ENDOSCOPIC PYLOROMYOTOMY FOR THE TREATMENT OF SEVERE AND REFRACTORY GASTROPARESIS: A PILOT, RANDOMIZED, SHAM-CONTROLLED TRIAL

Martinek Jan, Hustak Rastislav, Mares Jan, Vackova Zuzana, Spicak Julius, Kieslichova Eva, Buncova Marie, Pohl Daniel, Amin Sunil, Tack Jan

Supplementary Appendix

Martinek J, *et al. Gut* 2022;0:1–9. doi: 10.1136/gutjnl-2022-326904

Suppl Table S1. Table of contents

SUPPLEMENTARY TABLES

Suppl Table S2. List of Contributors

Name	Affiliation	City, Country
Jan Martinek	Department of Hepatogastroenterology, Institute for Clinical and Experimental Medicine (IKEM)	Prague, Czech Republic
Rastislav Hustak	Department of Internal Medicine, University Hospital Trnava	Trnava, Slovak Republic
Jan Mares	Department of IT and biostatistic, Institute for Clinical and Experimental Medicine (IKEM)	Prague, Czech Republic
Zuzana Vackova	Department of Hepatogastroenterology, Institute for Clinical and Experimental Medicine (IKEM)	Prague, Czech Republic
Julius Spicak	Department of Hepatogastroenterology, Institute for Clinical and Experimental Medicine (IKEM)	Prague, Czech Republic
Eva Kieslichova	Department of Anesthesiology and Intensive Care, Institute for Clinical and Experimental Medicine (IKEM)	Prague, Czech Republic
Marie Buncova	Department of Nuclear Medicine, Institute for Clinical and Experimental Medicine (IKEM)	Prague, Czech Republic
Daniel Pohl	Division of Gastroenterology and Hepatology, University Hospital Zurich	Zurich, Switzerland
Amin Sunil	Division of Digestive Health and Liver Diseases, University of Miami Miller School of Medicine	Miami, Florida, USA
Jan Tack	Department of Gastroenterology and Hepatology, University Hospitals Leuven	Leuven, Belgium
Jan Usak	Department of Internal Medicine, University Hospital Trnava	Trnava, Slovak Republic
Martin Janicko	First Department of Internal Medicine, PJ Safarik University	Kosice, Slovak Republic
Tereza Malkova	Department of Hepatogastroenterology, Institute for Clinical and Experimental Medicine (IKEM)	Prague, Czech Republic
Monika Horackova	Department of Hepatogastroenterology, Institute for Clinical and Experimental Medicine (IKEM)	Prague, Czech Republic
Helena Pitelkova	Department of Hepatogastroenterology, Institute for Clinical and Experimental Medicine (IKEM)	Prague, Czech Republic
Adela Kreckova	Department of Hepatogastroenterology, Institute for Clinical and Experimental Medicine (IKEM)	Prague, Czech Republic
Gabriela Petranova	Axon CRO Ltd.	Prague, Czech Republic
Emilia Oleksakova	Department of Internal Medicine, University Hospital Trnava	Trnava, Slovak Republic
Eva Kolarovicova	Department of Internal Medicine, University Hospital Trnava	Trnava, Slovak Republic

Suppl Table S3. Authors contribution	
Jan Martinek	study concept and study design, patient´s recruitment, performing all G-POEM procedures, follow-up endoscopies, drafting the manuscript
Rastislav Hustak	study design, data collection, assisting the procedures, follow-up endoscopies, drafting the manuscript
Jan Mares	statistical analysis, study design, drafting the manuscript
Zuzana Vackova	patient´s recruitment and selection, measurement and analysis of pyloric distensibility measurement, follow-up endoscopies, critical review of the manuscript
Julius Spicak	follow-up endoscopies, critical review of the manuscript
Eva Kieslichova	anesthesia during procedures, analysis of AEs, critical review of the manuscript
Marie Buncova	gastric emptying studies - analysis and critical review of the manuscript
Daniel Pohl	Endoflip measurement and analysis, study design
Sunil Amin	critical review of the manuscript, language corrections
Jan Tack	study concept, study design, data analysis and interpretation, critical review of the manuscript
Jan Usak	patient´s recruitment and selection
Martin Janicko	statistical advisor, study design
Tereza Malkova	endoscopic nurse, assisting the procedures
Monika Horackova	head nurse, post-operative care, follow-up visits
Helena Pitelkova	head nurse, post-operative care, follow-up visits
Adela Kreckova	Study administration, CRFs - data collection
Gabriela Petranova	study monitoring (Axon CRO Ltd.)
Emilia Oleksakova	endoscopic nurse, assisting the procedures
Eva Kolarovicova	study nurse, postoperative care, follow-up visits

Suppl Table S4a. Study design for patients in the active (G-POEM) group

	Baseline	*POD 0 – day of G-POEM*	*POD 1*	*3M*	*6M*	*12M*	*24M*	*36M*
Scintigraphy	✓			✓	-	✓		✓ (optional)
Endoscopy	✓	✓	✓ (optional)	✓	-	✓ (optional)		
GCSI + PAGI-SYM + PAGI-QoL	✓			✓	✓	✓	✓	✓
Blood tests	✓		✓					
Endoflip		✓ *Before and after G-POEM*		✓				

GCSI = Gastroparesis Cardinal Symptom Index (see also suppl table S5)
PAGI-SYM = Patient Assessment of Upper Gastrointestinal Symptom Severity Index (see also suppl table S9)
PAGI QoL = Patient Assessment of Upper Gastrointestinal Disorders-Quality of Life (see also suppl table S10)
POD = postoperative day
3M, 6M, 12M, 24M, 36M = 3 months, 6 months, 12 months, 24 months, and 36 months visit

Suppl Table S4b. Study design for patients randomized in the sham group*

	Baseline	*POD 0*	*POD 1*	*3M*	***6M****	*POD 0/* ***G-POEM***	***POD1***	*3M*	*6M*	*12M*	*24M*	*36M*
Scintigraphy	✓			✓	-			✓				✓ (optional)
Endoscopy	✓	✓	✓ (optional)	✓	-		✓ (optional)	✓		✓ (optional)		
GCSI + PAGI-SYM, PAGI-QoL	✓			✓	✓			✓	✓	✓	✓	✓
Blood tests	✓		✓				✓					
Endoflip		✓ *Before***		✓♦		✓ *After G-POEM*		✓				

***** *At 6 months, the patients will be offered cross-over G-POEM (if symptoms persist).*

****** *In patients having undergone Endoflip during the sham procedure, no Endoflip measurement will be repeated prior to G-POEM.*

♦ *sham measurement*

GCSI = Gastroparesis Cardinal Symptom Index (see also suppl table S5)
PAGI-SYM = Patient Assessment of Upper Gastrointestinal Symptom Severity Index (see also suppl table S9)
PAGI QoL = Patient Assessment of Upper Gastrointestinal Disorders-Quality of Life (see also suppl table S10)
POD = postoperative day
3M, 6M, 12M, 24M, 36M = 3 months, 6 months, 12 months, 24 months, and 36 months visit

Suppl Table S5. Gastroparesis Cardinal Symptom Index (GCSI)

	Symptoms	Score					
		0	1	2	3	4	5
1	Nausea	None	Very mild	Mild	Moderate	Severe	Very severe
2	Retching	None	Very mild	Mild	Moderate	Severe	Very severe
3	Vomiting	None	Very mild	Mild	Moderate	Severe	Very severe
4	Stomach fullness	None	Very mild	Mild	Moderate	Severe	Very severe
5	Not able to finish a normal sized meal	None	Very mild	Mild	Moderate	Severe	Very severe
6	Feeling extensively full after meals	None	Very mild	Mild	Moderate	Severe	Very severe
7	Loss of appetite	None	Very mild	Mild	Moderate	Severe	Very severe
8	Bloating	None	Very mild	Mild	Moderate	Severe	Very severe
9	Stomach or belly visibly larger	None	Very mild	Mild	Moderate	Severe	Very severe

The GCSI consists of nine items and three subscales to measure symptoms related to gastroparesis. The nausea/vomiting subscale consists of the following three items: nausea, retching, and vomiting. The postprandial fullness/early satiety subscale consists of the following four items: stomach fullness, inability to finish a normal-sized meal, feeling excessively full after meals, and loss of appetite. The bloating subscale consists of the following items: bloating and stomach or belly visibly larger. The GCSI total score is constructed as the average of the three symptom subscales. Its value ranges from zero meaning no symptoms to five indicating maximally severe symptomatology (see ref. No. 18-19 in the main article).

Calculation:
Total GSCI score = arithmetic mean of the three symptom subscales. Subscores = arithmetic means of (1-3), (4-7) and (8-9)

Suppl Table S6a. Inclusion criteria	
1	**Refractory** (> 6 months) and severe (based on a validated total GCSI = Gastroparesis Cardinal Symptom Index) gastroparesis, with confirmed gastric emptying based on a gastric emptying study: standardized protocol of **scintigraphy in all patients (**performed less than 6 months prior to enrolment (see ref. No. 13 in the main article), or confirmed by a **validated gastric emptying breath test** The total GCSI score must be >2.3. • Abnormal gastric emptying is defined as retention of Tc-99 m >60% at 2 h and/or ≥10% of residual activity at 4 h on a standardized sulphur colloid solid-phase gastric emptying study. • Radiolabelled liquids emptying study will be reserved as alternative technique for patients with poor tolerance of solids during scintigraphy. Abnormal gastric emptying will represent >50% retention of radiolabelled content (e.g. In-111) at 1 hour. • Abnormal gastric empyting breath test based on a solid or malrange determination for the test used (e.g. T1/2 > 109 min).
2	Severe refractory disease is defined as GCSI >2.3 and failure or recurrence in patients who received available optimal pharmacological therapies.
3	Persons 18 years or older at the time of signing the informed consent
4	Signed informed consent

Suppl Table S6b. Exclusion criteria	
1	No previous attempt with at least one prokinetic drug
2	No previous attempt to withdraw anticholinergic agents and glucagon like peptide - 1 (GLP-1) and amylin analogues* in patients treated with these substances (see ref. 1-2)
3	Active treatment with opioids or a history of treatment with opioids within 12 months before enrolment
4	Previous gastric surgery BI or II, esophagectomy, gastric pull-through
5	Previous pyloromyotomy or pyloroplasty
6	Known eosinophilic gastroenteritis
7	Organic pyloric (or intestinal) obstruction (fibrotic stricture, etc.)
8	Sever coagulopathy
9	Esophageal or gastric varices and /or portal gastropathy
10	Advanced liver cirrhosis (Child B or Child C)
11	Active peptic ulcer disease
12	Pregnancy or puerperium
13	Malignant or pre-malignant gastric diseases (dysplasia, gastric cancer, GIST): patients with a history of such disease after its cure are eligible for enrolment
14	Any other condition, which in the opinion of the investigator would interfere with study requirements
15	Uncontrolled diabetes mellitus
16	Diagnosis of rumination syndrome or “eating” disorder (mental anorexia, bulimia nervosa) **
17	Severe constipation without using laxatives
18	Inability to obtain informed consent

* Attempts to normalize glycaemic control using amylin analogues (e.g., pramlintide) or GLP-1 analogues (e.g., exenatide) may result in delayed gastric emptying.

** The presence of a rumination syndrome or eating disorders (anorexia nervosa, bulimia) is an exclusion criterion. In case of doubts, a psychiatric examination should be performed

GIST = Gastrointestinal Stromal Tumor

Suppl Table S7. Gastric emptying study protocol (GES)

Scintigraphy protocol in all patients (see ref. No. 15 in the main article, protocol endorsed by both American Neurogastroenterology and Motility Society and American Nuclear Medicine Society; 2008); less than 6 months prior to randomization. Test begun with patients under fasting conditions for a minimum of 6 hours. A radiolabelled meal was prepared by adding 0.75 mCi ^{99m}Tc-sulfur colloid into 2 the liquid egg whites. Eggs were cooked in a microwave or on a hot non-stick skillet, the eggs were stirred once or twice during cooking until firm – to the consistency of an omelette. Then, the bread was toasted and jelly spread on the toasted bread.

