

UC San Diego

UC San Diego Previously Published Works

Title

Interrater Reliability of Functional Lumen Imaging Probe Panometry and High-Resolution Manometry for the Assessment of Esophageal Motility Disorders.

Permalink

<https://escholarship.org/uc/item/80b8k93g>

Journal

The American Journal of Gastroenterology, 118(8)

Authors

Chen, Joan

Khan, Abraham

Chokshi, Reena

et al.

Publication Date

2023-08-01

DOI

10.14309/ajg.0000000000002285

Peer reviewed



Published in final edited form as:

Am J Gastroenterol. 2023 August 01; 118(8): 1334–1343. doi:10.14309/ajg.0000000000002285.

Interrater Reliability of Functional Lumen Imaging Probe (FLIP) Panometry and High-Resolution Manometry for Assessment of Esophageal Motility Disorders

Joan W Chen, MD, MS^{*,1}, Abraham Khan, MD^{*,2}, Reena V Chokshi, MD³, John O Clarke, MD⁴, Ronnie Fass, MD⁵, Jose M Garza, MD⁶, Milli Gupta, MD⁷, C Prakash Gyawali, MD⁸, Anand S Jain, MD⁹, Philip Katz, MD¹⁰, Vani Konda, MD¹¹, Adriana Lazarescu, MD¹², Kristle L Lynch, MD¹³, Felice Schnoll-Sussman, MD¹⁰, Stuart J Spechler, MD¹¹, Marcelo F Vela, MD¹⁴, Rena Yadlapati, MD, MS¹⁵, Jacob M Schauer, PhD¹⁶, Peter J Kahrilas, MD¹⁷, John E Pandolfino, MD, MS¹⁷, Dustin A Carlson, MD, MS¹⁷

¹Division of Gastroenterology & Hepatology, University of Michigan, Ann Arbor, MI, USA

²Center for Esophageal Health, Division of Gastroenterology & Hepatology, NYU Grossman School of Medicine, NYU Langone Health, New York, NY, USA

³Department of Medicine, Section of Gastroenterology and Hepatology, Baylor College of Medicine, Houston, TX, USA

Corresponding Author/ Article Guarantor: Joan W. Chen, Division of Gastroenterology & Hepatology, University of Michigan, 1500 E. Medical Center Dr., SPC 5362, Ann Arbor, MI 48109, USA, chenjoan@med.umich.edu.

*Indicates co-first authorship

Specific Author Contributions:

Study design/conception: DAC

Contribution to data collection: JWC, AK, RVC, JOC, JMG, MG, CPG, ASJ, PK, VK, AL, FSS, SJS, MFV, RY, JMS, PJK, JEP, DAC

Analysis of data: JMS, DAC

Interpretation of data: JWC, AK, JMS, DAC

Drafting of manuscript: JWC, AK, DAC

Critical revision of manuscript: All authors

Final approval of manuscript prior to submission: All authors

Disclosures:

JWC: Phathom Pharmaceuticals (Consulting)

AK: Medtronic (Consulting)

RVC: Medtronic (Speaking, Consulting)

RF: Takeda (Consulting, Speaking), Medtronic (Consulting), Phathom Pharmaceuticals (Consulting), GERDCare (Consulting),

Celexio (Consulting), Dexcal (Consulting), AztraZeneca (Speaking), Laborie/GI supply (Speaking), Eisai (Speaking), Johnson &

Johnson (Consulting, Speaking), Medicamenta (Speaking), Adock-Ingram (Speaking), Ginger Health (Stock options)

CPG: Medtronic, Ironwood, Takeda, Dexcel Pharma, Ardelyx, Johnson and Johnson (Consulting)

JMG: Medtronic (Speaking; Consulting); Abbott (Speaking); Biogaia (Speaking)

ASJ: AstraZeneca (Advisory board), Sanofi (Advisory board), Medtronic (Consulting)

VK: Ambu (Consulting), Castle (Consulting/Advisory board), Exact (Consulting/Advisory board), Medtronic (Consulting), Lucid

(Research)

KLL: Lucid (Consulting), Takeda (Consulting), Sanofi (Consulting)

RY: Medtronic (Institutional Consulting); Ironwood (institutional Consulting); Medstat Linx (Institutional Consulting);

RJSMediagnostix (Advisory Board); Phathom Pharmaceuticals (Consulting)

PJK: AstraZeneca (Consulting), Medtronic (License, shares patent with Medtronic for FLIP Panometry and receives \$75.00

per machine sold), Ironwood (Consulting), Reckitt (Consulting), Johnson and Johnson (Consulting), Phathom Pharmaceuticals

(Consulting)

JEP: Sandhill Scientific/Diversatek (Consulting, Speaking, Grant), Takeda (Speaking), Astra Zeneca (Speaking), Medtronic

(Speaking, Consulting, License, shares patent with Medtronic for FLIP Panometry and receives \$75.00 per machine sold), Torax

(Speaking, Consulting), Ironwood (Consulting)

DAC: Medtronic (Speaking; Consulting; License); Phathom Pharmaceuticals (Consulting)

⁴Division of Gastroenterology & Hepatology, Stanford University School of Medicine, Redwood City, CA, USA

⁵Division of Gastroenterology & Hepatology, MetroHealth Medical Center, Cleveland, OH, USA

⁶GI Care for Kids, Neurogastroenterology and Motility Program Children's Healthcare of Atlanta, Atlanta, GA, USA

⁷Division of Gastroenterology & Hepatology, University of Calgary, Calgary, Alberta, Canada

⁸Division of Gastroenterology, Washington University School of Medicine, St. Louis, MO, USA

⁹Division of Digestive Diseases, Emory University Department of Medicine, Atlanta GA, USA

¹⁰Department of Gastroenterology Weill Cornell Medical Center, New York, NY, USA

¹¹Division of Gastroenterology, Baylor University Medical Center, Dallas, TX, USA

¹²Division of Gastroenterology, University of Alberta, Edmonton, Alberta, Canada

¹³Division of Gastroenterology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

¹⁴Division of Gastroenterology, Mayo Clinic Arizona, Scottsdale, AZ, USA

¹⁵Division of Gastroenterology, University of California San Diego, La Jolla, CA, USA

¹⁶Department of Preventive Medicine, Division of Biostatistics, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

¹⁷Department of Medicine, Division of Gastroenterology and Hepatology, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

Abstract

Introduction: High-resolution manometry (HRM) and functional lumen imaging probe (FLIP) are primary and/or complementary diagnostic tools for evaluation of esophageal motility. We aimed to assess the interrater agreement and accuracy of HRM and FLIP interpretation.

Methods: Esophageal motility specialists from multiple institutions completed the interpretation of 40 consecutive HRM and 40 FLIP studies. Interrater agreement was assessed using intraclass correlation coefficient (ICC) for continuous variables and Fleiss' kappa statistics for nominal variables. Accuracies of rater interpretation were assessed using the consensus of three experienced raters as the reference standard.

