Upper Endoscopy for Gastroesophageal Reflux Disease & Upper Gastrointestinal Symptoms

Health Technology Assessment Program

Final Evidence Report

April 12, 2012

Upper Endoscopy for Gastroesophageal Reflux Disease (GERD) and Upper Gastrointestinal (GI) Symptoms

April 2012

Center for Evidence-based Policy

Oregon Health & Science University
3455 SW US Veterans Hospital Road
Mailstop SN-4N, Portland, OR 97239-2941
Phone: 503.494.2182
Fax: 503.494.3807
http://www.ohsu.edu/ohsuedu/research/policycenter/med/index.cfm

About the Center for Evidence-based Policy

The Center for Evidence-based Policy (Center) is recognized as a national leader in evidence-based decision making and policy design. The Center understands the needs of policymakers and supports public organizations by providing reliable information to guide decisions, maximize existing resources, improve health outcomes, and reduce unnecessary costs. The Center specializes in ensuring diverse and relevant perspectives are considered, and appropriate resources are leveraged to strategically address complex policy issues with high-quality evidence and collaboration. The Center is based at Oregon Health & Science University in Portland, Oregon.

Nature and Purpose of Technology Assessments

This technology assessment report is based on research conducted by a contracted technology assessment center, with updates as contracted by the Washington State Health Care Authority. This report is an independent assessment of the technology question(s) described based on accepted methodological principles. The findings and conclusions contained herein are those of the investigators and authors who are responsible for the content. These findings and conclusions may not necessarily represent the views of the HCA/Agency and thus, no statement in this report shall be construed as an official position or policy of the HCA/Agency.

The information in this assessment is intended to assist health care decision makers, clinicians, patients and policy makers in making sound evidence-based decisions that may improve the quality and cost-effectiveness of health care services. Information in this report is not a substitute for sound clinical judgment. Those making decisions regarding the provision of health care services should consider this report in a manner similar to any other medical reference, integrating the information with all other pertinent information to make decisions within the context of individual patient circumstances and resource availability.

Acknowledgements

This report was prepared by:

Robyn Liu, MD, MPH
Heidi Kriz, RD, MPH
Aasta Thielke, MPH
Shannon Vandegriff
Valerie King, MD, MPH

Center for Evidence-based Policy
Oregon Health and Science University

Suggested Citation:

Liu, R., Kriz, H., Thielke, A., Vandegriff, S., & King, V. (2012). Upper Endoscopy for Gastroesophageal Reflux Disease (GERD) and Upper Gastrointestinal (GI) Symptoms. Portland, OR: Center for Evidence-based Policy, Oregon Health and Science University.

Table of Contents

Executive Summary

Background

Gastroesophageal Reflux Disease

Gastroesophageal reflux disease (GERD) is the most common outpatient gastrointestinal diagnosis in the United States, with a prevalence of 10% to 58.3% and an annual incidence of 0.38% to 0.45% (Lacy 2010; Sobieraj 2011). The Montreal consensus panel, an international Consensus Group tasked with developing a global definition and classification of GERD, reached strong consensus in defining GERD as "a condition which develops when the reflux of stomach contents causes troublesome symptoms and/or complications" (Vakil 2006, p. 1903). Common symptoms of GERD include heartburn (defined as a burning sensation behind the breastbone), regurgitation and chest pain (Vakil 2006). Obesity; the presence of a hiatal hernia; and the use of estrogen, nitrates, anticholinergics, and tobacco products are considered risk factors for GERD (Lacy 2010). Gastroesophageal reflux disease can lead to a decreased quality of life and to more severe conditions such as esophagitis, Barrett's esophagus, and adenocarcinoma of the esophagus (University of Michigan Health System 2007).

Dyspepsia

Dyspepsia is estimated to range in prevalence in the United States from 2.9 to 34.4% (Sobieraj 2011). The Rome III Committee defines dyspepsia as having one or more of the following symptoms: epigastric pain or burning; postprandial fullness; and/or early satiety (Tack 2006). Other dyspeptic symptoms may include nausea and vomiting, upper abdominal bloating, heart burn, and regurgitation (Goswami 2012). Dyspepsia symptoms are distinguished from GERD as not being "troublesome" enough, referring to the Montreal definition of GERD; however, many authors have used the terms interchangeably.

Diagnostic Procedures

The signs and symptoms of GERD, dyspepsia, and other more severe conditions such as Barrett's esophagus, can be very similar, and diagnostic procedures can be used to establish a diagnosis and rule out other possible conditions. Diagnostic procedures for dyspepsia and GERD can include questionnaires, empiric therapeutic trial, pH monitoring, upper endoscopy, and/or double contrast barium swallow (Lacy 2010).

Key Questions

KQ1. What is the evidence of effectiveness for early treatment strategies that include upper endoscopy compared with empiric medical management?

KQ2. Are there clinical signs and symptoms useful to identify patients for whom early endoscopy is effective to improve health outcomes and/or disease management?

KQ3. For what diagnoses and within what time frames, is repeat endoscopy indicated versus other tests or no follow-up tests for surveillance of disease progression and/or treatment response? Does repeat endoscopy change treatment and outcome?