- Gamma camera images was obtained immediately after meal ingestion and then at 1, 2, 3 and 4 hours. The geometric mean of delay-corrected counts was used to estimate the proportion of ^{99m}Tc emptied at each time point. Diagnostic criterion for gastroparesis is defined as the percentage of gastric retention >60% at 2 h and equal to or greater than 10% at 4 h or both. Half-time (T1/2) emptying time was also be calculated. In case of poor tolerance of solids during gastric scintigraphy, radiolabelled liquids were used (see inclusion criteria, suppl table S6a). At least 72 hours before gastric emptying test, narcotics and other medications that can delay gastric emptying should be discontinued. Other alternative meals were used for patients with egg allergies or egg´s intolerance, patients with gluten-sensitive enteropathy, according to the local principles.
- Items needed for Egg Beaters Gastric Emptying Scintigraphy: 118 mL of liquid egg whites (Egg Beaters; egg substitute): 99% real eggs, cholesterol free, fat free, low calorie (120 g Egg Beater, 60 kcal, approx. two large eggs), 2 slices of wheat bread (120 kcal), Strawberry jam (30 g, 74 kcal), Water (120 mL), Technetium-99m 0.75 mCi. The subject completed the sandwich meal quickly, within max. 10 minutes. Generally, the fasting glucose in diabetic patients should be between 75 and 275 mg/dL (4.2 to 15.3 mmol/l). Diabetic patients administered their insulin with meal ingestion, generally ½ what they took normally. The nutritional composition of the meal was 69-72% carbohydrate, 22-24% protein, 2% fat and 2% fibre.

Suppl Table S8. Pyloric distensibility (Endoflip®) measurement protocol
See also **Figures S2a** and **S2b** (see ref. No. 16, 17)

The pyloric distensibility measurement was performed using the Endoflip™ 1.0 Impedance Planimetry System (Medtronic, Minneapolis, MN, USA). The Endoflip system consists of a 24 cm long 3mm outer diameter catheter with highly compliant balloon attached to its tip surrounding 16 paired impedance sensors mounted on the catheter and a solid-state pressure transducer on the distal end of the catheter within the balloon.

A single-use catheter EF-325N with 8 cm long balloon was used for all measurements. The catheter was attached to both the monitor and a syringe automatically filling the balloon with conductive fluid. Based on the principle of impedance planimetry, excitation electrodes at either end of the balloon emit a continuous low electric current, the voltage is measured across the paired impedance planimetry electrodes by leveraging Ohm's law to provide measurement of cross-sectional area at intervals based on electrode spacing. Cross-sectional area together with the pressure data from the intra-balloon pressure transducer allow to calculate resistance to distention, i.e. distensibility.

The catheter was introduced into the pylorus under direct endoscopic control, a snare or forceps were used to navigate the catheter through the pylorus if necessary. Once adequate position was achieved, with the balloon straddling the pylorus (an hourglass shape image on the Endoflip monitor), the balloon was automatically (but under direct visual supervision of the performing physician) filled with fluid from an 80mL syringe to three balloon filling volumes 30 mL, 40 mL, 50 mL. At each of these volumes the following parameters were recorded: distensibility index (mm^2/mmHg), cross-sectional area (mm^2), balloon diameter (mm), and intra-balloon pressure (mmHg). The measurements were performed in between the peristaltic waves driven by the motor migrating complex. The additional time to the procedure was approximately 10 minutes.

Suppl Table S9. PAGI-SYM score (Patient Assessment of Gastrointestinal Disorders Symptom Severity Index)

Symptoms		Score					
		0	1	2	3	4	5
		None	Very mild	Mild	Modarate	Severe	Very severe
1	Heartburn (burning pain rising in your chest or throat) during the day						
2	Regurgitation or reflux (fluid or liquid from your stomach coming up into your throat) during the day						
3	Heartburn (burning pain rising in your chest or throat) when lying down						
4	Regurgitation or reflux (fluid or liquid from your stomach coming up into your throat) when lying down						
5	Feeling of discomfort inside your chest during the day						
6	Bitter, acid or sour taste in your mouth						
7	Feeling of discomfort inside yourchest at night (during sleep time)						
8	Vomiting						
9	Nausea (feeling sick to your stomach as if you were going to vomit or throw up)						
10	Retching (heaving as if to vomit, but nothing comes up)						
11	Stomach fullness						
12	Not able to finish a normal-sized meal						
13	Feeling excessively full after meals						
14	Loss of appetite						
15	Bloating (feeling like you need to loosen your clothes)						
16	Stomach or belly visibly larger						
17	Upper abdominal (above the navel) discomfort						
18	Upper abdominal (above the navel) pain						
19	Lower abdominal (below the navel) pain						
20	Lower abdominal (below navel) discomfort						

(see ref. No. 20 in the main article)

Questionnaire was developed to measure specific symptoms of patients with upper gastrointestinal disorders. It records 20 symptoms (6 subscales) and assesses their severity within the 2 weeks prior to the test. Subscale scores are calculated by averaging across items comprising the subscale; scores vary from 0 (none or absent) to 5 (very severe). The PAGI-SYM subscale scores have good internal consistency and test-retest reliability (18).

1 - 7 = heartburn/regurgitation
8 - 10 = nausea/vomiting
11 - 14 = post-prandial fullness/early satiety
15 - 16 = bloating
17 - 18 = upper abdominal pain
19 - 20 = lower abdominal pain

Calculation:

Total PAGY-SIM score = arithmetic mean of the six symptom subscales. Subscores = arithmetic means of (1-7), (8-10), (11-14), (15-16), (17-18) and (19-20)

Suppl Table S10. PAGI – QoL questionnaire (Patient Assessment of Upper Gastrointestinal Disorders – Quality of Life)

Questions mostly related to previous 2 weeks.
Most desirable option: 5 points / Less desirable option: 0 points

Symptoms	Score					
	0	1	2	3	4	5
	None of the time	A little of the time	Some of the time	A good bit of the time	Most of the time	All of the time
During the past 2 weeks, because of your Gastrointestinal problems, how often …						
1	have you had to depend on others to do your daily activities?					
2	have you avoided performing your daily activities?					
3	have you had difficulty concentrating?					
4	has it taken you longer than usual to perform your daily activities?					
5	have you felt tired?					
6	have you lost the desire to participate in social activities such as visiting friends or relatives?					
7	have you been worried about having stomach symptoms in public?					
8	have you avoided performing physical activities or sports?					
9	have you avoided traveling?					
10	have you felt frustrated about not being able to do what you wanted to do?					
11	have you felt constricted in the clothes you wear?					
12	have you felt frustrated about not being able to dress as you wanted to?					
13	have you felt concerned about what you can and cannot eat?					
14	have you avoided certain types of foods?					
15	have you restricted eating at restaurant or at someone's home?					
16	have you felt less enjoyment in food than usual?					
17	have you felt concerned that a change in your food habits could trigger your symptoms?					
18	have you felt frustrated about not beingable to choose the food you wanted to?					
19	have you left frustrated about not being able to choose the type of beverage you wanted to?					
20	has your relationship with yours pouseor partner been disrupted?					
21	has your relationship with your children or relatives been disrupted?					
22	has your relationship with your friends been disrupted?					
23	have you been in a bad mood?					
24	have you felt depressed?					
25	have you felt anxious?					
26	have you felt angry?					
27	have you felt irritable?					
28	have you felt discouraged?					
29	have you been stressed?					
30	have you felt helpless?					

(see ref. No. 21 in the main article)

The PAGI-QoL contains 30 items with five subscales:

(1) daily activities (1 – 10)
(2) clothing (11 – 12)
(3) diet/food habits (13 – 19)
(4) relationship (20 – 22)
(5) psychological well-being and distress (23 – 30)

The PAGI-QoL questionare contains of 30 items with five subscales: (1) daily activities; (2) clothing; (3) diet/food habits; (4) relationship; and (5) psychological well-being and distress. Each items are scored on a 6-point Likert scale, with response options ranging from 0 (none) to 5 (severe problem all of the time). Subscale scores are calculated by averaging the item responses.

Calculation:

Total PAGI-QoL score = arithmetic mean of the five subscales. Subscores = arithmetic means of (1-10), (11-12), (13-19), (20-22) and (23-30)

Suppl Table S11. Detailed description of the statistical analysis

The intention to treat population and early study termination

All the main analyses were performed on the Intention To Treat (ITT) population as specified in the protocol. The ITT population includes all randomized patients and evaluates them as members of the groups to which they were originally allocated regardless of the actual treatment received or any other protocol deviations. Since some values were missing (including a complete follow-up of one patient in the sham group who withdrew consent before receiving any procedure), these values had to be imputed to recover the ITT population. The sample size for all the analyses on ITT is 41 patients as the trial was terminated early for efficacy of G-POEM in a planned interim analysis. As stated in the report of the Data and Safety Monitoring Board, the decision to stop the enrolment was adopted based on a combination of two factors: 1) The interim result was truly highly significant in favor of G-POEM with p=0.003 (the final p-value for the main outcome presented in the results is different since more follow-up data accumulated after the enrollment was stopped at the interim analysis. Unfortunately, no exact strategy for early termination was indicated in the study protocol. Therefore, the conservative Haybittle-Peto boundary was considered indicating to stop the trial at p=0.001 for any number of interim analyses. This boundary was almost reached. 2) The second factor was the risk of general anesthesia for patients undergoing the sham procedure.

Imputation of missing data and confidence intervals

The imputation of missing values was performed by the multiple imputation procedure with chained equations. We only imputed some missing values for the 41 patients enrolled in the study. We did not impute values for the remaining 45 potential patients, who were not enrolled in the study due to the early study termination. Although the amount of missing data was rather low – at most 3 values missing in any of the variables evaluated on the ITT basis – we decided on imputation to adhere to the protocol. The assumption of data missing at random (MAR) was considered plausible and given the low proportion of missing data even its partial violation would not pose a significant threat of biased results. To further prevent any suspicion that our result could be heavily influenced by the imputation, we also provide analysis of treatment success (primary outcome) on the per protocol population (1 patient with technical failure and 1 with missing GCSI excluded) and also the worst case imputation (1 technical failure in the G-POEM group as failure and 1 missing GCSI follow u in sham as success).

Multiple imputation in simple terms: The chained equations approach allows imputation of missing values using the information from the observed values. The estimates of missing values are updated iteratively which allows one to deal with missing values in all included variables. The process of imputation is random to some extent. This is further used in the multiple imputation approach. Here, multiple (e.g. 100) different random versions of imputed datasets are created. The desired analysis is performed on each realization of the dataset. Finally, the estimates of desired statistics (e.g. the median) are combined from all the imputations and their confidence intervals are constructed while reflecting variability originating both from the observed data itself and from the uncertainty of the imputation process. The resulting values are an aggregate of all the different realizations of the imputation. As a result, for example the treatment success in the sham group is 22% in 20 patients, so the value does not correspond to any of 4/20 or 5/20. This reflects the fact that the patient with missing follow-up GCSI values was assigned a treatment success in some imputations and treatment failure in the others.

Our imputation model included the following variables: age; gender; etiology of gastroparesis; baseline, 3 months, and 6 months values of total GCSI score, total PAGI-SYM score, and total QOL score; and baseline and 3 months values of 4h GES retention and GES retention halftime. We imputed the GCSI scores. The treatment success of the imputed patient was evaluated afterwards. The allocation of the patient was not used for imputation. Otherwise, the model would be strongly forced to impute high GCSI values for a patient just based on allocation into the sham group. We imputed data for the main part of the analysis separate from the cross-over data.

For the estimation of treatment success, we made 100 imputed datasets and on each used the Wilson method for construction of confidence intervals for proportions. Compared to the normal approximation approach, this method can result in non-symmetrical confidence intervals, which is very relevant for example for the GES halftime with a clearly skewed distribution. To combine the Wilson confidence intervals from all imputation datasets we used the method by Lott and Reiter (see ref. 24), which is particularly designed for this purpose.

As the primary statistics for the continuous variables we used the median since normality of the data was rejected by the Shapiro-Wilk test for at least one dataset among the compared groups and time points for each investigated variable. The confidence intervals were obtained as 2.5 % and 97.5 % quantiles from 20000 bootstrapping iterations. The same approach was used for the correlation coefficient between GCSI and gastric emptying.

Bootstrapping in simple terms: Bootstrapping is based on the idea that the distribution of observed values is the best available estimate of the actual distribution for the investigated population. Therefore, we resample the data many times (in our case 20000 times) to estimate confidence intervals for our statistic (e.g. the median). When we have N values in our sample, resampling means randomly choosing N of these values with replacement. We can imagine this as writing each of the values on a paper ticket and putting them into a hat. We randomly draw a ticket N times, but each chosen ticket is returned into the hat before another draw. As a result, the resampled dataset contains certain values multiple times and some other are not present at all. On this dataset we calculate our statistic (the median). We then take the dataset of statistics (medians) from all iterations of the resampling and estimate the confidence interval limits by discarding 2.5 % of the lowest and 2.5 % of the highest values (medians).

In our case, the process of bootstrapping had two additional steps:

1. As the number of observations in our dataset is relatively small, the median can be highly influenced by the middle values since the extreme values have no effect on the median. This can in some cases lead to underestimation of the width of the confidence intervals. We face this issue by smoothing with a Gaussian noise with sigma given by the inter quartile range of the sample divided by the square root of N.
2. We sampled the original dataset including missing values and after resampling we imputed the missing values. With this approach, both variability from the data and from the imputation are reflected in the final confidence interval.