Results: Fifteen raters completed the HRM and FLIP studies. Excellent interrater agreement was seen in supine median Integral Relaxation Pressure (ICC 0.96 (95% CI 0.95, 0.98)), and good agreement was seen with assessment of esophagogastric junction (EGJ) outflow, peristalsis, and assignment of a CCv4.0 diagnosis using HRM ($\kappa=0.71, 0.75, \text{ and } 0.70$). Excellent interrater agreement for EGJ-distensibility index and maximum diameter (0.91 (0.90, 0.94), 0.92 (0.89, 0.95)) was seen and moderate to good agreement was seen in assignment of EGJ opening classification, contractile response pattern, and motility classification ($\kappa=0.68, 0.56, 0.59$) on FLIP. Rater accuracy for CCv4.0 diagnosis on HRM was 82% (95% confidence interval 78–84%) and for motility diagnosis on FLIP Panometry was 78% (72–81%).

Conclusions: Our study demonstrates high levels of interrater agreement and accuracy in interpretation of HRM and FLIP metrics and moderate to high levels for motility classification in FLIP, supporting the use of these approaches for primary or complementary evaluation of esophageal motility disorders.

Keywords

Dysphagia; achalasia; lower esophageal sphincter; esophagogastric junction; peroral endoscopic myotomy; integrated relaxation pressure; distensibility index; contractile response

Introduction

The Chicago Classification (CC) has standardized categorization of esophageal motility disorders using esophageal high-resolution manometry (HRM), especially with version 4.0 (v4.0), the current version.(1–4) Studies which compared conventional line tracing with pressure topography plots consistently showed improved diagnostic accuracy, ease of interpretation and better interrater agreement with HRM.(5–7) As HRM gained acceptance worldwide, inter-rater and intra-rater agreement in classifying achalasia using CC was shown to be high,(8) although inter-rater agreement for diagnosing esophageal motility disorders overall was modest but acceptable.(9) The more expansive protocol of swallow positioning and provocative maneuvers with CCv4.0 has revised conclusive and inconclusive definitions for several motility disorders, with recommendation of supportive testing with a timed barium esophagram (TBE) or functional lumen imaging probe (FLIP) when diagnosis is inconclusive.(4, 10–15)

FLIP uses impedance planimetry technology with controlled volumetric distension of an infinitely compliant catheter mounted balloon during sedated endoscopy to measure lumen dimensions as well as distensibility along the length of the esophagus and esophagogastric junction (EGJ) and has burgeoning utility across several areas of clinical esophagology. (16) Standardization of the FLIP protocol to incorporate EGJ opening mechanics and the contractile response to distension provides evaluation of esophageal motility, termed FLIP Panometry.(17) Compared to esophageal HRM using CCv4.0, FLIP Panometry accurately identifies clinically relevant conclusive EGJ outflow obstruction including achalasia spectrum disorders. Further, classification of motor disorders by FLIP Panometry parallels HRM classification overall, including CCv4.0 disorders of peristalsis.(18, 19) Thus, FLIP Panometry serves as a well-tolerated method of evaluation of esophageal motility that can complement or in some cases be an alternative to HRM.

This study aimed to assess inter-rater agreement of metrics and disorder definitions of both HRM using CCv4.0 and FLIP Panometry. A secondary aim was to determine accuracy of interpretation using each diagnostic modality, compared to consensus gold standard interpretation by experienced raters.

Methods

Subjects

Of 17 esophageal motility specialists (all members of a FLIP working group) invited to participate in this study, 15 completed interpretation of identical sets of deidentified HRM and FLIP studies (Figure 1). Raters evaluated HRM and FLIP studies in randomized order (i.e. interpreted 40 HRM studies first or interpreted 40 FLIP Panometry studies first). To assess accuracy, interpretations by the study participants were compared against a group consensus interpretation by three experienced raters (JEP, PJK, and DAC) who created “reference diagnoses” for each HRM and FLIP study. Consensus on a reference diagnosis was able to be reached on 100% of studies; independent interpretation to facilitate formal agreement (kappa) was not assessed.

The deidentified HRM and FLIP studies were acquired from 40 consecutive, adult patients evaluated for dysphagia at the Esophageal Center of Northwestern, who were prospectively enrolled in the parent study (P01 DK117824). Patients with previous foregut surgery, mechanical esophageal obstruction (such as eosinophilic esophagitis, stricture, ring, or Los Angeles grade C-D erosive esophagitis), or hiatal hernia >3cm were excluded as these preclude application of CCv4.0 for diagnosis of primary esophageal motility disorders.(4) In each instance, HRM and FLIP studies were performed within 3 months of each other (36 of 40 patients had both on the same day). Raters were instructed that each patient was evaluated for dysphagia, had not had previous foregut surgery, and had completed upper endoscopy with no evidence of mechanical esophageal obstruction.

Following each study interpretation, raters were asked to select a management recommendation from a dropdown menu based on the HRM or FLIP findings, in the context of dysphagia, no previous foregut surgery, and a non-obstructive upper endoscopy. The menu includes common management options for esophageal conditions: obtain additional testing (e.g. FLIP and/or esophagram); reassurance, PPI, and/or GI-focused cognitive behavioral therapy; esophageal smooth muscle relaxant (e.g. off-label sildenafil); endoscopic dilation (20mm) and/or botulinum toxin injection; and POEM, laparoscopic Heller, or pneumatic dilation. Raters also reported the overall time spent to complete interpretation of each set of studies and assessed difficulty of interpretation on a scale of 1 (very easy) to 100 (very difficult).

Informed consent was obtained from raters and patients for study participation. The study protocol was approved by the Northwestern University Institutional Review Board.

HRM Study Protocol and Analysis

High-resolution impedance manometry (HRM) studies and interpretation were performed using a solid-state assembly with 36 circumferential pressure sensors at 1-cm intervals and 18 impedance sensors at 2-cm intervals (Medtronic Inc, Shoreview, MN) according to the CCv4.0.(4) After a minimum 6-hour fast, the HRIM assembly was placed transnasally and positioned to record from the hypopharynx to the stomach with approximately three intragastric pressure sensors. After a 2-minute baseline recording, the HRM protocol was performed with ten, 5-ml liquid swallows in the supine position and five, 5-ml

liquid swallows in the upright position. These swallows formed the basis of the rater impression per CCv4.0.(4) Additionally, two multiple-rapid swallow sequences (five 5-ml liquid swallows at 2–3 second intervals) and a rapid-drink challenge (RDC) involving rapid drinking of 200cc liquid were performed in the seated upright position as complementary components of the study. Manometry studies were analyzed using ManoView analysis software (Medtronic, Inc). Raters were provided with a summary of CCv4.0 criteria (Table S1) and the CCv4.0 associated publications.(4, 11–15) Raters recorded values for median supine and upright IRPs, classification of EGJ outflow, classification of primary peristalsis, and classification of esophageal motility.