KQ4. What are the potential harms of performing upper endoscopy in the diagnostic or treatment planning workup of adults with upper GI symptoms? What is the incidence of these harms? Include consideration of progression of treatment in unnecessary or inappropriate ways.

KQ5. What is the evidence that upper endoscopy has differential efficacy or safety issues in sub populations? Including consideration of:

   a. **Gender**
   b. **Age**
   c. **Psychological or psychosocial co-morbidities**
   d. **Other patient characteristics or evidence based patient selection criteria, especially comorbidities of diabetes, high BMI, and chronic ingestion of alcohol**
   e. **Provider type, setting or other provider characteristics**
   f. **Payer / beneficiary type: including worker's compensation, Medicaid, state employees?**

KQ6. What is the evidence of cost and cost-effectiveness of endoscopy compared to other treatment strategies when used in diagnostic or treatment planning workups of adults with upper GI symptoms?

Methods

A systematic review using best evidence methodology was used to search and summarize evidence for Key Questions #1 through #5 as outlined below:

- Completed search of the Medicaid Evidence-based Decisions Project primary evidence sources.

- Existing high quality systematic reviews (SRs) and technology assessments (TAs) summarized for each Key Question.

- If there were two or more comparable SRs or TAs identified and one was more recent, of better quality, or more comprehensive, then the other review(s) were excluded.

- Additional search of the MEDLINE® database completed to identify subsequently published studies. Individual studies published after the search dates of the last high quality review were appraised and synthesized with the results of the high quality SRs.

- If there were no high quality reviews identified, a search, appraisal, and summary of primary individual studies was completed for the last 10 years (January 2002 to January 2012).

Evidence – Inclusion Criteria and Quality Assessment
For this WA HTA report, a search was conducted to identify published SRs and individual studies (from January 2002 to January 2012) in the MEDLINE® database. An additional search using the Medicaid Evidence-based Decisions (MED) Project primary sources was completed to identify SRs and TAs (from January 2002 to January 2012).

Articles were included if the details on the study population incorporated abstractable information about adults presenting with initial complaints of upper gastrointestinal (GI) symptoms, dyspepsia or GERD. For Key Questions #1, #3, #4, and #5, SRs, TAs, meta-analyses, randomized controlled trials (RCTs), and controlled clinical trials or comparative observational studies were included. For Key Question #2, the same articles were included in addition to non-comparative cohort studies and case series. For Key Question #6, all relevant economic evaluations, cost-effectiveness analyses, and economic simulation models were included. Exclusion criteria includes long term treatment of GERD, confirmed Barrett's esophagus diagnosis, wireless capsule endoscopy, previous GI and anti-reflux surgeries, and studies exclusively containing Asian populations.

The methodological quality of the included studies was assessed using standard instruments developed and adapted by the Center for Evidence-based Policy and the MED Project that are modifications of the systems used by the National Institute for Health and Clinical Excellence (NICE) and the Scottish Intercollegiate Guidelines Network (SIGN) (NICE 2009; SIGN 2009). Each study was assigned a rating of good, fair, poor based on its adherence to recommended methods and potential for biases. The methodological quality of the economic studies was rated (good, fair, poor) using a standard instrument developed and adapted by the Center for Evidence-based Policy and the MED Project that is based on modifications of the BMJ (Drummond 1996), the Consensus on Health Economic Criteria list (Evers 2005), and the NICE economic evaluation checklist (NICE 2009). The overall strength of evidence was rated (high, moderate, low) using a modified version of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system (Guyatt 2008).

Guidelines
A search for relevant clinical practice guidelines was conducted using a list of predetermined high quality sources from the MED Project and additional relevant specialty organizations and associations. Guidelines included were limited to those published after 2006. The methodological quality of the guidelines was assessed using an instrument adapted from the Appraisal of Guidelines Research and Evaluation (AGREE) Collaboration (AGREE Next Steps Consortium 2009). Each guideline was assigned a rating of good, fair, poor based on the adherence to recommended methods and the potential for biases.

Policies
At the direction of the WA Health Technology Assesment (WA HTA) program, select payer policies were searched and summarized. Aetna, Blue Cross Blue Shield, GroupHealth, and Medicare National and Local Coverage Determinations (NCD and LCD, respectively) were searched using the payers' websites.

Findings

The MED Project primary sources identified two SRs and TAs and nine economic studies. For the Key Questions, the MEDLINE® search retrieved 1,316 citations, of which one review and seven articles representing seven studies were included.

KQ1. What is the evidence of effectiveness for early treatment strategies that include upper endoscopy compared with empiric medical management?

One good quality systematic review (Delaney 2005) included two separate meta-analyses, one of early endoscopy versus empiric PPI and one of early endoscopy versus test-and-treat for *H. pylori.* The first meta-analysis included five RCT's and found no difference in symptomatic cure at 12 months between endoscopy and PPI arms. The second meta-analysis, also including five RCT's of which one (Duggan 1999) was in the first MA, was first done by pooling trial-level data. This analysis found no difference in effect (RR 0.95, 95% CI 0.79 to 1.15), but a high degree of statistical heterogeneity. When an alternate analysis of these same five studies was done using individual patient data (IPD), there was no longer statistical heterogeneity and a small but statistically significant benefit to upper endoscopy emerged (Peto OR 0.75, 95% CI 0.58 to 0.96; RR 0.95, 95% CI 0.92 to 0.99).