P-values and multiple testing

As the protocol indicated a regression-based approach for the evaluation of the main outcome, we used logistic regression to calculate the only p-value presented in the manuscript for the only primary outcome (as previously specified in the protocol). All the remaining results are presented as point estimates (medians and hazard ratios) with 95 % confidence intervals with accordance to the CONSORT statement. We hope that this will prevent inadequate interpretations of the results

in terms of the multiple testing problem, which we consider likely to happen if we presented uncorrected p-values for all the other outcomes.

Technically, methods of multiple testing correction could be applied. Nevertheless, there are many strategies with different results. Primarily, the decision of which variables should be included into the analysis (defining the family of tests over which the false positive rate is to be controlled) is of major importance. The multiple testing corrections are suited for situations, where many tests are performed without a pre-defined primary hypothesis or for situations where multiple primary hypotheses are aimed to be tested simultaneously in a single trial.

We are convinced that presenting uncorrected confidence intervals for the secondary outcomes is the best option as they both show the uncertainty of the actual presented value and allow the reader to judge the single-test statistical significance. Whenever a confidence interval for a difference lies entirely below or entirely above zero, this corresponds a to statistically significant decrease or increase. As no correction for multiple testing is applied (as it is a common standard), there is 95% confidence for each individual interval to contain the true value of the population statistic (e.g. the median), but not 95% confidence that all the intervals contain their respective true values. This is presumably understandable to the reader. In contrast, by presenting all the p-values a less statistically experienced reader could be tempted to just interpret any p-value below 5 % as a clear indication of a proven effect.

Suppl Table S12. Patients treated by a respective Trial Center (in and out of the trial) and Number of Monitoring Visits

Centre No.	Randomized patients	Patients underwent G-POEM	Patients underwent sham	Patients underwent cross-over G-POEM	Patients treated outside the trial during trial period (G-POEM)	Number of monitoring visits
IKEM	33	17	15	9	7	18
Trnava	8	3	4	3	3	3
Total	41	20	19	12	10	21

G-POEM = Gastric Per Oral Endoscopic Pyloromyotomy

Suppl Table S13. Screened and excluded (not enrolled) patients

Centre No.	Screened patients	Patients uderwent GES	Patients with positive GES	Patients did not fulfill inclusion criteria	Patients fulfilled at least one exclusion criterium
IKEM	147	136	57	8	15
Trnava	42	42	18	7	4
Total	189	178	75	15	19

GES = Gastric Emptying Study (scintigraphy)

Suppl Table S14. Definition of Adverse event (AE) / Serious Adverse Event (SAE)

An adverse event (AE) is any undesirable, unintentional or unanticipated event that occurs during use of the investigational device, whether or not considered related to the therapy. A serious adverse event (SAE) is an event that is: fatal, life-threatening, results in persistent or significant disability/incapacity, requires or prolongs inpatient hospitalization, requires an intervention (endoscopy, radiology, surgery, etc.) postoperatively. Abdominal pain requiring analgetics without a need for prolongation of hospitalization was not considered as adverse event. SAE had to be reported within 24 hours to the Prague study center and the Ethics Committees / IRB if applicable. AE/AES were documented on designated CRF forms.

Report of a Adverse Event Form

Hospital visits due to follow up visits are not considered to be SAE.

□ Initial report

□ Consecutive report

Date AE start: _____ / _____ / _____ (DDMMYY)

□ Expected event □ Unexpected event

Event related to G-POEM / SHAM procedure

□ No □ Possibly □ Yes

Complication: □ Perforation □ Bleeding □ Infection □ Other

Please describe complication:

……

……

Intervention required: □ No □ Yes

Please describe intervention

……

……

Medication required: □ No □ Yes

Medication(s):

……

……

Report of a Serious Adverse Event

Hospitalization or prolongation of hospital stay required (SAE):

□ Yes □ No

If yes, please report within 24 hours to the Prague study center and Ethics Committee/IRB if applicable!

Date of hospitalisation/ - prolongation _____ (DDMMYY)

Date hospital discharge _____ (DDMMYY)

……

……

□ Event resolved □ Event ongoing

□ Long term sequela □ Death □ Unknown

Description / comment: ……

Date AE stop: _____ / _____ / _____ (DDMMYY)

Suppl Table S15. Overall incidence of adverse events

Patient	Serious / non-serious	G-POEM / Sham / cross over G-POEM	Time of AEs occurrence after the allocated procedure	Adverse Event / Serious Adverse Event	Related to procedure
01-04	Serious	G-POEM	1 month	Hospitalisation due to vomiting (not related to gastroparesis), probably food toxin	no
01-08	Non-serious	G-POEM	4 months	Mild abdominal pain without need for analgetics	no
01-10	Serious	cross over G-POEM	POD 1	Sever abdominal pain, deep ulcer of the pylorus, prolonged hospitalisation for 6 days	yes
01-11	Non-serious	G-POEM	POD 0	small periprocedural perforation of duodenal mucosae without need for intervention, no need for prolonged hospitalisation	yes
01-26	Non-serious	cross over G-POEM	POD 0	Hyperglycemia (24 mmol/L) with mild metabolic acidosis	no
01-26	Non-serious	cross over G-POEM	POD 0	Small gastric serosal perforation during G-POEM without need for intervention without sequelae	yes
01-26	Non-serious	cross over GPOEM	6 months	Non-complicated Hp- positive gastric ulcer of stomach, eradication of Hp	no
01-28	Serious	Sham	3 months	Need for hospitalisation due to severe mycotic esophagitis not allowing to eat and newly diagnosed achalasia, pneumatic dilation of achalasia, NG tube placement for feeding, prolonged hospitalisation for 23 days	no
01-28	Non-serious	cross over G-POEM	3 months	Decompensation of achalasia, mycotic esophagitis, prolonged hospitalisation for 20 days	no
01-30	Serious	G-POEM	1 month	Vomiting, need for 3 days hospitalisation, temporary nasojejunal tube placement, mycotic esophagitis	no
01-30	Non-serious	G-POEM	4 months	Feeding intolerance, hyponutrition	no
01-31	Non-serious	G-POEM	3 months	Hypoglycemia, no dumping syndrome	no
01-32	Non-serious	Sham	2 months	Recurent abdominal pain, need for opioids	no

01-32	Non serious	Sham	5 months	Vomiting, abdominal pains, administration of prokinetics and opioids	no
01-32	Serious	Sham	5 months	Hospitalisation due to vomiting for 2 days, feeding intolerance, need for nasogastric tube placement and enteral nutrition	no
01-32	Non-serious	Sham	5 months	Nausea, diarrhea, feeding intolerance, need for painkillers (opioids)	no
01-32	Non-serious	cross-over G-POEM	1 month	Severe nausea, feeding intolerance, need for administration of parenteral prokinetics	no
01-32	Serious	cross-over G-POEM	4 months	Abdominal pains, weightloss, feeding intolerance, nasogastric tube placement, hospitalisation for 6 days, acute urinary retention, pains of ears.	no
01-35	Serious	G-POEM	POD 0	During G-POEM mucosal injury, prolonged hospitalisation for precautionary reasons for 7 days, no need for intervention or specific treatment	yes
01-38	Serious	G-POEM	3 months	Hospitalisation for 30 days due to hypocalcemia, examination before transplantation, diarrhea, hypoglycemia - confirmed dumping syndrome	yes
01-40	Serious	Sham	3 months	Hospitalisation for 1 day, abdominal pains, nausea	no
01-41	Serious	G-POEM	4 months	Hospitalisation for 6 days, due to intestinal infection – gastroenteritis	no

POD = postoperative day
Hp = Helicobacter pylori

Suppl Table S16. Summary of adverse events (AE)			
	G-POEM	**Sham**	**Cross-over G-POEM**
Serious AE – n			
Hospitalisation (required or prolonged) related to procedure	2	0	1
Need for additional endoscopic, radiological or surgical intervention	0	0	0
Hospitalisation (required or prolonged) not related to procedure	3	3	1
Live-threatening events	0	0	0
Death	0	0	0
Overall	**5**	**3**	**2**
Overall SAEs related to procedure	**2**	**0**	**1**
Overall SAEs not related to procedure	**3**	**3**	**1**
Non-serious AE – n			
Abdominal pain (not related to procedure)	1	1	0
Periprocedural serosal perforation	1	0	1
Nausea or vomiting, feeding intolerance (not related to procedure)	1	2	1
Decompansation of achalasia with mycotic esophagitis	0	0	1
Hypoglycemia/hyperglycemia	1	0	1
Gastric ulcer	0	0	1
Overall	**4**	**3**	**5**
Overall AEs related to procedure	**1**	**0**	**1**
Overall AEs not related to procedure	**3**	**3**	**4**

Suppl Table S17. Need for analgesics administration after G-POEM, sham procedure or cross-over G-POEM

	G-POEM	Sham	Cross-over
Number n (%)	10 (41%)	2 (10%)	4(33%)
Total number of procedures	21	20	12

Postprocedural pain necessitating administration of analgesics on postoperative days 0 or 1 was not considered as adverse event but rather a standard part of the postoperative course like with other similar procedures.

Suppl Table S18. Procedure details. The analysis was performed on the available data, one procedure length was missing. There was one technical failure of G-POEM procedure, which is included into the analysis but no closure was used in this case.

Procedure length	**G-POEM** n=21	**Cross-over G-POEM** n=12	**Sham** n=19
Mean	76 min	58 min	55 min
Standard deviation	41 min	17 min	9 min
Median	61 min	56 min	55 min
Minimal	35 min	40 min	40 min
Maximal	185 min	91 min	76 min
Length of myotomy	**G-POEM**	**Cross-over G-POEM**	**Sham**
Mean	27 mm	27 mm	-
Standard deviation	7 mm	4 mm	-
Median	30 mm	30 mm	-
Minimal	25 mm	20 mm	-
Maximal	30 mm	30 mm	-
Hospitalization after procedure	**G-POEM**	**Cross-over G-POEM**	**Sham**
Mean	1.9 days	2.4 days	1 day
Standard deviation	1.4 days	1.3 days	0 days
Median	1.5 days	2 days	1 day
Minimal	1 day	1 day	1 day
Maximal	7 days	6 days	1 day
Technical success	95% (20/21)	100% (12/12)	NA
Closure with endoclips (n)	9	8	NA
Closure with endoscopic suturing system, (n)	11	4	NA
Need for capnoperitoneum puncture	No	No	No
Other gas related adverse events	No	No	No
Anesthesia related adverse events	No	No	No

NA = not applicable

Suppl. Table S19. Treatment success for the primary outcome, sensitivity analysis and etiology subgroups. In this trial, treatment success was defined as a reduction by 50% from baseline GCSI for the primary G-POEM and sham procedure and as a reduction by 50% from the 6 months visit (after the sham procedure) for the cross-over G-POEM. In addition, table shows treatment success rates if the treatment success had been defined as a decrease of GCSI by 1 point (a common definition of treatment success). Subgroup analysis in different etiologies of gastroparesis after cross-over G-POEM was not performed because of small numbers of patients. For the ITT population, one of the 41 values (2 %) was multiply imputed (in the sham group). For the worst case scenario, we assumed treatment failure for the G-POEM patient with technical failure and treatment success in the sham patient with missing GSCI data.

Treatment success rate [%] (95% CI) at 6 months	**G-POEM**	N	**Sham**	N	**Cross-over G-POEM**	N
ITT population, GCSI reduction by 50 %	71 (50 – 86)	21	22 (8 – 47)	20	75 (47 – 91)	12
PP population, GCSI reduction by 50 %	70 (48 – 85)	20	21 (9 – 43)	19	75 (47 – 91)	12
Worst case scenario, GCSI reduction by 50 %	67 (45 – 83)	21	25 (11 – 47)	20	75 (47 – 91)	12
Diabetic etiology (ITT, reduction by 50 %)	89 (56 – 98)	9	17 (3 – 57)	8	Not applicable	
Post-surgical etiology (ITT, reduction by 50 %)	50 (18 – 82)	6	29 (7 – 67)	7	Not applicable	
Idiopathic etiology (ITT, reduction by 50 %)	67 (30 – 90)	6	20 (3 – 67)	5	Not applicable	
ITT population, GCSI reduction by 1 point	95 (76 – 99)	21	37 (19 – 60)	20	75 (47 – 91)	12
Treatment success rate [%] (95% CI) at 3 months	**G-POEM**	N	**Sham**	N	**Cross-over G-POEM**	N
ITT population, GCSI reduction by 50 %	57 (36 – 76)	21	22 (8 – 47)	20	58 (32 – 81)	12
PP population, GCSI reduction by 50 %	55 (34 – 74)	20	21 (9 – 43)	19	58 (32 – 81)	12
Worst case scenario, GCSI reduction by 50 %	52 (32 – 72)	21	25 (11 – 47)	20	58 (32 – 81)	12
Diabetic etiology (ITT, reduction by 50 %)	67 (35 – 88)	9	17 (3 – 57)	8	Not applicable	
Post-surgical etiology (ITT, reduction by 50 %)	33 (9 – 72)	6	43 (15 – 76)	7	Not applicable	
Idiopathic etiology (ITT, reduction by 50 %)	67 (30 – 90)	6	0 (0 – 43)	5	Not applicable	
ITT population, GCSI reduction by 1 point	76 (55 – 89)	21	42 (23 – 64)	20	67 (39 – 86)	12

ITT – intention to treat population (all patients evaluated according to their allocation, missing data multiply imputed)

PP – per protocol population (only patients following the study protocol)

GCSI – gastroparesis cardinal symptom index

N – number of patients in a given group

Suppl. Table S20. Evolution of variables in time. The table presents estimates of medians of various quantities at different time points in the study and differences between time points. The differences are calculated on a single patient level. The confidence intervals (CI) are not corrected for multiple testing. The analysis was performed on the ITT population with 21, 20, and 12 patients in the G-POEM, sham, and cross-over G-POEM groups, respectively. In total, 2 GCSI values (1 %), 3 PAGI-SYM values (2 %), 7 PAGI-QOL values (5 %) and 10 GES values (5 %) were imputed across all groups and time points.

