasive and uncomfortable for the patient. It is not widely available and not used routinely. In this study, we looked at IMD^0 as a potential measure of FA and correlated the finding of abnormal IMD^0 with patient symptoms.

Our study used visual assessment by trained readers of solid meal GES to determine normal or abnormal FA. Using Kappa analysis and Kundel's mixture distribution analysis, we found that FA can be assessed visually during routine GES with fair to moderate pairwise agreement and high panel consistency (84%) among trained readers. For quantitative IMD, semi-automated MatLab® software was developed not only to semiautomate the gastric region of interest outlines but also to divide the stomach into proximal and distal halves based on the midpoint of the longitudinal axis of the stomach. The relative amount in the proximal half of the stomach can then be automatically calculated. Using this analysis, a ratio of <0.568 for IMD^0 was an optimum cutoff point to correlated with the visual determination of impaired FA.

In a large group patients with symptoms of gastroparesis from the NIH Gastroparesis Registry (most with delayed GES and some with normal GES), impaired IMD^0 was present in 25 of 177 patients

(14%) with symptoms of gastroparesis. However, of the 47 patients with normal GE, nine (19%) had low IMD^0 yielding new diagnostic information on abnormal gastric motility. Of the 130 patients with delayed GE, 16 (12%) had low IMD^0.

Early satiety was more severe in patients with abnormal IMD^0 compared to those with normal IMD^0. This agrees with the prior study of Tack using GES in functional dyspepsia patients (13), as well as studies using the gastric barostat (5). Abnormal IMD^0 was also significantly associated with loss of weight and low BMI, and predominantly seen in nondiabetic patients.

Solid-meal GES is a clinical test that is widely available and currently performed typically to only measure total GE. This study shows that GES can not only measure global GE, but can also be used to assess IMD, as an indirect measure of FA. This approach can add additional diagnostic information on gastric motility assessing GE as well as FA. Several agents such as sumatriptan and buspirone have been shown to help improve fundic accommodation and the symptom of early satiety (9,10,11).

Our study shows that there are different pathophysiological bases associated with different symptoms in patients with symptoms of gastroparesis. Gastric retention on GES was associated primarily with vomiting. Abnormal IMD^0 was associated with early satiety. Abnormal water load test was associated with nausea. Thus, different treatment targeting these distinct gastric motility defects may be indicated for different symptoms.

We recognize several limitations of this study. First, the population studied had a high percentage of delayed GE likely because they were a highly selected group from the NIH gastroparesis consortium. Our results may not be typical of what would occur with screening a more diverse patient population with dyspepsia and suspected gastroparesis. This may explain why only 14% of our patients had impaired FA compared to the work of Tack et al in patients with functional dyspepsia where impaired FA was around 30% (3). We recognize that a potential limitation of utilizing routine GES imaging is that if the patient takes a long time eating the meal, there may be some progression of the radiolabeled meal into the antrum when the first image is obtained. This could lead to an impression of abnormal IMD^0. Under our study protocol, all patients completed the meal in less than 10 minutes. Another potential

limitation of GES is if the patient is not able to consume the entire meal, the lower volume and calories ingested will affect FA and gastric emptying.

In conclusion, FA can be visually assessed during routine solid-meal GES with moderate pairwise agreement and high consistency among trained readers. For quantitative analysis of FA, semi-automated software has been developed using a simple division of the stomach into proximal and distal halves along the long axis of the stomach to measure IMD. This approach adds further information during routine solid-meal GES. Abnormal IMD^0 was significantly associated with early satiety. This physiologic and quantitative assessment of FA can improve our understanding of the relationship of symptoms to gastric dysmotility.

REFERENCES

1. Stanghellini V, Tack J. Gastroparesis: separate entity or just a part of dyspepsia? Gut. 2014;63:1972-8.

2. Pathikonda M, Sachdeva P, Malhotra N, Fisher RS, Maurer AH, Parkman HP. Gastric emptying scintigraphy: is four hours necessary? J Clin Gastroenterol 2012;46:209-15.

3. Kelly KA. Gastric emptying of liquids and solids: roles of proximal and distal stomach. Am J Physiol 1980;239:G71-6.

4. Troncon LE, Bennett RJ, Ahluwalia NK, Thompson DG. Abnormal intragastric distribution of food during gastric emptying in functional dyspepsia patients. Gut 1994;35:327-32.

5. Tack J, Piessevaux H, Coulie B, et al. Role of impaired gastric accommodation to a meal in functional dyspepsia. Gastroenterology 1998;115:1346–1352.

6. Ang D. Viewpoint: Measurement of gastric accommodation response: a reappraisal of conventional and emerging modalities. Neurogastroenterol Motil 2011;23:287- 291.

7. Kindt S, Tack J. Impaired gastric accommodation and its role in dyspepsia. Gut 2006;55:1685-1691.

8. Abell TL, Camilleri M, Donohoe K, et al. Consensus Recommendations for Gastric Emptying Scintigraphy. Am J Gastro 2008;103:753-763.

9. Tack J, Coulie B, Wilmer A, Andrioli A, Janssens J. Influence of sumatriptan on gastric fundus tone and on the perception of gastric distension in man. Gut 2000;46:468-73.

10. Sekino Y, Vamada E, Sakai E, Ohkubo H, Higurashi T, et al. Influence of sumatriptain on gastric accommodation and on antral contraction in healthy subjects assessed by ultrasonography. Neurogastroenterol Motil 2012;24:1083-1089.

11. Tack J, Janssen P, Masaoka T, Farré R, Van Oudenhove L. Efficacy of buspirone, a fundus-relaxing drug, in patients with functional dyspepsia. Clin Gastroenterol Hepatol 2012;10:1239-45

12. 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;95:1456-62.

13. Donohoe KJ, Maurer AH, Ziessman HA, et al. Procedure guideline for adult gastric emptying study 3.0. J Nucl Med Technol 2009; 37; 196-200.

14. Kundel HL, Polansky M. Mixture Distribution and Receiver Operating Characteristic Analysis of Bedside Chest Imaging with Screen-Film and Computed Radiography. Acad Radiol 4:1-7. 1997

15. Piessevaux H, Tack J, Walrand S, Pauwels S, Geubel A. Intragastric distribution of a standardized meal in health and functional dyspepsia: correlation with specific symptoms. Neurgastroenterol Motil 2003;15:447-455.

16. Tomita T, Okugawa T, Yamasaki T, Kondo T, Toyoshima F, Sakurai J, Oshima T, Fukui H, Daimon T, Watari J, Kashiwagi T, Matsumoto T, Miwa H. Use of scintigraphy to evaluate gastric accommodation and emptying: comparison with barostat. J Gastroenterol Hepatol 2013;28:106–111.