FLIP Study Protocol and Analysis

The FLIP study using 16-cm FLIP (EndoFLIP® EF-322N; Medtronic, Inc, Shoreview, MN) was performed during sedated endoscopy as previously described.(20) The FLIP study included stepwise 10-ml FLIP distensions (each stepwise distension volume being maintained for 60 seconds) with the FLIP catheter positioned across the EGJ (1–3 intragastric channels). Interpretation focused on the 50ml, 60ml, and 70ml fill volumes (based on the previously described FLIP classification schemes). The FLIP Panometry studies were reviewed by raters as deidentified recordings of real-time FLIP Panometry studies (FLIP 2.0, Medtronic, Inc).(18, 21, 22) Raters were instructed to view the recordings as they would during an endoscopic encounter, i.e. continuous forward play of the recording with pause of the recording allowed to review and/or record measures. Raters were provided with a summary of FLIP Panometry motility classification criteria (Table S2), a 20 minute video tutorial, and the associated publications.(18, 21, 22) Raters recorded values for EGJ-distensibility index (DI) during 60ml FLIP fill volume and maximum EGJ diameter (60–70ml FLIP fill volume), classification of EGJ opening, classification of contractile response pattern, and classification of esophageal motility.

Statistical Analysis

Interrater agreement was assessed using intraclass correlation coefficient (random rater-random patient model) for continuous variables and with Fleiss' kappa statistic for nominal variables. Degree of agreement was interpreted based on kappa values as poor (0–0.2), fair (0.21–0.4), moderate (0.41–0.6), good (0.61–0.8), and excellent (0.81–1). Accuracy was calculated for each rater as the percent agreement with the reference diagnoses; group accuracy was then described as the mean (95% confidence interval (CI)) among the raters. For additional description of accuracy, confusion matrices of all rater responses relative to reference were pooled per test, i.e. 600 responses (15×40) for both HRM and FLIP. Results of continuous variables were reported as mean (standard deviation; SD), or median (interquartile range; IQR) depending on data distribution. Paired t-tests were two-sided as appropriate and were considered statistically significant at $P < 0.05$.

Results

Raters

Among the 15 raters that completed the study, all reported previous experience of interpreting >500 HRM studies (Table 1). 5 (33%) raters had previous experience of

interpreting >500 FLIP Panometry cases, 7 (47%) reported 100–499 previous cases, and 3 (20%) reported 25–100 previous cases. Differences were not detected between raters reported interpretation time (mean (SD) 263 (73) minutes for HRM vs 221 (75) for FLIP; $P=0.136$) or interpretation difficulty (mean (SD) rating 31 (29) for HRM vs 29 (18) for FLIP; $P=0.589$). The three *experienced* raters (JEP, PJK, and DAC) had each completed >500 HRM and >500 FLIP cases.

Interrater reliability

Table 2 includes the intra-rater agreement data for HRM and FLIP. For HRM interpretation, inter-rater agreement (intraclass correlation coefficient (95% confidence interval)) was excellent for supine median IRP (0.96 (0.95, 0.98)) as well as for upright (0.97 (0.96, 0.98)). Excellent agreement was seen whether the raters completed the HRM or FLIP studies first. Fleiss' kappa indicated good inter-rater agreement in assessing normality of esophagogastric outflow ($\kappa=0.71$), assessment of peristalsis ($\kappa=0.75$), and assigning a Chicago Classification diagnosis ($\kappa=0.70$) on HRM. FLIP interpretation also demonstrated excellent inter-rater agreement for EGJ-DI at 60mL balloon volume (0.91 (0.90, 0.94)) and maximum EGJ diameter (0.92 (0.89, 0.95)); again, whether raters first completed HRM or FLIP studies. On FLIP studies, agreement in EGJ opening classification was good ($\kappa=0.68$) and moderate for contractile response pattern and motility classification ($\kappa=0.56$ and 0.59, respectively). Figure 2 shows examples of HRM and FLIP studies with rater diagnosis patterns.

Accuracy of HRM interpretation

The most common diagnosis by the raters for HRM and FLIP studies for each patient is given in Table S3 (along with experienced rater consensus agreement diagnosis). Accuracy of the raters, when comparing to the reference diagnosis was 0.75 (95% CI 0.71, 0.78) for interpretation of EGJ outflow on HRM. The raters had an accuracy of 0.83 (0.80, 0.86) for interpreting peristalsis pattern (absent, premature, hypercontractile, ineffective, or normal). The raters had an accuracy of 0.82 (0.78, 0.85) for esophageal motility diagnosis (CCv4.0). Among the 111 "inaccurate" diagnoses totaled across HRM studies by the entire group, a type II achalasia or EGJOO diagnosis was involved in 84 (76%); Table 3. There were 15 rater diagnoses of type II achalasia in studies with reference diagnosis of type III achalasia or EGJOO. There were zero inaccuracies that involved rater diagnosis of normal motility in studies with reference diagnosis of achalasia (all subtypes) or absent contractility. Similarly, there were zero rater diagnoses of achalasia or absent contractility in studies with reference diagnosis of normal motility (Table 3).

Accuracy of FLIP interpretation

Rater accuracy of interpreting EGJ opening using FLIP was 0.82 (0.79, 0.85) and for assessing contractile response pattern was 0.74 (0.71, 0.78). Rater accuracy for FLIP motility diagnosis was 0.78 (0.74, 0.81). Among the 134 'inaccurate' FLIP motility diagnoses, an 'inconclusive' or spastic-reactive diagnosis was involved in 93% (125/134); Table 4. There was 1 rater diagnosis of obstruction with weak CR among studies with a reference diagnosis of 'weak' and 12 rater diagnoses of obstruction with weak CR among studies with a reference diagnosis of 'inconclusive' (Table 4). There were zero inaccuracies that involved rater diagnosis of obstruction with weak contractile response or

spastic-reactive in studies with reference diagnosis of normal. Similarly, there were zero rater diagnoses of normal motility in studies with reference diagnoses of obstruction with weak contractile response or spastic-reactive (Table 4).

Management Decisions

Raters' responses regarding management recommendations based on test results included a recommendation to obtaining additional testing after 42% of HRM interpretations and 70% of FLIP interpretations. Additional testing was recommended with 96% of EGJ outflow diagnoses on HRM and with 90% of interpretations of inconclusive FLIP. Reassurance and conservative management with a trial of proton pump inhibitor or neuromodulator was recommended after 27% of HRM cases compared to 17% of FLIP cases. When the most common HRM diagnosis by raters is achalasia (all subtypes), management recommendation most commonly (74%) involved pneumatic dilation (PD), laparoscopic Heller myotomy (LHM), or peroral endoscopic myotomy (POEM). Diagnoses of obstruction with weak contractile response on FLIP; however, triggered a recommendation for additional testing in 70%; the other 30% were associated with a recommendation for PD, LHM, or POEM. Normal HRM interpretations lead to reassurance with 67% of diagnoses and request for additional information with 32% of diagnoses, while normal FLIP interpretations lead to reassurance with 74% of diagnoses and request for additional testing with 24% of diagnoses.

