Published in final edited form as:

Neurogastroenterol Motil. 2017 April ; 29(4): . doi:10.1111/nmo.12971.

Benchmarks for the Interpretation of Esophageal High Resolution Manometry

Rena Yadlapati¹, Rajesh N. Keswani¹, Kerry B. Dunbar², Andrew J. Gawron³, C. Prakash Gyawali⁴, Peter J. Kahrilas¹, Philip O. Katz⁵, David Katzka⁶, Stuart J. Spechler², Roger Tatum⁷, and John E. Pandolfino¹

¹Division of Gastroenterology and Hepatology, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

²University of Texas Southwestern Medical Center and the Dallas VA Medical Center, Department of Internal Medicine, Division of Gastroenterology and Hepatology, Dallas, TX, USA

³Division of Gastroenterology, University of Utah, Salt Lake City, IL, USA

⁴Division of Gastroenterology, Washington University School of Medicine, Saint Louis, Missouri, USA

⁵Thomas Jefferson University, Philadelphia, Pennsylvania; Division of Gastroenterology, Albert Einstein Medical Center, Philadelphia, PA, USA

⁶Division of Gastroenterology, Mayo Clinic, Rochester, MN, USA

⁷Department of Surgery, University of Washington, Seattle, WA, USA

Abstract

Background—Competent interpretation of esophageal high-resolution manometry (HRM) is integral to a quality study. Currently, methods to assess physician competency for the interpretation of esophageal HRM do not exist. The aim of this study was to use formal techniques to 1) develop an HRM interpretation exam, and 2) establish minimum competence benchmarks for HRM interpretation skills at the trainee, physician interpreter and master level.

Methods—A total of 29 physicians from 8 academic centers participated in the study: 9 content experts separated into 2 study groups - expert test-takers (n=7) and judges (n=2), and 20 HRM inexperienced trainees (“trainee test-taker”; n=20). We designed the HRM interpretation exam based on expert consensus. Expert and trainee test-takers (n=27) completed the exam. According

Corresponding Author: Rena Yadlapati, MD, Division of Gastroenterology & Hepatology, Northwestern University, 676 North St. Clair St. Suite 1400 Chicago, IL 60611, Office: (312) 695-4065; Fax: (312) 695-3999, rena.yadlapati@northwestern.edu.

Disclosures/Conflicts of Interests: RNK: Consults for Boston Scientific and Cook Endoscopy; JEP: Consults for Covidien, Sandhill Scientific, and Given. The authors have no other reported disclosures, competing interests, or conflicts of interest.

Writing Assistance:
None

Author Contributions:

Study concept & design: RY, RNK, JEP; Acquisition of data: RY, RNK, KBD, AJG, CPG, PJK, POK, DK, SS, RT, JEP; Analysis and interpretation of data: RY, JEP; Drafting of manuscript: RY, RNK, JEP; Critical revision of manuscript for important intellectual content: RY, RNK, KBD, AJG, CPG, PJK, POK, DK, SS, RT, JEP; Approval of final manuscript: RY, RNK, KBD, AJG, CPG, PJK, POK, DK, SS, RT, JEP

to the modified Angoff method, the judges reviewed the test-taker performance and established minimum competency cut scores for HRM interpretation skills.

Results—The HRM interpretation exam consists of 22 HRM cases with 8 HRM interpretation skills per case: identification of pressure inversion point, hiatal hernia > 3cm, integrated relaxation pressure, distal contractile integral, distal latency, peristaltic integrity, pressurization pattern, and diagnosis. Based on the modified Angoff method, minimum cut scores for HRM interpretation skills at the trainee, physician interpreter, and master level ranged from 65-80%, 85-90% (with the exception of peristaltic integrity), and 90-95%, respectively.

Conclusions—Using a formal standard setting technique, we established minimum cut scores for 8 HRM interpretation skills across interpreter levels. This examination and associated cut scores can be applied in clinical practice to judge competency.