17. Arasu S, Parkman HP, Maurer AH. Assessment of Gastric Accommodation during Gastric Emptying Scintigraphy: Relationship to Symptoms. Gastroenterology 2014;146:S787 (abstract).
18. Parkman HP, Hallinan EK, Hasler WL, Farrugia G, Koch KL, Calles J, Snape WJ, Abell TL, Sarosiek I, McCallum RW, Nguyen L, Pasricha PJ, Clarke J, Miriel L, Lee L, Tonascia J, Hamilton F; NIDDK Gastroparesis Clinical Research Consortium (GpCRC). Nausea and vomiting in gastroparesis: similarities and differences in idiopathic and diabetic gastroparesis. Neurogastroenterol Motil 2016;28:1902-1914.
19. Rentz AM, Kahrilas P, Stanghellini V, et al. Development and psychometric evaluation of the patient assessment of upper gastrointestinal symptom severity index (PAGI-SYM) in patients with upper gastrointestinal disorders. Qual Life Res 2004;13:1737-49.
20. Revicki DA, Rentz AM, Dubois D, et al. Development and validation of a patient-assessed gastroparesis symptom severity measure: the Gastroparesis Cardinal Symptom Index. Aliment Pharmacol Ther 2003;18:141-50.
21. Koch KL, Hong SP, Xu L**.** Reproducibility of gastric myoelectrical activity and the water load test in patients with dysmotility-like dyspepsia symptoms and in control subjects. J Clin Gastroenterol 2000;31:125-9.
22. Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the EM algorithm. J Royal Stat Soc 1977;B39:1-38.
23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–174.
24. Altman DG. 1991, Practical Statistics for Medical Research. London: Chapman and Hall.
25. Allison PD. Logistic Regression Using SAS: Theory and Application, Second Edition. 2012.
26. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. 2005. Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models.
27. Hosmer DW, Jr., Lemeshow S, Sturdivant RX. 2013. Applied Logistic Regression, 3rd ed. John Wiley & Sons, Inc., Hoboken, New Jersey.
28. Youden, WJ. Index for rating diagnostic tests. Cancer 1950;3:32–35.
29. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cut-points using two ROC based criteria. Am J Epidemiology 2006;163:670–675.
30. Greiner M. Two-graph receiver operating characteristic (TG-ROC): a Microsoft-EXCEL template for the selection of cut-off values in diagnostic tests. J Immunological Methods 1995;185,145–146.
31. Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control 1974;19:716-723.
32. Hosmer D, Lemeshow S. Applied Logistic Regression. Second ed. New York: John Wiley & Sons, Inc.; 2000.

Figure 1. Examples of normal and abnormal fundic accommodation assessed using gastric emptying scintigraphy. Figure 1A demonstrates normal fundic accommodation with the majority of radiolabeled solids in the proximal stomach immediately post meal ingestion (t = 0 min). Over time there is progression of the solids into the distal stomach. Figure 1B demonstrates abnormal fundic accommodation with the majority of radiolabeled solid appearing in the distal stomach at t = 0 min.

Figure 1A:

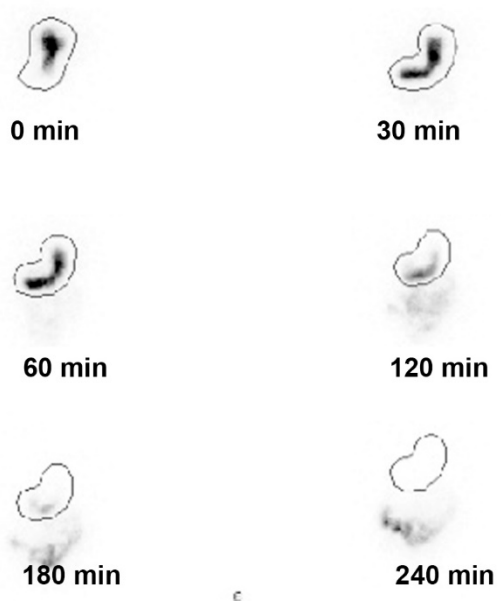


Figure 1B:

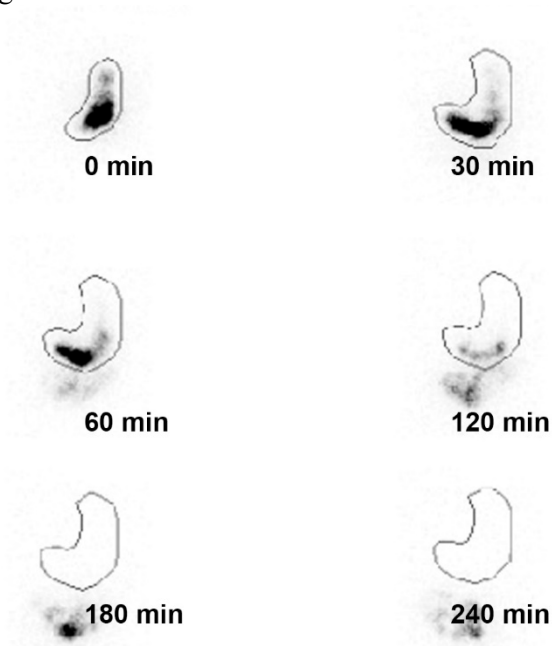


Figure 2. Three methods to divide the stomach into proximal and distal portions. Figure 2A shows how the computer generated ROIs for the proximal and distal stomachs (solid lines) are defined by dividing the stomach at ½ the distance along the long axis of the stomach (dotted line). Figure 2B similarly demonstrates how the computer generated ROIs are generated by selecting equal 1/3 divisions along the long axis of the stomach. Figure 2C. The stomach incisura angularis is the site of acute angle formation located on the lesser curvature (arrow) which forms a localized "notch". The location of the incisura will vary depending on the degree of gastric distention and is therefore more difficult to consistently localize.

