Original Article

Factors Determining the Inter-observer Variability and Diagnostic Accuracy of High-resolution Manometry for Esophageal Motility Disorders

Ji Hyun Kim,[1] Sung Eun Kim,[2] Yu Kyung Cho,[3]* Chul-Hyun Lim,[3] Moo In Park,[2] Jin Won Hwang,[1] Jae-Sik Jang,[1] and Minkyung Oh[4];
Motility Study Club of Korean Society of Neurogastroenterology and Motility

[1]*Department of Internal Medicine, Busan Paik Hospital, Inje University College of Medicine, Busan, Korea;* [2]*Department of Internal Medicine, Kosin University College of Medicine, Busan, Korea;* [3]*Department of Internal Medicine, College of Medicine, The Catholic University of Korea, Seoul, Korea; and* [4]*Clinical Trial Center, Inje University College of Medicine, Busan, Korea*

Background/Aims
Although high-resolution manometry (HRM) has the advantage of visual intuitiveness, its diagnostic validity remains under debate. The aim of this study was to evaluate the diagnostic accuracy of HRM for esophageal motility disorders.

Methods
Six staff members and 8 trainees were recruited for the study. In total, 40 patients enrolled in manometry studies at 3 institutes were selected. Captured images of 10 representative swallows and a single swallow in analyzing mode in both high-resolution pressure topography (HRPT) and conventional line tracing formats were provided with calculated metrics.

Results
Assessments of esophageal motility disorders showed fair agreement for HRPT and moderate agreement for conventional line tracing ($\kappa = 0.40$ and 0.58, respectively). With the HRPT format, the κ value was higher in category A (esophagogastric junction [EGJ] relaxation abnormality) than in categories B (major body peristalsis abnormalities with intact EGJ relaxation) and C (minor body peristalsis abnormalities or normal body peristalsis with intact EGJ relaxation). The overall exact diagnostic accuracy for the HRPT format was 58.8% and rater's position was an independent factor for exact diagnostic accuracy. The diagnostic accuracy for major disorders was 63.4% with the HRPT format. The frequency of major discrepancies was higher for category B disorders than for category A disorders (38.4% vs 15.4%; *P* < 0.001).

Conclusions
The interpreter's experience significantly affected the exact diagnostic accuracy of HRM for esophageal motility disorders. The diagnostic accuracy for major disorders was higher for achalasia than distal esophageal spasm and jackhammer esophagus.
(J Neurogastroenterol Motil 2018;24:58-69)

Key Words
Diagnosis; Esophageal motility disorders; Manometry; Observer variability

*Correspondence: Yu Kyung Cho, MD, PhD
Department of Internal Medicine, College of Medicine, The Catholic University of Korea, 222 Banpodae-ro, Seocho-gu, Seoul 06591, Korea
Tel: +82-2-590-2471, Fax: +82-2-590-2387, E-mail: ykcho@catholic.ac.kr
Ji Hyun Kim and Sung Eun Kim contributed to this paper equally.

ⓒ 2018 The Korean Society of Neurogastroenterology and Motility
J Neurogastroenterol Motil, Vol. 24 No. 1 January, 2018
www.jnmjournal.org

58

Introduction

High-resolution manometry (HRM) has the advantages of compact, intuitive esophageal pressure tomography generated from a closely spaced sensor of 1 cm from the pharynx to the stomach. This technique seems to provide detailed data on low esophageal sphincter (LES) relaxation and body peristalsis, facilitating the interpretation of manometric diagnosis of esophageal motility disorders. The Chicago classification scheme for esophageal motility disorders was also introduced and recently revised as version 3.0, which classifies motility disorders into 3 criteria: achalasia and esophagogastric junction (EGJ) outflow obstruction, major disorders of peristalsis and minor disorders of peristalsis.[1] HRM data can be more valuable than conventional manometry from a diagnostic perspective and for making decisions and predicting medical or interventional treatment, including up-to-date therapeutic options, such as peroral endoscopic myotomy (POEM).[2-5]

However, according to the Chicago classification version 3.0, a hierarchical approach considering multiple novel metrics, such as integrated relaxation pressure (IRP), contractile deceleration point, distal contractile integral (DCI), and distal latency (DL), is necessary for accurate diagnosis and is more complex than conventional manometry. In addition, the total number of differential diagnoses with HRM was as many as nine, which was also more than that with conventional manometry.[6] Combining these issues, despite the advantage of visual intuitiveness, the interpretation of HRM data seems to require an appropriate understanding of diagnostic metrics, and a certain level of training is important for accurate diagnosis. Although early comparisons between HRM and conventional manometry showed higher diagnostic yields with HRM diagnosis,[7,8] a controlled study assessing the degree of agreement and the clinical factors associated with diagnostic accuracy based on the updated classification system found only sparse data.

Our aims are to evaluate the accuracy of HRM diagnosis based on Chicago classification version 3.0 for esophageal motility disorders by comparing the high-resolution pressure topography (HRPT) and conventional line tracing (CLT) formats.

Materials and Methods

Manometry Data, Outcomes, and Sample Size

Fifty-one patients who underwent manometry studies performed between March 2012 and August 2015, accompanied by clinical information, including sex, age, chief complaint, duration of symptoms, esophagography, and prognosis, were enrolled from 3 institutions (the Catholic University of Korea Seoul St. Mary's Hospital, Inje University Busan Paik Hospital, and Kosin University Gospel Hospital). All patients suffered from esophageal symptoms, such as dysphagia, chest pain, and heartburn. All studies were conducted using a solid-state HRM system (Sierra Scientific, Los Angeles, California, USA) with 36 circumferential sensors at 1 cm intervals. Dedicated software (ManoView; Sierra Scientific) was used to visualize the plots and calculate the metrics. Captured images of 10 representative swallows and a single swallow in the analyzing mode in both HRPT format and the corresponding CLT format, consisting of 7 stacked tracings (1 in the upper esophageal sphincter, 4 in the esophageal body [3, 8, 13, and 18 cm from the LES], 1 in the LES, and 1 gastric) without using an electric-sleeve sensor, were provided with calculated metrics (HRPT format: median IRP, DCI, DL, IRP, and contractile front velocity [CFV] with each swallow; CLT format: basal LES pressure, residual LES pressure, distal esophageal amplitude, and onset velocity with each swallow) (Fig. 1).