Keywords

Angoff Method; Competency; Standard Setting; Esophageal manometry

Introduction

Most US gastroenterology fellowship programs do not provide formal training in gastrointestinal motility diagnostics such as esophageal high-resolution manometry (HRM).¹ Consequently, despite the high prevalence and health care burden of esophageal motility disorders, gastroenterologists are inadequately trained to interpret esophageal HRM.^{1,2} Experts recognize this practice gap and its negative impact upon patient outcomes. As such, interpreter competency for HRM is considered a high priority quality measure in the performance and interpretation of HRM.³ Unfortunately, a reliable measure of competency is lacking.

Currently, the US Gastroenterology (GI) Core Curriculum recommends completion of 50 esophageal manometry studies to achieve proficiency. However, this recommendation is based only on expert opinion.^{1,4} To assess proficiency for HRM according to these recommendations, we recently conducted a study in which 20 HRM naive GI trainees interpreted 50 standardized HRM cases. All trainees viewed video-based introductory educational materials and subsequently received feedback after every 2 cases. Despite this, we found that learning curves for acquiring proficient HRM interpretation skills varied tremendously.⁵ Trainees who achieved competency did so at differing case volumes. Moreover, the majority of trainees failed to demonstrate competency for HRM even after interpreting 50 HRM cases.⁵ These results suggest that using a minimal case volume to assess esophageal HRM competency is inaccurate and, furthermore, most trainees need to interpret more than 50 cases to achieve competency. Thus, a standardized assessment of competency for HRM interpretation is necessary. The aims of this study were to: 1) develop an HRM interpretation exam, and 2) establish minimum competence benchmarks for specific interpretation skills at the trainee, physician interpreter, and master level.

Methods

Study Design

This study was approved by the Northwestern University Institutional Review Board. To establish minimum competence cut scores, we adapted the modified Angoff method of standard setting for performance examinations in health education.^{6,7} Nine content experts and 20 inexperienced trainees from 8 academic medical centers participated. Content experts were all board certified physicians and included 8 gastroenterologists and 1 gastrointestinal surgeon. Importantly, all experts interpret esophageal HRM and work with resident and/or fellow trainees. Inexperienced trainees were all GI fellows or gastrointestinal surgery residents with minimal (0-5 HRM cases) HRM experience.

HRM Interpretation Exam Development

Based on the group's prior work in defining quality measures for esophageal manometry, the experts identified 8 requisite skills for HRM interpretation.³ The study team proposed question and answer formats for each skill to the experts. For instance, in assessing Integrated Relaxation Pressure (IRP) interpretation skills, the question could be to either determine the numerical value of the IRP with an acceptable range of ± 3.0 mmHg or to judge the IRP to be abnormally high or not. Similarly, the corresponding answer format could be numerical or yes/no. Using REDCap (Research Electronic Data Capture), a secure web-based data collection application, housed at Northwestern University,⁸ experts provided their expert opinion on question and answer format for each skill. Expert agreement determined the final question and answer format for the HRM interpretation exam. The study team (RY, JEP) compiled a set of 25 HRM cases to include 2 to 3 cases of each diagnostic type according to Chicago Classification v3.0⁹, examples of technically limited studies, and five cases including hiatal hernias.

Test-Takers

Seven experts (KD, AG, CPG, POK, JEP, SS, RT) and 20 inexperienced trainees participated as test-takers. The test-takers interpreted the set of 25 HRM cases and provided answers for each case. Test-takers were remunerated for their participation. Individual test-taker, aggregate by level of test-taker, and cumulative performance to the HRM Interpretation exam was calculated. The answer-key for the 25 HRM cases was based on response agreement among the experts.

