



Published in final edited form as:

*Clin Gastroenterol Hepatol.* 2015 May ; 13(5): 874–83.e2. doi:10.1016/j.cgh.2014.11.012.

## Development of Quality Measures for the Care of Patients with Gastroesophageal Reflux Disease

Rena Yadlapati<sup>1</sup>, Andrew Gawron<sup>2</sup>, Karl Bilimoria<sup>3,4</sup>, Rajesh N. Keswani<sup>1</sup>, Kerry B. Dunbar<sup>5</sup>, Peter J. Kahrilas<sup>1</sup>, Philip Katz<sup>6</sup>, Joel Richter<sup>7</sup>, Felice Schnoll-Sussman<sup>8</sup>, Nathaniel Soper<sup>9</sup>, Marcelo F. Vela<sup>10</sup>, and John Pandolfino<sup>1</sup>

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

<sup>2</sup> Division of Gastroenterology, Hepatology, Nutrition, University of Utah, Salt Lake City, UT, United States

<sup>3</sup> Center for Healthcare Studies, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

<sup>4</sup> Surgical Outcomes and Quality Improvement Center, Department of Surgery, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

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

Disclosures/Conflicts of Interest:

Rena Yadlapati: None Andrew J Gawron: None Karl Bilimoria: None Rajesh N. Keswani: None Kerry B. Dunbar: None Peter J. Kahrilas: Dr. Kahrilas has done consulting for: AstraZeneca, Pfizer, Glaxo SmithKline, Trimedyne Inc., and Reckitt Benckiser.

Philip Katz: Dr. Katz received an honorarium for lectures with Takeda; Consultant for Pfizer Consumer Health, Torax.

Joel Richter: None

Felice Schnoll-Sussman: None

Nathaniel Soper: None

Marcelo Vela: Dr. Vela serves on the Covidien Advisory Board.

John Pandolfino: Dr. Pandolfino has done consulting for Given Imaging, is a speaker for Given imaging, and has grant support from Given imaging; is a speaker for Astra Zeneca; is a speaker for Takeda.

Author Contributions:

Rena Yadlapati: Study concept & design; acquisition of data; analysis and interpretation of data; drafting of manuscript; critical revision of the manuscript for important intellectual content; statistical analysis

Andrew Gawron: Study concept & design; acquisition of data; analysis and interpretation of data; drafting of manuscript; critical revision of the manuscript for important intellectual content; statistical analysis

Karl Bilimoria: Study concept & design; acquisition of data; analysis and interpretation of data; drafting of manuscript; critical revision of the manuscript for important intellectual content; statistical analysis; study supervision

Rajesh N. Keswani: Study concept & design; acquisition of data; analysis and interpretation of data; drafting of manuscript; critical revision of the manuscript for important intellectual content; statistical analysis

Kerry B. Dunbar: Acquisition of data; drafting of manuscript; critical revision of the manuscript for important intellectual content

Peter J. Kahrilas: Acquisition of data; drafting of manuscript; critical revision of the manuscript for important intellectual content

Philip Katz: Acquisition of data; drafting of manuscript; critical revision of the manuscript for important intellectual content

Joel Richter: Acquisition of data; drafting of manuscript; critical revision of the manuscript for important intellectual content

Felice Schnoll-Sussman: Acquisition of data; drafting of manuscript; critical revision of the manuscript for important intellectual content

Nathaniel Soper: Acquisition of data; drafting of manuscript; critical revision of the manuscript for important intellectual content

Marcelo F. Vela: Acquisition of data; drafting of manuscript; critical revision of the manuscript for important intellectual content

John Pandolfino: Study concept & design; acquisition of data; analysis and interpretation of data; drafting of manuscript; critical revision of the manuscript for important intellectual content; statistical analysis; study supervision

<sup>5</sup> University of Texas Southwestern Medical Center and the Dallas VA Medical Center, Department of Internal Medicine, Division of Gastroenterology and Hepatology, Dallas, TX, United States

<sup>6</sup> Thomas Jefferson University, Philadelphia, Pennsylvania; Division of Gastroenterology, Albert Einstein Medical Center, Philadelphia, PA, United States

<sup>7</sup> Department of Digestive Diseases and Nutrition, Joy McCann Culverhouse Center for Swallowing Disorders, University of South Florida Morsani College of Medicine, Tampa, FL, United States

<sup>8</sup> Division of Gastroenterology and Hepatology, Weill Cornell Medical College, New York, NY, United States

<sup>9</sup> Department of Surgery, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

<sup>10</sup> Division of Gastroenterology, Mayo Clinic Arizona, Scottsdale, AZ, United States

## Abstract

**Background & Aims**—Gastroesophageal reflux disease (GERD) is a common and costly disorder. Symptoms attributed to GERD have a wide spectrum of presentations and complications that have led to complex diagnostic and management algorithms. As such, there is considerable variation in clinical approaches to GERD. In contrast to multiple published guidelines for the management of GERD, there are few validated GERD quality measures. The objective of this study was to use a well-described, formal methodology to develop valid, physician-led, quality measures for all aspects of care for patients with GERD.

**Methods**—Quality measures were identified from the literature, consensus guidelines, and GERD experts. Eight clinical experts ranked potential measures for validity on the basis of the RAND/University of California, Los Angeles Appropriateness Methodology (RAM).

**Results**—Of the 52 proposed quality measures, 24 were rated as valid and 1 new measure was developed. These valid measures were related to initial diagnosis and management (9), monitoring (3), further diagnostic testing (4), proton pump inhibitor refractory symptoms (2), symptoms of chest pain (1), erosive esophagitis (3), esophageal stricture or ring (1), and surgical therapy (2). Fifteen of these measures were ranked with the highest validity. Twenty-seven measures were determined to be equivocal; 89% of these were extracted from guidelines based on low or moderate level evidence.

**Conclusion**—We used RAM to develop quality measures for GERD care. By examining performance on these valid, formally developed quality measures, clinical practices and individual providers can assess their adherence with them and direct quality improvement efforts accordingly.