The primary outcome was the assessment of the diagnostic accuracy of HRM for esophageal motility disorders by comparing the HRPT and CLT formats. The secondary outcome was to evaluate the inter-observer variability and disease- and interpreter-related factors associated with the accuracy of HRM diagnosis for esophageal motility disorders.
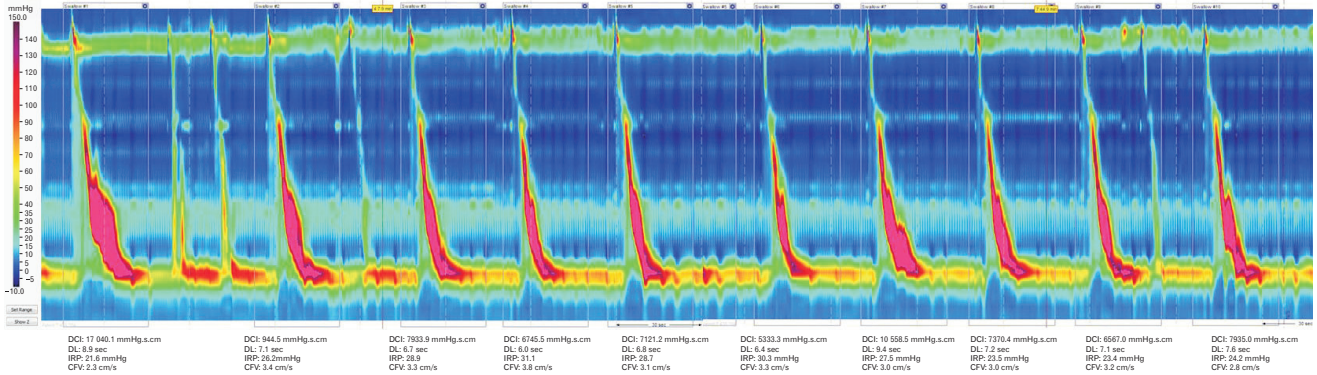
A total of 40 cases of manometry was an adequate number for evaluating the diagnostic accuracy of HRM, assuming an exact diagnostic accuracy of 75% using 14 raters, as well as an α of 0.05 and a β of 0.15.[9]

Subjects and Study Design

Six gastroenterology staff members who had experienced at least 300 cases of esophageal manometry interpretation and 8 trainees who were in the first or second year of a gastroenterology fellowship at 3 university hospitals (the Catholic University of Korea Seoul St. Mary's Hospital, Inje University Busan Paik Hospital, and Kosin University Gospel Hospital) were recruited for the study.

All the raters were given a 60-minute tutorial focusing on the interpretation of HRM and conventional esophageal manometry by the same instructor. HRM manometry interpretation was performed based on the Chicago classification version 3.0, consisting of 9 disease entities and 3 disease categories. Conventional manometry interpretation was also performed with a hierarchical scheme derived from that for HRM manometry interpretation, comprising

A
Median IRP: 26.6 mmHg

| DCI: 17 040.1 mmHg.s.cm | DCI: 944.5 mmHg.s.cm | DCI: 7933.9 mmHg.s.cm | DCI: 6745.5 mmHg.s.cm | DCI: 7121.2 mmHg.s.cm | DCI: 5333.3 mmHg.s.cm | DCI: 10 558.5 mmHg.s.cm | DCI: 7370.4 mmHg.s.cm | DCI: 6567.0 mmHg.s.cm | DCI: 7935.0 mmHg.s.cm |

C
Basal LESP: 45.3 mmHg
Mean residual LESP: 15.9 mmHg
Mean distal amplitude: 260.8 mmHg

B

IRP: 27.5 mmHg
DCI: 10 558.5 mmHg·sec·cm
DL: 8.9 s, CFV: 3.0 cm/sec

D

Residual LESP: 9.7 mmHg
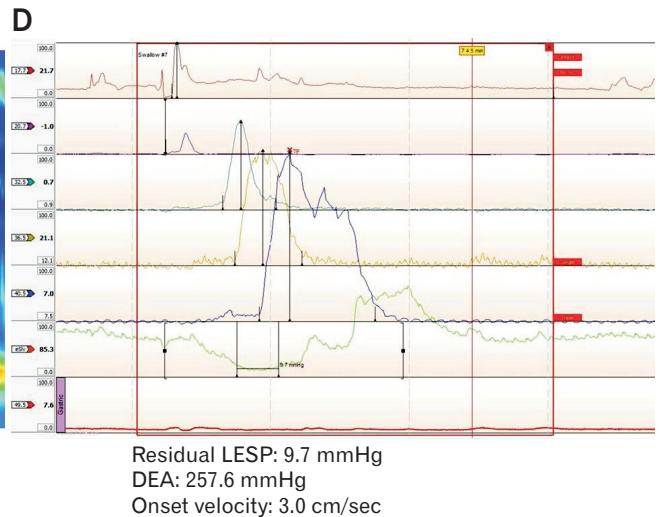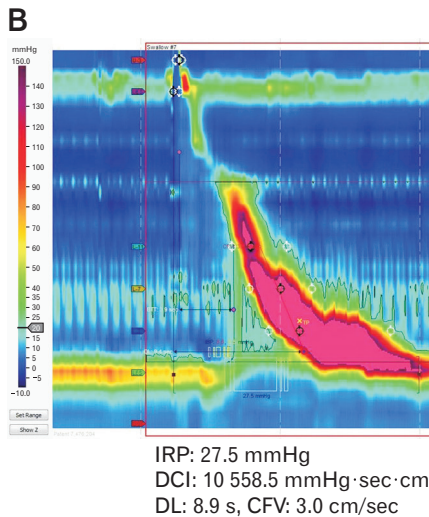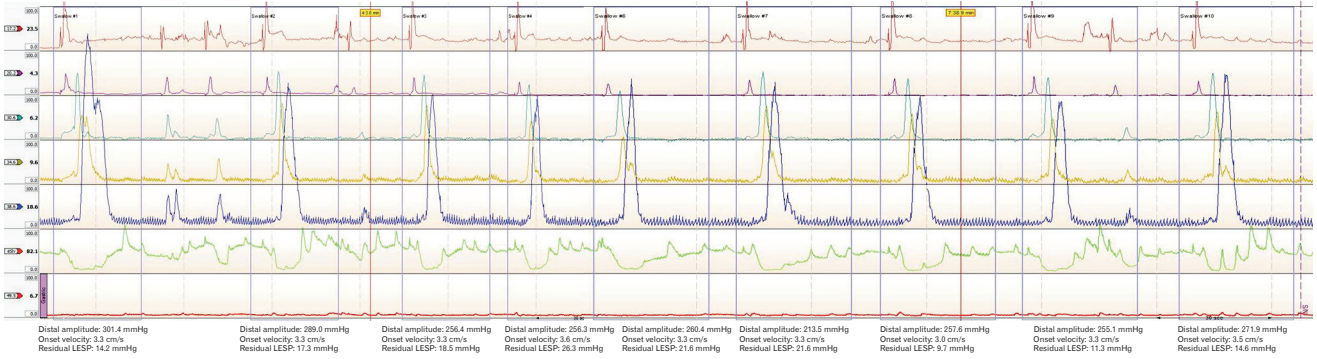DEA: 257.6 mmHg
Onset velocity: 3.0 cm/sec

Figure 1. Captured images of 10 swallows (A) and a single swallow image in analysis mode (B) with the high-resolution pressure topography (HRPT) format. Calculated metrics, such as median integrated relaxation pressure (IRP), distal contractile integral (DCI), distal latency (DL), and contractile front velocity (CFV), were also provided. Corresponding images of 10 swallows (C) and a single swallow image (D) in conventional line tracing (CLT) format. Calculated metrics, such as basal lower esophageal sphincter pressure (LESP), residual LESP, distal esophageal amplitude (DEA), onset velocity, mean residual LESP, and mean distal amplitude, were provided.