Discussion

In this study evaluating reliability of HRM and FLIP interpretation among esophageal motility specialists, we demonstrated that raters extracted key HRM and FLIP metrics with moderate to almost perfect inter-rater agreement, and high accuracy compared to reference standard interpretations. Formal interpretation of HRM using CCv4.0 had accuracy of 82%, and FLIP had accuracy of 78%. Importantly, there were no FLIP diagnoses of normal EGJ opening in patients with established achalasia, and no FLIP diagnosis of achalasia in patients with normal EGJ opening and contractility; similar results were demonstrated using HRM. Notably, there was disagreement even among esophageal specialists in non-obstructive motor disorders, but raters frequently indicated that they would request alternate confirmatory testing prior to invasive management, conforming to current guidelines and recommendations. We conclude that esophageal specialists follow similar interpretation paradigms using HRM and FLIP and recognize the complementary value of these tests when findings are inconclusive.

Previous studies assessing interobserver and intraobserver agreement of HRM have used earlier versions of the Chicago Classification and included only a small number of raters.(8, 9, 24, 25) Rater reliabilities in HRM (using CCv4.0) and FLIP interpretation had not been assessed. Our study included a group of esophageal motility specialists with extensive HRM experience and varying experiences with FLIP interpretation to assess rater reliability. There was no difference reported by the raters in the time commitment or perceived difficulty in HRM vs FLIP interpretations. Our study shows substantial to almost perfect interrater agreement in assessment of esophagogastric outflow, peristalsis and a CCv4.0 diagnosis on HRM, and similar substantial to almost perfect agreement in EGJ distensibility, maximum

EGJ diameter, and EGJ opening classification using FLIP. Agreement was moderate for contractile response pattern and motility classification using FLIP. The high agreement in HRM interpretation is reassuring noting this cohorts' experience (likely thousands of HRM studies interpreted) with HRM and the Chicago Classification evolution over the past 10+ years where HRM has been considered the primary diagnostic tool for esophageal motility. On the other hand, FLIP Panometry is not currently broadly used as a first-line diagnostic tool and the majority of our raters had less than 500 cases of experience. Thus already, the FLIP Panometry approach to assessing esophageal motility performed at a highly functional stage and likely will improve with additional evolution and experience of its users. Similar to the progression and standardization with the Chicago Classification of HRM, we expect future iterations of a FLIP Panometry classification will allow for improved agreement and diagnostic confidence amongst users.

There has been an increasingly recognized importance to establish accurate esophageal motility diagnoses in order to provide appropriate and necessary treatments, while also avoiding inappropriate invasive therapies. This was highlighted by the latest iteration of the Chicago Classification which provided both revised algorithms for diagnoses of esophageal motility disorders as well as recommendations for further testing during instances in which a conclusive motility diagnosis is not reached.(4) In our study, accuracy for interpretation of EGJ opening and peristalsis/contractile response pattern was high when comparing to the consensus agreement of experienced raters using HRM or FLIP. Of note, none of the patients with achalasia on HRM, regardless of subtype, were diagnosed as normal by the raters, and none of the normal peristalsis patients were given a diagnosis of achalasia by a rater. Furthermore, a FLIP Panometry accurate diagnosis of normal was never assessed as obstruction with weak contractile response (suggesting achalasia physiology) by a rater, while none of the obstruction with weak contractile response patients were labeled as normal by a rater as well. Thus, the overall clinical relevance of many of the 'inaccurate' interpretations was minimal as they unlikely would relate to major alterations in management decisions. However, several cases with variation in interpretation even among this group of experienced providers have occurred, e.g. Figure 2B with interpretations of both achalasia and IEM and thus significantly different management approaches. Additionally, panesophageal pressurization seen in type II achalasia could potentially have been confused with simultaneous esophageal contractions seen in type III achalasia on HRM (e.g. Figure 2A), and spastic reactive features may be variable on FLIP (e.g. Figure 2D). These; however, may also reflect scenarios that would be circumvented by application of complementary test results.

The CCv4.0 provided recommendations based on the use of HRM as the primary diagnostic tool for esophageal motility and has included FLIP as a complementary test in cases of inconclusive HRM diagnoses, such as EGJ outflow obstruction. Timed barium esophagram was also recommended and offers an alternative complementary test for inconclusive initial evaluations. However, FLIP has also been suggested as a potential first-line diagnostic test in patients with dysphagia if performed during an index upper endoscopy, as this can both avoid treatment delay as well as potentially replace or determine the need for HRM and the discomfort associated with an intranasal catheter.(17, 23) FLIP may be particularly useful in identifying patients with normal motility. In our study, FLIP led to a normal

motility diagnosis in 1 of 5 studies overall. Thus, using FLIP as a first-line tool on index endoscopy in patients with nonobstructive dysphagia may eliminate need for HRM in some patients, such as if FLIP is normal. In this study, when the most common rater diagnosis was normal, the reference FLIP diagnosis was also normal in all cases, and this correlated with normal HRM diagnosis in 75% of cases (two cases of normal FLIP correlated with EGJOO on HRM, though both had normal clearance on TBE; Table S3). However, if FLIP is abnormal, this can alternatively emphasize the need to pursue additional esophageal motility testing, where the FLIP findings will still provide a potentially useful complement (e.g. an inconclusive HRM diagnosis such as EGJOO).

When we assessed the potential impact on management recommendations, we overall observed that the most frequent recommendation from this group after interpretation with HRM or FLIP was to obtain additional testing. This aligns with recommendations from CCv4.0 related to obtaining complementary testing when diagnostic uncertainty persists after an initial test. Further, a recommendation for additional testing was more frequent in FLIP than HRM, supporting that FLIP was still often recommended as a complementary test in this group. However, reassurance/PPI/CBT was the recommendation associated with 74% of normal FLIP Panometry diagnoses, again, supporting that a normal FLIP on initial investigation can obviate the need for HRM in these cases. Overall, with the goal of an esophageal motility evaluation being to reach a confident diagnosis to guide treatment decisions, it was notable that utilization of more than a single motility test was often the recommendation from the individuals of this group. This also likely relates to the growing recognition that there is not a single ‘gold standard’ test for the diagnosis of esophageal motility disorders, but that instead a comprehensive evaluation may be necessary to achieve a diagnosis.