## Keywords

Gastroesophageal Reflux Disease (GERD); measure of quality of care; RAND/UCLA Appropriateness Methodology (RAM); Proton pump inhibitor (PPI)

## INTRODUCTION

Gastroesophageal reflux disease (GERD) is a prevalent chronic disease, which inflicts a substantial economic burden on our healthcare system. It is the leader in terms of gastrointestinal (GI) diagnosis in ambulatory care (10-20% of adult outpatient visits), GI discharge diagnoses, and indication for upper endoscopy in the United States (US). GERD accounts for nine million hospital visits and \$10 billion in healthcare costs.<sup>1,2</sup> GERD manifests as a spectrum of syndromes including typical disease, extra-esophageal symptoms, erosive esophagitis, esophageal stricture, Barrett's esophagus, and esophageal adenocarcinoma. As a result, the diagnostic and management algorithm for GERD is complex and there is substantial heterogeneity in clinical approach.<sup>3-5</sup>

Several consensus guidelines developed by professional societies for the care of GERD exist. For instance, the most recent recommendations by the American Gastroenterological Association (AGA), American College of Gastroenterology (ACG), and the American Society for Gastrointestinal Endoscopy (ASGE) encompass over 80 different recommendations involving the care of GERD.<sup>6,7</sup>

With the healthcare shift from volume based to value based practice and legislative mandates such as the Affordable Care Act, there has been an increased emphasis to develop quality measures.<sup>8</sup> The National Quality Forum has introduced an initiative to establish a framework for quality measures to ensure that they are scientifically acceptable, usable and feasible.<sup>9</sup> Quality measures have been used to reduce the discrepancy in care for other diseases, such as colorectal cancer screening, inflammatory bowel disease and pancreatic cancer. Quality measures are held to a higher standard than guidelines, and non-adherence to quality measures is considered suboptimal care. There are currently five measures for GERD endorsed by the Agency for Healthcare Research and Quality (AHRQ).<sup>10</sup> These measures encompass the initial diagnostic evaluation and follow-up activities for GERD, however, do not address further diagnostic testing, surgical options, or the management of patients with proton pump inhibitor (PPI) refractory symptoms, non-cardiac chest pain, erosive esophagitis, or strictures/rings. Comprehensive quality measures for burdensome and prevalent disorders, such as GERD, are needed.

The objective of this study was to use a well-described, formal methodology to develop valid quality measures across the spectrum of GERD care. The hope is that these measures can be utilized to reduce variation in the management of GERD and offer a method to assess, monitor, and standardize GERD care.

## METHODS

This study was approved by the Northwestern University Institutional Review Board.

We used the RAND/University of California, Los Angeles Appropriateness Methodology (RAM) to develop quality measures for GERD care. RAM is a modified Delphi method, unique from the original Delphi in that it provides panelists with the opportunity to discuss their judgments between rating rounds.<sup>11</sup> It is a well-described method used to develop quality-of-care measures, and has been applied across a broad range of disease processes

such as resuscitation in cardiac arrest or surgical oncology.<sup>12-14</sup> First, members of the recruited expert panel independently rank potential quality measures for appropriateness. Next, the expert panel convenes for an in-person discussion focused on areas of disagreement, which is followed by a second round of independent rankings by each expert. Analysis of the measures for appropriateness (median ranking) and agreement (dispersion of rankings) generates quality measures, which have been shown to have face, construct and predictive validity (Figure 1).<sup>14-24</sup>

### **Compilation of Potential Quality Measures**

Potential GERD quality measures were identified by the authors (RY, AG) through an extensive systematic literature review, assessment of guidelines endorsed by professional societies (eg, ACG, AGA, and ASGE), and existing quality measures. Literature review included detailed study of forty-two scientific papers including large randomized controlled trials, cohort studies, and systematic reviews. The candidate measures encompassed initial diagnosis and management, monitoring, further diagnostic testing, surgical therapy, non-cardiac chest pain, erosive esophagitis, and strictures/rings.

### **Recruitment of the Expert Panel**

The main selection criteria in nominating the expert panel included leadership in the field of esophagology, geographic diversity and diversity of practice setting. According to RAM the expert panel should include 7 to 15 members, so as to be large enough to permit diversity of representation while still being small enough to allow all members to be involved in the group discussion. As such, an expert panel of 14 physicians was nominated, and ultimately 8 physicians accepted the nominations and participated in all processes. The panel was comprised of 2 female and 6 male clinicians and researchers in the field of GI and GI surgery from 6 academic institutions across the country with a mean of 26.5 years (range 13-39) experience in the management of GERD.

### **Round 1: Initial ranking of potential GERD quality measures**

The list of potential measures with specific instructions for ranking was sent to the expert panel members via electronic mail for the first round of rankings. A measure is considered appropriate if adherence is critical to providing quality care to patients with GERD, regardless of cost or feasibility of implementation. Rankings are based on the panelists' personal judgment and not on what they thought other experts believed. In addition, the measure should apply to the average patient who presents to the average physician at an average hospital. Finally, measures may not always provide benefit to an individual patient but should be beneficial to the overall care of patients with GERD.

Each measure was ranked on a nine-point interval scale in which a score of 1 signified definitely inappropriate, 5 signified uncertain/equivocal appropriateness, and 9 signified definitely appropriate. Panelists were also given the opportunity to suggest wording modifications to improve the clarity and potential validity of the measure or to suggest a new measure. Summary statistics were calculated for each individual potential quality measure, and the measures were assessed for agreement. Agreement for a panel of eight members was defined by seven or more rankings falling in the same three-point range (ie,

1-3, 4-6, or 7-9) whereas disagreement was defined by two or more rankings falling in separate ranges.

A literature search specific to the ten measures with disagreement was performed via PubMed by the authors (RY, AG). Studies conducted after 1990 relevant to the specific measure including large randomized control trials, cohort trials or systematic reviews, as well as the most recent guidelines for GERD management from the ACG, AGA & ASGE were reviewed and summarized. For each of these specific measures an overview document was created that included: the proposed measure; deidentified group rankings with median score; a detailed review of the methods and results of pertinent trials; and guideline with level of evidence if the measure was derived from a guideline.<sup>6,7,10,25-41</sup>

## Round 2: Discussion of potential GERD quality indicators and re-ranking

At a face-to-face meeting of all panelists (May 2014 in Chicago, IL), a packet of information summarizing the round 1 rankings, the RAM process, and measure-specific overview documents (as described above) was provided to each member of the expert panel. This packet was also sent via electronic mail to each panel member two weeks prior to the round 2 meeting for review. The quality measures with disagreement were discussed amongst the panel to identify opportunities to improve wording and review the evidence. New measures could also be proposed. After the expert panel discussion of each measure, the panelists independently re-ranked each measure for appropriateness. The rankings from round 2 were used as the final assessment of validity. The rankings were compiled, and summary statistics were again calculated for each individual measure.