7 disease entities and 3 disease categories.

Within 1 week of the tutorial, all raters received randomized data from 40 cases of manometry in both HRPT and CLT formats and an answer sheet providing specific diagnoses and a baseline experience questionnaire assessing the number of previous manometry interpretations. Each rater was asked to analyze the set of 40 patients' manometry studies in both HRPT and CLT formats, separated in time by 2 weeks with a week of time constraint (Fig. 2).

None of the raters were involved in collecting the sample cases, and they were blinded to the clinical information of the patients, with the exception of the chief complains and the manometry data and to the responses of the other participants.

This study was approved by the ethics committee of the Catholic University of Korea Seoul St. Mary's Hospital (Approval No. KC17RISI0405), Inje University Busan Paik Hospital (Approval No. 17-0261), and Kosin University Gospel Hospital (Approval No. 2017-12-014). Written consent was obtained from all individuals before each procedure.

Reference Diagnosis, Types of Motility Abnormality Analysis of Diagnostic Accuracy

Levels of difficulty, reference standards and case selection were determined by 2 authors (J.H.K. and Y.K.J.) who did not participate in the study as raters. Reference diagnoses for each manometry format were determined by assessing other clinical information and the agreement of 2 authors according to the diagnostic scheme. The level of difficulty was classified into 3 grades (low, mid, and high) in all cases. A high level of difficulty was defined as a case that received a high grade by both authors.

Eleven of 51 cases were excluded considering the accordance between manometry and other clinical information (3 cases), similarity of the data, and the level of difficulty in interpretation (8 cases). One to 3 cases with a high level of difficulty were included in each HRPT diagnosis of Type I, II, and III achalasia (1, 2, and 2 cases, respectively); esophagogastric junction outflow obstruction (EGJOO) (3 cases); jackhammer esophagus (2 cases); and distal esophageal spasm (DES) (2 cases). Finally, 40 cases in both HRPT and CLT formats were included in the study.

Based on the types of motility abnormalities, all disease entities were classified into 1 of 3 categories: category A (EGJ/LES relaxation abnormality), category B (major body peristalsis abnormalities with intact EGJ/LES relaxation), and category C (minor body peristalsis abnormalities or normal body peristalsis with intact EGJ/LES relaxation). Category A included Type I, II, and III achalasia and EGJOO with an HRPT diagnosis and classic and vigorous achalasia and atypical disorders of LES relaxation with a CLT diagnosis. Category B included DES, jackhammer esophagus and absent peristalsis with an HRPT diagnosis, and diffuse esophageal spasm and nutcracker esophagus with a CLT diagnosis. Category C included ineffective esophageal motility (IEM), fragmented peristalsis and normal status with an HRPT diagnosis, and IEM and normal status with a CLT diagnosis.

Regarding diagnostic accuracy, 2 diagnostic accuracy indicators were analyzed: (1) exact accuracy and (2) diagnostic accuracy for major disorders.

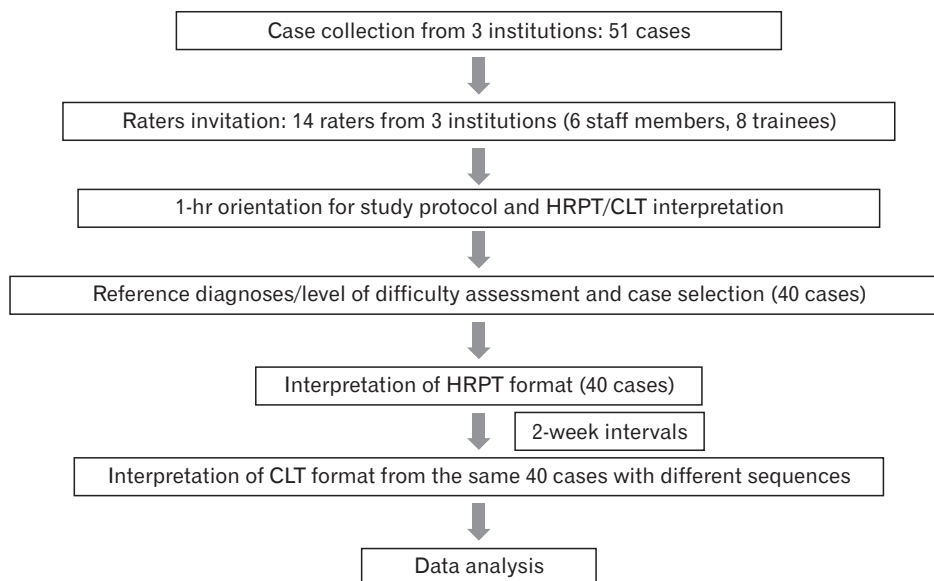To determine exact accuracy, the correct answer was defined

Case collection from 3 institutions: 51 cases

↓

Raters invitation: 14 raters from 3 institutions (6 staff members, 8 trainees)

↓

1-hr orientation for study protocol and HRPT/CLT interpretation

↓

Reference diagnoses/level of difficulty assessment and case selection (40 cases)

↓

Interpretation of HRPT format (40 cases)

↓ 2-week intervals

Interpretation of CLT format from the same 40 cases with different sequences

↓

Data analysis

Figure 2. Study scheme. All raters were asked to analyze a set of 40 patients' manometry studies in both high-resolution pressure topography (HRPT) and conventional line tracing (CLT) formats that were separated in time by 2 weeks with a week of time constraint.

as the absolute agreement between the reference diagnoses and the raters' answers for each disease entity; otherwise, the answers were considered incorrect.