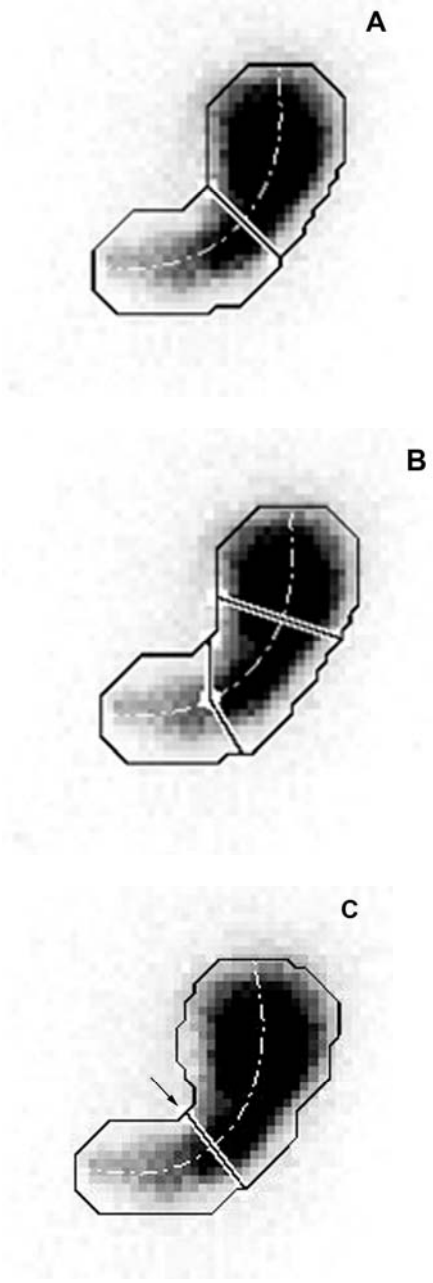


Figure 3. Intragastric Meal Distribution (IMD) over Time after meal ingestion. Shown are the values for IMD over time for normal volunteers (Figure 3A) and for abnormal patients as assessed by the readers' assessment of fundic accommodation (Figure 3B). The means are shown, along with +/- 1 standard deviation at each recorded time.

Figure 3A.

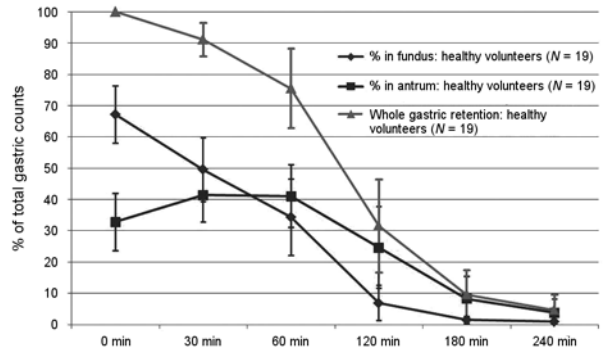


Figure 3B.

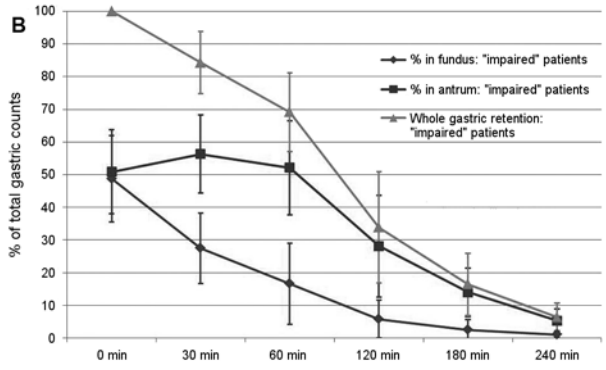
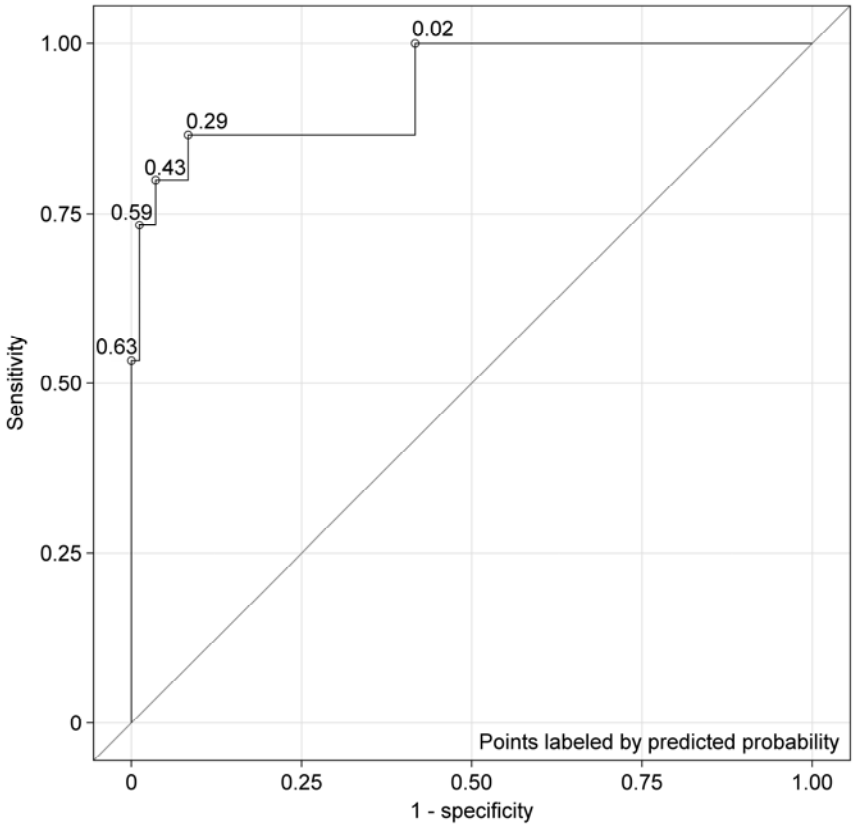


Figure 4. Logistic Regression and ROC Curve Based on 99 Test Subjects Using the IMD^0 or Percent Retention at Baseline as A Predictor for Abnormal/Impaired Fundic Accommodation‡



‡The AUC/c-statistic for this ROC curve is 0.934, which implies that the corresponding logistic regression model offers an excellent fit to the data (25).

Table 1. Agreement of assessment of fundic accommodation among the four Nuclear Medicine/Radiology physicians readers evaluating 99 subjects †

Entire Panel Judgment§	Number of Readers Giving a Positive Report				
	4	3	2	1	0
Positive (impaired/abnormal)	7	8	0	0	0
Negative (normal)	0	0	11	20	53
Overall: n (%)	7 (7.1%)	8 (8.1%)	11 (11.1%)	20 (20.2%)	53 (53.5%)

§Images that were classified as abnormal/impaired/positive by at least 3 out of the four Nuclear Medicine/Radiology Physicians readers were considered true positives; all others were considered true normal/negatives.