This study was the first to establish a baseline interrater agreement analysis for both the updated CCv4.0 and the latest FLIP Panometry recommended protocol. The results support that the present classification schemes for both testing modalities were associated with reliable interpretation among this group of esophageal motility specialists. A strength of the study was that we used consecutive HRM and FLIP studies, simulating the real-life scenario and including many non-classic and borderline cases. This is in contrast to the previous study assessing interrater agreement and accuracy of manometry interpretation in which non-consecutive studies were selected to represent each motility diagnosis for comparison of esophageal pressure topography vs. conventional line tracing that has the obvious drawback of being less generalizable to real life interpretation.⁽⁶⁾ However, a limitation of the present study was that this patient cohort was from an esophageal referral center, particularly noting the high rate of patients with achalasia. While this may somewhat limit generalizability to other practices, this patient cohort did provide the sufficient sample size of the most actionable diagnosis in the esophageal motility evaluation – achalasia. Another limitation was the lack of patients with distal esophageal spasm or hypercontractile esophagus cases, which reflects the rarity of these conditions. In the current study, our raters were all motility specialists with a high degree of familiarity with HRM and FLIP. As a result, we were unable to observe and compare the learning curve and requirements for competency in HRM and FLIP interpretations as those previously assessed in HRM.^(26, 27) However, doing so is an intended future direction. Further, our study that focused on interrater agreement required

independent review of each test, as opposed to complementary application of other clinical data (including FLIP, HRM, and often TBE as performed sequentially in a patient with borderline/inconclusive motility findings) that would often impact management decisions. For reference diagnoses, we used the consensus agreement of three experienced raters (who served integral roles in the development of the diagnostic schemes utilized for both HRM and FLIP) as reference diagnosis when determining accuracy. However, noting the absence of a true gold standard diagnostic tool, this may be viewed as a limitation.

In conclusion, interrater agreement for esophageal motility diagnosis using HRM and CCv4.0 and FLIP Panometry was moderate to good in our group of motility specialists. Raters achieved high levels of accuracy using both modalities for motility diagnosis when compared to the experienced rater consensus diagnoses. While these results support that HRM diagnoses per CCv4.0 and the novel FLIP Panometry motility classifications can be reliably interpreted by esophageal motility specialists, it also demonstrates that inconsistency may still occur. This highlights the importance of recognizing uncertainty in esophageal motility diagnoses and the need to seek additional complementary testing in order to improve confidence in an accurate clinical diagnosis – particularly prior to pursuing invasive therapies. Noting that studies including non-expert GI providers are needed for assessment of the learning curve of each testing modality, this study also highlights the need for development of esophageal motility learning modules and analytic tools, such as development of computer-assisted diagnostic support tools to further improve the reliability and accuracy of study interpretation. Overall, ongoing development of robust and reliable management algorithms based on complementary application of esophageal motility tests will continue to improve the standardization of care in patients with esophageal motility disorders.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Grant support:

This work was supported by P01 DK117824 (JEP) from the Public Health service.

References

1. Bredenoord AJ, Fox M, Kahrilas PJ, Pandolfino JE, Schwizer W, Smout AJ, et al. Chicago classification criteria of esophageal motility disorders defined in high resolution esophageal pressure topography. *Neurogastroenterol Motil.* 2012;24 Suppl 1:57–65. [PubMed: 22248109]
2. Kahrilas PJ, Bredenoord AJ, Fox M, Gyawali CP, Roman S, Smout AJ, et al. The Chicago Classification of esophageal motility disorders, v3.0. *Neurogastroenterol Motil.* 2015;27(2):160–74. [PubMed: 25469569]
3. Pandolfino JE, Fox MR, Bredenoord AJ, Kahrilas PJ. High-resolution manometry in clinical practice: utilizing pressure topography to classify oesophageal motility abnormalities. *Neurogastroenterol Motil.* 2009;21(8):796–806. [PubMed: 19413684]
4. Yadlapati R, Kahrilas PJ, Fox MR, Bredenoord AJ, Prakash Gyawali C, Roman S, et al. Esophageal motility disorders on high-resolution manometry: Chicago classification version 4.0(c). *Neurogastroenterol Motil.* 2021;33(1):e14058. [PubMed: 33373111]

5. Soudagar AS, Sayuk GS, Gyawali CP. Learners favour high resolution oesophageal manometry with better diagnostic accuracy over conventional line tracings. *Gut*. 2012;61(6):798–803. [PubMed: 21997554]
6. Carlson DA, Ravi K, Kahrilas PJ, Gyawali CP, Bredenoord AJ, Castell DO, et al. Diagnosis of Esophageal Motility Disorders: Esophageal Pressure Topography vs. Conventional Line Tracing. *Am J Gastroenterol*. 2015;110(7):967–77; quiz 78. [PubMed: 26032151]
7. Roman S, Huot L, Zerbib F, Bruley des Varannes S, Gourcerol G, Coffin B, et al. High-Resolution Manometry Improves the Diagnosis of Esophageal Motility Disorders in Patients With Dysphagia: A Randomized Multicenter Study. *Am J Gastroenterol*. 2016;111(3):372–80. [PubMed: 26832656]
8. Hernandez JC, Ratuapli SK, Burdick GE, Dibaise JK, Crowell MD. Interrater and intrarater agreement of the Chicago classification of achalasia subtypes using high-resolution esophageal manometry. *Am J Gastroenterol*. 2012;107(2):207–14. [PubMed: 22008895]
9. Fox MR, Pandolfino JE, Sweis R, Sauter M, Abreu YAAT, Anggiansah A, et al. Inter-observer agreement for diagnostic classification of esophageal motility disorders defined in high-resolution manometry. *Dis Esophagus*. 2015;28(8):711–9. [PubMed: 25185507]
10. Fox MR, Sweis R, Yadlapati R, Pandolfino J, Hani A, Defilippi C, et al. Chicago classification version 4.0(c) technical review: Update on standard high-resolution manometry protocol for the assessment of esophageal motility. *Neurogastroenterol Motil*. 2021;33(4):e14120. [PubMed: 33729668]
11. Bredenoord AJ, Babaei A, Carlson D, Omari T, Akiyama J, Yadlapati R, et al. Esophagogastric junction outflow obstruction. *Neurogastroenterol Motil*. 2021;33(9):e14193. [PubMed: 34120375]
12. Chen JW, Savarino E, Smout A, Xiao Y, de Bortoli N, Yadlapati R, et al. Chicago Classification Update (v4.0): Technical review on diagnostic criteria for hypercontractile esophagus. *Neurogastroenterol Motil*. 2021;33(6):e14115. [PubMed: 33729642]
13. Gyawali CP, Zerbib F, Bhatia S, Cisternas D, Coss-Adame E, Lazarescu A, et al. Chicago Classification update (V4.0): Technical review on diagnostic criteria for ineffective esophageal motility and absent contractility. *Neurogastroenterol Motil*. 2021;33(8):e14134. [PubMed: 33768698]
14. Khan A, Yadlapati R, Gonlachanvit S, Katzka DA, Park MI, Vaezi M, et al. Chicago Classification update (version 4.0): Technical review on diagnostic criteria for achalasia. *Neurogastroenterol Motil*. 2021;33(7):e14182. [PubMed: 34190376]
15. Roman S, Hebbard G, Jung KW, Katz P, Tutuiian R, Wong R, et al. Chicago Classification Update (v4.0): Technical review on diagnostic criteria for distal esophageal spasm. *Neurogastroenterol Motil*. 2021;33(5):e14119. [PubMed: 33666299]
16. Savarino E, di Pietro M, Bredenoord AJ, Carlson DA, Clarke JO, Khan A, et al. Use of the Functional Lumen Imaging Probe in Clinical Esophagology. *Am J Gastroenterol*. 2020;115(11):1786–96. [PubMed: 33156096]
17. Carlson DA, Kahrilas PJ, Lin Z, Hirano I, Gonsalves N, Listernick Z, et al. Evaluation of Esophageal Motility Utilizing the Functional Lumen Imaging Probe. *Am J Gastroenterol*. 2016;111(12):1726–35. [PubMed: 27725650]
18. Carlson DA, Gyawali CP, Khan A, Yadlapati R, Chen J, Chokshi RV, et al. Classifying Esophageal Motility by FLIP Panometry: A Study of 722 Subjects With Manometry. *Am J Gastroenterol*. 2021;116(12):2357–66. [PubMed: 34668487]
19. Carlson DA, Prescott JE, Baumann AJ, Schauer JM, Krause A, Donnan EN, et al. Validation of Clinically Relevant Thresholds of Esophagogastric Junction Obstruction Using FLIP Panometry. *Clin Gastroenterol Hepatol*. 2022;20(6):e1250–e62. [PubMed: 34216821]
20. Carlson DA, Baumann AJ, Donnan EN, Krause A, Kou W, Pandolfino JE. Evaluating esophageal motility beyond primary peristalsis: Assessing esophagogastric junction opening mechanics and secondary peristalsis in patients with normal manometry. *Neurogastroenterol Motil*. 2021:e14116. [PubMed: 33705590]
21. Carlson DA, Baumann AJ, Prescott JE, Donnan EN, Yadlapati R, Khan A, et al. Validation of secondary peristalsis classification using FLIP panometry in 741 subjects undergoing manometry. *Neurogastroenterol Motil*. 2021:e14192. [PubMed: 34120383]