Judging per the Angoff Method

The Angoff method is a judge-based standard setting method that has been used across a spectrum of educational settings.¹⁰ Angoff estimates performance standards at the pass-fail level in an effort to distinguish the competent from non-competent candidate. In this process, each judge estimates the proportion of minimally competent examinees that would give a correct answer to each of the items. Each judge's estimates are summed across items, with the average sums determining the skill cut score.¹⁰ The modified Angoff includes a second round of judgements after the judges have seen their peers' judgements.¹¹

Two experts (PJK, DK) served as judges and participated in the modified Angoff method. The judges reviewed the blinded test-taker performance at both ends of the experience spectrum: (inexperienced trainee and expert) and individually determined the cut scores at the trainee, minimally competent physician interpreter, and master interpreter level for each skill. Per the Angoff method, judges were instructed to provide ratings no less than 25% and no greater than 95%. Ratings were compiled and discussed including a specific discussion for outlier high and low scores. Judges were given the opportunity to adjust their scores if desired and individually submitted another round of scores. The mean judge scores were calculated to determine the final cut scores for HRM interpretation skills.

Results

Determination of Question and Answer Format for HRM Interpretation Skills

Table 1 depicts the 8 HRM interpretation skills the experts considered essential in the interpretation of HRM. Expert agreement for the question and answer format ranged between 56 to 100% for the following 5 skills: recognition of pressure inversion point (PIP), IRP, hiatal hernia, distal latency, and diagnosis. For the remaining 3 skills (distal contractile integral (DCI), peristaltic integrity, and pressurization pattern), the study team generated only one question and answer format. Of note, experts considered recognition of a 3cm hiatal hernia to be more clinically significant than distinguishing EGJ morphology type, particularly type I versus type II. For both PIP and IRP experts agreed that identification of an abnormal study (e.g. absence of PIP, elevated IRP) was more relevant than was accuracy of the discrete value.

Test-taker Performance

The 7 expert test-takers completed the set of 25 HRM cases. Experts disagreed on the final diagnosis for 3 cases (Ineffective Esophageal Motility vs. Fragmented Peristalsis, Type II vs. Type III Achalasia, Normal vs. Technically Limited); these 3 cases were excluded from data analysis. Twenty trainee test-takers also completed the set of 25 HRM cases.

Cut Scores for HRM Interpretation

The 2 expert judges reviewed the individual and aggregate test-taker performance (aggregate results presented in Table 2). Per the Angoff method, judges based their estimations for minimum cut scores on the performance data and their expert opinions (Table 3). Figure 1 depicts the real-life cumulative test-taker performance and judge estimation by interpreter level for each HRM interpretation skill. For the most part, cut scores for the individual skills were consistent across levels (Figure 2). For instance, the cut scores for the physician interpreter generally ranged from 85 to 90% whereas the cut scores for masters were 95% and for the trainee ranged from 75 to 80%. However, scores were notably lower for peristaltic integrity and pressurization pattern. With regard to specific diagnoses, cut scores for identification of a major motility disorder were high across all three levels. On the other hand, while cut scores for the identification of a technically limited study were high for clinical interpreters (physician interpreter – 80% and master – 95%), the cut score for a minimally competent trainee was only 50%.

Discussion

In this study, we used a modified evidence-based standard setting method to determine minimum cut scores for HRM interpretation skills across three interpreter levels. The results suggest that a trainee with minimal proficiency in interpreting HRM should generally achieve between 65-80% accuracy for individual HRM skills and a master HRM interpreter should achieve at least 90-95% accuracy. Importantly, the physician interpreting HRM in unsupervised practice should achieve 85-90% accuracy for individual HRM skills, with the exception of judging peristaltic integrity, for which 65% was acceptable.

This work has important implications for training programs and physician practice.

Training Programs

We created a formal HRM interpretation exam in this study, inclusive of 22 HRM cases, each with 8 questions. This exam can be used in training programs to augment the deficiency in competency assessment for gastrointestinal motility.^{1,12} Our intention is that this exam will be a mechanism to familiarize trainees with esophageal HRM, provide meaningful and personalized self-assessment, and prompt focused improvement. The cut scores are meant to represent benchmarks for assessment toward the end of training.