## Analyzing Measures for Validity

Analysis of the measures was performed using the scoring definitions delineated by RAM. Validity was determined based on median rankings (appropriateness) and the dispersion of rankings (agreement). Agreement for a panel of eight members was defined by seven or more panelists' rankings falling in the same three-point range (ie, 1-3, 4-6, or 7-9). Agreement was further separated into strict agreement in which all panel members' rankings fell in the same three-point range whereas relaxed agreement indicated that all but one of the members' rankings fell in the same three-point range. If two or more of the rankings were in disparate categories this was considered to be indicative of disagreement. A measure was deemed to have high validity if there was strict agreement for rankings in the range of 7-9 for a measure, and was deemed to be of moderate validity if all but one of the rankings were in this range. Measures were equivocal if the median ranking was in the range of 4-6 or if there was overall disagreement for the measure. If the median ranking was in the range of 1-3 and there was agreement amongst the panel, the measure was deemed to be invalid (Table 1).<sup>11</sup>

## RESULTS

Review of the literature and consideration of consensus guidelines generated 52 potential quality measures (Supplementary Table).<sup>6,7,10</sup> On the basis of the final expert panel rankings, 25 of these measures (48%) were found to be valid and two were combined to

develop one valid measure, yielding a total of 24 valid measures. These valid measures were related to initial diagnosis and management (9), monitoring (2), further diagnostic testing (4), PPI-refractory symptoms (2), symptoms of chest pain (1), erosive esophagitis (3), esophageal stricture or ring (1), and surgical therapy (2). Additionally, one new valid measure related to monitoring was developed during round 2, which stated “IF a patient with GERD is prescribed an initial empiric trial of PPI, THEN the patient should have scheduled follow-up within 4-12 weeks.” Fifteen of these measures were determined to have high-validity (Table 2) and 10 were ranked with moderate-validity (Table 3). Of the measures rated as valid, 22 could potentially be derived from patient medical records and 3 could be extracted from procedure reports.

There were 27 measures (52%) that were determined to have equivocal validity. Panelists commented that there was insufficient data available to translate these guidelines into measures of quality of care. Panelists also felt that these potential measures were based on guidelines that practitioners should individualize on a case-by-case basis, and that the guidelines did not necessarily apply to the average patient. As such, they did not feel that the measure defined appropriateness of care. Additionally, some potential measures were not felt to be feasible as a quality measure. Of the 27 measures that were ranked as equivocal, 9 (33%) were derived from guidelines based on low level of evidence, 14 (52%) were based on moderate level evidence, and 4 (15%) were based on high level of evidence (Table 4). None of the proposed quality measures were ranked as invalid.

## DISCUSSION

GERD is a prevalent and costly disorder, with substantial variability in clinical care.<sup>4</sup> In an attempt to develop comprehensive quality measures for GERD through application of RAM, an expert panel evaluated 52 potential measures. After two rounds of ranking, a total of 25 measures were determined to be valid.

Of these, there was a subset of 15 quality measures identified by the experts to have highest validity. We believe that adherence to this core group of measures is necessary in the management of GERD. Combining this core group of quality measures to yield a composite quality measure for GERD may be of value, as composite measures have been found to ultimately be a more reliable assessment of overall care.<sup>12</sup> In addition, future directions should involve assessing the performance of these measures at the clinical practice and individual provider level to identify areas for quality improvement. The ability to accurately measure and report the quality of GERD care from electronic medical records will require advances in data extraction techniques so that quality measurement and reporting are not burdensome.

Four of the 15 measures ranked with high validity and 1 of the 10 ranked with moderate validity were derived from the five existing AHRQ performance measures. The overlap between existing AHRQ measures and our study highlights the reliability of our methodology and expert panel rankings (Table 5). In addition to agreeing with the AHRQ's existing five performance measures, our study offers a more comprehensive group of measures, which better encompasses the wide spectrum of GERD and its complications.

Through this process, 27 measures were determined to be equivocal. For example, experts felt that in patients with non-cardiac chest pain attributed to GERD the next step should be individualized based on symptoms and previous workup. When discussing whether patients with Schatzki rings should be treated with a PPI following dilation, there was disagreement regarding the relationship between a Schatzki ring and reflux. Interestingly, 23 (85%) of the equivocal measures were derived from guidelines based on low or moderate level of evidence. Only 4 equivocal measures were based on high level of evidence. Two of these related to surgical therapy, and in both these scenarios experts agreed that the decision to recommend surgical therapy as the next step should be individualized on a case-by-case basis. Disagreement additionally existed for the proposed measure discussing concomitant clopidogrel and PPI use. Although recent high level evidence did not reveal adverse cardiovascular events in the setting of clopidogrel and PPI use, there was disagreement amongst the panel for whether this should be a measure that all providers should be expected to comply with given the previous “Black Box” warnings.<sup>42</sup> This process highlights the fact that not all guidelines are appropriate quality measures. Through application of RAM we examined the abundance of expert opinions and recommendations in the published literature, and identified guidelines that are not valid quality measures. Determination of potentially invalid or equivocal measures is an important process, particularly for common disorders with complex algorithms for which varying opinions may exist.

This is the first report of using RAM to develop quality measures for GERD care. This structured process expands on previous reports and existing measures that focused on individual aspects of GERD care, but did not address the wide spectrum of syndromes related to GERD. This is not an attempt to promote or create specific practice guidelines, but rather to provide baseline quality measures by which payers, institutions, physicians and patients can assess GERD care. While several guidelines make recommendations for GERD care on the basis of the best available evidence, quality measures are held to higher standards. Quality measures must be measurable, reportable, scientifically acceptable, usable, and logistically feasible, and non-adherence to a measure is considered suboptimal care. The development and utilization of valid measures to improve quality and reduce variability in healthcare has been endorsed and shown to improve care.<sup>10</sup>

In conclusion, we developed a physician-led concise and comprehensive group of valid quality measures in GERD care by using a formal, well-described methodology. Our intent was to develop measures of quality of care that physicians could use for self-assessment to identify quality initiatives for improving GERD care. This work is an initial, but important, step in developing GERD quality measures. As the US healthcare system transitions financially from volume based to value and quality based, it is critical that quality measures are rigorously vetted and evaluated.