Diagnostic accuracy for major disorders was evaluated with Type I, II, III achalasia, DES and jackhammer esophagus with HRPT diagnosis and classic achalasia, vigorous achalasia, diffuse esophageal spasm, and nutcracker esophagus with CLT diagnosis. The raters' responses were subclassified into 3 categories: major discrepancy, minor discrepancy, and no discrepancy. Major discrepancy was defined as disagreements between reference diagnoses and the raters' answers failing in (1) differentiating achalasia from other disorders (reference diagnosis achalasia and rater's answer EGJOO, category B or C) or (2) differentiating premature/simultaneous contraction and hypercontractility from failed/weak and normal contraction (reference diagnosis DES/diffuse esophageal spasm or jackhammer esophagus/nutcracker esophagus and rater's answer EGJOO/atypical disorders of LES relaxation, absent peristalsis or category C) (Fig. 3). Minor discrepancy was defined as disagreements between reference diagnoses and the raters' answers failing in (1) differentiating subtypes of achalasia or (2) differentiation
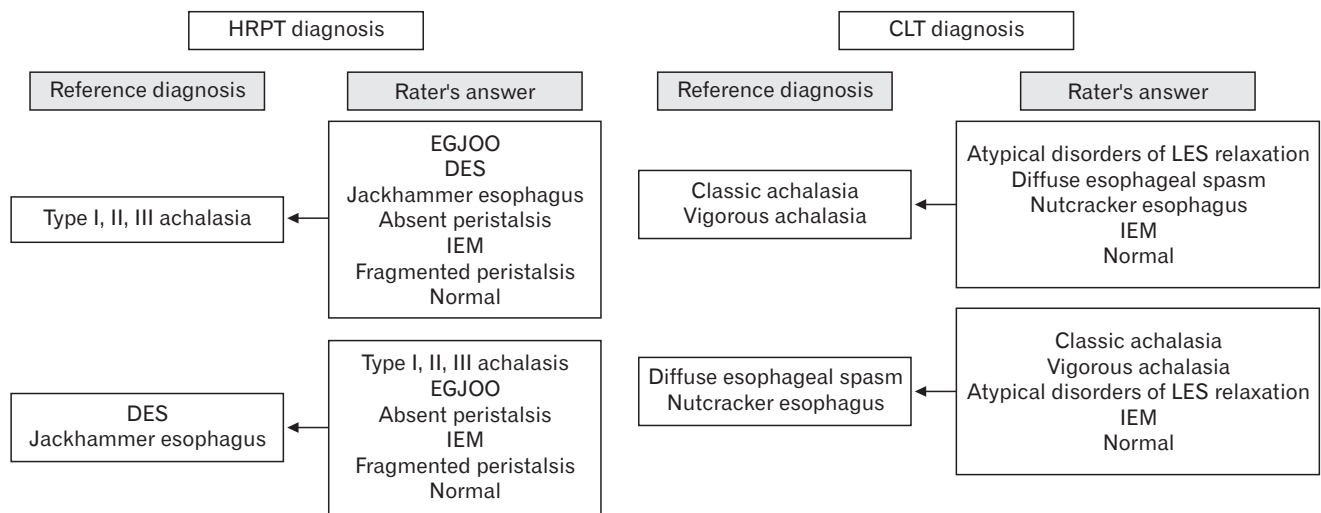

Figure 3. Scheme explaining major discrepancies of diagnostic accuracy for major disorders. HRPT, high-resolution pressure topography; CLT, conventional line tracing; EGJOO, esophagogastric junction outflow obstruction; DES, distal esophageal spasm; LES, low esophageal sphincter; IEM, ineffective esophageal motility.

Table 1. Reference Diagnosis and Categorical Classification

High-resolution pressure topography				Conventional line tracing			
Reference diagnosis	n (%)	Categorical classification	n (%)	Reference diagnosis	n (%)	Categorical classification	n (%)
Type I achalasia	4 (10.0)	Category A	24 (60.0)	Classic achalasia	10 (25.0)	Category A	19 (47.5)
Type II achalasia	8 (20.0)			Vigorous achalasia	4 (10.0)		
Type III achalasia	4 (10.0)			Atypical disorders of LES relaxation	5 (12.5)		
EGJOO	8 (20.0)						
Distal esophageal spasm	5 (12.5)	Category B	11 (27.5)	Diffuse esophageal spams	6 (15.0)	Category B	11 (27.5)
Jackhammer esophagus	5 (12.5)			Nutcracker esophagus	5 (12.5)		
Absent peristalsis	1 (2.5)						
Fragmented peristalsis	1 (2.5)	Category C	5 (12.5)	Ineffective esophageal motility	4 (10.0)	Category C	10 (10.0)
Ineffective esophageal motility	1 (2.5)						
Normal	3 (7.5)			Normal	6 (15.0)		
Total	40			Total	40		

EGJOO, esophagogastric junction outflow obstruction; LES, lower esophageal sphincter.

between premature/simultaneous contraction and hypercontractility (reference diagnosis DES/diffuse esophageal spasm and rater's answer jackhammer esophagus/nutcracker esophagus, and vice versa). No discrepancy was defined as absolute agreement between the reference diagnoses and the raters' answers for each disease entity. Answers with no discrepancies were considered correct, whereas answers with minor or major discrepancies were considered incorrect.

Statistical Methods

Univariate analysis for exact accuracy and diagnostic accuracy for major disorders was performed using Pearson's chi-square test. A multivariate analysis of the association and dependency of factors for exact diagnostic accuracy according to manometry formats was performed using generalized estimating equations. Variables for generalized estimating equations analysis included rater's position, type of motility abnormality, and interaction between rater's position and type of motility abnormality. Inter-observer agreement between raters was calculated by κ coefficient. The strength of agreement was graded as follows: 0-0.2 (poor), 0.21-0.40 (fair), 0.41-0.60 (moderate), 0.61-0.80 (substantial), and 0.81-1.00 (almost complete). A P-value < 0.05 was considered statistically significant.

Results

Reference Diagnosis of Manometry

Among 40 cases of HRPT diagnoses, 24 cases (60.0%) were achalasia, and 8 cases (12.5%) were EGJOO. Category B diagnoses comprised 11 cases (27.5%), with DES, jackhammer's esophagus and absent peristalsis numbering 5 cases (12.5%), 5 cases (12.5%), and 1 case (2.5%), respectively.

Regarding CLT diagnoses, category A and B disorders comprised 47.5% (19 cases) and 27.5% (11 cases) of all cases, respectively (Table 1).

Nine cases showed discordance between the HRPT and CLT reference diagnoses. One case of Type I achalasia according to the HRPT format was classified as IEM with the CLT format. One case of Type II achalasia with the HRPT format was classified as diffuse esophageal spasm with the CLT format. Two cases of EGJOO with the HRPT format were classified as normal with the CLT format. One case of EGJOO with the HRPT format was classified as nutcracker esophagus with the CLT format. Two cases of DES and JH with the HRPT format were classified as normal with the CLT format. One case classified as normal with

Table 2. Inter-observer Agreement Between All Raters According to Esophageal Motility Disorders and Categories