A single fair quality prospective cohort study (Madan 2005) of 70 patients found that 24-hour pH monitoring is the most accurate single diagnostic test for GERD, when a concordance of three separate tests is taken as the gold standard. However, the authors note that there are barriers to its widespread use including invasiveness, cost, and availability. A serial application of an omeprazole challenge test, endoscopy, and finally histopathology achieves a sensitivity of 100% for GERD diagnosis.

Overall, the evidence does not point to a clinically relevant benefit of prompt upper endoscopy over test and treat strategies or empiric PPI therapy for uninvestigated GERD symptoms in the primary care setting.

 Overall strength of evidence: High

KQ2. Are there clinical signs and symptoms useful to identify patients for whom early endoscopy is effective to improve health outcomes and/or disease management?

One good quality systematic review of 57,363 patients in 17 prospective cohort studies (Vakil 2006) found that alarm symptoms, clinical opinion, and computer modeling programs based on symptom questionnaires were all unreliable predictors of gastrointestinal malignancy. Sensitivity ranged from 0% to 83% while specificity varied from 40% to 98%. A good quality prospective cohort study (Marmo 2005) found cancer in 0.9% of patients presenting with uncomplicated dyspepsia (i.e. without alarm symptoms) and the findings suggest that risk is correlated with age >35 for males and >57 for females. A fair quality prospective cohort study

(Rossi 2002) determined that ASGE guideline criteria were poorly correlated with clinically relevant endoscopic findings, although their presence does marginally increase the pre-test probability of endoscopy (from 45% to 47%) and their absence lowers it (from 45% to 29%). A second fair-quality prospective cohort study (Bowrey 2005) in the setting of open-access endoscopy found that 15% of the patients with esophagogastric carcinoma did not present with alarm symptoms and may have suffered delayed diagnosis without early endoscopy; however, there was an unusually high prevalence (3%) of cancer in the study population. Finally, a fair-quality prospective cohort study of primary care patients with uninvestigated dyspepsia (Veldhuyzen van Zanten 2006) found that Barrett's esophagus was most likely in patients who were male, >50 years old, had symptoms of at least 5-10 years duration, and suffered predominantly from reflux.

Vakil and colleagues (2006) suggest that, in the absence of compelling predictors, the concept of "alarm symptoms" should not be abandoned at this time. They suggest age greater than 55 as "the most logical alternative strategy… because the incidence of upper GI malignancy is negligible in Western populations at younger ages and only rises in prevalence above the age of 55 years" (p. 398). Marmo and colleagues (2005) suggest that age should be lower (35) for males and could be higher (57) for females.

 Overall strength of evidence: Moderate

KQ3. For what diagnoses and within what time frames, is repeat endoscopy indicated versus other tests or no follow-up tests for surveillance of disease progression and/or treatment response? Does repeat endoscopy change treatment and outcome?

Only one study, a prospective cohort study of good quality (Westbrook 2005), addressed the question of repeat endoscopy in patients who initially presented with dyspeptic symptoms and had non-malignant endoscopic findings. About a third of these patients underwent a subsequent endoscopy within nine years of the index study. The results of these later endoscopies are not known; however, patients who had further endoscopy were neither more nor less likely than other patients to be symptomatic eight to nine years after the index study (χ^2=0.6, df=1, P > 0.05). Overall, evidence is insufficient to suggest repeat endoscopy to any patients with initial dyspepsia who have non-malignant findings on their index endoscopy.

 Overall strength of evidence: Low

KQ4. What are the potential harms of performing upper endoscopy in the diagnostic or treatment planning workup of adults with upper GI symptoms? What is the incidence of these harms? Include consideration of progression of treatment in unnecessary or inappropriate ways.

Unfortunately there were no studies that could be included for this Key Question. All but one of the systematic reviews, meta-analyses, and economic evaluations (Ford 2005) neglected to factor harms of endoscopy into their reports. According to the authors of one economic evaluation (Spiegel 2002), most harms of endoscopy are cardiorespiratory in nature; that is, related to the procedure sedation rather than the endoscope itself. These authors used a 0.02% incidence of severe harms and modeled their economic assumptions on the surgical repair of perforation.

Our search identified no data on harms associated with empiric acid-suppression or *H. pylori* test-and-treat.