## Acknowledgments

Donald Castell, David Katzka, Marco Patti, Nicholas Shaheen, Michael Vaezi

**Grant Support:** Dr. Kahrilas was supported by grant #DK056033 from the public health service. Dr. Pandolfino was supported by grants #DK07659 and #DK092217 from the Public Health Service.

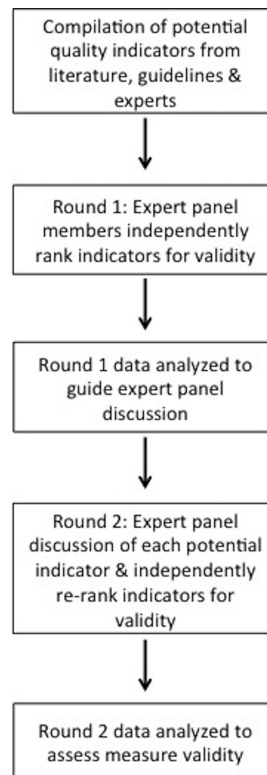
## REFERENCES

1. Peery AF, Dellon ES, Lund J, et al. Burden of gastrointestinal disease in the United States: 2012 update. *Gastroenterology*. 2012; 143:1179–87 e1-3. [PubMed: 22885331]
2. Shaheen NJ, Hansen RA, Morgan DR, et al. The burden of gastrointestinal and liver diseases, 2006. *Am J Gastroenterol*. 2006; 101:2128–38. [PubMed: 16848807]
3. Ronkainen J, Agreus L. Epidemiology of reflux symptoms and GORD. *Best Pract Res Clin Gastroenterol*. 2013; 27:325–37. [PubMed: 23998972]
4. Stefanidis D, Hope WW, Kohn GP, et al. Guidelines for surgical treatment of gastroesophageal reflux disease. *Surg Endosc*. 2010; 24:2647–69. [PubMed: 20725747]
5. Ip, S.; Chung, M.; Moorthy, D., et al. Comparative Effectiveness of Management Strategies for Gastroesophageal Reflux Disease: Update. Comparative Effectiveness Review No. 29. AHRQ Publication No. 11-EHC049-EF. Agency for Healthcare Research and Quality; Rockville, MD: Sep. 2011 Available at:[www.effectivehealthcare.ahrq.gov/reports/final.cfm](http://www.effectivehealthcare.ahrq.gov/reports/final.cfm)
6. Kahrilas PJ, Shaheen NJ, Vaezi MF, American Gastroenterological Association I, Clinical P, Quality Management C. American Gastroenterological Association Institute technical review on the management of gastroesophageal reflux disease. *Gastroenterology*. 2008; 135:1392–413. 413, e1–5. [PubMed: 18801365]
7. Katz PO, Gerson LB, Vela MF. Guidelines for the diagnosis and management of gastroesophageal reflux disease. *Am J Gastroenterol*. 2013; 108:308–28. quiz 29. [PubMed: 23419381]
8. Manchikanti L, Caraway DL, Parr AT, Fellows B, Hirsch JA. Patient Protection and Affordable Care Act of 2010: reforming the health care reform for the new decade. *Pain physician*. 2011; 14:E35–67. [PubMed: 21267047]
9. Werner RM, Asch DA. The unintended consequences of publicly reporting quality information. *JAMA*. 2005; 293:1239–44. [PubMed: 15755946]
10. Agency for Healthcare Research and Quality. [March 27, 2014] National Quality Measures Clearinghouse. 2014. Retrieved from <http://www.qualitymeasures.ahrq.gov/search/search.aspx?term=gerd>
11. Fitch, KBS.; Aguilar, MD.; Burnand, B., et al. The RAND/UCLA appropriateness method user's manual. RAND; Santa Monica: 2001.
12. Halverson AL, Sellers MM, Bilimoria KY, et al. Identification of Process Measures to Reduce Postoperative Readmission. *J Gastrointest Surg*. 2014 Epub 2014/06/11. doi: 10.1007/s11605-013-2429-5.
13. Bilimoria KY, Bentrem DJ, Lillemoe KD, Talamonti MS, Ko CY. Pancreatic Cancer Quality Indicator Development Expert Panel ACoS. Assessment of pancreatic cancer care in the United States based on formally developed quality indicators. *J Natl Cancer Inst*. 2009; 101:848–59. [PubMed: 19509366]
14. Bilimoria KY, Raval MV, Bentrem DJ, Wayne JD, Balch CM, Ko CY. National assessment of melanoma care using formally developed quality indicators. *J Clin Oncol*. 2009; 27:5445–51. [PubMed: 19826131]
15. Aguilar MD, Fitch K, Lazaro P, Bernstein SJ. The appropriateness of use of percutaneous transluminal coronary angioplasty in Spain. *Int J Cardiol*. 2001; 78:213–21. discussion 21-3. [PubMed: 11376822]
16. Anderson SD, Lambert S, Brannan JD, et al. Laboratory protocol for exercise asthma to evaluate salbutamol given by two devices. *Med Sci Sports Exerc*. 2001; 33:893–900. [PubMed: 11404653]
17. Bernstein SJ, Lazaro P, Fitch K, Aguilar MD, Kahan JP. Effect of specialty and nationality on panel judgments of the appropriateness of coronary revascularization: a pilot study. *Med Care*. 2001; 39:513–20. [PubMed: 11317099]
18. Brook RH, McGlynn EA, Shekelle PG. Defining and measuring quality of care: a perspective from US researchers. *Int J Qual Health Care*. 2000; 12:281–95. [PubMed: 10985266]
19. Lawson EH, Gibbons MM, Ko CY, Shekelle PG. The appropriateness method has acceptable reliability and validity for assessing overuse and underuse of surgical procedures. *J Clin Epidemiol*. 2012; 65:1133–43. [PubMed: 23017632]