Practicing Physicians

Moreover, this work may importantly impact the practice of unsupervised HRM interpretation. Currently, a method to assure physician readiness and competency for unsupervised HRM interpretation does not exist. The cut scores developed in this study provide minimum standards that a physician performing unsupervised HRM interpretation should achieve. Thus, these minimum cut scores may offer a benchmark to identify physicians who are competent and importantly, those who require further personalized and/or supervised training. Potentially, these benchmarks could be incorporated into a certification process for HRM interpretation.

This study also highlights that content experts view identification of major motility disorders and recognition of proper catheter placement (PIP) as priority skills. In addition, the disparity between high estimated cut scores for recognition of hiatal hernia > 3cm and low real-life performance suggests that although a challenging skill, experts consider recognition of hiatal hernia to be important.

Our study has several strengths. Our panel of experts included gastrointestinal surgeons and gastroenterologists from different clinical sites, improving the generalizability of our results. The cut scores were based on the modified Angoff method; evidence-based standard setting approaches are useful for setting fair and reliable performance cuts. In addition, our approach considered actual performance across two spectrums of interpreter level. Thus, our study synthesized an evidence-based approach to standard setting and data from real-life experience.

There are also limitations of these results. The predictive validity of performance against these standards to clinical performance and outcomes has yet to be studied. While these

skills assess the technical components of HRM interpretation, they do not examine decision making capabilities or knowledge of esophageal physiology. Although the cut scores are derived from performance to a specific set of 22 HRM cases, the cases were chosen to represent a broad range of real-life scenarios inclusive of various motility disorders, diaphragmatic defects and technical limitations. Additionally, the cases vary across difficulty level. As HRM metrics and classification schemes evolve, these benchmarks may require updating.

In conclusion, our study provides minimum cut scores for HRM interpretation across interpreter levels based on a formal standard setting process. These results can be used to conduct HRM interpretation exams amongst trainees, technicians and physicians, and to assess physician readiness for unsupervised HRM interpretation. Our study presents a novel approach to standard setting and addresses the current gap in competency-based training and assessment for the interpretation of esophageal HRM. This process can be used to develop benchmarks and improve the quality of interpretation across the spectrum of gastrointestinal motility diagnostics.

Acknowledgments

Grant Support: This work was supported by a grant from the National Institutes of Health, USA (T32 DK101363-02 to RY), a grant from the Digestive Health Foundation to RY, and a grant from the Eleanor Wood Prince Initiative to RY. No other grant support or funding.

Abbreviations

DCI	Distal contractile integral
DL	Distal latency
EGJ	Esophagogastric junction
GI	Gastroenterology
HRM	High-resolution manometry
IRP	Integrated relaxation pressure
PIP	Pressure inversion point
REDCap	Research Electronic Data Capture

References

1. Rao SS, Parkman HP. Advanced training in neurogastroenterology and gastrointestinal motility. *Gastroenterology*. 2015; 148:881–5. [PubMed: 25805422]
2. Carlson DA, Ravi K, Kahrilas PJ, et al. Diagnosis of Esophageal Motility Disorders: Esophageal Pressure Topography vs. Conventional Line Tracing. *Am J Gastroenterol*. 2015; 110:967–77. quiz 78. [PubMed: 26032151]
3. Yadlapati R, Gawron AJ, Keswani RN, et al. Identification of Quality Measures for Performance of and Interpretation of Data From Esophageal Manometry. *Clin Gastroenterol Hepatol*. 2015
4. American Gastroenterological Association; American Association for the Study of Liver Diseases; American College of Gastroenterology; American Society for Gastrointestinal Endoscopy. The