Overall strength of evidence: Insufficient

KQ5. What is the evidence that upper endoscopy has differential efficacy or safety issues in sub populations? Including consideration of:

   g. **Gender**
   h. **Age**
   i. **Psychological or psychosocial co-morbidities**
   j. **Other patient characteristics or evidence based patient selection criteria, especially comorbidities of diabetes, high BMI, and chronic ingestion of alcohol**
   k. **Provider type, setting or other provider characteristics**
   l. **Payer / beneficiary type: including worker's compensation, Medicaid, state employees?**

The search uncovered no evidence related to most of the subpopulations named in this Key Question. Age was the only factor associated with differential effectiveness in one good quality meta-analysis (Ford 2005). The authors of this study performed subgroup analyses based on age, gender, predominant symptom, and presence of *H. pylori*. There was a small but statistically significant benefit of endoscopy in patients 50 years of age and older (RR=0.90, 95% CI 0.82 to 1.00, p < 0.05); no other associations were found. A good-quality prospective cohort study (Marmo 2005) found that patients with malignancy were on average 20 years older than patients without malignancy (p < 0.001). A fair-quality prospective cohort study (Bowrey 2005) also found increasing prevalence of malignancy with rising age. In a good quality economic evaluation simulation model (Barton 2008), relative effectiveness of interventions was the same in hypothetical 30 year olds as in hypothetical 60 year olds. A poor quality retrospective chart review of VA patients (Connor 2004) failed to find any correlation between significant endoscopic findings (Barrett's esophagus and/or erosive esophagitis) and age, gender, race, or NSAID use.

Overall strength of evidence: Moderate (Age), Insufficient (All others)

KQ6. What is the evidence of cost and cost-effectiveness of endoscopy compared to other treatment strategies when used in diagnostic or treatment planning workups of adults with upper GI symptoms?

With the exception of empiric therapy for US 30 year olds, all five good quality studies (Barkun 2010; Barton 2008; Makris 2003; Spiegel 2002; You 2006), one of two fair quality studies (Duggan 2008), and one of three poor quality studies (Klok 2005) favored *H. pylori* test-and-treat as the most cost-effective strategy for adults with uninvestigated symptoms of dyspepsia and/or GERD.

Only two studies, a second-order simulation model and a decision analysis (Barton 2008; Spiegel 2002, respectively), both of good quality, evaluated the cost-effectiveness of different management strategies for new upper gastrointestinal symptoms in a US population. In the Barton study (2008), empiric PPI was the strategy of choice for 30 year old patients, and test-and-treat for *H. pylori* was the most cost-effective intervention for 60 year olds. Spiegel and colleagues (2002) looked only at patients less than 45 years of age, and determined that adding a 6-week trial of PPI to the test-and-treat strategy improved its cost-effectiveness. A good quality economic evaluation of Canadian individual patient data (Barkun 2010) concluded that no one strategy was the most clearly cost-effective, but at a clinically relevant willingness-to-pay threshold of CAN$30,000 to 70,000 per quality adusted life year (QALY), omeprazole treatment based on the CanDys protocol (which incorporates test-and-treat for those without heartburn or reflux as the predominant symptom) was the most cost-effective. Two other good quality models (Makris 2003; You 2006) also favored the test-and-treat approach, along with one fair and one poor quality RCT (Duggan 2008; Klok 2005, respectively).

One fair quality decision analysis (García-Altés 2005) favored a screening questionnaire followed by prompt endoscopy for high-risk patients. Two poor quality RCTs (Giannini 2008; Kjeldsen 2007) found empiric PPI to be the most cost-effective alternative, but did not include comparison to *H. pylori* testing and treatment. There were no economic studies that found prompt endoscopy to be the most cost-effective intervention.

Overall strength of evidence: Moderate

Guidelines

Three guidelines (American Gastroenterological Association (AGA) 2008; American Society for Gastrointestinal Endoscopy (ASGE) 2007a; ASGE 2007b) were identified in relation to the role of endoscopy for the diagnosis and early management of dyspepsia and GERD. An additional guideline specifically related to modifications in endoscopic practice for the elderly was identified from the ASGE (2006).

<u>Role of Endoscopy in the Diagnosis and Management of GERD</u>
The AGA (2008) guideline, rated as good quality, recommends endoscopy to evaluate patients who have not responded to empirical trial of twice-daily PPI therapy and who have suspected esophageal GERD symptoms. The AGA (2008) recommends against routine endoscopy for

patient with GERD for assessment of disease progression, and finds insufficient evidence for routine upper endoscopy to reduce mortality from esophageal cancer.

The ASGE (2007a) guideline, rated as fair quality, recommends that GERD can be diagnosed based on typical symptoms alone without the need for endoscopy, except in patients with alarm symptoms. In addition, the ASGE (2007a) recommends endoscopy for evaluation of patients for the screening of Barrett's esophagus as clinically indicated, for patients with suspected extra-esophageal manifestations of GERD, or with recurrent symptoms after endoscopic or surgical antireflux procedures.

The ASGE (2006) guideline, rated as poor quality, recommends that endoscopy only be conducted in an elderly population when the results will influence clinical management or outcomes. The guideline states that endoscopy preparation does not differ for geriatric populations, and that intensified monitoring may be appropriate for many elderly patients.

Role of Endoscopy in the Diagnosis and Management of Dyspepsia
The ASGE (2007b) guideline, rated as fair quality, recommends that patients between the age 45 to 55 years who have new onset dyspepsia and those who have alarm features should undergo an endoscopy. For patients without alarm features for which there is clinical suspicion of malignancy, the ASGE (2007b) guideline recommends endoscopy should be considered. For patients that are younger than 50 years old and are *H pylori* negative, the ASGE (2007b) guideline recommends an initial endoscopy or a short trial of PPI acid suppression.