Table 2. Kappa values and their confidence intervals (CIs) between readers for assessing fundic accommodation on test subjects (n=99)

Summary of Weighted Kappas Using the Five-Level Scale Assessment for Fundic Accommodation

Pair of Readers from	*Weighted Kappa (95% CI)*
TUH vs. Wake	0.230 (0.122 - 0.338)
TUH vs. Stanford	0.410 (0.267 - 0.554)
TUH vs. JHH	0.423 (0.246 - 0.600)
Wake vs. Stanford	0.438 (0.325 - 0.551)
Wake vs. JHH	0.293 (0.190 - 0.396)
Stanford vs. JHH	0.483 (0.350 - 0.616)
Overall Average:	***0.380 (0.250 – 0.509)***

Summary of Simple Kappas Using the Two-Level Scale Assessment for Fundic Accommodation

Pair of Readers from	*Simple Kappa (95% CI)*
TUH vs. Wake	0.230 (0.097 - 0.364)
TUH vs. Stanford	0.457 (0.252 - 0.661)
TUH vs. JHH	0.476 (0.230 - 0.722)
Wake vs. Stanford	0.525 (0.364 - 0.685)
Wake vs. JHH	0.381 (0.223 - 0.539)
Stanford vs. JHH	0.521 (0.320 - 0.722)
Overall Average:	***0.432 (0.248 – 0.616)***

Appendix 1. Reader Training Interpretive Guide. The following were provided to train the readers how to visually assess fundic accommodation from routine solid-meal gastric emptying scintigraphy.

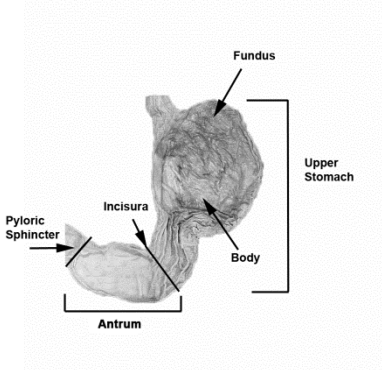


Figure 1: Anatomy. Anatomic and functional correlates used to define gastric regions of interest for fundic accommodation (FA). The proximal stomach includes the fundus to the acute angle of the incisura. After meal ingestion and initial FA of solids the majority of solids (> 50%) should appear in the upper stomach. For confirmation of upper from distal stomach, one should see progression of the solids past the incisura into the antrum in later gastric emptying images.

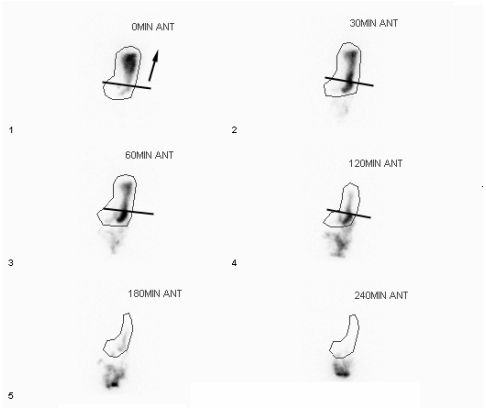


Figure 2: Normal accommodation with normal gastric emptying. Anterior views are shown to demonstrate normal FA with predominant (> 50%) visualization of the radiolabeled solids in the proximal stomach (arrow). The heavy straight line shows where the visualized lower level of the upper stomach is taken using the incisura as a reference. With increasing time (30 min, 60 min, and 120 min), the solids can be seen to progress distally into the antrum which is below the incisura. The thin black line is included to show the large region of interest drawn around the entire stomach used to calculate the total gastric emptying which was normal.

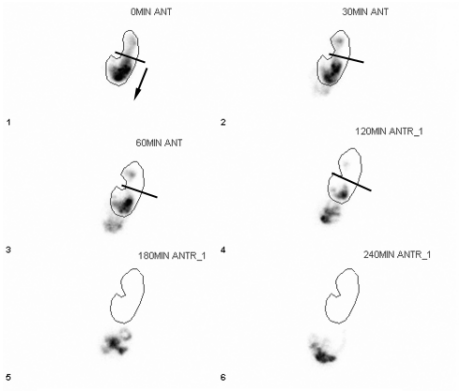


Figure 3: Impaired gastric accommodation with normal gastric emptying. The thin black line is the total gastric region of interest, the thick black line is visual reference to define the proximal stomach and the black arrow now demonstrates that > 50% of the solids are initially (t=0 min) in the distal stomach consistent with lack of normal FA. Measured total gastric emptying was normal.

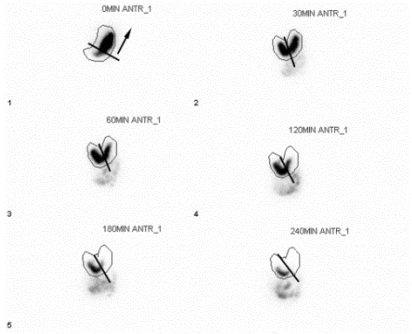


Figure 4: Normal accommodation with delayed gastric emptying. The thin black line is total gastric region of interest, thick black line is visual reference to define the proximal stomach and the black arrow demonstrates normal FA with > 50% of the solids immediately (t=0 min) in the proximal stomach consistent with normal FA. There is normal progression of solids into the antrum but measured total gastric emptying; however, was delayed (70% retained at 2 hours and 18% at 4 hours).

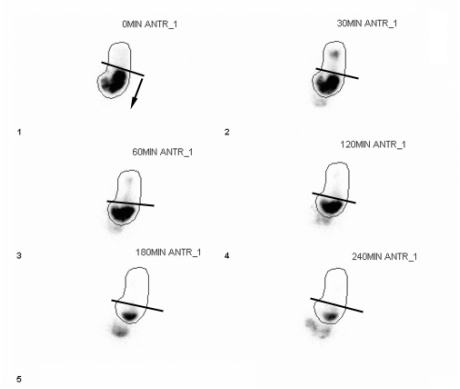


Figure 5: Abnormal accommodation with delayed gastric emptying. The thin black line is total gastric region of interest, the thick black line is visual reference to define proximal stomach and the black arrow demonstrates abnormal FA with >50% of the solids initially (t=0 min) in the distal stomach consistent with abnormal FA response. There is persistent retention of solids in the antrum and measured total gastric emptying was delayed (80% retained at 2 hours and 22% at 4 hours).

Appendix 2. Description of computer software

The software developed to compute IMD combines and analyzes the gastric images from all time points and uses MatLab® (MathWorks®, Natick, Massachusetts), a licensed software development tool. There are four fundamental steps that the software takes to process the scintigraphic images. Step one is opening, selecting, and coregistering the images. Step two is finding the gastric region of interest (ROI) boundaries by thresholding and modifying boundaries manually if necessary. Step three is calculating the longitudinal axis and separating the whole-stomach region of interest (ROI) into separate sections. Finally, step four calculates the gastric counts in each region images obtained at the different times including options at: 0, 30, 60, 120, 180, 240 minutes.

The first step is opening, selecting, and coregistering a composite image set for image alignment. The graphic user interface (GUI) displays all the study images on the screen. The user selects images and can rotate each image if the images are angled askew to optimize alignment. The user drags each image in succession to line up correctly with the 0-minute image. This ensures that the boundaries and regions drawn are positioned appropriately for all images in the set.