20. Maggard MA, McGory ML, Shekelle PG, Ko CY. Quality indicators in bariatric surgery: improving quality of care. *Surg Obes Relat Dis.* 2006; 2:423–9. discussion 9-30. [PubMed: 16925372]
21. McGory ML, Shekelle PG, Ko CY. Development of quality indicators for patients undergoing colorectal cancer surgery. *J Natl Cancer Inst.* 2006; 98:1623–33. [PubMed: 17105985]
22. McGory ML, Shekelle PG, Rubenstein LZ, Fink A, Ko CY. Developing quality indicators for elderly patients undergoing abdominal operations. *J Am Coll Surg.* 2005; 201:870–83. [PubMed: 16310690]
23. Shekelle P. The appropriateness method. *Med Decis Making.* 2004; 24:228–31. [PubMed: 15090107]
24. Shekelle PG, Park RE, Kahan JP, Leape LL, Kamberg CJ, Bernstein SJ. Sensitivity and specificity of the RAND/UCLA Appropriateness Method to identify the overuse and underuse of coronary revascularization and hysterectomy. *J Clin Epidemiol.* 2001; 54:1004–10. [PubMed: 11576811]
25. Armstrong D, Marshall JK, Chiba N, et al. Canadian Consensus Conference on the management of gastroesophageal reflux disease in adults - update 2004. *Can J Gastroenterol.* 2005; 19:15–35. [PubMed: 15685294]
26. Ates F, Vaezi MF. New Approaches to Management of PPI-Refractory Gastroesophageal Reflux Disease. *Curr Treat Options Gastroenterol.* 2014; 12:18–33. [PubMed: 24430334]
27. Campos GM, Peters JH, DeMeester TR, et al. Multivariate analysis of factors predicting outcome after laparoscopic Nissen fundoplication. *J Gastrointest Surg.* 1999; 3:292–300. [PubMed: 10481122]
28. Castell DO, Kahrilas PJ, Richter JE, et al. Esomeprazole (40 mg) compared with lansoprazole (30 mg) in the treatment of erosive esophagitis. *Am J Gastroenterol.* 2002; 97:575–83. [PubMed: 11922549]
29. Cicala M, Gabbrielli A, Emerenziani S, et al. Effect of endoscopic augmentation of the lower oesophageal sphincter (Gatekeeper reflux repair system) on intraoesophageal dynamic characteristics of acid reflux. *Gut.* 2005; 54:183–6. [PubMed: 15647177]
30. Clayton SB, Rife CC, Singh ER, Kalbfleisch JH, Castell DO. Twice-daily proton pump inhibitor therapy does not decrease the frequency of reflux episodes during nocturnal recumbency in patients with refractory GERD: analysis of 200 patients using multichannel intraluminal impedance-pH testing. *Dis Esophagus.* 2012; 25:682–6. [PubMed: 22292567]
31. Devault KR, Johanson JF, Johnson DA, Liu S, Sostek MB. Maintenance of healed erosive esophagitis: a randomized six-month comparison of esomeprazole twenty milligrams with lansoprazole fifteen milligrams. *Clin Gastroenterol Hepatol.* 2006; 4:852–9. [PubMed: 16682260]
32. Fass R, Murthy U, Hayden CW, et al. Omeprazole 40 mg once a day is equally effective as lansoprazole 30 mg twice a day in symptom control of patients with gastro-oesophageal reflux disease (GERD) who are resistant to conventional-dose lansoprazole therapy-a prospective, randomized, multi-centre study. *Aliment Pharmacol Ther.* 2000; 14:1595–603. [PubMed: 11121907]
33. Freston JW, Jackson RL, Huang B, Ballard ED 2nd. Lansoprazole for maintenance of remission of erosive oesophagitis. *Drugs.* 2002; 62:1173–84. [PubMed: 12010078]
34. Galindo G, Vassalle J, Marcus SN, Triadafilopoulos G. Multimodality evaluation of patients with gastroesophageal reflux disease symptoms who have failed empiric proton pump inhibitor therapy. *Dis Esophagus.* 2013; 26:443–50. [PubMed: 22862422]
35. Hanna S, Rastogi A, Weston AP, et al. Detection of Barrett's esophagus after endoscopic healing of erosive esophagitis. *Am J Gastroenterol.* 2006; 101:1416–20. [PubMed: 16863541]
36. Hatlebakk JG, Berstad A. Pharmacokinetic optimisation in the treatment of gastro-oesophageal reflux disease. *Clin Pharmacokinet.* 1996; 31:386–406. [PubMed: 9118586]
37. Kahrilas PJ, Hughes N, Howden CW. Response of unexplained chest pain to proton pump inhibitor treatment in patients with and without objective evidence of gastro-oesophageal reflux disease. *Gut.* 2011; 60:1473–8. [PubMed: 21508423]
38. Lauritsen K, Deviere J, Bigard MA, et al. Esomeprazole 20 mg and lansoprazole 15 mg in maintaining healed reflux oesophagitis: Metropole study results. *Aliment Pharmacol Ther.* 2003; 17(Suppl 1):24. discussion 5-7. [PubMed: 12614304]

39. Marshall JB, Kretschmar JM, Diaz-Arias AA. Gastroesophageal reflux as a pathogenic factor in the development of symptomatic lower esophageal rings. *Arch Intern Med.* 1990; 150:1669–72. [PubMed: 2383161]
40. Roorda AK, Marcus SN, Triadafilopoulos G. Algorithmic approach to patients presenting with heartburn and epigastric pain refractory to empiric proton pump inhibitor therapy. *Dig Dis Sci.* 2011; 56:2871–8. [PubMed: 21512760]
41. Sgouros SN, Vlachogiannakos J, Karamanolis G, et al. Long-term acid suppressive therapy may prevent the relapse of lower esophageal (Schatzki's) rings: a prospective, randomized, placebo-controlled study. *Am J Gastroenterol.* 2005; 100:1929–34. [PubMed: 16128935]
42. Bhatt DL, Cryer BL, Contant CF, et al. Clopidogrel with or without omeprazole in coronary artery disease. *N Engl J Med.* 2010; 363:1909–17. [PubMed: 20925534]



**Figure 1.** Overview of the RAND/UCLA Appropriateness Methodology used to develop quality measures for GERD.<sup>10</sup>

**Table 1**

Scoring system to determine validity of quality measures.