Policy Considerations

At the direction of WA HTA, this review searched for Medicare, Aetna, Regence BCBS, and GroupHealth policies addressing coverage of upper endoscopy for patients with symptoms of GERD.   A Medicare NCD for "endoscopy" allows coverage of "endoscopic procedures when reasonable and necessary for the individual patient" (Centers for Medicare and Medicaid Services (CMS) 2012). Medicare contractor LCDs may further define criteria constituting reasonable and necessary use of the procedure. However, there are no relevant LCDs applicable to Washington or the Northwest Region (CMS Region X). Among private payers searched, Aetna is the single payer that outlines coverage criteria for upper endoscopy. Aetna's policy sets forth detailed clinical indications for the use of upper endoscopy in the following categories: high risk screening, diagnostic, therapeutic, and sequential or periodic surveillance. The policy excludes coverage of upper endoscopy as experimental and investigational in several explicit circumstances.

Overall Summary

Evidence

There are a variety of options for initiating workup and treatment of patients presenting with uninvestigated dyspepsia and/or GERD symptoms. A good quality systematic review (Delaney 2005) and a fair quality prospective cohort study (Madan 2005) show that non-invasive strategies, such as an empiric trial of PPI or *H. pylori* test and treat, are equally as effective as prompt endoscopy for achieving symptom improvement. A 24-hour esophageal pH study might

be the gold standard for GERD diagnosis according to the Madan study (2005), but its clinical usefulness is limited by invasiveness, cost, and availability.

There is wide acceptance of the use of "alarm symptoms" such as anemia, dysphagia, and unintentional weight loss to determine patients' need for prompt endoscopy. Patients with these symptoms are excluded from trials of endoscopy for GERD or dyspepsia, on the grounds that they represent a population with a higher-than-normal risk of malignancy. Clinical guidelines invoke these alarm features as indications to bypass empiric treatment or non-invasive testing and move straight to endoscopy. One good quality meta-analysis (Vakil 2006) and one fair-quality prospective case series (Rossi 2002) both agreed that alarm symptoms, as well as clinical opinion, are poor predictors of gastrointestinal malignancy. However, at this time there is no compelling evidence as to what should replace them. Vakil and colleagues (2006) do point out the very low incidence of GI malignancy in Western populations below the age of 55, and note that an age cutoff may be appropriate in formulating a strategy for use of endoscopy.

One meta-analysis that included prespecified subgroup analyses did show that there was a small but statistically significant effect in favor of endscopy for patients aged 50 years and older. Other subgroup analyses based on gender, predominant symptom, and presence of *H. pylori* showed no difference in the effectiveness of endoscopy between groups. A poor quality retrospective cohort study failed to demonstrate any significant associations between meaningful endoscopic findings and patient demographics (e.g., age, race, or gender) or NSAID use.

Patients with findings of malignancy or other serious pathology on endoscopy will be followed up appropriately. But for those whose endoscopic diagnosis was nothing more serious than esophagitis and/or peptic ulcer disease, is there an indication to perform a follow-up endoscopy? One good quality prospective cohort study (Westbrook 2005) followed patients who had presented initially with dyspepsia for eight to nine years after their index endoscopy. More than half of patients had persistent symptoms, and those who had undergone repeat endoscopy (31%) were neither more nor less likely to be symptomatic than those who had not. The study did not, however, assess the findings of these subsequent endoscopies.

There is very little recent data on the harms of upper endoscopy when performed for dyspepsia and/or GERD. The author of one economic evaluation noted that complications are commonly cardiorespiratory (related to sedation), and for purposes of the model used an incidence of severe harms of 0.02%. We found no studies reporting harms associated with empiric acid-suppressing medication or *H. pylori* test-and-treat.

There have been several studies of varying quality that have attempted to determine the most cost-effective means of managing the uninvestigated patient with dyspepsia. Five good quality economic evaluations, along with one of fair quality and one poor quality study, have identified an *H. pylori* test-and-treat strategy as the most cost-effective option. The one exception is a US study that looked at a hypothetical population of 30 year olds and preferred empiric PPI for this younger age group. Two poor quality RCTs also recommended empiric PPI as a more cost-

effective choice than endoscopy. There were no studies that demonstrated prompt endoscopy to be the most cost-effective option.

Guidelines

Four guidelines (one good, two fair, and one poor quality) discuss the role endoscopy in the diagnosis and early management of dyspepsia and GERD. One guideline (AGA 2008) recommends endoscopy to evaluate patients who have not responded to PPI therapy and have suspected GERD symptoms. A fair quality guideline (ASGE 2007a) recommends the use of endoscopy for the screening of Barrett's esophagus as clinically indicated, in patients with recurrent symptoms after endoscopic or surgical antireflux procedures, and/or patients with suspected extra-esophageal manifestations of GERD. One poor quality guideline (ASGE 2006) recommends endoscopy only be conducted in an elderly population when the results will influence clinical management or outcomes, that endoscopy preparation does not differ for geriatric populations, and that intensified monitoring may be appropriate for many elderly patients. The AGA (2008) recommends against routine endoscopy for patient with GERD for assessment of disease progression, and finds insufficient evidence for routine upper endoscopy to reduce mortality from esophageal cancer.