The second step defines the boundaries to determine the whole-stomach ROI. First, the software performs a first iteration of the boundaries of the stomach by finding the boundary around the portion of the stomach visible in each image. The software automatically creates a region of interest (ROI) around each individual image using a threshold of 0.25 based on the maximum gastric counts in each image. This can then be reviewed and adjusted for a best fit, if needed, using the graphic user interface (GUI). After creating these ROIs, the software smooths the edges to simplify the boundaries. The software then combines all images and ROIs to reconstruct a final total gastric ROI that contains all the gastric activity seen from time 0 to 4 hours (Figure 1 of Appendix 2). The operator can manually modify the calculated boundary if needed by smoothing the edges, expanding it in all directions, or in specified locations, called “bumping” the boundary in the GUI.

The third step calculates the longitudinal axis, and separates the whole-stomach ROI into equal halves. To produce the long axis, the software finds the two points on the stomach boundary which are most distal, and most proximal. Moving from the most distal point, the computer calculates the midpoints between the boundaries on the combined gastric image (Figure 2 of Appendix 2). Using this method, the computer plots approximately fifty midpoints which are located close to the radial center of the stomach, from the most distal point to the most proximal point of the stomach. A best-fit third degree polynomial curve is calculated along these points forming the longitudinal axis. The stomach is divided by taking the midpoint of the longitudinal axis. The line used for separation is perpendicular to the longitudinal axis at its midpoint, dividing the stomach into two regions: a proximal region and a distal region (Figure 3 of Appendix 2).

Step four is calculating the gastric counts in each region. The proximal and distal ROIs, are overlaid onto the original, aligned images at each time, and the number of counts in the proximal region and the number of counts in the distal region are obtained, for each image. The ratio of the gastric count in the proximal region to the total gastric count in the whole stomach ROI at the 0-minute image is defined as the intragastric meal distribution (IMD^0).

Appendix 2 Figures

Figure 1. The software creates a region around each individual image (multiple colored ROIs) using thresholding, combines the images, then expands or contracts the boundaries with user input to create a single (white ROI) for the final composite image.

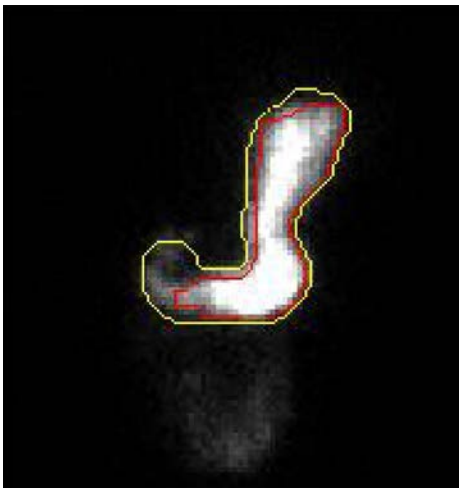


Figure 2. The computer software constructs the long axis by finding midpoints(red dots) between a line drawn between the gastric boundaries (x,x) on the composite image (Figure 1).

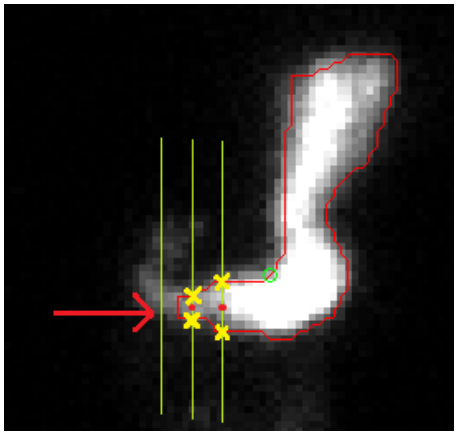
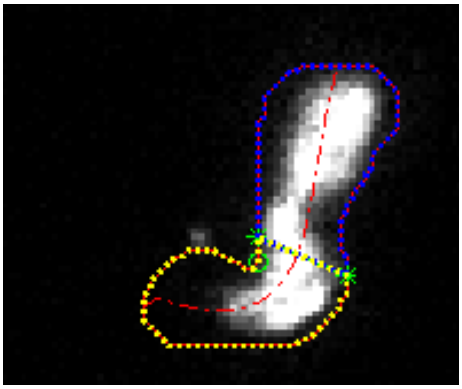


Figure 3. The software then computes the length of the long axis and draws a line perpendicular to the long axis (between green x-x) where the stomach is separated into two equal halfs. After the stomach is segmented into proximal and distal halfs, the counts in each region are recorded and IMD calculated.



Appendix 3. Mixture Distribution Analysis Statistical Analysis Methods

Mixture distribution analysis is a numerically intense, probabilistic approach to estimate the "positivity" and its "reliability" or panel consistency of image readings based on a sample of patients' images by a group of expert readers, where independent verification of the imaging diagnosis does not exist. Agreement is considered in the context of one reader agreeing with a panel of readers rather than just another reader. Fundamental to this approach is that the 'true' reading is not known. Within this context, the reading made by a panel of readers is the best reading possible. Therefore, mixture distribution analysis assesses the positivity and its panel consistency (i.e., whether a large group of readers could make a positive or negative diagnosis consistently in a reproducible manner) rather than accuracy (i.e., whether a reader's diagnosis is correct as measured against a known truth or gold standard). For this approach, panel consistency is expressed as the relative percentage "correctly" or consistently diagnosed.

Specifically, images are conceptually categorized by how easy they are to read (i.e., easy or hard) and whether those images are disease-positive or disease-negative. Thus, images can be easy disease-positive, hard disease-positive, hard disease-negative, and easy disease-negative. Based on this categorization, Kundel and Polansky (14) utilize the following parameterization in the MDA analysis that is used to define the mixture distribution on images and readers: p_i is the proportion of images in each group in the target population, where

- p_1 is the proportion of easy disease-positive images,
- p_2 is the proportion of hard disease-positive and hard disease-negative images, and
- p_3 the proportion of easy disease-negative images.