22. Carlson DA, Prescott JE, Baumann AJ, Schauer JM, Krause A, Donnan EN, et al. Validation of Clinically Relevant Thresholds of Esophagogastric Junction Obstruction Using FLIP Panometry. *Clin Gastroenterol Hepatol*. 2021. Jun;20(6):e1250–e1262 [PubMed: 34216821]
23. Donnan EN, Pandolfino JE. EndoFLIP in the Esophagus: Assessing Sphincter Function, Wall Stiffness, and Motility to Guide Treatment. *Gastroenterol Clin North Am*. 2020;49(3):427–35. [PubMed: 32718562]
24. Nayar DS, Khandwala F, Achkar E, Shay SS, Richter JE, Falk GW, et al. Esophageal manometry: assessment of interpreter consistency. *Clin Gastroenterol Hepatol*. 2005;3(3):218–24. [PubMed: 15765440]
25. Sweis R, Anggiansah A, Wong T, Kaufman E, Obrecht S, Fox M. Normative values and inter-observer agreement for liquid and solid bolus swallows in upright and supine positions as assessed by esophageal high-resolution manometry. *Neurogastroenterol Motil*. 2011;23(6):509–e198. [PubMed: 21342362]
26. Yadlapati R, Keswani RN, Ciolino JD, Grande DP, Listerick ZI, Carlson DA, et al. A System to Assess the Competency for Interpretation of Esophageal Manometry Identifies Variation in Learning Curves. *Clin Gastroenterol Hepatol*. 2017;15(11):1708–14 e3. [PubMed: 27473627]
27. DeLay K, Pandolfino JE, Gyawali CP, Frye J, Kaizer A, Menard-Katcher P, et al. Esophageal Manometry Competency Program Improves Gastroenterology Fellow Performance in Motility Interpretation. *Am J Gastroenterol*. 2020;115(9):1453–9. [PubMed: 32453055]

WHAT IS KNOWN:

- The latest Chicago Classification version 4.0 (CCv4.0) for interpretation of high-resolution manometry (HRM) has revised conclusive and inconclusive definitions of motility disorders and incorporated supportive testing when diagnosis is inconclusive.
- Functional lumen imaging probe (FLIP) Panometry uses impedance planimetry technology to measure lumen opening mechanics and contractile response to distension and has been implicated as a primary or complementary diagnostic modality for evaluation of esophageal motility.
- Interrater agreement and accuracy of both HRM using CCv4.0 and FLIP Panometry have not been previously assessed.

WHAT IS NEW HERE:

- Interrater agreement for esophageal motility diagnosis using HRM per CCv4.0 and FLIP Panometry was moderate to good among esophageal motility specialists.
- High levels of accuracy in motility diagnosis were achieved by raters using both modalities when compared to the reference diagnosis based on the consensus agreement of experienced raters.
- Inconsistency in interpretation of HRM and FLIP nevertheless exists among esophageal specialists, highlighting the importance of complementary testing to improve confidence in diagnostic accuracy, especially prior to pursuing invasive therapies.

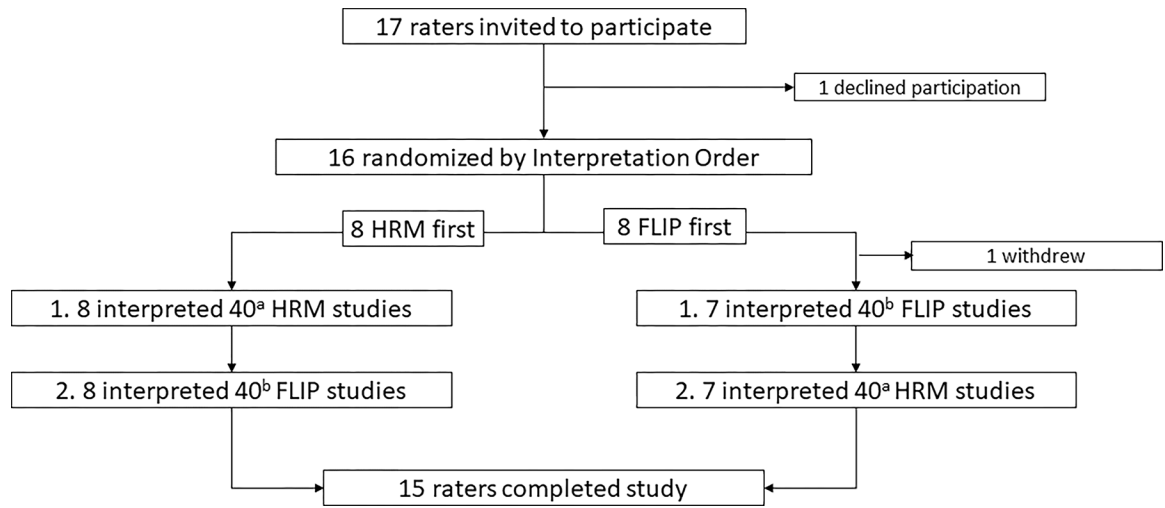


Figure 1. Study scheme.
Same HRM^a or same FLIP^b studies.