One fair quality guideline (ASGE 2007b) specifically recommends endoscopy in patients between the age 45 to 55 years who have new onset dyspepsia and those who have alarm features should undergo an endoscopy, or for patients without alarm features for whom there is clinical suspicion of malignancy. The ASGE (2007b) guideline recommends either endoscopy or a short trial of PPI acid suppression for patients who are younger than 50 years old and are *H. pylori* negative.

Policies

This review identified two payers, Medicare and Aetna, with policies addressing coverage of upper endoscopy for patients with symptoms of GERD. Medicare has issued a general NCD for "endoscopy" allowing coverage of "endoscopic procedures when reasonable and necessary for the individual patient" (CMS 2012). There are no LCDs applicable to Washington or the Northwest Region that further define criteria constituting "reasonable and necessary" use of the procedure. Among private payers, Aetna has issued a policy setting forth detailed clinical indications for the use of upper endoscopy in the following categories: high risk screening, diagnostic, therapeutic, and sequential or periodic surveillance. The policy excludes coverage of upper endoscopy as experimental and investigational in several explicit circumstances.

Discussion and Limitations of the Evidence

Upper endoscopy for diagnosis of GERD and other upper gastrointestinal symptoms is a thorny topic, subject to many sources of imprecision and potential bias. First, there is the problem of defining which symptoms are indicative of gastro-esophageal reflux disease, and which are dyspepsia. In a primary care office setting, patients are rarely clear-cut members of one category or the other. Second, there is the question of practice setting. Some studies look only at patients in primary care, while others include patients in a specialty referral setting such as an endoscopy center. Depending upon the health care

system, patients may or may not be able to self-refer into these specialty centers. Therefore it becomes unclear whether patients in a primary care setting and those in a specialty center are in fact comparable populations.

There is not a consensus on how outcomes should be measured in patients who are treated for dyspepsia or GERD. Several symptom scoring tools exist, some of which are validated. When data are pooled into meta-analyses, these symptom scores are necessarily dichotomized into "cured" versus "not cured," or "improved" versus "not improved." A cost-utility analysis that converts these symptom scores into QALYs is one further step removed from the actual patient experience.

Economic modeling studies and cost-effectiveness analyses in this report came to a consensus around one type of intervention as being generally the most cost-effective (test-and-treat), and initial endoscopy as being less cost-effective. Using incremental cost-effectiveness ratios (ICERs) allows for some degree of comparison across multiple nations whose health care costs may be defined in radically different ways.

Background

Gastroesophageal reflux disease
Gastroesophageal reflux disease (GERD) is the most common outpatient gastrointestinal diagnosis in the United States, with a prevalence of 10% to 58.3% and an annual incidence of 0.38% to 0.45% (Lacy 2010; Sobieraj 2011). Over 60 million Americans experience GERD-related heartburn once a month and over 25 million experience GERD-related heartburn on a daily basis (University of Michigan Health Systems (UMHS) 2007). The Montreal consensus panel, an international Consensus Group tasked with developing a global definition and classification of GERD, reached strong consensus in defining GERD as "a condition which develops when the reflux of stomach contents causes troublesome symptoms and/or complications" (Vakil 2006, p. 1903). Under this definition, symptoms, as determined by the patient in clinical practice, are considered troublesome when they adversely affect an individual's well-being; reflux symptoms that are not troublesome by the patient should not be diagnosed as GERD (Vakil 2006).

Common symptoms of GERD include heartburn (defined as a burning sensation behind the breastbone), regurgitation and chest pain (Vakil 2006). Physiologic causes of GERD include an increased number of transient lower esophageal sphincter relaxations, ineffective esophageal motility, and reduced lower esophageal sphincter tone (Lacy 2010). Obesity; the presence of a hiatal hernia; and the use of estrogen, nitrates, anticholinergics, and tobacco products are considered risk factors for GERD (Lacy 2010). Gastroesophageal reflux disease can lead to a decreased quality of life and to more severe conditions such as esophagitis, Barrett's esophagus, and adenocarcinoma of the esophagus (UMHS 2007).

Dyspepsia
Dyspepsia is estimated to range in prevalence in the United States from 2.9 to 34.4% (Sobieraj 2011). The Rome III Committee defines dyspepsia as having one or more of the following symptoms: epigastric pain or burning; postprandial fullness; and/or early satiety (Tack 2006). Other dyspeptic symptoms may include nausea and vomiting, upper abdominal bloating, heart burn, and regurgitation (Goswami 2012). Dyspepsia symptoms are often distinguished from GERD as not being "troublesome" enough, referring to the Montreal definition of GERD; however, many authors have used the terms interchangeably.

Diagnostic Procedures
The signs and symptoms of GERD, dyspepsia, and other more severe conditions such as Barrett's esophagus, can be very similar, and diagnostic procedures can be used to establish a diagnosis and rule out other possible conditions. Diagnostic procedures for GERD vary according to specificity and sensitivity of testing accuracy. Additionally, some diagnostic procedures may allow for additional testing, such as a biopsy, if warranted. This section gives an overview of the different diagnostic procedures possible for patients with symptoms of GERD.

> ***Questionnaires*** – questionnaires have been used to improve the accuracy of diagnosing GERD symptoms. Many of the questionnaires have limited sensitivity and specificity,

and tend to lack cross-cultural validity. Currently, there is no "gold standard" questionnaire (Lacy 2010).