Furthermore, m_i represents the probability of a large group of readers agreeing that the image is positive given the image is from group *i* corresponding to p_i. Intuitively, if 80% of the readers conclude that disease is present (i.e., $m_1 = 0.8$), then the image is classified as easy disease-positive; if 50% of the readers conclude that disease is present (i.e., $m_2 = 0.5$), then the image is classified as either hard disease-positive or hard disease-negative--in this case, note that there is no definitive "radiologic truth" because it is impossible to distinguish positive from negative; and if only 20% of the readers conclude that disease is present (i.e., $m_3 = 0.2$), then the image is classified as easy disease-negative since 80% of the readers conclude that the disease is absent. Thus, the probability that exactly *r* out of *n* readers will classify an image as positive assuming it is randomly selected from all the possible target images rather than specifically from one of the three groups (or if it is not known which group the image comes from) is:

$$\frac{n!}{r!\,(n-r)!}\sum_{i=1}^{3} p_i m_i r(1-m_i)^{n-r}$$

The likelihood function therefore can be written down based on a set of observed values of r among n readers for a random sample of images taken from the target population. The maximum likelihood estimates of the *m*'s and *p*'s and their 95% confidence intervals can then be found using the expected maximization (EM) algorithm. Furthermore, overall proportion correct (% of "correct" or consistent diagnosis) can be calculated as per Kundel, et. al.:

$$p_1 m_1 + p_2(1 - m_2) + p_3(1 - m_3)$$

Kappa analysis (23,24) was also employed to assess the agreement in assessments of fundic accommodation by any two readers at a time using the scoring 1=definitely abnormal accommodation, 2=probably abnormal accommodation, 3=possibly abnormal accommodation, 4=probably normal accommodation, and 5=definitely normal accommodation. We also simplified this approach by using the condensed scoring of abnormal/impaired/positive accommodation (scores 1, 2, 3) and normal/negative accommodation (scores 4, 5). Both weighted or simple Kappas and their 95% confidence intervals were reported for these two scoring methods. The average of all the pairwise Kappas provided some evidence on the strength of pairwise agreement the four readers evaluated the images.

Note that agreement and consistency refer to slightly different "concepts" here: agreement simply means how agreeable the FA assessments by any two different readers are; whereas consistency refers to the degree a group of readers reach their "correct" consensus under the framework of the Kundel's Mixture Distribution Analysis Approach (i.e., % correct). In essence, Kappa measures the pairwise agreement while Kundel's MDA yields information on panel consistency for the readers' scoring. Kappa is a well-known summary statistic while Kundel's concept is less known. Therefore, Kappa analyses have been used for making the main findings with consistency as a supplement.

Abnormal FA as defined by the expert panel consensus was then compared to computer-derived IMD data via the use of logistic regression and ROC analyses to estimate an optimal cutpoint of IMD for the purposes of diagnosing impaired FA. In these analyses, images that were classified as abnormal/impaired/positive by at least 3 out of the four readers were considered true positives; all others were considered true normal/negatives. Logistic regression model performance can be evaluated based on the area under the ROC curve, also known as the AUC or the c-statistic. The AUC or c-statistic is a measure of classification "accuracy" or "discrimination", that is, the ability of the model to correctly discriminate cases (i.e., positive reads) from non-cases (i.e., negative reads), or the predictive power of the model (25,26,27). Three approaches were used to identify an optimal cutpoint of IMD for classifying impaired/abnormal FA: 1) Maximum Youden's index J (i.e., max{Sensitivity + Specificity – 1}) (Table 2). Under this approach, the optimal cutpoint is at the point of the maximum vertical distance from the diagonal line to the ROC curve (28); 2) The closest-to-(0,1) criterion – The optimal cutpoint is defined as the point of the minimum distance from the ROC curve to the point where the diagnosis is perfect, i.e.,

sensitivity=1 and specificity=1 (the upper-left corner of the ROC plot) (min{sqrt((1-sensitivity)^2+(1-specificity)^2)}) (29); and 3) Sensitivity and Specificity Equality criterion (min{abs(Sensitivity - Specificity)}) – the optimal cutpoint is chosen to be at the point of minimum absolute difference between sensitivity and specificity under this criterion (30). As these approaches are all defined in terms of sensitivity and specificity, corresponding IMD cutpoints (i.e., % retention at time 0) associated with these optimal choices were calculated using the coefficients from the logistic regression model used to derive the ROC curve to facilitate impaired fundic accommodation diagnosis.

Statistical Results Tables

Table 1 of Appendix 3. Estimated proportion of images in each group (Mixing Proportions) and the probability that a large group of readers concludes "Positive" impaired fundic accommodation (Point Distributions) and CIs using Kundel's Mixture Distribution analysis (n=99 Test Subjects)†

	Mixing Proportions			Point Distributions			Proportion Correct
	p_1	p_2	p_3	m_1	m_2	m_3	
Mean	0.109	0.267	0.624	0.872	0.410	0.053	0.844
95% CI	(0.030, 0.223)	(0.067, 0.578)	(0.363, 0.753)	(0.602, 1.000)	(0.119, 0.734)	(0.000, 0.113)	(0.779, 0.894)

†Table entries are the means and 95% bootstrap confidence intervals calculated with the EM algorithm. CI = confidence interval.

Table 2 of Appendix 3. Optimal cutpoint for the IMD^0 at baseline by approach based on data from 99 test subjects$

Approach	*Optimal Cutpoint*	*Sensitivity*	*Specificity*	*No. of Correctly Predicted Events*	*No. of Correctly Predicted Nonevents*	*No. of Nonevents Predicted as Events*	*No. of Events Predicted as Nonevents*
Youden's Index J	0.568	0.867	0.917	13	77	7	2
The closest-to-(0,1)	0.568	0.867	0.917	13	77	7	2
Sensitivity and Specificity Equality	0.590	0.867	0.869	13	73	11	2

$Based on these analyses, the optimal cutpoint for the IMD^0 or % retention at time 0 is 0.568 since there is no obvious advantage with regard to sensitivity or specificity by selecting 0.590 as the cutpoint.

Appendix 4. Results of IMD analysis in 177 patients in the NIH Gastroparesis Registry

Figure 1 of Appendix 4. Distribution of the IMD^0 for 177 patients of the NIH Gastroparesis Registry