<b>Validity</b>	<b>Appropriateness + Agreement</b>
Highly Valid	Median Ranking of 7-9 <i>and</i> Strict Agreement (all 8 panelists ranked the measure as 7, 8, or 9)
Moderately Valid	Median Ranking of 7-9 <i>and</i> Relaxed Agreement (7 of the 8 panelists ranked the measure as 7, 8, or 9)
Equivocally Valid	Median Ranking of 4-6 <i>and/or</i> Disagreement (6 or less of the 8 panelists ranked the measure within the same three-point range (1-3, 4-6, or 7-9))
Invalid	Median Ranking of 1-3 <i>and</i> Agreement (7 or more of the 8 panelists ranked the measure as 1, 2, or 3)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

## High-validity GERD Care Quality Measures.

	Median Rankings (Individual panelist rankings)
<b>Initial Diagnosis &amp; Management (n=7)</b>	
IF a patient has typical GERD, THEN an initial trial of empiric PPI therapy, H <sub>2</sub> RA, or antacid is appropriate. <sup>†</sup>	Round 1: 8.0 (8, 7, 9, 9, 9, 7, 8, 8)
IF PPI therapy is initiated, THEN it should be at once a day dosing before the first meal of the day (or before an evening meal for patients with predominant nighttime symptoms). <sup>†</sup>	Round 1: 9.0 (9, 9, 9, 9, 9, 8, 8, 9) Round 2: 9.0 (9, 9, 9, 9, 9, 8, 8, 9)
IF a patient with a diagnosis of GERD is seen for initial evaluation, THEN the patient should be assessed for the presence or absence of the following alarm symptoms: involuntary weight loss, dysphagia, and GI bleeding.*	Round 1: 8.5 (8, 8, 8, 8, 9, 9, 9, 9)
IF a patient with a diagnosis of GERD has at least one alarm symptom, THEN upper endoscopy should be performed.* <sup>†</sup>	Round 1: 9.0 (8, 9, 9, 7, 9, 9, 9, 9)
IF a patient with GERD is overweight or obese, THEN weight loss should be advised.	Round 1: 8.0 (9, 8, 8, 9, 8, 8, 8, 9)
IF a patient with GERD has an endoscopy report that indicates a suspicion of Barrett's esophagus, THEN suspicious areas should be biopsied.* <sup>†</sup>	Round 1: 9.0 (8, 9, 8, 9, 9, 9, 9, 9)
IF a patient with GERD has troublesome dysphagia, THEN endoscopy with biopsy should be performed. Biopsies should target any areas of suspected metaplasia, dysplasia, or in the absence of visual abnormalities, normal mucosa (4 biopsies from both proximal and distal esophagus to rule out eosinophilic esophagitis). <sup>†</sup>	Round 1: 9.0 (8, 8, 8, 9, 9, 9, 9, 9)
<b>Monitoring (n=2)</b>	
IF a patient with GERD is prescribed chronic PPI or H <sub>2</sub> RAs, THEN the patient should receive an assessment of their GERD symptoms within 12 months.*	Round 1: 7.5 (8, 6, 7, 7, 2, 9, 9, 8) Round 2: 8.0 (8, 7, 7, 7, 8, 8, 9, 9)
IF a patient with GERD is prescribed an initial empiric trial of PPI, THEN the patient should have scheduled follow-up within 4 to 12 weeks. <sup>†</sup>	Round 1: Measure did not exist Round 2: 8.0 (9, 8, 7, 8, 9, 9, 8, 9)
<b>PPI-Refractory Symptoms (n=1)</b>	
IF a patient has refractory typical GERD symptoms despite twice daily PPI and adherence to PPI, THEN an upper endoscopy should be performed to exclude non-GERD etiologies. <sup>†</sup>	Round 1: 9.0 (9, 9, 9, 9, 9, 9, 9, 9)
<b>Chest Pain (n=1)</b>	
IF a patient has chest pain, THEN a cardiac cause should be excluded before the commencement of a gastrointestinal evaluation.	Round 1: 9.0 (9, 9, 7, 9, 7, 8, 9, 9)
<b>Erosive Esophagitis (n=3)</b>	
IF erosive esophagitis is seen on endoscopy, THEN findings should be classified according to the Los Angeles (LA) classification system. <sup>†</sup>	Round 1: 9.0 (9, 9, 8, 9, 9, 9, 7, 9)
	See supplemental table measures #41 & #42. (Combination of proposed measures).
IF a patient has LA grade B or greater erosive esophagitis, THEN at least an 8-week course of PPI is the therapy of choice for symptom relief and healing. <sup>†</sup>	Round 1: 9.0 (9, 9, 8, 1, 9, 9, 9, 8) Round 2: 9.0 (9, 9, 9, 8, 9, 9, 9, 9)
IF a patient has LA grade C or D erosive esophagitis, THEN repeat endoscopy should be performed after a course of antisecretory therapy to exclude underlying Barrett's esophagus. <sup>†</sup>	Round 1: 7.0 (7, 7, 7, 6, 9, 8, 8, 7) Round 2: 8.5 (9, 8, 8, 8, 9, 8, 9, 9)
<b>Stricture/Ring (n=1)</b>	
IF a patient has a peptic stricture, THEN maintenance PPI therapy is recommended following stricture dilation to reduce the need for repeated dilations.	Round 1: 8.0 (8, 8, 8, 8, 9, 9, 9, 8)

GERD: Gastroesophageal Reflux Disease; PPI Proton Pump Inhibitor; H<sub>2</sub>RA: H<sub>2</sub> Receptor Antagonist.