	Values at visits			**Decrease from baseline***	
Variable – median (95% CI)	**Baseline**	**3 months**	**6 months**	**to 3 months**	**to 6 months**
GCSI – G-POEM	3.5 (3.2 – 3.7)	1.4 (0.9 – 1.9)	1.1 (0.5 – 1.5)	2.3 (1.3 – 2.6)	2.4 (2.0 – 2.8)
GCSI – sham	3.2 (2.8 – 3.4)	2.5 (1.9 – 3.1)	2.5 (1.9 – 3.2)	0.8 (0.1 – 1.2)	0.7 (0.0 – 1.2)
PAGI-SYM – G-POEM	2.7 (2.0 – 3.0)	0.9 (0.7 – 1.4)	0.7 (0.5 – 1.2)	1.5 (1.0 – 1.9)	1.5 (1.2 – 2.0)
PAGI-SYM – sham	2.8 (2.5 – 3.0)	2.0 (1.5 – 2.8)	2.0 (1.5 – 2.6)	0.7 (0.1 – 1.1)	0.5 (0.1 – 1.1)
PAGI-QoL – G-POEM	2.1 (1.7 – 2.5)	1.6 (0.9 – 2.5)	0.8 (0.6 – 1.5)	0.3 (-0.5 – 0.9)	1.1 (0.1 – 1.6)
PAGI-QoL – sham	2.5 (1.5 – 2.9)	1.9 (1.2 – 2.7)	1.7 (1.2 – 2.4)	0.4 (-0.2 – 0.7)	0.4 (-0.1 – 0.8)
BMI [kg/m2] – G-POEM	22 (19 – 26)	22 (20 – 25)	22 (21 – 26)	-0.4 (-1.2 – 0.5)	-0.7 (-1.8 – 0.2)
BMI [kg/m2] – sham	26 (21 – 28)	24 (21 – 27)	24 (21 – 28)	-0.4 (-1.0 – 0.4)	-0.7 (-1.2 – 0.4)
GES 4h retention [%] – G-POEM	22 (17 – 31)	12 (5 – 22)		12 (3 – 19)	
GES 4h retention [%] – sham	26 (18 – 39)	24 (11 – 35)		6 (-7 – 19)	
GES ret. halftime [min] – G-POEM	152 (127 – 185)	95 (77 – 118)		53 (5 – 94)	
GES ret. halftime [min] – sham	157 (128 – 263)	110 (82 – 158)		49 (-3 – 144)	

	Values at visits			**Decrease from baseline ***	
Variable – median (95% CI)	**Baseline (= 6 months visit after sham)‡**	**3 months after cross-over**	**6 months after cross-over**	**to 3 months after cross-over**	**to 6 months after cross-over**
GCSI – cross-over G-POEM	2.8 (2.5 – 3.7)	1.1 (0.7 – 1.9)	1.0 (0.6 – 1.7)	1.9 (1.1 – 2.4)	2.1 (1.3 – 2.6)
PAGI-SYM – cross-over G-POEM	2.2 (1.9 – 3.0)	0.8 (0.6 – 1.6)	0.5 (0.4 – 1.8)	1.3 (1.0 – 2.0)	1.6 (0.8 – 2.2)
PAGI-QoL – cross-over G-POEM	2.2 (1.3 – 3.3)	1.8 (0.9 – 2.6)	1.6 (0.7 – 2.3)	0.5 (-0.1 – 1.5)	0.3 (-0.1 – 1.6)
BMI [kg/m2] – cross-over G-POEM	22 (19 – 26)	22 (19 – 27)	22 (20 – 27)	0.0 (-1.0 – 0.9)	-0.2 (-1.1 – 0.5)
	Baseline (= 3 months visit after sham)‡	**3 months after cross-over**		**to 3 months after cross-over**	
GES 4h ret. [%] – cross-over G-POEM	24 (11 – 38)	7 (1 – 14)		13 (5 – 23)	
GES ret. halftime [min] – cross-over G-POEM	138 (83 – 178)	66 (32 – 154)		80 (29 – 179)	

* The table presents a decrease, so positive values indicate reduction of the score/measurement.
‡ For the cross-over procedure, values obtained at 6 months visit (at 3 months in case of gastric emptying study) after the sham procedure are considered as baseline value
GCSI = Gastroparesis Cardinal Symptom Index (see also suppl table S5)
PAGI-SYM = Patient Assessment of Upper Gastrointestinal Symptom Severity Index (see also suppl table S9)
PAGI QoL = Patient Assessment of Upper Gastrointestinal Disorders-Quality of Life (see also suppl table S10)
GES = Gastric Emptying Study
ITT = Intention To Treat

Suppl. Table S21. Evolution of GCSI sub-scores in time. Means of GCSI subscales are presented at different time points in the study and differences between time points. The differences are calculated on a single patient level. The Nausea / vomiting subscale comprises of the questions 1 to 3, Fullness of questions 4 to 7 and Bloating of questions 8 and 9 (see Table S5). The confidence intervals (CI) are not corrected for multiple testing. The analysis was performed on the available data basis with N=21 for G-POEM, N=19 for sham, and N=12 for cross-over G-POEM.

	Values at visits			**Decrease from baseline***	
Variable – mean (95% CI)	**Baseline**	**3 months**	**6 months**	**to 3 months**	**to 6 months**
G-POEM					
Nausea / vomiting	3.3 (2.8 – 3.7)	1.3 (0.7 – 1.9)	0.8 (0.4 – 1.2)	2.1 (1.5 – 2.6)	2.5 (2.1 – 3.0)
Fullness	3.6 (3.4 – 3.9)	1.7 (1.3 – 2.0)	1.2 (0.8 – 1.7)	2.0 (1.6 – 2.4)	2.4 (2.0 – 2.9)
Bloating	3.5 (3.0 – 4.0)	1.5 (0.9 – 2.0)	1.4 (0.9 – 1.9)	2.0 (1.5 – 2.6)	2.1 (1.6 – 2.6)
Sham					
Nausea / vomiting	3.0 (2.5 – 3.4)	2.0 (1.4 – 2.4)	1.8 (1.4 – 2.4)	1.1 (0.7 – 1.7)	1.2 (0.6 – 1.8)
Fullness	3.4 (3.1 – 3.7)	3.0 (2.6 – 3.4)	2.9 (2.5 – 3.4)	0.4 (-0.2 – 0.9)	0.5 (0.1 – 1.0)
Bloating	3.3 (2.6 – 3.8)	2.6 (1.9 – 3.3)	3.0 (2.3 – 3.7)	0.7 (0.4 – 1.1)	0.3 (-0.3 – 1.0)
	Values at visits			**Decrease from 6 months after sham***	
Variable – mean (95% CI)	**Baseline (= 6 months visit after sham)‡**	**3 months after cross-over**	**6 months after cross-over**	**to 3 months after cross-over**	**to 6 months after cross-over**
Cross-over G-POEM					
Nausea / vomiting	2.2 (1.6 – 2.8)	0.7 (0.3 – 1.3)	0.6 (0.2 – 1.1)	1.5 (1.0 – 2.1)	1.6 (1.0 – 2.3)
Fullness	3.5 (3.1 – 3.9)	1.4 (0.8 – 2.0)	1.3 (0.9 – 1.6)	2.1 (1.4 – 2.8)	2.3 (1.8 – 2.7)
Bloating	3.5 (2.7 – 4.2)	1.9 (1.4 – 2.4)	1.5 (0.9 – 2.2)	1.7 (1.2 – 2.0)	2.0 (1.4 – 2.7)

* The table presents decrease, so positive values indicate reduction of the score/measurement.

‡ For the cross-over procedure, values obtained at 6 months visit after the sham procedure were considered as baseline values.

GCSI – gastroparesis cardinal symptom index

Suppl. Table S22. Endoflip® measurements – primary G-POEM and cross-over G-POEM combined. Means of pyloric distensibility are presented at different time points in the study and differences between time points. The differences are calculated on a single patient level. The confidence intervals (CI) are not corrected for multiple testing. The table presents only available data; the imputation model was not used as over 50% of data is missing because the measurement of pyloric distensibility was added after beginning of the trial.
Note that at pre-procedure, post-procedure, and follow-up time points there were 16, 17, and 15 (14 for filling volume 50 mL) values available. There were 14 values for the pre vs. post treatment difference and 12 values for the pre vs. follow-up difference.

	Values at visits			**Increase from pre-procedure** ‡	
Variable – mean (95% CI)	**Pre-G-POEM**	**Post-G-POEM**	**3 months follow-up**	**to post-G-POEM**	**to 3 months follow-up**
DI [mm2/mmHg] 30 mL filling	6.8 (5.2 – 8.4)	12.6 (10.3 – 14.9)	10.2 (8.6 – 11.8)	7.4 (6.0 – 9.0)	5.2 (3.4 – 7.5)
DI [mm2/mmHg] 40 mL filling	7.6 (6.0 – 9.3)	12.7 (11.4 – 14.3)	13.1 (11.3 – 15.7)	5.4 (3.7 – 7.0)	8.0 (5.5 – 10.2)
DI [mm2/mmHg] 50 mL filling	9.1 (6.5 – 12.4)	11.6 (9.5 – 14.1)	10.3 (8.2 – 12.4)	2.6 (0.7 – 4.4)	3.6 (1.2 – 6.0)
CSA [mm2] 30 mL filling	91 (75 – 107)	128 (114 – 142)	142 (111 – 176)	50 (35 – 64)	35 (10 – 58)
CSA [mm2] 40 mL filling	144 (125 – 165)	199 (177 – 219)	206 (185 – 234)	66 (36 – 99)	64 (37 – 83)
CSA [mm2] 50 mL filling	216 (180 – 247)	291 (267 – 319)	279 (246 – 306)	92 (63 – 120)	66 (18 – 110)

‡ The estimates of change are based only on cases where both relevant values were available. Therefore, the expected median difference does not have to correspond to the difference in medians for the corresponding visits. Please note that increase of both DI and CSA at all three filling volumes are significant.