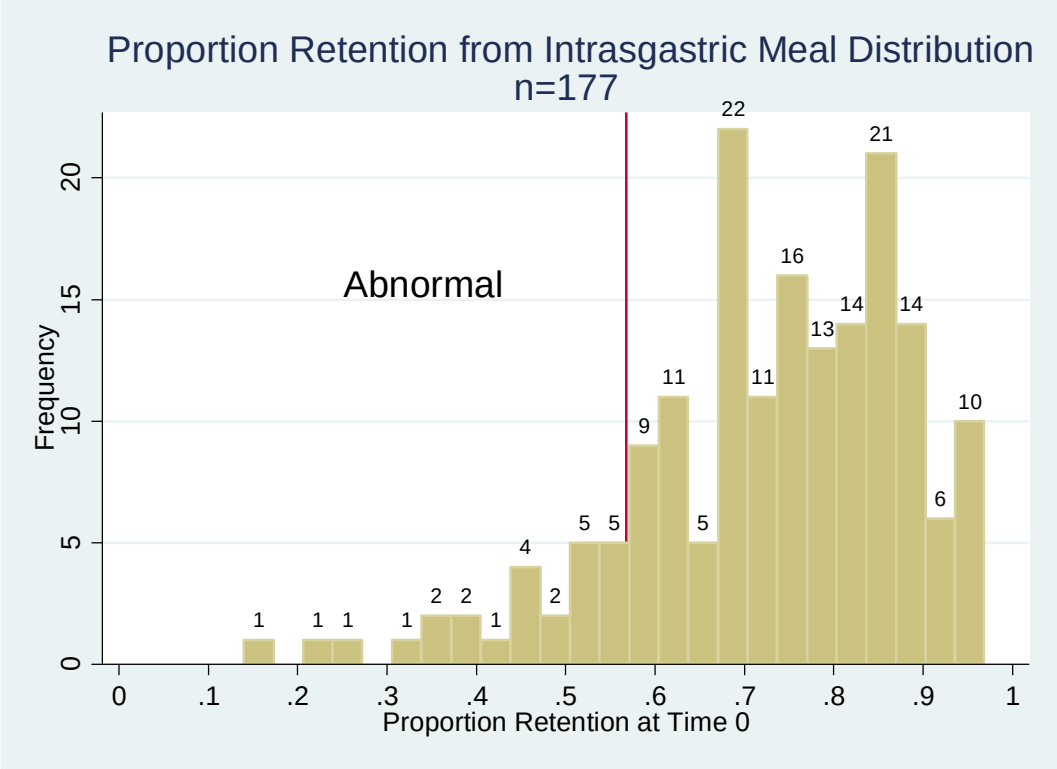


Table 1 of Appendix 4. Association of intragastric meal distribution status (normal vs. borderline vs. impaired) with characteristics of patients with gastroparesis at screening visit

	Intragastric Meal Distribution‡				
	Normal (n=132)	**Borderline* (n=20)**	**Impaired* (n=25)**	**Total (n=177)**	**p-value**
Demographics					
Age – yrs	42 (14)	45 (18)	47 (17)	43 (15)	0.10
Male gender	14%	0%	16%	13%	0.73
Anthropometric					
BMI – kg/m2	27 (7)	26 (8)	23 (6)	27 (7)	0.006
Weight change since Gp dx					0.06
Decrease	47%	50%	72%	51%	
Same	3%	10%	4%	4%	
Increase	50%	40%	24%	45%	

Weight change in past 6 mo					0.24
Decrease	42%	35%	56%	43%	
Same	26%	45%	16%	27%	
Increase	33%	20%	28%	31%	
Metabolic					
HbA1c - %	6.5 (1.9)	6.0 (1.7)	6.0 (1.6)	6.4 (1.8)	0.15
Gp characteristics					
Etiology					0.37
Diabetes	33%	25%	16%	30%	
Idiopathic	64%	75%	80%	67%	
Fundoplication	3%	0%	4%	3%	
Acute onset of symptoms	35%	40%	36%	36%	0.81
Initial infectious prodrome	23%	25%	12%	22%	0.27
Duration – yrs	6.6 (6.9)	4.3 (4.0)	8.7 (9.7)	6.7 (7.2)	0.42
Gp symptom severity - (0-5)	3.0 (0.6)	2.8 (0.5)	2.9 (0.5)	3.0 (0.5)	0.46
Gp severity - % severe	15%	10%	8%	14%	0.29
Predominant symptom					0.13
Nausea / vomiting	31%	50%	56%	37%	
Abdominal pain	23%	15%	12%	21%	
Other	45%	35%	32%	42%	
Co-morbidities					
Diabetes	39%	25%	16%	34%	0.01
Type I	58% (30/52)	40% (2/5)	50% (2/4)	56% (34/61)	0.56
Post Nissan Fundoplication	5%	0%	8%	5%	0.83
Gastric emptying					
Delayed Gastric Emptying†	77%	65%	64%	73%	0.13
1 hr solid gastric retention - %	78 (16)	75 (12)	67 (18)	76 (16)	0.001
2 hr solid gastric retention - %	58 (22)	57 (19)	43 (20)	56 (22)	0.002
4 hr solid gastric retention - %	26 (20)	21 (24)	18 (14)	24 (20)	0.05
1 hr liquid gastric retention - %	48 (17)	46 (14)	44 (21)	47 (17)	0.29
PAGI-SYM at screening visit					
GCSI					
Nausea (0-5)	3.2 (1.5)	3.1 (1.7)	3.5 (1.0)	3.2 (1.4)	0.39
Retching (0-5)	1.5 (1.7)	1.0 (1.2)	1.5 (1.5)	1.4 (1.6)	0.64
Vomiting (0-5)	1.5 (1.8)	0.6 (1.3)	1.5 (1.7)	1.4 (1.7)	0.49
Nausea subscore (0-15)	6.1 (4.1)	4.7 (3.5)	6.4 (3.1)	6.0 (3.9)	0.86
Stomach fullness (0-5)	3.6 (1.3)	4.0 (1.4)	3.9 (0.9)	3.7 (1.3)	0.19
Not able to finish meal (0-5)	3.4 (1.5)	3.8 (1.7)	4.1 (1.0)	3.5 (1.5)	0.02
Feeling excessively full (0-5)	3.8 (1.3)	3.8 (1.6)	4.1 (1.0)	3.8 (1.3)	0.34
Loss of appetite (0-5)	2.8 (1.5)	2.7 (1.8)	3.0 (1.2)	2.8 (1.5)	0.53
Fullness subscore (0-20)	13.6 (4.6)	14.3 (5.6)	15.1 (3.1)	13.9 (4.6)	0.12
Bloating (0-5)	3.2 (1.6)	3.7 (1.5)	3.7 (1.2)	3.3 (1.5)	0.10
Stomach visibly larger (0-5)	2.9 (1.8)	3.4 (2.0)	3.3 (1.4)	3.0 (1.7)	0.16