\* Denotes overlap with existing Agency for Healthcare Research and Quality Performance Measures

† Initial proposed measure was reworded during the two rounds

‡ New measure developed by the expert panel in Round 2

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3**

## Moderate-Validity GERD Care Quality Measures.

	Median Rankings (Individual panelist rankings)
<b>Initial Diagnosis &amp; Management</b> (n=2)	
IF a patient with non-erosive GERD experiences heartburn relief with H <sub>2</sub> RA therapy, THEN an H <sub>2</sub> RA can be used as a maintenance option.	Round 1: 8.0 (5, 8, 7, 9, 8, 8, 9, 9)
IF a patient has suspected GERD without dysphagia, THEN a barium radiograph should not be used as a diagnostic test. <sup>*,†</sup>	Round 1: 8.0 (8, 9, 9, 8, 8, 7, 9, 2)
<b>Monitoring</b> (n=1)	
IF PPIs have proven clinically effective for patients with GERD, THEN PPIs should be used long-term. <sup>†</sup>	Round 1: 8.0 (8, 6, 8, 8, 8, 8, 9, 8)
<b>PPI-Refractory Symptoms</b> (n=1)	
IF a patient has refractory GERD symptoms despite standard PPI therapy, THEN the first step in management is optimization of PPI therapy. <sup>†</sup>	Round 1: 8.0 (9, 9, 9, 1, 9, 9, 9, 8)
<b>Further Diagnostic Testing</b> (n=4)	
IF a patient with suspected troublesome GERD has not responded to empirical trial of PPI therapy, has normal findings on endoscopy, and has no major abnormalities on manometry, THEN ambulatory reflux monitoring off of PPI therapy for 7 days should be performed. <sup>†</sup>	Round 1: 8.0 (8, 9, 8, 7, 9, 9, 9, 5)
IF a patient has suspected GERD with disease refractory to PPI therapy and no findings of erosive disease on endoscopy, THEN ambulatory esophageal reflux monitoring off of PPI therapy for 7 days is indicated before consideration of endoscopic or surgical therapy.	Round 1: 9.0 (9, 9, 9, 9, 9, 9, 9, 6)
IF planning to perform reflux monitoring off of anti-reflux medication, THEN either pH or impedance-pH monitoring are sufficient to establish a GERD diagnosis. <sup>†</sup>	Round 1: 8.0 (8, 8, 8, 6, 9, 9, 9, 7)
IF planning to perform reflux monitoring on anti-reflux medication, THEN impedance-pH monitoring should be performed to enable measurement of persistent acid or nonacid reflux. <sup>†</sup>	Round 1: 8.5 (8, 9, 8, 6, 9, 9, 9, 7)
<b>Surgical Therapy</b> (n=2)	
IF a patient with GERD is refractory to medical therapy and has objective evidence of ongoing reflux as the cause of symptoms, THEN consideration should be given to anti-reflux surgery. <sup>†</sup>	Round 1: 8.0 (6, 8, 8, 8, 9, 9, 9, 8)
IF anti-reflux surgery and PPI therapy are judged to offer similar efficacy in a patient with an esophageal GERD syndrome, THEN PPI therapy should be recommended as initial therapy because of superior safety and long-term efficacy. <sup>†</sup>	Round 1: 8.0 (6, 8, 8, 9, 7, 8, 9, 8)

GERD: Gastroesophageal Reflux Disease; PPI Proton Pump Inhibitor; H<sub>2</sub>RA: H<sub>2</sub> Receptor Antagonist.

\* Denotes overlap with existing Agency for Healthcare Research and Quality Performance Measures

<sup>†</sup> Initial proposed measure was reworded during the two rounds

Table 4

GERD Quality Measures ranked with equivocal validity.

Proposed Quality Measure	Median Rankings (Individual panelist rankings)	Level of evidence
<b>Initial Diagnosis &amp; Management</b>		
IF a patient has typical symptoms of heartburn and regurgitation, THEN a presumptive diagnosis of GERD can be made without the need for diagnostic testing, including endoscopy.	Round 1: 8.0 (6, 7, 9, 9, 8, 6, 8, 6)	Moderate
IF antisecretory drugs are being use for treatment of GERD, THEN it should be understood that PPIs are more effective than H <sub>2</sub> RAs, which are more effective than placebo.	Round 1: 8.5 (9, 6, 9, 9, 9, 8, 9, 6)	High
IF traditional delayed release PPIs are used, THEN they should be administered 30 to 60 minutes before a meal for maximal pH control.	Round 1: 8.5 (9, 8, 8, 9, 9, 6, 9, 6)	Moderate
IF a patient has partial but inadequate response to heart burn symptoms on once daily PPI therapy, THEN the next step should be increase the dose to twice daily. <sup>†</sup>	Round 1: 7.0 (7, 4, 8, 7, 5, 7, 8, 7) Round 2: 5.5 (7, 4, 4, 3, 8, 7, 6, 5)	Low
IF a patient on PPI therapy has objective evidence of nighttime reflux, THEN bedtime H <sub>2</sub> RA therapy can be added if needed, but may be associated with the development of tachyphylaxis after several weeks of use.	Round 1: 7.0 (7, 6, 6, 3, 7, 8, 7, 6)	Low
IF a patient is on concomitant clopidogrel, THEN PPI therapy does not need to be altered as there does not appear to be an increased risk for adverse cardiovascular events.	Round 1: 7.0 (5, 8, 7, 7, 6, 5, 9, 7)	High
IF a patient has GERD, THEN therapy for GERD other than acid suppression, including prokinetic therapy and/or baclofen, should not be used in GERD patients without diagnostic evaluation.	Round 1: 8.0 (6, 8, 8, 8, 8, 9, 2)	Moderate
IF a patient has GERD, THEN lifestyle modifications including elevation of the head of the bed, avoiding late meals, avoiding specific foods, or avoiding specific activities should be tailored to the circumstances of the individual patient rather than broadly advocated for all patients.	Round 1: 8.0 (8, 8, 6, 9, 9, 8, 7, 6)	Moderate
IF a non-pregnant patient has GERD, THEN there is no role for sucralfate in the management of GERD.	Round 1: 8.0 (8, 8, 6, 8, 8, 5, 8, 7)	Moderate
IF a patient has GERD, then screening for <i>Helicobacter pylori</i> infection is not recommended.	Round 1: 8.0 (5, 9, 7, 8, 8, 6, 9, 9)	Low
IF a patient has suspected GERD, THEN routine biopsies from the distal esophagus are not recommended specifically to diagnose GERD.	Round 1: 8.0 (8, 8, 8, 1, 9, 6, 9, 8)	Moderate
IF a patient has uncomplicated GERD, THEN endoscopic anti-reflux therapy may be considered for selected patients after careful discussion with the patient regarding potential side effects, benefits, and other available therapeutic options.	Round 1: 6.0 (3, 3, 7, 1, 9, 6, 9, 8)	Low
<b>Monitoring</b>		
IF a patient with GERD continues to have symptoms after a PPI is discontinued, THEN maintenance PPI therapy should be administered.	Round 1: 7.0 (7, 7, 8, 7, 5, 7, 9, 6)	Moderate
IF a patient has LA grade C or D erosive esophagitis, THEN at least daily PPI dosing should be used. <sup>†</sup>	Round 1: 6.0 (7, 8, 8, 8, 2, 5, 8, -) Round 2: 4.0 (5, 2, 2, 1, 9, 7, 6, 3)	Low
IF a patient has erosive or nonerosive reflux disease, THEN routine endoscopy to assess for disease progression should not be performed.	Round 1: 8.0 (7, 8, 8, 8, 2, 5, 8, -)	Low
<b>Further Diagnostic Testing</b>		
IF a patient with suspected GERD has not responded to empirical trial of twice-daily PPI and has normal findings on endoscopy, THEN manometry should be performed.	Round 1: 6.5 (8, 5, 7, 7, 5, 6, 9, 1)	Moderate
IF short- or long-segment Barrett's esophagus is present, THEN ambulatory reflux monitoring is not needed to establish a diagnosis of GERD.	Round 1: 6.5 (6, 6, 8, 1, 5, 9, 9, 7)	Moderate