High-resolution pressure topography				Conventional line tracing			
Reference diagnosis	κ value (95% CI)	Categorical classification	κ value (95% CI)	Reference diagnosis	κ value (95% CI)	Categorical classification	κ value (95% CI)
Type I achalasia	0.34 (0.27-0.41)	Category A	0.67 (0.48-0.87)	Classic achalasia	0.62 (0.54-0.71)	Category A	0.68 (0.53-0.83)
Type II achalasia	0.53 (0.44-0.62)			Vigorous achalasia	0.42 (0.35-0.49)		
Type III achalasia	0.39 (0.32-0.46)			Atypical disorders of LES relaxation	0.56 (0.48-0.64)		
EGJOO	0.47 (0.39-0.55)						
Distal esophageal spasm	0.27 (0.20-0.34)	Category B	0.45 (0.34-0.57)	Diffuse esophageal spasm	0.46 (0.39-0.54)	Category B	0.66 (0.54-0.77)
Jackhammer esophagus	0.54 (0.46-0.61)			Nutcracker esophagus	0.78 (0.70-0.86)		
Absent peristalsis	0.23 (0.17-0.29)						
Fragmented peristalsis	0.24 (0.18-0.29)	Category C	0.41 (0.33-0.50)				
Ineffective esophageal motility	0.11 (0.03-0.40)			Ineffective esophageal motility disorders	0.66 (0.58-0.73)	Category C	0.64 (0.53-0.76)
Normal	0.53 (0.47-0.59)			Normal	0.60 (0.52-0.68)		
Overall	0.40 (0.39-0.42)		0.54 (0.49-0.59)	Overall	0.58 (0.56-0.59)		0.66 (0.63-0.69)

EGJOO, esophagogastric junction outflow obstruction.

the HRPT format was classified as nutcracker esophagus with the CLT format.

Inter-observer Agreement According to All Motility Disorders

Agreement for esophageal motility disorders among all raters was fair with the HRPT format and moderate with the CLT format (κ = 0.40 and 0.58, respectively). Among the diseases classified as major disorders, Type II achalasia, and jackhammer esophagus showed moderate agreement (κ = 0.53 and 0.54, respectively); in contrast, DES and absent peristalsis showed fair agreement (κ = 0.27 and 0.23, respectively) in the HRPT format. In the CLT format, nutcracker esophagus showed substantial agreement (κ = 0.78), whereas diffuse esophageal spasm showed moderate agreement (κ = 0.46) (Table 2).

When inter-observer agreement was assessed by the rater's position, agreement between staff members was greater than among trainees in both the HRPT and CLT formats. The differences in κ values were prominent in DES and EGJOO with the HRPT format and in vigorous achalasia with the CLT format (Tables 3 and 4).

Inter-observer Agreement According to Types of Motility Abnormality

The level of agreement among all raters according to types of motility abnormality was moderate in the HRPT format (κ = 0.54), and substantial in the CLT format (κ = 0.66) (Table 1).

A higher level of agreement was observed in staff members than in trainees in all HRPT disease categories (range of κ values: 0.43-0.76 and 0.40-0.64, respectively). The κ value was higher in category A than in categories B and C among both staff members

Table 3. Inter-observer Agreement Among Trainees and Staff Members According to Esophageal Motility Disorders and Categories With High-resolution Pressure Topography Format

Diagnosis	κ value (95% CI)		Categorical classification	κ value (95% CI)	
	Trainees	Staff members		Trainees	Staff members
Type I achalasia	0.30 (0.18-0.41)	0.42 (0.28-0.57)	Category A	0.64 (0.39-0.89)	0.76 (0.41-1.10)
Type II achalasia	0.42 (0.29-0.55)	0.67 (0.49-0.84)			
Type III achalasia	0.32 (0.21-0.43)	0.51 (0.36-0.66)			
EGJOO	0.31 (0.20-0.42)	0.67 (0.50-0.84)			
Distal esophageal spasm	0.18 (0.07-0.29)	0.47 (0.32-0.61)	Category B	0.42 (0.26-0.58)	0.55 (0.35-0.75)
Jackhammer esophagus	0.49 (0.38-0.60)	0.63 (0.47-0.79)			
Absent peristalsis	0.16 (0.06-0.30)	0.37 (0.21-0.53)			
Fragmented peristalsis	0.19 (0.09-0.30)	0.22 (0.05-0.39)	Category C	0.40 (0.27-0.54)	0.43 (0.28-0.58)
Ineffective esophageal motility	0.12 (0.02-0.23)	0.19 (0.00-0.40)			
Normal	0.43 (0.33-0.54)	0.61 (0.46-0.76)			
Overall	0.31 (0.29-0.33)	0.56 (0.52-0.59)		0.51 (0.44-0.57)	0.62 (0.51-0.72)

EGJOO, esophagogastric junction outflow obstruction.

Table 4. Inter-observer Agreement Among Trainees and Staff Members According to Esophageal Motility Disorders and Categories With Conventional Line Tracing Format

Diagnosis	κ value (95% CI)		Categorical classification	κ value (95% CI)	
	Trainees	Staff members		Trainees	Staff members
Classic achalasia	0.59 (0.47-0.71)	0.68 (0.51-0.85)	Category A	0.64 (0.44-0.85)	0.71 (0.45-0.97)
Vigorous achalasia	0.32 (0.20-0.43)	0.54 (0.39-0.69)			
Atypical disorders of LES relaxation	0.49 (0.37-0.61)	0.64 (0.49-0.79)			
Diffuse esophageal spasm	0.43 (0.31-0.54)	0.55 (0.39-0.70)	Category B	0.62 (0.46-0.79)	0.71 (0.51-0.91)
Nutcracker esophagus	0.70 (0.58-0.82)	0.88 (0.72-1.04)			
Ineffective esophageal motility	0.55 (0.44-0.67)	0.80 (0.65-0.95)	Category C	0.65 (0.48-0.82)	0.63 (0.44-0.83)
Normal	0.57 (0.45-0.70)	0.63 (0.47-0.79)			
Overall	0.51 (0.48-0.0.53)	0.67 (0.64-0.70)		0.64 (0.59-0.69)	0.69 (0.62-0.69)

LES, lower esophageal sphincter.

and trainees (Table 3).

In the analysis of the inter-observer agreement according to CLT disease categories, a higher κ value was observed among staff members than among trainees in categories A (0.71 and 0.64, respectively) and B (0.71 and 0.62, respectively). The κ value was similar between categories A and B among staff members and trainees (Table 4).

Exact Accuracy and Diagnostic Accuracy for Major Disorders

Exact diagnostic accuracy among all raters was significantly higher with the CLT format (70.5%) than with the HRPT format (58.8%) ($P < 0.001$). In further analysis, exact diagnostic accuracy among staff members was significantly greater than that among trainees with HRPT format but not with CLT format (Fig. 4).
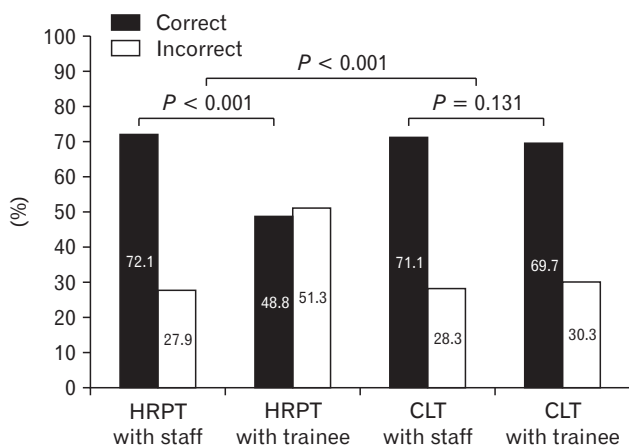
When data were analyzed according to categorical criteria, there was a tendency toward lower exact accuracy in category B disorders than in category A or C disorders with the HRPT format, but the trend was not statistically significant. There were no significant differences in the exact accuracy among category A, B, and C disorders with the CLT format (Fig. 5).