- Gastroenterology Core Curriculum, Third Edition. Gastroenterology. 2007; 132:2012–8. [PubMed: 17484892]
5. Yadlapati R, Keswani RN, Ciolino JD, et al. A System to Assess the Competency for Interpretation of Esophageal Manometry Identifies Variation in Learning Curves. Clin Gastroenterol Hepatol. 2016
 6. Angoff, WH. Scales, norms and equivalent scores. Washington DC: 1971.
 7. Cusimano MD. Standard setting in medical education. Acad Med. 1996; 71:S112–20. [PubMed: 8940953]
 8. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. Journal of biomedical informatics. 2009; 42:377–81. [PubMed: 18929686]
 9. Kahrilas PJ, Bredenoord AJ, Fox M, et al. The Chicago Classification of esophageal motility disorders, v3.0. Neurogastroenterol Motil. 2015; 27:160–74. [PubMed: 25469569]
 10. Angoff W. Test reliability and effective test length. Psychometrika. 1953; 18:1–14.
 11. Kardong-Edgren S, Mulcock PM. Angoff Method of Setting Cut Scores for High-Stakes Testing: Foley Catheter Checkoff as an Exemplar. Nurse Educ. 2016; 41:80–2. [PubMed: 26402909]
 12. Yadlapati R, Keswani RN, Pandolfino JE. Competency based medical education in gastrointestinal motility. Neurogastroenterol Motil. 2016 [Epub ahead of print].

Key Points

1. There is a deficiency of competency based training and assessment for gastrointestinal motility diagnostics such as high-resolution esophageal manometry.
2. We used a formal standard setting technique with a group of experts to develop a high-resolution esophageal manometry exam and to establish minimum cut scores for specific interpretation skills.
3. These results can be used to assess competency for high resolution esophageal manometry interpretation. Additionally, this method can be applied to developing benchmarks across the spectrum of gastrointestinal motility.

Figure 1. Test-taker cumulative performance and final cut score determined by judges according to the Angoff method for each HRM interpretation skill

Each panel is a scatterplot for each HRM Interpretation skill. The x-axis represents the percentage correct score and the y-axis represents test-taker cumulative frequency (percentage of cohort achieving that percentage correct score or lower). Curve colors correspond to which test-taker type achieved that percentage correct score (blue indicates trainee, yellow indicates expert, and green indicates trainee & expert). Orange circles correspond to judge determination of cut score based on Angoff method for each level. For each skill test-taker performance improved with experience level. In general, trainees (blue curve) achieved a broad range of lower scores, a mix of trainee and experts (green curve) achieved higher scores, and experts (yellow curve) achieved the highest percentage correct scores. [To editorial board – each panel uploaded as an individual figure. Intention is to display a 2×4 panel image]

Figure 2. Final HRM Interpretation Skill Cut Score by Interpreter Level
 Pressure inversion point (PIP); Hiatal hernia (HH); Integrated relaxation pressure (IRP);
 Distal contractile integral (DCI); Distal latency (DL)

Table 1
HRM Interpretation skills and answer options

The bolded question format represents the option with majority agreement amongst the panel.