DI – distensibility index
CSA – cross-sectional area

Supp. Table S23. List of pre-specified primary and secondary outcomes and post-hoc outcomes and other analyses with references	
Primary outcome	
Treatment success at 6 months	Table S19, Figures 2 (main document), S3
Secondary outcomes	
Treatment success at 3 months	Table S19, Figure S4
Treatment success in per-protocol population	Table S19, Figures 2 (main document), S3, S4
Treatment success in etiology sub-groups	Table S19, Figures 2 (main document), S3, S4
Treatment success predictors	Table S2 (main document)
GCSI score	Tables 1 (main document), S20, S21, Figures 3 (main document), S5, S6
PAGI-SYM score	Table S20, Figure S7
PAGI-QoL score	Tables 1 (main document), S20, Figure S8
GES 4h retention	Tables 1 (main document), S20, Figures 4 (main document), S9
GES retention halftime	Table S20, Figure S10
BMI	Tables 1 (main document), S20
Endoflip® DI and CSA (pyloric distensibility)	Table S22, Figure S11
Adverse events	Tables S14, S15, S16
Need for analgetics (pain analysis)	Table S17
Post hoc analyses	
GCSI by sub-scores	Table S21, Figure S6
GCSI and GES correlation at 3 months	Figure S12
Other analyses	
Baseline Demographic and Clinical characteristics	Table 1 (main document)
Procedure details	Table S18
Patients treated in centres in and out of the Trial	Tables S12
Screened and enrolled patients	Tables S13

GCSI = Gastroparesis Cardinal Symptom Index

PAGI-SYM = Patient Assessment of Gastrointestinal Disorders Symptom Severity Index

PAGI-QoL = Quality of Life Questionnaire

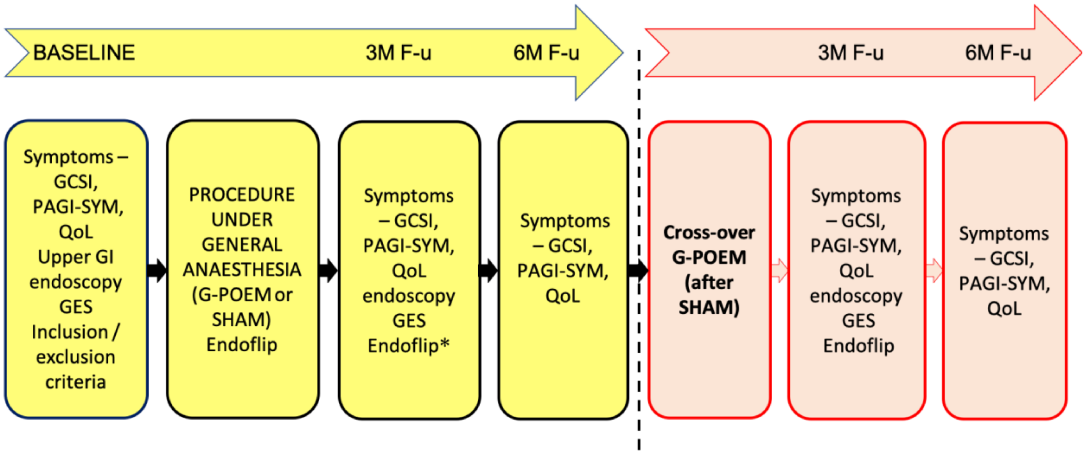
GES = Gastric Emptying Study

BMI = Body Mass Index

DI = distensibility

CSA= Cross-Sectional Area

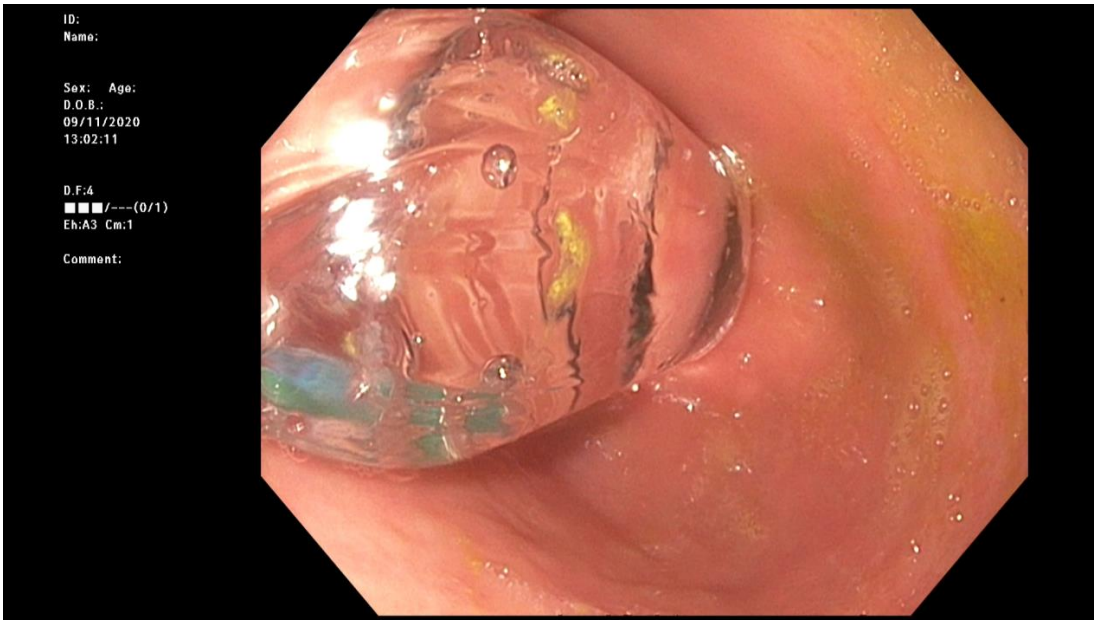
SUPPLEMENTARY FIGURES



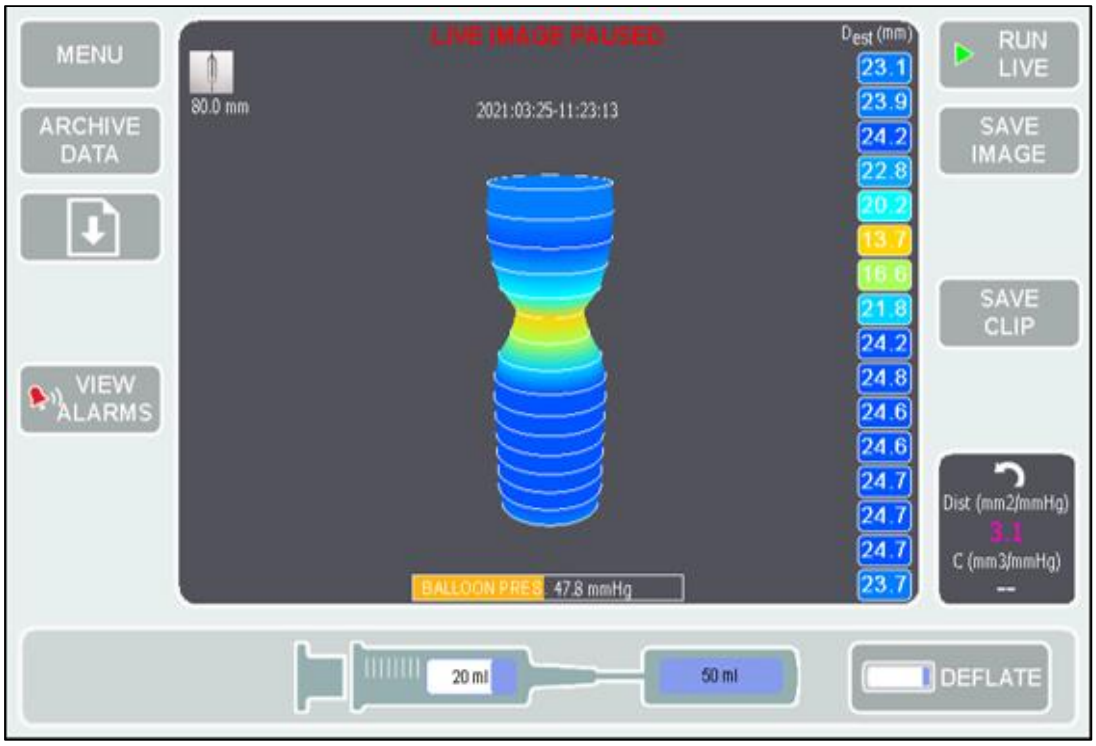
Suppl Figure S1. Study design showing the study course after randomisation (yellow boxes) and for patients who underwent cross-over G-POEM

* *In patients having undergone Endoflip during the sham procedure, no Endoflip measurement was repeated prior to G-POEM.*

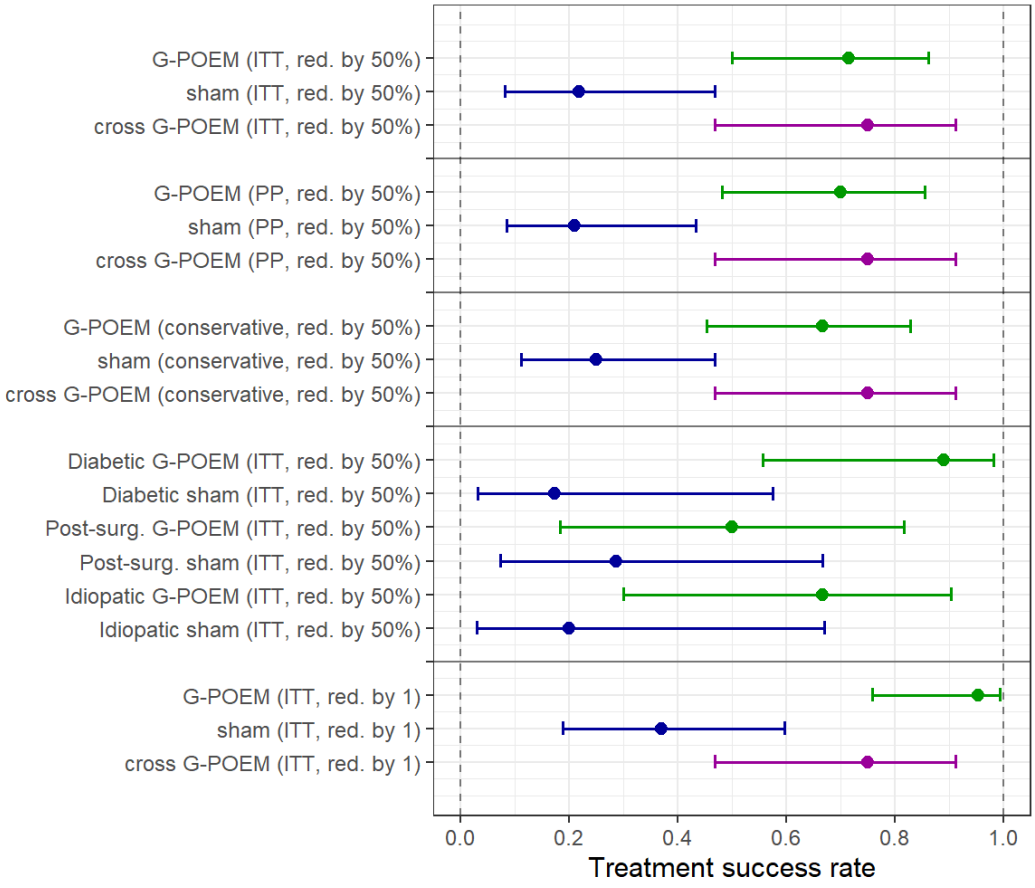
GCSI = Gastroparesis Cardinal Symptom Index (see also suppl table S5)
PAGI-SYM = Patient Assessment of Upper Gastrointestinal Symptom Severity Index (see also suppl table S9)
PAGI QoL = Patient Assessment of Upper Gastrointestinal Disorders-Quality of Life (see also suppl table S10)
GES = Gastric Emptying Study
3M, 6M = 3 months, 6 months visit



Suppl Figure S2a. Measurement of pyloric distensibility. A balloon is introduced through the pylorus under endoscopic control and inflated automatically. Figure shows endoscopic image during measurement



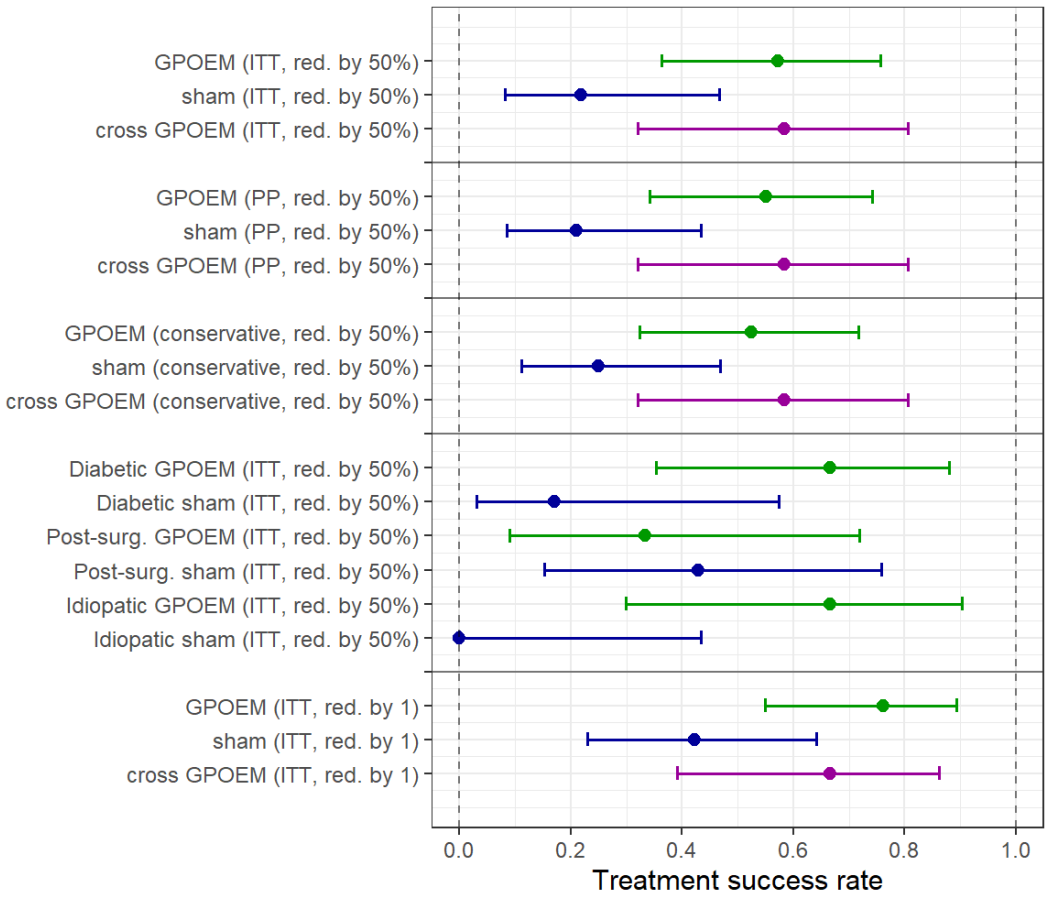
Suppl Figure S2b. Measurement of pyloric distensibility. An hourglass shape image on the Endoflip monitor during measurement. The narrowed place in the picture points to a pylorus. In the right down corner a value shows pyloric distensibility (3.1mm^2/mmHg)