In the multivariate analysis, the rater's position was an independent factor significantly associated with exact diagnostic accuracy with the HRPT format. However, there was no significant association between the type of motility abnormality and exact diagnostic accuracy with either HRP or CLT formats. In addition, there was no significant interaction between the rater's position and the type of motility abnormality for exact diagnostic accuracy with HRPT or CLT formats (Table 5).

Regarding the diagnostic accuracy for major disorders, the overall diagnostic accuracy was 63.4% with HRPT format and 68.3% with CLT format. The diagnostic accuracy was significantly higher among staff members than among trainees for both
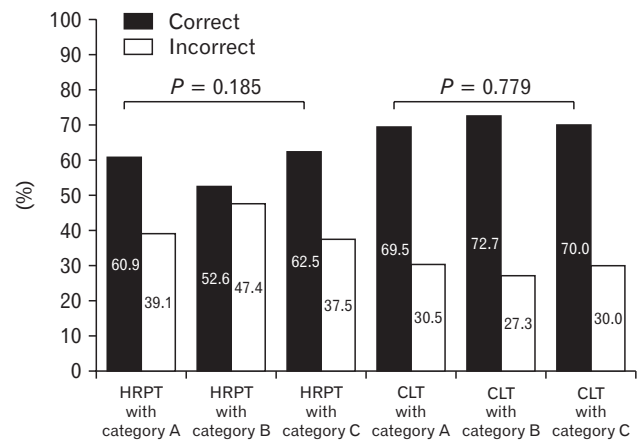
Figure 4. Comparison of exact diagnostic accuracy according to the rater's position and manometry format. Exact diagnostic accuracy among all raters was significantly higher with the conventional line tracing (CLT) format than with the high-resolution pressure topography (HRPT) format ($P < 0.001$). Exact diagnostic accuracy among staff members was significantly greater than that among trainees with the HRPT format ($P < 0.001$) but not with the CLT format ($P = 0.131$).

Figure 5. Diagnostic accuracy according to types of motility abnormality in the high-resolution pressure topography (HRPT) format and conventional line tracing (CLT) format. The frequency of correct answers was not different according to types of motility abnormality in either the CLT or HRPT format.

Table 5. Factors Associated With Exact Diagnostic Accuracy: Multivariate Analysis

Factors	High-resolution pressure topography		Conventional line tracing
	P-value	Multivariate OR (95% CI)	P-value
Rater's position	< 0.001	Trainee: reference Staff members: 2.16 (1.02-4.56)	0.936
Type of motility abnormality	0.630		0.581
Interaction between rater's position and type of motility abnormality	0.077		0.191

the HRPT (73.1% vs 57.6%, $P = 0.002$) and CLT (80.0% vs 61.3%, $P < 0.001$) formats. The diagnostic accuracy was higher for category A disorders (55.4%, 73.6% and 70.3% with Type I, II, and III achalasia, respectively) than for category B disorder (43.6% with DES and 67.9% with jackhammer esophagus) ($P = 0.013$) with the HRPT format. By contrast, the diagnostic accuracy was higher for category B disorders (60.0% with DES and 91.1% with nutcracker esophagus) than for category A disorders (64.4% with classic achalasia and 62.5% with vigorous achalasia) with the CLT format ($P = 0.028$). The overall frequency of major discrepancies was 24.2% with the HPRT format and 22.9% with the CLT format. Detailed analysis showed the highest frequency in DES (52.6%), followed by Type I achalasia (26.2%), jackhammer esophagus (24.7%), and Type II achalasia (8.0%), in order (Fig. 6). Major discrepancy was more frequently found in DES and jack-hammer esophagus than in achalasia (38.4% vs 15.4%; $P < 0.001$) and among trainees than among staff members (29.2% vs 16.0%;

Figure 6. Diagnostic accuracy for major disorders according to types of motility abnormality in the high-resolution pressure topography (HRPT) format and conventional line tracing (CLT) format. Major discrepancies were more frequently found in distal esophageal spasm and jackhammer esophagus than in Type I, II and III achalasia ($P < 0.001$). There was no significant difference in frequency of major discrepancy between classic/vigorous achalasia and diffuse esophageal spasm/nutcracker esophagus ($P = 0.959$).

$P = 0.002$). The frequency of major discrepancies was the highest in diffuse esophageal spasm (36.8%) with the CLT format (Fig. 6). The frequency of major discrepancies was also higher among trainees than among staff members with the CLT format (27.8% vs 14.7%; $P = 0.002$). However, there was no significant difference in the frequency of major discrepancies between classic/vigorous acha-lasia and diffuse esophageal spasm/nutcracker esophagus (22.8% vs 23.0%; $P = 0.959$).

Discussion

Currently, the κ value of overall HRPT diagnosis was lower than that of CLT diagnosis (0.54 and 0.66, respectively), and the differences were more remarkable with category B and C disorders than with category A disorders. Interestingly, the level of agreement was lower in Type I, II, and III achalasia (κ value were 0.4, 0.53, and 0.39, respectively) than in cases reported in existing studies.[9-11] κ values slightly increased among staff members; however, the level of agreement remained moderate with Type I and Type III achala-sia (κ values were 0.42 and 0.51, respectively). The unique method of the present study might explain those results. First, a large proportion of the raters (8/14) were trainees who had experienced fewer than 50 cases of manometry interpretation. Second, cases with a high level of difficulty comprised 34.2% of the total cases of cat-egory A and B disorders.