Skill	Outcome	Question and Answer (Q&A) Format	Expert Agreement for Q&A Format
1. Pressure Inversion Point	Confirm whether the catheter traverses the esophagogastric junction	<ul style="list-style-type: none"> Numerical value within 2cm range Is a PIP present? (Yes/No) 	56%
2. Integrated Relaxation Pressure	Ensure that outflow disorders are recognized	<ul style="list-style-type: none"> Numerical value within a 3.0 mmHg range Is the IRP elevated? (Yes/No) 	56%
3. Hiatal Hernia (HH)	Identify clinically significant HHs	<ul style="list-style-type: none"> Is an HH > 5cm present? (Yes/No) Is an HH > 3cm present? (Yes/No) What is the EGJ morphology type? (I, II, III) 	78%
4. Distal Contractile Integral	Recognize clinically significant contractility patterns	<ul style="list-style-type: none"> Multiple choice: < 450, 450-8000, > 8000 mmHg.s.cm 	One Q&A format option
5. Distal Latency	Identify spastic motility disorders	<ul style="list-style-type: none"> Multiple choice: < 4.5, > 4.5 s Multiple Choice: < 4.5, 4.5-5.5, > 5.5 s 	78%
6. Peristaltic Integrity	Assess for clinically significant peristaltic breaks	<ul style="list-style-type: none"> Multiple Choice: Normal or Fragmented 	One Q&A format option
7. Pressurization Pattern	Identify clinically significant pressurization	<ul style="list-style-type: none"> Multiple Choice: No pressurization, Panesophageal pressurization, Compartmentalized pressurization 	One Q&A format option
8. Diagnosis (According to Chicago Classification v3.0 ⁹).	Identify clinically significant patterns of esophageal motility (see 8a and 8b)	<ul style="list-style-type: none"> Subject must correctly identify cases of absent contractility, hypercontractile esophagus, distal esophageal spasm, and normal esophageal motility 	One Q&A format option
8a. In the case of elevated IRP	Determine clinically significant disorders of outflow obstruction	<ul style="list-style-type: none"> Subject must correctly differentiate between type I achalasia, type II achalasia, type III achalasia, or EGJ Outflow Obstruction Either type I or type II achalasia acceptable if absent peristalsis Any diagnosis with EGJ obstruction acceptable 	100%
8b. In the case of a minor disorder		<ul style="list-style-type: none"> Subject must correctly differentiate between IEM or Fragmented peristalsis Either IEM or Fragmented peristalsis acceptable 	67%

Esophagogastric junction (EGJ); Ineffective esophageal motility (IEM).

Table 2

Aggregate Results of Trainee and Expert Test-Taker Accuracy for HRM Interpretation Skills

	Trainee Performance, Mean (SD)	Content Expert Performance, Mean (SD)
Presence of Pressure Inversion Point	89% (8%)	98% (3%)
Abnormal Integrated Relaxation Pressure	61% (14%)	86% (6%)
Presence of Hiatal Hernia > 3cm	72% (13%)	89% (9%)
Distal Contractile Integral Range	82% (6%)	97% (4%)
Distal Latency	80% (7%)	94% (2%)
Peristaltic Integrity	73% (7%)	91% (5%)
Pressurization Pattern	70% (14%)	84% (9%)
Diagnosis (per Chicago Classification v3.0)	67% (15%)	95% (5%)
Identification of Major Motility Disorder of Outflow Obstruction	83% (15%)	98% (4%)
Identification of Major Motility Disorder	87% (11%)	98% (3%)
Identification of Technically Limited Study	25% (44%)	86% (38%)

Major Motility Disorder of Outflow Obstruction include: Type I Achalasia, Type II Achalasia, Type III Achalasia, Esophagogastric Junction Outflow Obstruction. In addition to these, Major Motility Disorders include Absent Contractility, Hypercontractile Esophagus, and Distal Esophageal Spasm.

Table 3
Minimum Cut Scores for HRM Interpretation Skills Based on Angoff Method

Major Motility Disorder of Outflow Obstruction include: Type I Achalasia, Type II Achalasia, Type III Achalasia, Esophagogastric Junction Outflow Obstruction. Major Motility Disorder of Outflow Obstruction include: Type I Achalasia, Type II Achalasia, Type III Achalasia, Esophagogastric Junction Outflow Obstruction. In addition to these, Major Motility Disorders include Absent Contractility, Hypercontractile Esophagus, and Distal Esophageal Spasm.

	Trainee	Unsupervised Practice	Master
Presence of Pressure Inversion Point	80%	90%	95%
Abnormal Integrated Relaxation Pressure	80%	90%	95%
Presence of Hiatal Hernia > 3cm	80%	85%	95%
Distal Contractile Integral Range	80%	85%	95%
Distal Latency	65%	90%	95%
Peristaltic Integrity	65%	65%	90%
Pressurization Pattern	65%	85%	90%
Diagnosis (per Chicago Classification v3.0)	75%	90%	95%
Major Disorder of Outflow Obstruction	80%	90%	95%
Major Disorder	80%	90%	95%
Technically Limited Study	50%	80%	95%