Proposed Quality Measure	Median Rankings (Individual panelist rankings)	Level of evidence
<b>Chest Pain</b>		
IF a patient has suspected reflux chest pain syndrome and a cardiac etiology has been carefully considered, THEN twice-daily PPI therapy should be started as an empiric trial.	Round 1: 7.0 (8, 4, 8, 8, 8, 6, 3, 6)	Moderate
IF a patient has non-cardiac chest pain suspected due to GERD, THEN diagnostic evaluation should be performed before institution of therapy. <sup>†</sup>	Round 1: 6.5 (6, 7, 4, 1, 3, 7, 7, 9) Round 2: 3.5 (5, 4, 1, 1, 3, 6, 4, 1)	Moderate
<b>Erosive Esophagitis</b>		
IF a patient has LA Grade A esophagitis, THEN further testing should be performed to confirm the presence of GERD.	Round 1: 4.0 (5, 4, 4, 1, 3, 6, 8, 1)	Low
<b>Stricture/Ring</b>		
IF a patient has refractory, complex peptic strictures, THEN injection of intralesional corticosteroids can be used.	Round 1: 7.0 (7, 7, 7, 7, 6, 8, 6, 7)	Low
IF a patient has lower esophageal (Schatzki) rings, THEN treatment with a PPI is suggested following dilation.	Round 1: 7.0 (7, 5, 7, 1, 9, 8, 7, -)	Low
<b>Surgical Therapy</b>		
IF a patient with an esophageal GERD syndrome is responsive to, but intolerant of, acid suppressive therapy, THEN anti-(-)reflux surgery should be recommended as an alternative.	Round 1: 7.5 (6, 8, 7, 9, 7, 9, 9, 6)	High
IF a patient with esophageal GERD syndrome is non-responsive to PPI therapy, THEN surgical therapy should generally not be recommended.	Round 1: 6.0 (5, 6, 9, 1, 5, 7, 8, -)	High
<b>PPI-Refractory Symptoms</b>		
IF a patient with suspected GERD does not respond to empiric trial of twice daily PPI therapy, THEN endoscopy with biopsy of suspicious areas should be performed. <sup>†</sup>	Round 1: 7.0 (6, 7, 7, 1, 8, 9, 7, 9) Round 2: 5.5 (6, 5, 3, 1, 9, 8, 4, 7)	Moderate

<sup>†</sup> Initial proposed measure was reworded during the two rounds; (-) indicates that a score was not submitted; GERD: Gastroesophageal Reflux Disease; PPI Proton Pump Inhibitor; H<sub>2</sub>RA: H<sub>2</sub> Receptor Antagonist.

**Table 5**

GERD quality measures developed via RAM encompass the five existing AHRQ performance measures for GERD.

AHRQ Performance Measures for GERD	Quality Measure Developed by RAM	Validity of Quality Measure Determined by RAM Process (Median Ranking; Dispersion of Rankings)
Percentage of patients seen for an initial evaluation, who were assessed for the presence or absence of the following alarm symptoms: involuntary weight loss, dysphagia, and GI bleeding	IF a patient with a diagnosis of GERD is seen for initial evaluation, THEN the patient should be assessed for the presence or absence of the following alarm symptoms: involuntary weight loss, dysphagia, and GI bleeding.	High Validity (Median ranking 8.5; Strict Agreement)
Percentage of patients seen for an initial evaluation of GERD with at least one alarm symptom who were either referred for upper endoscopy or had an upper endoscopy performed	IF a patient with a diagnosis of GERD has at least one alarm symptom, THEN upper endoscopy should be performed.	High Validity (Median ranking 9.0; Strict Agreement)
Percentage of patients with a diagnosis of GERD or heartburn whose endoscopy report indicates a suspicion of Barrett's esophagus who had a forceps esophageal biopsy performed	IF a patient with GERD has an endoscopy report that indicates a suspicion of Barrett's esophagus, THEN suspicious areas should be biopsied.	High Validity (Median ranking 9.0; Strict Agreement)
Percentage of patients seen for an initial evaluation of GERD who did not have a Barium swallow test ordered	IF a patient has suspected GERD without dysphagia, THEN a barium radiograph should not be used as a diagnostic test.	Moderate Validity (Median ranking 8.0; Relaxed Agreement)
Percentage of patients who have been prescribed continuous PPI or H <sub>2</sub> RA therapy who received an assessment of their GERD symptoms within 12 months	IF a patient with GERD is prescribed maintenance PPI or H <sub>2</sub> RAs, THEN the patient should receive a follow-up assessment of their GERD symptoms at least every 12 months.	High Validity (Median ranking 8.0; Strict Agreement)

Agency for Healthcare Research & Quality (AHRQ); RAND/UCLA Appropriateness Methodology (RAM); Gastroesophageal Reflux Disease (GERD); Proton Pump Inhibitor (PPI); H<sub>2</sub> Receptor Antagonist (H<sub>2</sub>RA).