An existing validation study for the diagnosis of primary achalasia using conventional manometry showed substantial inter-observer agreement, with a κ value of 0.68, and the degree of inter-observer agreement was higher among experienced interpreters than among low-experience interpreters. Conversely, the κ value was as low as 0.27 for the diagnosis of motility disorders with body peristalsis abnormalities.[10] In a study using HRM, the κ value for the diagnosis of achalasia ranged between 0.48 and 0.60; in con-trast, the κ value for the diagnosis of body peristalsis disorders, such as DES and hypertensive dysmotility, was as low as 0.21-0.29.[12] Our study also showed a lower level of inter-observer agreement for the diagnosis of disorders included in categories B and C, in which the disorders are combined with abnormal velocity and intensity of body peristalsis without EGJ/LES relaxation abnormalities, than that for the diagnosis of disorders in category A. The difference was observed among staff as well as among trainees and was more prominent with the HRPT format than with the CLT format. Along with previous studies, our results suggest that the diagnosis of body peristalsis abnormalities made based on comprehensive consideration of metrics, such as DCI, DL, and CFV, was more
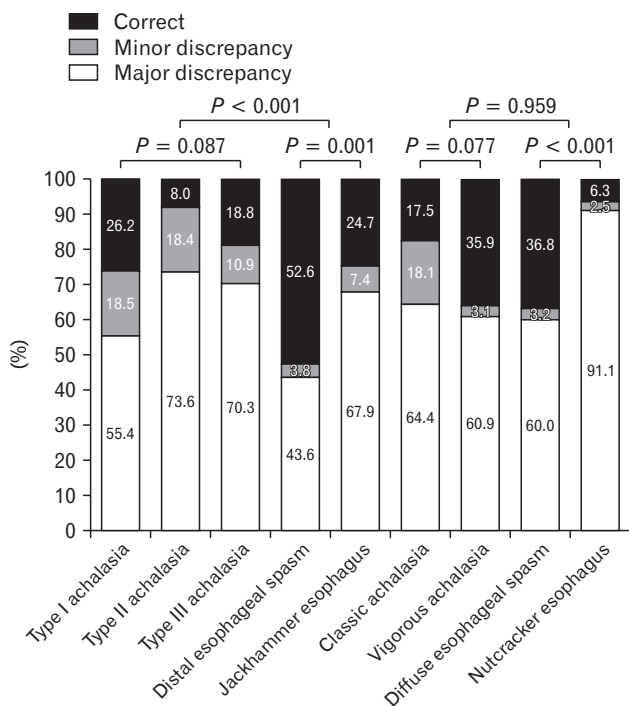
complicated than that of EGJ relaxation abnormality based on the single IRP metric.

The present study showed the lowest level of inter-observer agreement ($\kappa = 0.27$ among all raters) and a prominent difference in κ value between staff members and trainees (0.47 and 0.18, respectively) for the diagnosis of DES with the HRPT format. These results could be explained by several assumptions about the difficulty in diagnosing simultaneous contractions. First, simultaneous contraction was determined by measuring the DL, according to the Chicago classification version 3.0, in which a precise understanding of the esophageal contraction onset and contractile deceleration point is essential.[1,13] Second, although CFV was excluded from the key metrics for the diagnosis of motility disorders in the updated classification scheme,[1] a certain proportion of cases with clinical features suggestive of DES, such as typical symptoms and simultaneous contraction on barium esophagography, showed rapid contraction (CFV > 9 cm/sec) with normal DL ($\geq$ 4.5 seconds).[14-16] Finally, in cases in which an automatically calculated DL by commercial software was not available, especially cases combined with bolus transport disturbances, the interpreters were obliged to calculate the DL by themselves. All these factors together may have contributed to a low level of inter-observer agreement for the diagnosis of DES with HRPT format images, especially among less experienced interpreters, in the present study.

With regard to diagnostic accuracy, the overall exact diagnostic accuracy with HRPT diagnosis was 58.6%, which was significantly lower than that with CLT diagnosis. The difference in diagnostic accuracy between the HRPT and CLT formats was remarkable among low-experience interpreters (trainees). We also suggested several explanations for the results that were not in accordance with those of previous studies reporting the advantages of HRM over conventional manometry in diagnostic accuracy.[9,17] First, according to the Chicago classification version 3.0, the number of disease entities with the HRPT format was greater than that in the CLT format (9 vs 6) in the present study. Second, providing the manometry data as captured images from 10 significant swallows with calculated essential metrics, including IRP, DCI, DL, and CFV for the HRPT format and baseline and residual LES pressures, onset velocity and distal contractile amplitude for the CLT format may have facilitated the interpretation of the manometry data, especially in the CLT format. Finally, the modified hierarchical scheme used in the present study, which was based on the HRM used in the Chicago classification version 3.0, may have reduced the frequency of overlapping diagnoses that would have been regarded as wrong answers in previous studies. Taken together, all these factors may

have contributed to the unexpected results of lower exact diagnostic accuracy with the HRPT format than with the CLT format in the present study. The multivariate analysis also demonstrated that the rater's position was an independent factor associated with exact diagnostic accuracy with the HRPT format but not with the CLT format.

The Chicago classification version 3.0 defined major disorders of peristalsis, other than achalasia or EGJOO, as motility patterns that are not encountered in normal subjects. This finding was based on spastic esophageal disorders, including achalasia (especially Type III achalasia), jackhammer esophagus and DES, sharing common pathophysiologic characteristics, such as loss of inhibitory ganglionic neuron function and excess cholinergic drive in the distal esophagus, resulting in simultaneous contraction or hypercontraction of the distal esophagus and incomplete deglutitive EGJ relaxation.[11,15,18-21] Although the clinical significance remains under evaluation, this classification provided a theoretical basis for the importance of early diagnosis and an active management strategy, such as endoscopic or surgical treatment for major motility disorders such as achalasia, DES and jackhammer esophagus.[2-4,22,23] Nevertheless, few existing studies are available that have analyzed the reliability and diagnostic accuracy for distinguishing major esophageal motility disorders from minor motility disorders or a normal state or differential diagnosis among major disorders.[9] Our study is meaningful because we analyzed the exact accuracy and diagnostic accuracy for distinguishing major esophageal motility disorders and disease- or interpreter-associated factors influencing the diagnostic accuracy of the HRM.

A recent study by Carlson et al[9] reported that the odds ratio of an incorrect diagnosis of a major motility disorder among all raters was 3.4 times greater with the CLT format than with the HRPT format and that the odds ratio of an incorrect diagnosis was similar according to the interpreter's experience. Currently, in contrast to these results, the overall frequency of major discrepancies was not different between the HRPT and CLT formats. Instead, major discrepancies were more frequently found with low-experience interpreters and in diseases with body peristalsis abnormalities without EGJ/LES relaxation abnormality, especially with the HRPT format. In other words, the diagnostic accuracy of HRM for distinguishing major disorders was accentuated by experienced interpreters and in disorders with EGJ relaxation abnormality, such as Type I, II, and III achalasia. These results were concordant with the inter-observer agreement results with the HRPT format in the current study, showing higher κ values with Type A category diseases than with Type B and C category diseases and a higher κ value among staff members than among trainees.

There were several methodological controversies of this study.

Unlike previous studies, we provided HRPT and CLT manometry data as captured images with automatically calculated metrics for raters instead of analyzing software and raw data. In doing so, we attempted to eliminate bias caused by unfamiliarity with the handling of the commercial analyzing software. Considering that the adequacy of esophageal manometry is greatly affected by the performer's technical experience and patient compliance, this method of providing manometry data in the present study may be arguable due to concern over whether the selected swallowing manometry images represented all manometry data. To overcome this contradiction, the selection of 10 individual swallows, as well as a single typical swallow, was performed by a clinician who examined and managed the patient and who interpreted all manometry data initially; therefore, the manometry data were collected in the most pertinent manner.