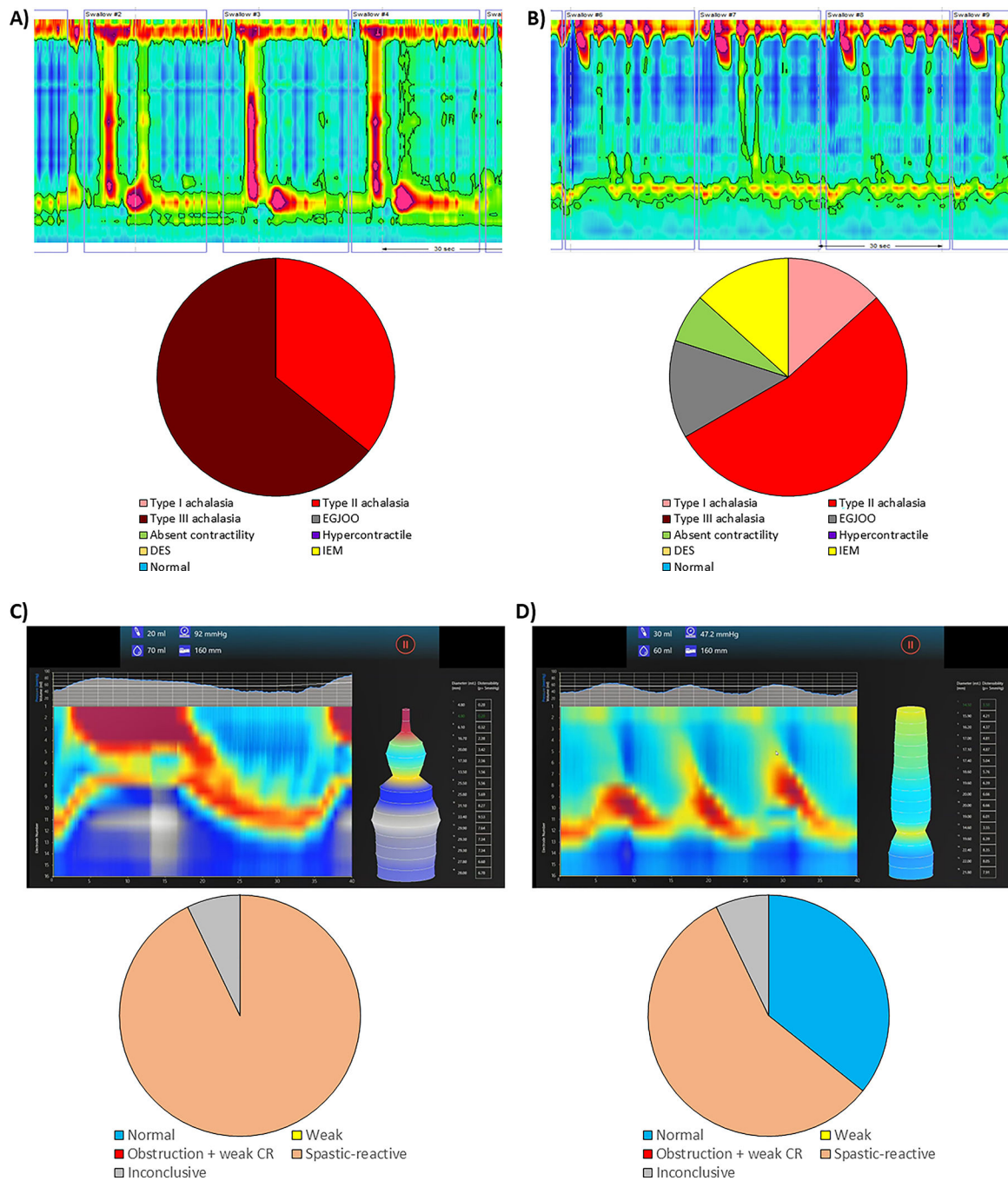


Figure 2. Examples of individual HRM and FLIP studies with rater diagnostic patterns. A) shows an HRM study that was interpreted as type III achalasia by majority of raters, while a minority interpreted it as type II achalasia (reference diagnosis: type III achalasia). B) was an HRM study with high rater discrepancy where about half the raters interpreted the study as type II achalasia, and the other half interpreted as type I, EGJ outflow obstruction, IEM, and absent contractility (reference diagnosis: EGJOO). In panel C), majority of raters interpreted the FLIP study as spastic-reactive, while the others interpreted it as obstruction with weak

contractile response and inconclusive (reference diagnosis: spastic reactive). Majority of raters interpreted the FLIP study in panel D) as spastic reactive, while a minority interpreted as normal (reference diagnosis: normal).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