Empiric Therapeutic Trial – Proton pump inhibitor (PPI) response is often used by clinicians in detecting the presence or absence of GERD (Lacy 2010). A trial of PPI typically includes twice daily dosing for four weeks (Kahrilas 2008). A daily dose of 40 to 80mg of omeprazole is the most common PPI regimen used in clinical empiric therapy studies (Lacy 2010). The sensitivity and specificity of this PPI test ranges from 62 to 92% and 36 to 100%, respectively (Lacy 2010).

pH Monitoring – pH monitoring as a diagnostic procedure refers to a broad range of tests that monitor the reflux of acid between the stomach and the esophagus through the peristalsis and contractile pressures in the esophagus (Lacy 2010). This procedure places a thin plastic catheter through the nostril and down the back of the throat, into the esophagus (Lacy 2010). The catheter, a sixteenth of an inch in diameter, has an acid sensor at its tip and is placed directly above the lower esophageal sphincter (Lacy 2010). Several other types of pH monitoring exist including ambulatory pH probes; ambulatory impedance monitoring; and capsule (wireless) pH monitoring (Lacy 2010).

Upper Endoscopy – this diagnostic procedure, also called esophagogastroduodenoscopy (EGD), uses a thin scope with a light and camera at its tip to look inside of the upper digestive tract. Endoscopy tends to have a lower sensitivity and specificity than pH monitoring, although has a high specificity (95%) for diagnosing Barrett's esophagus (Lacy 2010). During the endoscopic procedure, tissue samples of the digestive tract can be obtained through a biopsy (Lacy 2010). The use of newer types of endoscopy for diagnosing GERD, such as narrow band imaging, chromoendoscopy, confocal endomicroscopy, capsule endoscopy, and ultra-thin, unsedated transnasal endoscopy care are considered controversial due to the lack of comparison with other validated tests (Lacy 2010).

Double Contrast Barium Swallow – this diagnostic procedures scans for signs of reflux during the examination and looks for morphologic evidence of reflux esophagitis (Lacy 2010). Double-contrast barium swallow is not considered a primary diagnostic procedure for GERD although it may be useful to define the anatomy of the esophagus or to identify complications of gastroesophageal reflux (Lacy 2010).

Washington State Agency Utilization Data

Upper Endoscopies (UE) were captured for all outpatient and Ambulatory Surgery Centers (ASCs) using the CPT codes 43200-43259. Only charges directly related by code were captured. For context, selected cases were reviewed: an average primary payer cost for just the upper endoscopy codes was $1,268, with all "day of procedure" charges totaling $1,521. From this, it appears that our conservative data extraction method underestimates the UE charges by about 20%. When all data for "day of procedure" was evaluated, about 25% of charges related to additional procedures (colonoscopies, mammograms, etc).

Identification of UE for Gastroesophageal Reflux Disease (GERD): diagnosis codes potentially related to GERD and GI symptoms were reviewed by the Washington State Agency Medical Directors, and a set of codes was agreed upon. The GERD/GI symptom diagnosis codes were further divided into two groups, one for more objective diagnoses, and the other for more general symptoms. These codes and categorizations are listed at the end of the State Agency data section.

Figure 1a: PEB Upper Endoscopies Compared to Total Population

	2007	2008	2009	2010
PEB Total Population	172,009	204,804	210,501	213,487
% of Total Population w/ a GERD Dx	14.0%	13.9%	14.0%	13.6%
% of Total Population having UE	2.7%	2.7%	2.9%	2.8%
% of Total Pop. having UE for GERD	1.5%	1.5%	1.5%	1.4%

Figure 1b: PEB GERD Endoscopy Patient Counts Compared to All Endoscopy Counts

PEB Patient counts	2007	2008	2009	2010	4 year overall*
All patients w/ GERD	24,035	28,529	29,546	29,050	76,900
All GERD Dx Upper Endoscopies	2,531	2,997	3,196	3,077	11,801
% of all GERD patients receiving UE	10.5%	10.5%	10.8%	10.6%	15.3%

*4 year overall patient counts show unique patients over 4 years, and therefore are not the total of annual counts.

Figure 1c: Medicaid Upper Endoscopies Compared to Total Population

	2007	2008	2009	2010
Medicaid Total FFS Population	378,915	392,808	416,871	424,230
% of Total Population w/ GERD Dx	15.1%	15.1%	15.3%	15.1%
% of Total Population having UE	2.1%	2.0%	2.3%	2.7%

% of Total Pop. having UE for GERD	1.1%	1.1%	1.2%	1.4%

Figure 1d: Medicaid GERD Endoscopy Patient Counts Compared to All Upper Endoscopy Counts

Medicaid Patient Counts	2007	2008	2009	2010	4 year overall
All patients w/ GERD	57,332	59,268	63,851	63,994	17,5561
All GERD Dx Upper Endoscopies	4,093	4,199	5,016	6,031	19,339
% of all GERD patients receiving UE	7.1%	7.1%	7.9%	9.4%	11.0%

*4 year overall patient counts show unique patients over 4 years, and therefore are not the total of annual counts.