Bloating subscore (0-10)	6.1 (3.2)	7.1 (3.4)	7.0 (2.4)	6.3 (3.1)	0.11
Total score (0-45)	25.8 (9.1)	26.1 (10.9)	28.5 (6.2)	26.2 (9.0)	0.20
Upper abdominal pain (0-10)	6.0 (2.9)	5.4 (3.5)	5.9 (2.6)	5.9 (2.9)	0.68
Lower abdominal pain (0-10)	3.9 (3.2)	4.2 (3.5)	4.7 (2.6)	4.1 (3.1)	0.28
GERD (0-35)	11.9 (9.6)	10.8 (9.3)	14.6 (7.6)	12.2 (9.3)	0.31
Constipation (0-5)	2.8 (1.7)	2.6 (2.0)	2.9 (1.5)	2.8 (1.7)	0.86
Diarrhea (0-5)	1.6 (1.6)	1.4 (1.7)	1.5 (1.8)	1.6 (1.6)	0.63
Satiety testing					
Water (mL)	368 (199)	313 (112)	413 (248)	368 (200)	0.53
Nutrient bar - % consumed	88 (22)	91 (16)	87 (23)	88 (22)	0.90
Medication use					
Prokinetics	30%	45%	20%	30%	0.66
Narcotics	33%	25%	28%	32%	0.48
Source of nutrition					0.34
Enteral	2%	0%	0%	2%	
Parenteral	0%	0%	0%	0%	
Oral	98%	100%	100%	98%	
Treatments					
Botox	25%	35%	44%	29%	0.04
G tube	2%	0%	0%	2%	0.34
J Tube	2%	0%	0%	2%	0.34
Central line	2%	0%	0%	2%	0.34
Gastric stimulator	8%	15%	0%	8%	0.33

‡Table entry = mean (SD) or %.

*Normal, borderline, and impaired intragastric meal distribution defined as (proximal counts / total time-0 ROI counts) at baseline at Gastric Emptying 0 min from >0.642, 0.568 to 0.642, and <0.568, respectively.

†Delayed Gastric Emptying defined as gastric retention >60% at 2 hours or >10% at 4 hours.

Appendix 5. Relationships of gastric emptying and water load testing to symptoms.

Table 1 of Appendix 5. Association of Gastric Retention at 4 hours with PAGI-SYM items (n=172)†

	Classification Based on Gastric retention at 4 hours‡			
PAGI-SYM items	**Very Abnormal >35% (n=43)**	**Moderately Abnormal >10-35% (n=80)**	**Normal ≤10% (n=49)**	**p-value***
GCSI				
Nausea (0-5)	3.4 (1.6)	3.1 (1.5)	3.2 (1.3)	0.57
Retching (0-5)	1.5 (1.8)	1.5 (1.6)	1.2 (1.6)	0.43
Vomiting (0-5)	1.9 (2.0)	1.3 (1.7)	1.1 (1.5)	0.03
Nausea subscore (0-15)	6.8 (4.3)	5.9 (3.9)	5.5 (3.5)	0.13
Stomach fullness (0-5)	3.8 (1.4)	3.6 (1.2)	3.7 (1.2)	0.80
Not able to finish meal (0-5)	3.5 (1.6)	3.5 (1.5)	3.5 (1.4)	1.00
Feeling excessively full (0-5)	4.0 (1.4)	3.8 (1.3)	3.8 (1.3)	0.58
Loss of appetite (0-5)	3.0 (1.4)	2.7 (1.5)	2.8 (1.5)	0.63
Fullness subscore (0-20)	14.2 (5.0)	13.5 (4.5)	13.8 (4.5)	0.71
Bloating (0-5)	3.5 (1.7)	3.1 (1.6)	3.4 (1.4)	0.85
Stomach visibly larger (0-5)	3.1 (1.8)	2.9 (1.6)	3.1 (1.8)	0.95
Bloating subscore (0-10)	6.6 (3.3)	6.1 (3.0)	6.5 (3.1)	0.90
Total score (0-45)	27.6 (9.8)	25.5 (9.2)	25.9 (8.2)	0.37
Upper abdominal pain (0-10)	6.3 (3.1)	5.5 (2.8)	5.9 (3.0)	0.53
Lower abdominal pain (0-10)	4.2 (3.4)	4.0 (3.0)	4.0 (3.2)	0.86
GERD (0-35)	13.1 (9.2)	11.7 (9.6)	11.5 (9.0)	0.43
Constipation (0-5)	3.3(1.7)	2.5 (1.7)	2.8 (1.7)	0.19
Diarrhea (0-5)	1.4 (1.7)	1.6 (1.7)	1.4 (1.5)	0.99

*Based on linear regression of each PAGI-SYM item on ordered classification categories of gastric retention at 4 hours.

†5 patients missing gastric retention at 4 hours data.

‡Table entry = mean (SD) severity of the PAGI-SYM item.

Table 2 of Appendix 5. Association of Satiety Water Load with PAGI-SYM items (n=173)†

	Classification Based on Satiety Water Load‡			
PAGI-SYM items	**Very Abnormal ≤ 240 mL (n=51)**	**Moderately Abnormal 241-500 mL (n=94)**	**Normal 501+ mL (n=28)**	**p-value***
GCSI				
Nausea (0-5)	3.5 (1.4)	3.1 (1.5)	2.9 (1.4)	0.05
Retching (0-5)	1.4 (1.6)	1.3 (1.7)	1.6 (1.6)	0.71
Vomiting (0-5)	1.6 (1.7)	1.3 (1.7)	1.4 (1.7)	0.43
Nausea subscore (0-15)	6.6 (3.8)	5.7 (4.0)	5.9 (3.9)	0.37
Stomach fullness (0-5)	3.7 (1.2)	3.7 (1.3)	3.5 (1.1)	0.75
Not able to finish meal (0-5)	3.7 (1.4)	3.6 (1.5)	3.2 (1.7)	0.23
Feeling excessively full (0-5)	3.9 (1.3)	3.9 (1.3)	3.6 (1.4)	0.31
Loss of appetite (0-5)	3.2 (1.4)	2.6 (1.5)	2.7 (1.5)	0.06
Fullness subscore (0-20)	14.5 (4.7)	13.8 (4.5)	13.0 (5.0)	0.17
Bloating (0-5)	3.4 (1.6)	3.3 (1.5)	3.1 (1.4)	0.42
Stomach visibly larger (0-5)	3.1 (1.7)	3.0 (1.8)	2.8 (1.7)	0.35
Bloating subscore (0-10)	6.5 (3.2)	6.3 (3.2)	5.8 (2.9)	0.36
Total score (0-45)	27.6 (8.7)	25.8 (9.1)	24.8 (9.5)	0.16
Upper abdominal pain (0-10)	6.5 (2.8)	5.6 (3.0)	5.5 (2.8)	0.09
Lower abdominal pain (0-10)	5.4 (2.9)	3.9 (3.1)	2.4 (2.9)	<0.0001
GERD (0-35)	12.9 (9.8)	11.4 (.1)	13.2 (9.2)	0.89
Constipation (0-5)	3.0 (1.6)	2.7 (1.8)	2.5 (1.8)	0.17
Diarrhea (0-5)	2.1 (1.6)	1.3 (1.6)	1.2 (1.7)	0.01

*Based on linear regression of each PAGI-SYM item on ordered classification categories of satiety water load.
†4 patients missing satiety water load data.
‡Table entry = mean (SD) severity of the PAGI-SYM item.