Suppl Figure S3. Treatment success at 6 months after procedure, from top to bottom:

- the main outcome on the intention to treat (ITT) population with treatment success defined as reduction of the total GCSI score by 50% form baseline,
- treatment success evaluated on the per-protocol (PP) population (for cross-over the ITT population and PP population are the same),
- treatment success evaluated with the most conservative approach (worst case scenario), where the patient with technical failure of G-POEM is assigned failure and the sham patient who withdraw consent is assigned success (note, that overlap of confidence intervals does not exclude significant difference, which is 42% with 95% CI: 9% to 74% not containing zero),
- treatment success in sub-groups defined by etiology of gastroparesis,
- treatment success on the ITT population defined as reduction of the total GCSI score by 1 point from baseline.

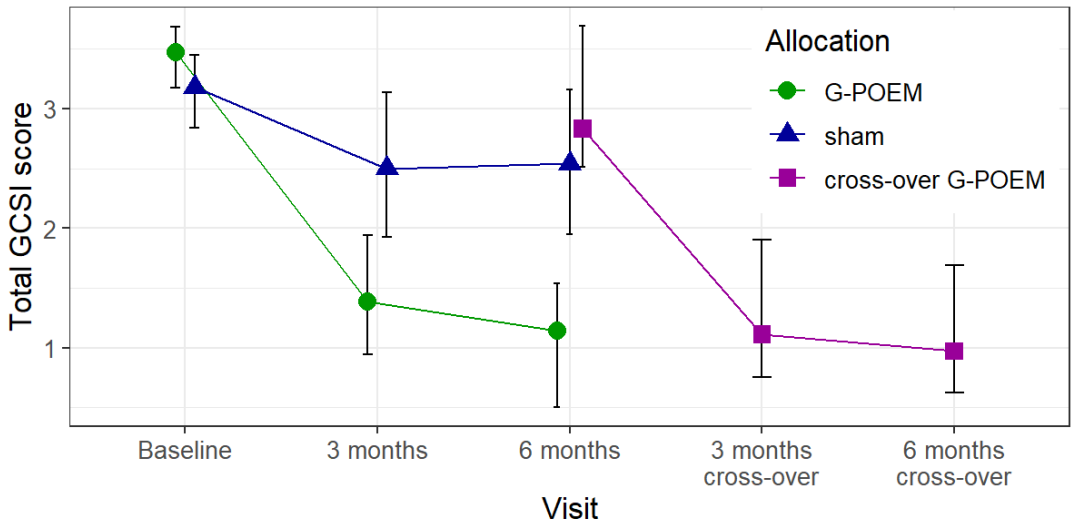
The results analyzed on the intention to treat (ITT) population (N=41, $N_{Di\text{-}G\text{-}POEM}$=9, $N_{Di\text{-}Sham}$=8, $N_{PS\text{-}G\text{-}POEM}$=6, $N_{PS\text{-}Sham}$=7, $N_{Id\text{-}G\text{-}POEM}$=6, $N_{Id\text{-}Sham}$=5, 1 value (2 %) imputed in diabetic GP patient in the sham group) are supplemented by the main outcome analysis on the per protocol (PP) population (N=39).



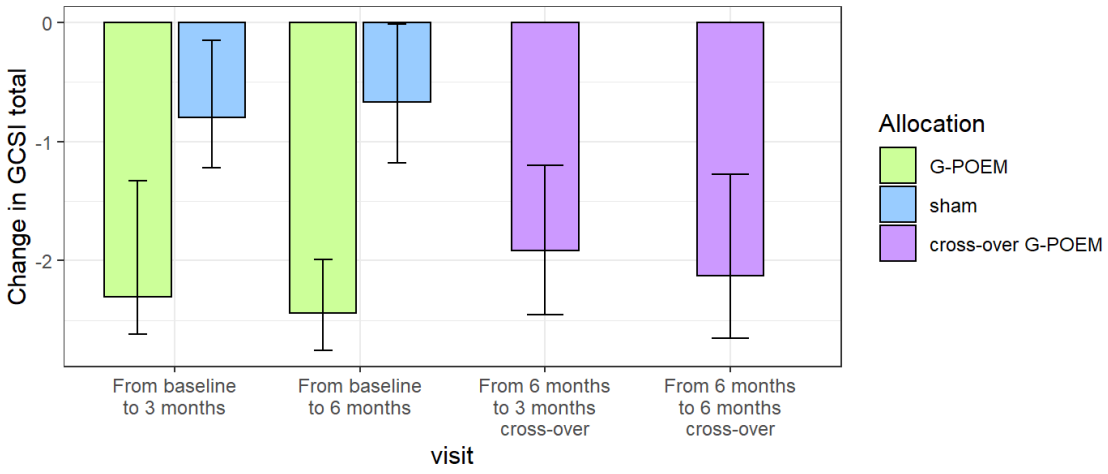
Suppl Figure S4. Treatment success 3 months after procedure, from top to bottom:

- treatment success in the G-POEM, sham and cross-over arms on the intention to treat (ITT) population with treatment success defined as reduction of the total GCSI score by 50% form baseline,
- treatment success evaluated on the per-protocol (PP) population,
- treatment success evaluated with the most conservative approach (worst case scenario), where the patient with technical failure of G-POEM is assigned failure (despite having success) and the sham patient who withdraw consent is assigned success,
- treatment success in sub-groups defined by etiology of gastroparesis (not evaluated for cross-over G-POEM due to low number of patients in groups),
- treatment success on the ITT population defined as reduction of the total GCSI score by 1 point form baseline.

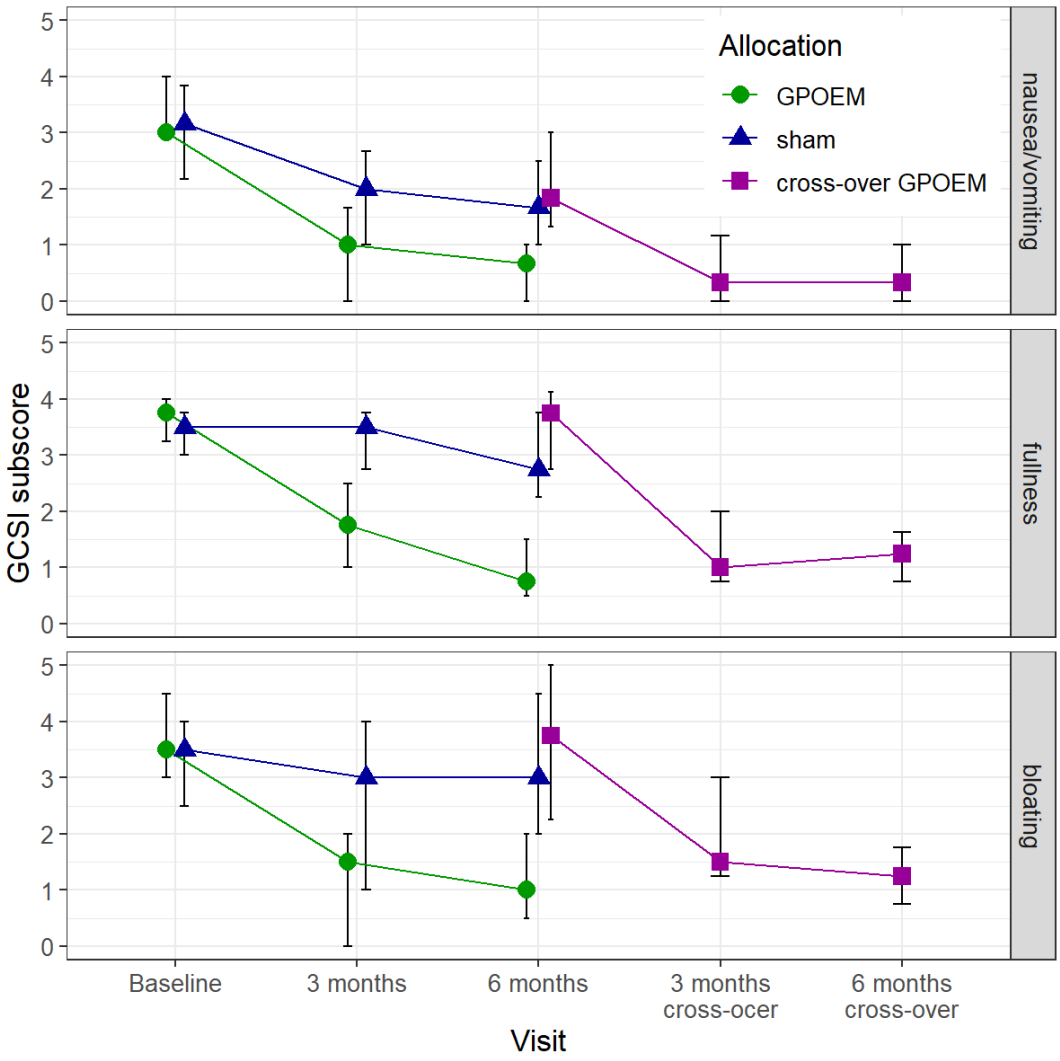
The results analyzed on the intention to treat (ITT) population (N=41, $N_{Di\text{-}G\text{-}POEM}$=9, $N_{Di\text{-}Sham}$=8, $N_{PS\text{-}G\text{-}POEM}$=6, $N_{PS\text{-}Sham}$=7, $N_{Id\text{-}G\text{-}POEM}$=6, $N_{Id\text{-}Sham}$=5, 1 value (2 %) imputed in diabetic GP patient in the sham group) are supplemented by the analysis on the per protocol (PP) population (N=39).



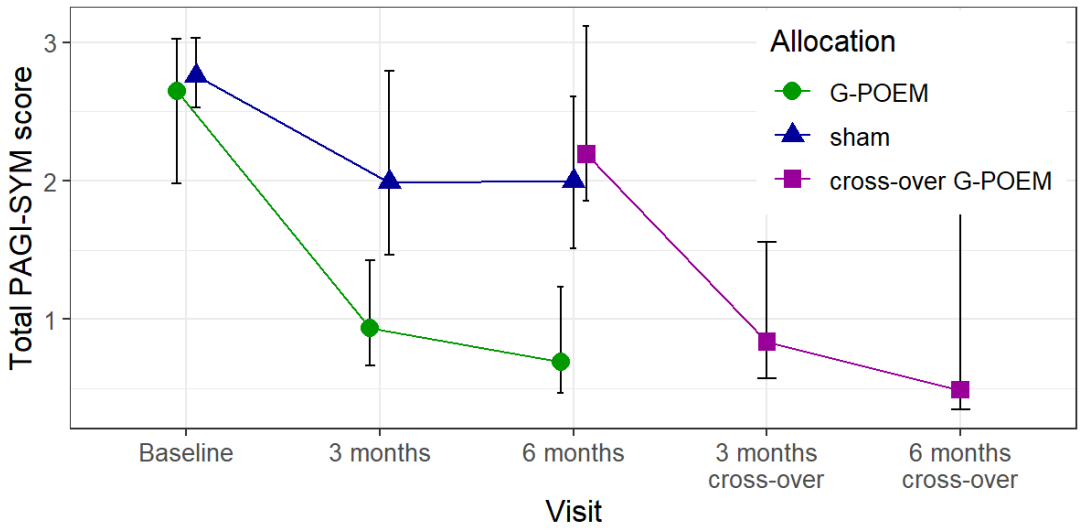
Suppl Figure S5a. Evolution of the GCSI total score. Point estimates of medians with 95% confidence intervals calculated on the ITT population are shown for patients after the G-POEM procedure (green circles, N=21), sham procedure (blue triangles, N=20, imputed 1 value (5 %) for 3 months and 1 value (5 %) for 6 moths), and cross-over G-POEM procedure (purple squares, N=12). For the cross-over G-POEM group the value at 6 months reflects only the data for the patients in this group (who subsequently underwent the cross-over G-POEM procedure). Points are connected for visual aid.