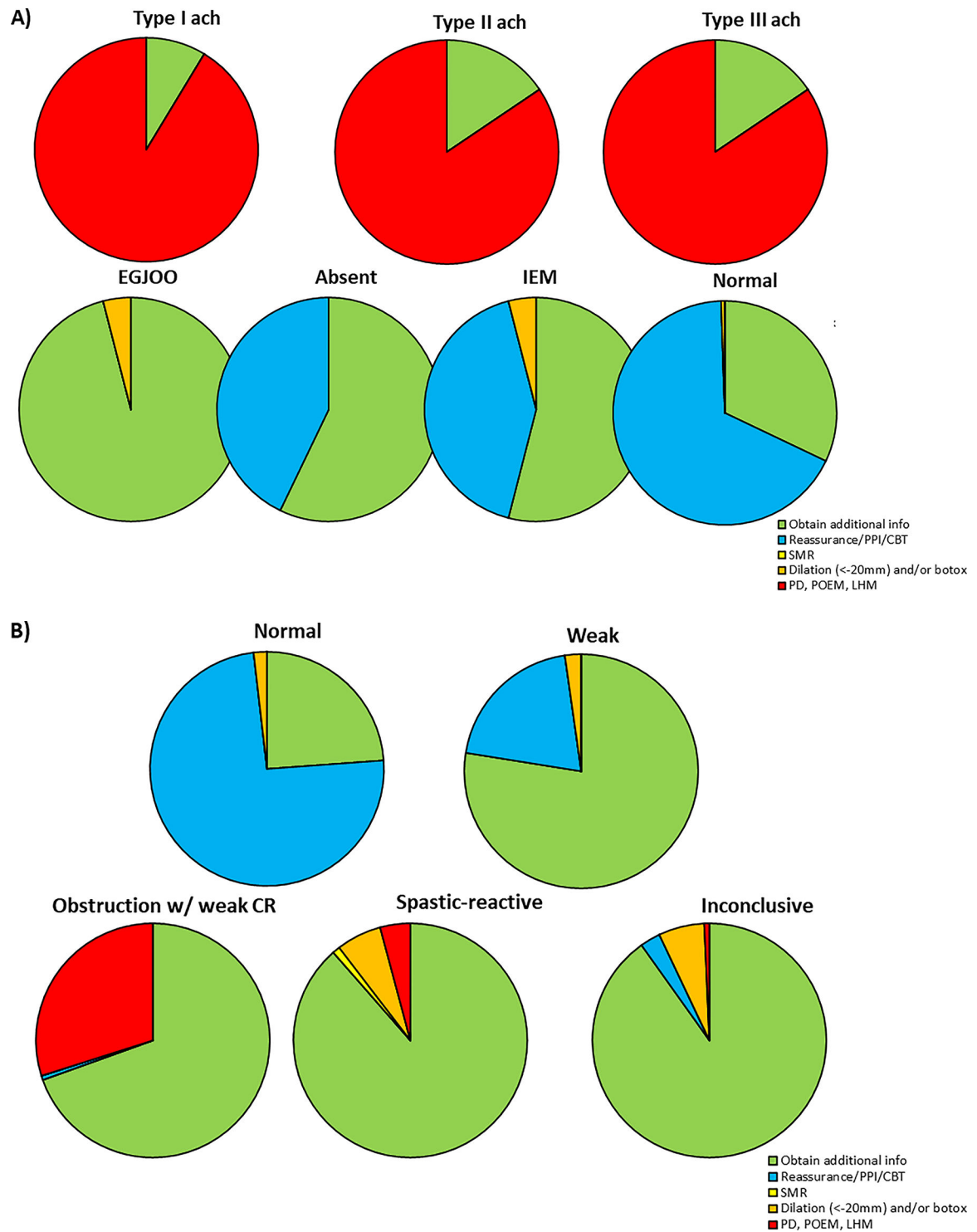


Figure 3. Management decision based on diagnosis. A) shows management decision by the raters based on HRM diagnosis and B) shows management decision by raters based on FLIP diagnosis. Ach, achalasia; EGJOO, esophagogastric outflow obstruction; IEM, ineffective esophageal motility; PPI, proton pump inhibitor; CBT, cognitive behavioral therapy; SMR,

systemic muscle relaxant; PD, pneumatic dilation; POEM, peroral endoscopic myotomy; LHM, laparoscopic Heller myotomy; CR, contractile response

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Rater characteristics.

| | Raters; n (%) |
|--------------------------|---------------|
| Years in practice | |
| 1–5 | 1 (7) |
| 6–10 | 5 (33) |
| 11–15 | 3 (20) |
| >15 | 6 (40) |
| HRM experience | |
| 25–100 | 0 |
| 100–499 | 0 |
| >500 | 15 (100) |
| FLIP experience | |
| 25–100 | 3 (20) |
| 100–499 | 7 (47) |
| >500 | 5 (33) |

Reported experience among the 15 raters that completed the study are reported. High-resolution manometry (HRM) or functional lumen imaging probe (FLIP) experience reflects numbers of previous studies interpreted.

Table 2.

Interrater agreement in HRM and FLIP interpretation.

| HRM | ICC (95% CI) |
|-------------------------|---------------------|
| Supine IRP | 0.96 (0.95, 0.98) |
| Upright IRP | 0.97 (0.96, 0.98) |
| | <i>kappa</i> |
| EGJ outflow | 0.71 |
| Peristalsis | 0.75 |
| CCv4.0 classification | 0.70 |
| FLIP | ICC (95% CI) |
| EGJ-DI @60mL | 0.91 (0.87, 0.94) |
| Max EGJ diameter | 0.92 (0.89, 0.95) |
| | <i>kappa</i> |
| EGJ classification | 0.68 |
| Contractile Response | 0.56 |
| Motility | 0.59 |
| Normal/abnormal | 0.66 |
| Achalasia/non-achalasia | 0.66 |

ICC – intraclass coefficient. EGJ-DI – esophagogastric junction distensibility index.

Table 3.

Accuracy of high-resolution manometry / Chicago Classification v4.0 diagnoses.

| | | Reference diagnoses | | | | | | | | |
|------------------------|----------------------|---------------------|-------------------|--------------------|----------|----------------------|------------------|-----|----------|-----------|
| | | Type I achalasia | Type II achalasia | Type III achalasia | EGJOO | Absent contractility | Hypercontractile | DES | IEM | Normal |
| <i>Total</i> | | 4 | 7 | 4 | 6 | 3 | 0 | 0 | 3 | 13 |
| Rater diagnoses | Type I achalasia | 45 (75%) | 0 | 0 | 2 | 4 | 0 | 0 | 0 | 0 |
| | Type II achalasia | 14 | 98 (93%) | 15 | 15 | 0 | 0 | 0 | 0 | 0 |
| | Type III achalasia | 0 | 1 | 41 (68%) | 8 | 0 | 0 | 0 | 0 | 0 |
| | EGJOO | 0 | 3 | 4 | 57 (63%) | 0 | 0 | 0 | 2 | 9 |
| | Absent contractility | 1 | 2 | 0 | 1 | 38 (84%) | 0 | 0 | 0 | 0 |
| | Hypercontractile | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 3 |
| | DES | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | IEM | 0 | 1 | 0 | 2 | 3 | 0 | 0 | 36 (80%) | 8 |
| | Normal | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 7 | 174 (89%) |

Numbers represent the number of studies with pooled from all rater diagnoses relative to the experienced rater consensus diagnoses. Shaded boxes indicate exact agreement, i.e. 'accurate diagnosis'. EGJOO – esophagogastric junction outflow obstruction. DES – distal esophageal spasm. IEM – ineffective esophageal motility.

Table 4.

Accuracy of FLIP Panometry diagnoses.

| | | Reference diagnosis | | | | |
|------------------------|-------------------------------|---------------------|----------|------------------------|------------------|--------------|
| | | Normal | Weak | Obstruction w/ weak CR | Spastic-reactive | Inconclusive |
| | <i>Total</i> | 9 | 7 | 11 | 7 | 6 |
| Rater diagnoses | Normal | 102 (76%) | 7 | 0 | 0 | 0 |
| | Weak | 0 | 79 (75%) | 1 | 1 | 8 |
| | Obstruction w/ weak CR | 0 | 1 | 143 (87%) | 8 | 12 |
| | Spastic-reactive | 10 | 0 | 2 | 78 (74%) | 6 |
| | Inconclusive | 23 | 18 | 19 | 18 | 64 (71%) |

Numbers represent the number of studies with pooled from all rater diagnoses relative to the internal reference diagnoses. Shaded boxes indicate exact agreement, i.e. 'accurate diagnosis'. CR – contractile response.