We applied hierarchical classified CLT diagnosis in 6 disease entities and 3 categories, which were modified from the traditional classification system to avoid overlapping diagnoses.[6,9] The definitions of major disorders with CLT diagnosis, which were composed of 5 disease entities (classic achalasia, vigorous achalasia, atypical disorders of LES relaxation, nutcracker esophagus, and diffuse esophageal spasm), also came from the Chicago classification version 3.0. Although the classification was arbitrary and has not been validated by existing research, it was essential for the rational comparison of diagnostic accuracy between the HRPT and CLT formats.

There was concern that the uneven sample sizes among the different disease entities and categories may have influenced the degree of inter-observer agreement assessed by κ statistics. This inevitable bias was introduced by handling by the authors during the study process. There was a need for various cases with different levels of difficulty, especially weighted sample sizes with category A and B disorders. Because the diagnostic accuracy of manometry could be dependent on the difficulty of the case, there is also the possibility of selection bias induced by arbitrary case selection by the authors of the present study. Alterations in diagnosis also occurred during the process of reference diagnosis after sample collection for a significant number of cases.

To analyze diagnostic accuracy, we performed reference diagnoses of all 40 cases for each manometry format by assessing the agreement of 2 experts considering clinical information and manometry findings. Concern about the certainty of the reference diagnoses in the present study could also have been one of the main methodological limitations of the study. We attempted to dispel this controversy by excluding cases of disagreement in reference diagnoses between the 2 experts and cases showing a lack of relevance between the manometry findings and clinical aspects, such as symptoms and radiologic findings.

In conclusion, we demonstrated that the exact diagnostic accuracy of HRM for esophageal motility disorders was significantly affected by the interpreter's experience and that diagnostic accuracy for major disorders was higher in achalasia than DES and jackhammer esophagus. In addition, there was a higher level of inter-observer agreement and diagnostic accuracy for major disorders involving motility disorders combined with EGJ/LES relaxation abnormality than for those with body peristalsis abnormality and intact EGJ/LES relaxation with the HRPT format.

Conflict of interest: None.

Author contributions: Ji Hyun Kim and Sung Eun Kim collected the cases and analyzed and interpreted the data and wrote a manuscript; Yu Kyung Cho designed, organized, and mediated the present study design, and collected cases and supervised the manuscript; Chul-Hyun Lim, Moo In Park, and Jin Won Hwang participated as a rater; Minkyung Oh calculated sample size and analyzed the data; and Jae-Sik Jang analyzed and interpreted the data.

References

1. Kahrilas PJ, Bredenoord AJ, Fox M, et al. The Chicago Classification of esophageal motility disorders, v3.0. Neurogastroenterol Motil 2015;27:160-174.

2. Pandolfino JE, Kwiatek MA, Nealis T, Bulsiewicz W, Post J, Kahrilas PJ. Achalasia: a new clinically relevant classification by high-resolution manometry. Gastroenterology 2008;135:1526-1533.

3. Vanuytsel T, Bisschops R, Farré R, et al. Botulinum toxin reduces dysphagia in patients with nonachalasia primary esophageal motility disorders. Clin Gastroenterol Hepatol 2013;11:1115-1121, e2.

4. Hoppo T, Thakkar SJ, Schumacher LY, et al. A utility of peroral endoscopic myotomy (POEM) across the spectrum of esophageal motility disorders. Surg Endosc 2016;30:233-244.

5. Aziz Q, Fass R, Gyawali CP, Miwa H, Pandolfino JE, Zerbib F. Functional esophageal disorders. Gastroenterology 2016;150:1368-1379.

6. Spechler SJ, Castell DO. Classification of oesophageal motility abnormalities. Gut 2001;49:145-151.

7. Clouse RE, Staiano A, Alrakawi A, Haroian L. Application of topographical methods to clinical esophageal manometry. Am J Gastroenterol

2000;95:2720-2730.

8. Fox M, Hebbard G, Janiak P, et al. High-resolution manometry predicts the success of oesophageal bolus transport and identifies clinically important abnormalities not detected by conventional manometry. Neurogastroenterol Motil 2004;16:533-542.

9. Carlson DA, Ravi K, Kahrilas PJ, et al. Diagnosis of esophageal motility disorders: esophageal pressure topography vs. conventional line tracing. Am J Gastroenterol 2015;110:967-977.

10. Nayar DS, Khandwala F, Achkar E, et al. Esophageal manometry: assessment of interpreter consistency. Clin Gastroenterol Hepatol 2005;3:218-224.

11. 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;107:207-214.

12. Fox MR, Pandolfino JE, Sweis R, et al. Inter-observer agreement for diagnostic classification of esophageal motility disorders defined in high-resolution manometry. Dis Esophagus 2015;28:711-719.

13. Pandolfino JE, Roman S, Carlson D, et al. Distal esophageal spasm in high-resolution esophageal pressure topography: defining clinical phenotypes. Gastroenterology 2011;141:469-475.

14. Bredenoord AJ, Fox M, Kahrilas PJ, Pandolfino JE, Schwizer W, Smout AJ. Chicago classification criteria of esophageal motility disorders defined in high resolution esophageal pressure topography. Neurogastroenterol Motil 2012;24(suppl 1):57-65.

15. Roman S, Kahrilas PJ. Distal esophageal spasm. Curr Opin Gastroenterol 2015;31:328-333.

16. De Schepper HU, Ponds FA, Oors JM, Smout AJ, Bredenoord AJ. Distal esophageal spasm and the Chicago classification: is timing everything? Neurogastroenterol Motil 2016;28:260-265.

17. Soudagar AS, Sayuk GS, Gyawali CP. Learners favour high resolution oesophageal manometry with better diagnostic accuracy over conventional line tracings. Gut 2012;61:798-803.

18. Kahrilas PJ, Boeckxstaens G. The spectrum of achalasia: lessons from studies of pathophysiology and high-resolution manometry. Gastroenterology 2013;145:954-965.

19. Sodikoff JB, Lo AA, Shetuni BB, Kahrilas PJ, Yang GY, Pandolfino JE. Histopathologic patterns among achalasia subtypes. Neurogastroenterol Motil 2016;28:139-145.

20. Jung HY, Puckett JL, Bhalla V, et al. Asynchrony between the circular and the longitudinal muscle contraction in patients with nutcracker esophagus. Gastroenterology 2005;128:1179-1186.

21. Roman S, Kahrilas PJ. Management of spastic disorders of the esophagus. Gastroenterol Clin North Am 2013;42:27-43.

22. Pratap N, Kalapala R, Darisetty S, et al. Achalasia cardia subtyping by high-resolution manometry predicts the therapeutic outcome of pneumatic balloon dilatation. J Neurogastroenterol Motil 2011;17:48-53.

23. Vaezi MF, Pandolfino JE, Vela MF. ACG clinical guideline: diagnosis and management of achalasia. Am J Gastroenterol 2013;108:1238-1249.