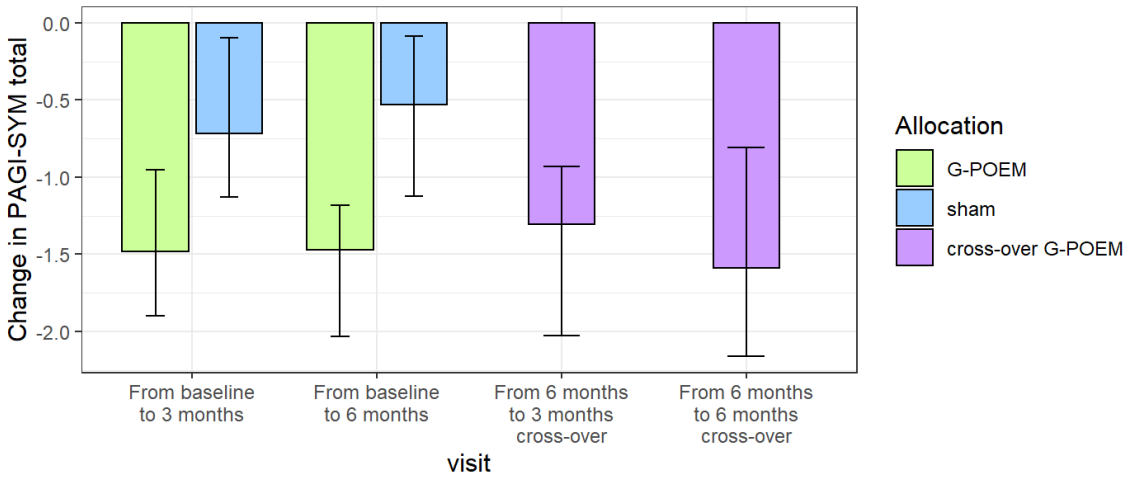
Suppl Figure S5b. Changes of the GCSI total score between visits. Point estimates of medians of differences between the specified visits with 95% confidence intervals calculated on the ITT population are shown for patients after the G-POEM procedure (green, N=21), sham procedure (blue, N=20, imputed 1 value (5 %) for 3 months and 1 value (5 %) for 6 months), and cross-over G-POEM procedure (purple, N=12) (patients who subsequently underwent the cross-over G-POEM procedure).



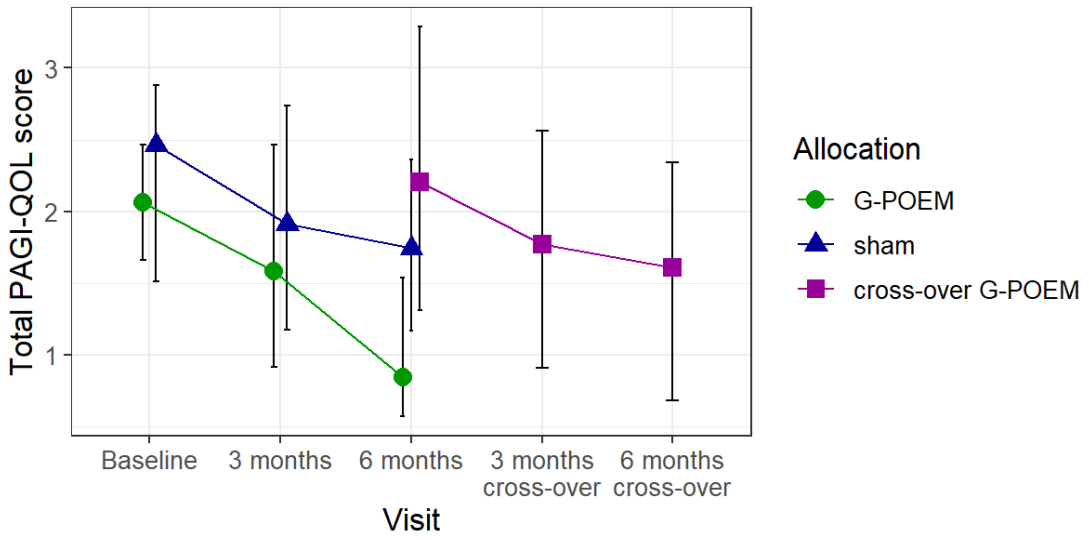
Suppl Figure S6. Evolution of the GCSI sub-scores. Point estimates of medians with 95% confidence intervals calculated on the available data are shown for patients after the G-POEM procedure (green circles, N=21), sham procedure (blue triangles, N=19), and cross-over G-POEM procedure (purple squares, N=12). For the cross-over G-POEM group the value at 6 months reflects only the data for the patients in this group (patients who subsequently underwent the cross-over G-POEM procedure). Points are connected for visual aid. The nausea / vomiting subscale comprises of the questions 1 to 3, Fullness of questions 4 to 7 and Bloating of questions 8 and 9, see Table S5.



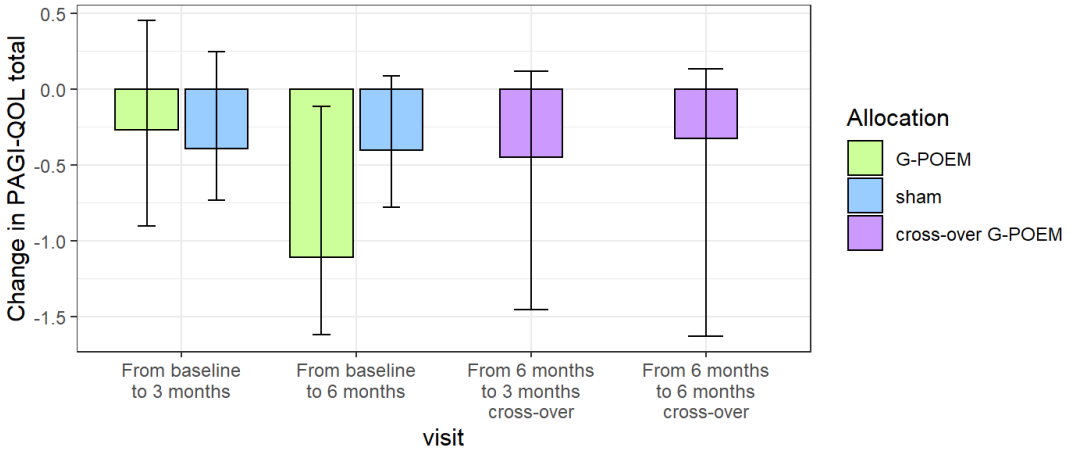
Suppl Figure S7a. Evolution of the PAGI-SYM total score. Point estimates of medians with 95% confidence intervals calculated on the ITT population are shown for patients after the G-POEM procedure (green circles, N=21), sham procedure (blue triangles, N=20, imputed 1 value (5 %) for 3 months and 1 value (5 %) for 6 months), and cross-over G-POEM procedure (purple squares, N=12). For the cross-over G-POEM group the value at 6 months reflects only the data for the patients in this group (patients who subsequently underwent the cross-over G-POEM procedure). Points are connected for visual aid.



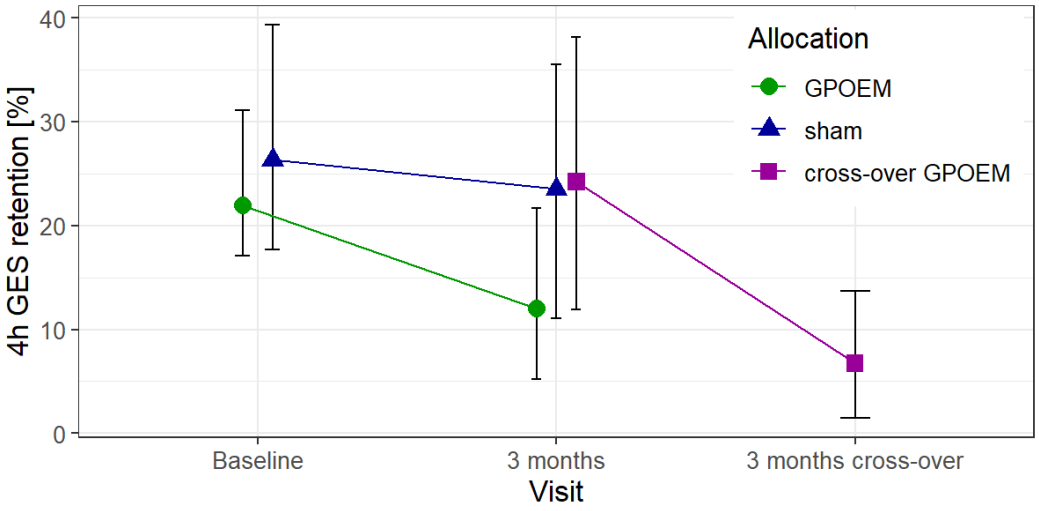
Suppl Figure S7b. Changes of the PAGI-SYM total score between visits. Point estimates of medians of differences between the specified visits with 95% confidence intervals calculated on the ITT population are shown for patients after the G-POEM procedure (green, N=21), sham procedure (blue, N=20, imputed 1 value (5 %) for 3 months and 1 value (5 %) for 6 months), and cross-over G-POEM procedure (purple, N=12) (patients who subsequently underwent the cross-over G-POEM procedure).



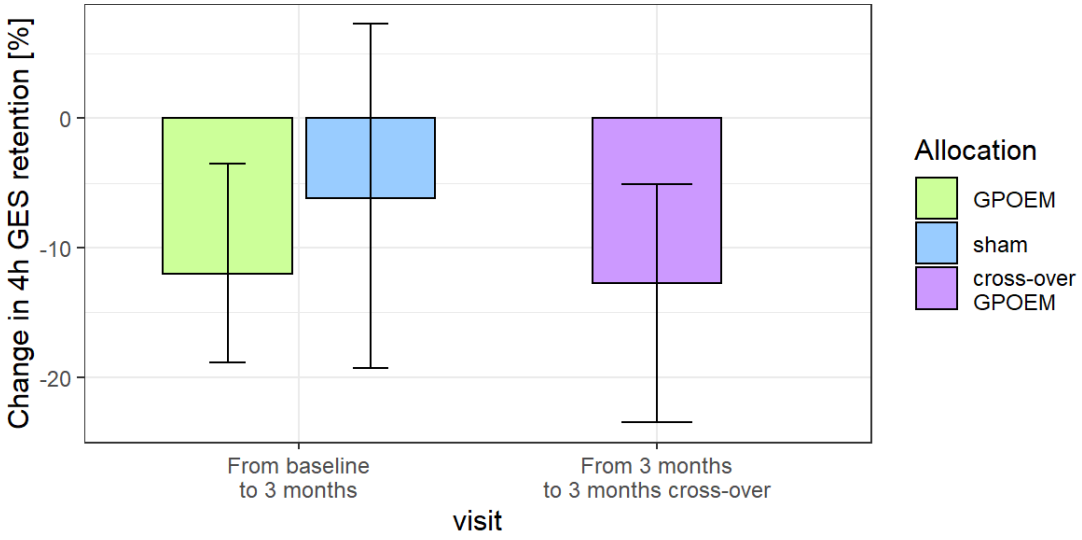
Suppl Figure S8a. Evolution of the PAGI-QoL total score. Point estimates of medians with 95% confidence intervals calculated on the ITT population are shown for patients after the G-POEM procedure (green circles, N=21, imputed 1 value (5 %) for 3 months and 1 value (5 %) for 6 months), sham procedure (blue triangles, N=20, imputed 1 value (5 %) for baseline, 1 value (5 %) for 3 months , and 1 value (5 %) for 6 months), and cross-over G-POEM procedure (purple squares, N=12). For the cross-over G-POEM group the value at 6 months reflects only the data for the patients in this group (patients who subsequently underwent the cross-over G-POEM procedure). Points are connected for visual aid.



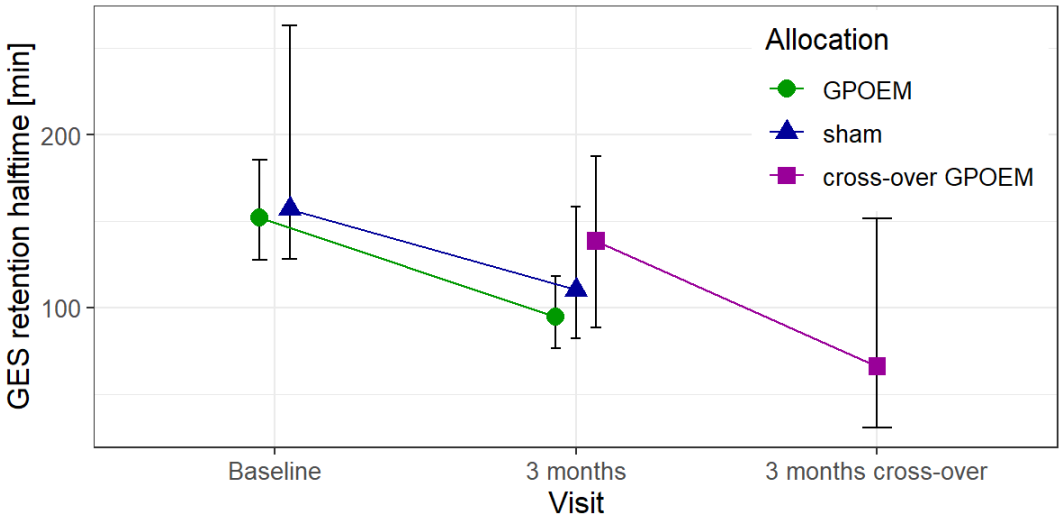
Suppl Figure S8b. Changes of the PAGI-QoL total score between visits. Point estimates of medians of differences between the specified visits with 95% confidence intervals calculated on the ITT population are shown for patients after the G-POEM procedure (green, N=21, imputed 1 value (5 %) for 3 months and 1 value (5 %) for 6 months), sham procedure (blue, N=20, imputed 1 baseline value (5 %), 1 value (5 %) for 3 months, and 1 value (5 %) for 6 months), and cross-over G-POEM procedure (purple, N=12) (patients who subsequently underwent the cross-over G-POEM procedure).



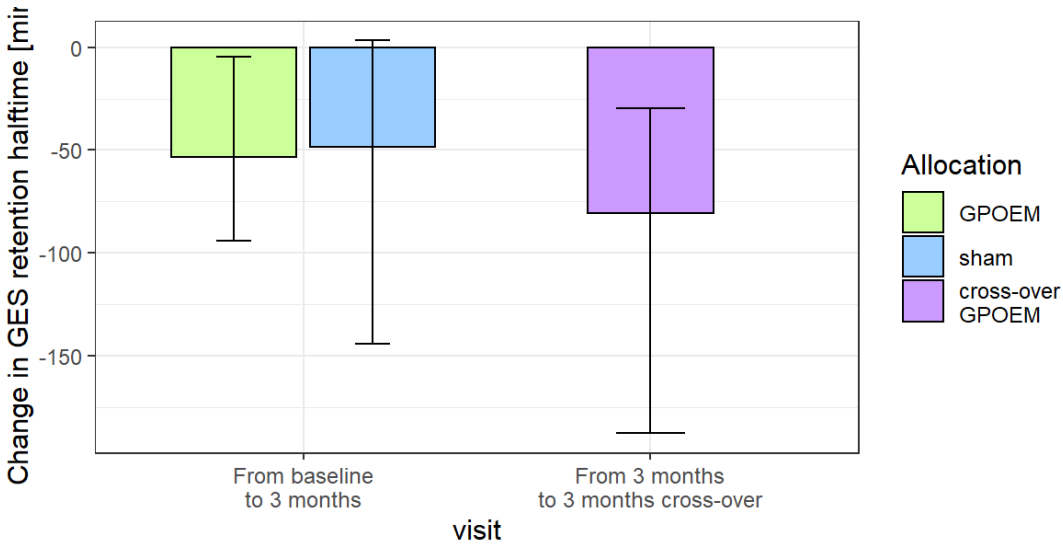
Suppl Figure S9a. Evolution of the GES 4h retention. Point estimates of medians with 95% confidence intervals calculated on the ITT population are shown for patients after the G-POEM procedure (green circles, N=21, imputed 2 values (10 %) for 3 months), sham procedure (blue triangles, N=20, imputed 1 value (5 %) for 3 months), and cross-over G-POEM procedure (purple squares, N=12). For the cross-over G-POEM group the value at 3 months reflects only the data for the patients in this group (patients who subsequently underwent the cross-over G-POEM procedure). Points are connected for visual aid.



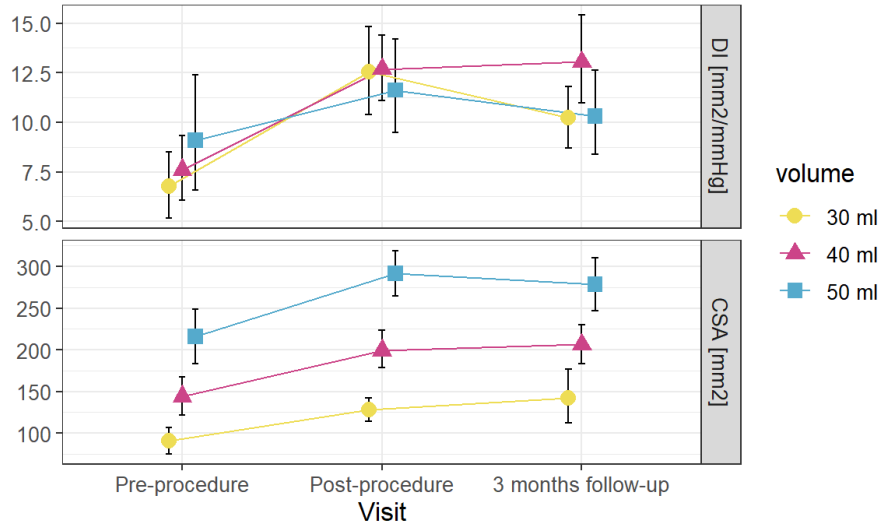
Suppl Figure S9b. Changes of GES 4h retention between visits. Point estimates of medians of differences between the specified visits with 95% confidence intervals calculated on the ITT population are shown for patients after the G-POEM procedure (green, N=21, imputed 2 values (10 %) for 3 months), sham procedure (blue, N=20, imputed 1 value (5 %) for 3 months), and cross-over G-POEM procedure (purple, N=12) (patients who subsequently underwent the cross-over G-POEM procedure).



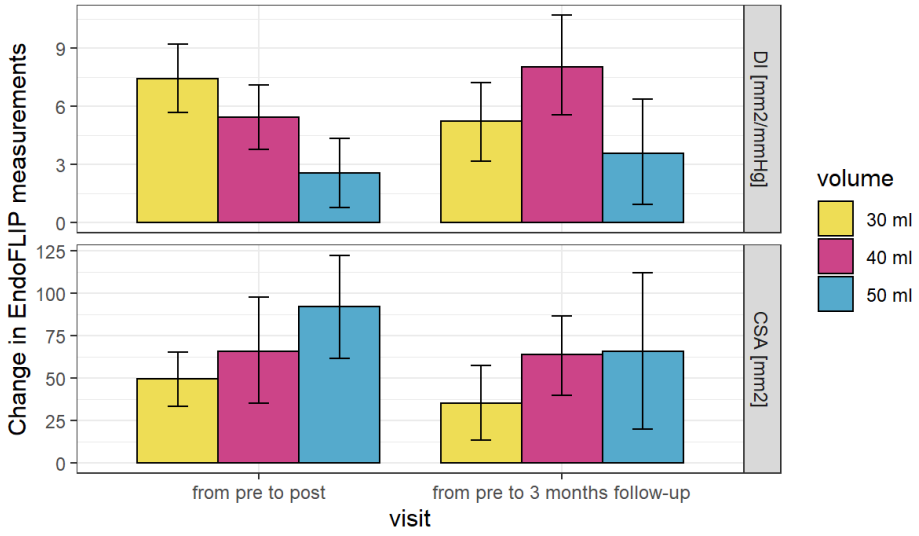
Suppl Figure S10a. Evolution of GES retention halftime. Point estimates of medians with 95% confidence intervals calculated on the ITT population are shown for patients after the G-POEM procedure (green circles, N=21, imputed 1 value (5 %) for 3 months), sham procedure (blue triangles, N=20, imputed 1 baseline value (5 %) and 1 value (5 %) for 3 months), and cross-over G-POEM procedure (purple squares, N=12). For the cross-over G-POEM group the value at 3 months reflects only the data for the patients in this group (patients who subsequently underwent the cross-over G-POEM procedure). Points are connected for visual aid.



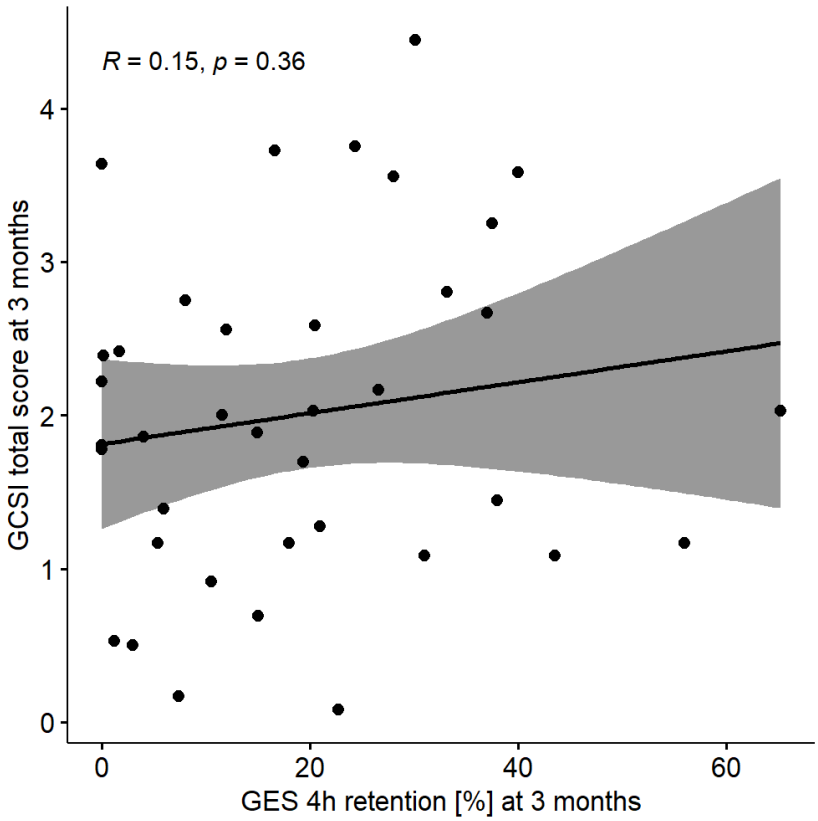
Suppl Figure S10b. Changes of GES retention halftime between visits. Point estimates of medians of differences between the specified visits with 95% confidence intervals calculated on the ITT population are shown for patients after the G-POEM procedure (green, N=21, imputed 1 value (5 %) for 3 months), sham procedure (blue, N=20, imputed 1 baseline value (5 %) and 1 value (5 %) for 3 months), and cross-over G-POEM procedure (purple, N=12) (patients who subsequently underwent the cross-over G-POEM procedure).



Suppl Figure S11a. Evolution of pyloric distensibility measurements (Endoflip) for different filling volumes – primary G-POEM and cross-over G-POEM combined**.** Point estimates of means for distensibility index (DI, top panel) and cross-sectional area (CSA, bottom panel) with 95% confidence intervals are shown for 30 mL (yellow circles), 40 mL (magenta triangles), and 50 mL (cyan squares) balloon fillings. The figure presents only available data; the imputation model was not used as for pre-procedure, post-procedure, and follow-up time points a total of 16, 17, and 15 (14 for 50 mL) values were available - this measurement was added after beginning of the trial. Points are connected for visual aid.



Suppl Figure S11b. Changes of measurements of pyloric distensibility by Endoflip for different filling volumes between visits – primary G-POEM and cross-over G-POEM combined**.** Point estimates of means of differences between the specified visits for distensibility index (DI, top panel) and cross-sectional area (CSA, bottom panel) with 95% confidence intervals are shown for 30 mL (yellow circles), 40 mL (magenta triangles), and 50 mL (cyan squares) balloon fillings. The figure presents only available data; the imputation model was not used as only 14 values were available for pre vs. post treatment difference and 12 for the pre vs. follow-up difference. The measurement of pyloric distensibility was added after beginning of the trial)



Suppl Figure S12. Correlation between GCSI total score and GES 4h retention at 3 months. All points for available data are plotted (no imputation performed) along with the linear regression line (black) and the corresponding confidence interval area (gray). The fact that also a decreasing line can be placed into the gray area indicates that there is no significant correlation. Correlation at 6 months can not be shown as GES was not measured at 6 months.

SUPPLEMENT

Finantial Support

a) The trial was finantially supported by a Grant 17-28797A from the Czech Ministry of Health
b) Medtronic provided for free Endoflip balloons for measurement of pyloric distensibility