Figure 1e: L&I GERD Endoscopy Patient Counts Compared to All Upper Endoscopy Counts

L&I Patient Counts	2007	2008	2009	2010	4 year overall*
All patients w/ GERD	1,234	1,163	1,099	1,039	4,220
% of GERD Diagnosis Claimants having UE	7.05%	7.48%	8.37%	6.83%	32.44%
All GERD Dx Upper Endoscopies	46	46	51	32	175
% of all GERD patients receiving UE	3.73%	3.96%	4.64%	3.08%	4.15%

*4 year overall patient counts show unique patients over 4 years, and therefore are not the total of annual counts.

Figure 2: All Agency Summary: Upper Endoscopy with GERD/Upper GI Symptoms

PEB Upper Endoscopies with GERD Diagnoses	2007	2008	2009	2010	4 year overall*
Total Payments	$1,576,355	$2,058,633	$2,363,815	$2,277,442	$8,276,245
Patient Count	2,578	3,087	3,366	3,335	12,366
Max per Procedure	$4,896	$4,677	$4,964	$6,030	$6,030
Average per Procedure†	$611	$667	$702	$683	$669
Average per Procedure (primary payer only)‡	$872	$912	$978	$953	$933
Minimum per Procedure (primary payer only)	$18	$23	$29	$4	$4
Median Payment	$631	$723	$748	$661	$711
Standard Deviation	$482	$546	$592	$601	$563

Medicaid Upper Endoscopies with GERD Diagnoses	2007	2008	2009	2010	4 year overall*
Total Payments	$1,215,982	$1,297,634	$1,640,671	$1,772,311	$5,926,598
Patient Count	4,093	4,199	5,016	6,031	19,339
Max per Procedure	$3,221	$4,896	$3,469	$3,604	$4,896
Average per Procedure‡	$297	$309	$327	$294	$306
Minimum per Procedure	$0	$0	$0	$0	$0
Median Payment	$276	$286	$300	$220	$276
Standard Deviation	$267	$276	$292	$303	$287

L&I Upper Endoscopies with GERD Diagnoses	2007	2008	2009	2010	4 year overall*
Total Payments	$34,577	$33,466	$36,548	$20,837	$125,429
Patient Count	46	46	51	32	175
Max per Procedure	$3,407	$1,606	$3,139	$1,679	$3,407
Average per Procedure‡	$752	$728	$717	$651	$717
Minimum per Procedure	$37	$39	$114	$236	$37
Median Payment	$799	$855	$813	$511	$813
Standard Deviation	$591	$365	$506	$409	$479

*Four year patient counts are unique patients over the period and are therefore lower than the sum of annual patient counts..

†Includes procedures where PEB was the secondary payer.

‡Based on UE CPT code charges only, so approximately 20% lower than actual average payment per procedure.

Figure 3a: PEB Payments and Patient Counts by Age and Gender

	Payments by Age and Gender					Patient Counts by Age and Gender				
Age	**2007**	**2008**	**2009**	**2010**	**Grand Total**	**2007**	**2008**	**2009**	**2010**	**4 year overall***
0-17	$36,575	$58,964	$69,086	$97,838	$262,463	46	62	60	81	241
18-34	$135,941	$181,758	$221,201	$202,233	$741,132	149	184	209	204	715
35-49	$364,960	$470,361	$520,055	$488,514	$1,843,890	393	509	503	503	1,776
50-64	$849,704	$1,090,858	$1,252,106	$1,162,832	$4,355,501	971	1,186	1,298	1,210	4,212
65-79	$166,438	$229,230	$270,091	$296,909	$962,668	706	802	921	970	2,947
80+	$22,686	$27,272	$30,823	$28,963	$109,744	145	171	166	174	557
Total	**$1,576,305**	**$2,058,442**	**$2,363,363**	**$2,277,288**	**$8,275,398**	**2,410**	**2,914**	**3,157**	**3,142**	**1,0448**
% Female	**2007**	**2008**	**2009**	**2010**	**Grand Total**	**2007**	**2008**	**2009**	**2010**	**4 year overall***
0-17	59.6%	62.0%	52.3%	51.8%	55.3%	54.3%	59.7%	53.3%	59.3%	58.1%
18-34	62.5%	70.8%	61.5%	60.4%	63.6%	62.4%	67.9%	61.7%	60.8%	63.8%
35-49	63.6%	62.8%	63.3%	62.4%	63.0%	63.1%	62.5%	63.0%	64.2%	63.1%
50-64	60.6%	59.3%	61.1%	61.0%	60.5%	61.2%	60.0%	61.0%	61.2%	61.1%
65-79	46.3%	49.1%	45.2%	47.3%	47.0%	53.4%	58.2%	53.1%	56.3%	55.7%
80+	41.5%	51.1%	41.2%	57.0%	47.9%	44.8%	50.3%	50.6%	52.9%	51.0%
Total	**59.7%**	**60.0%**	**59.3%**	**59.0%**	**59.4%**	**58.2%**	**59.9%**	**58.4%**	**59.6%**	**59.5%**

Figure 3b: Medicaid Payments and Patient Counts by Age and Gender

	Payments by Age and Gender					Patient Counts by Age and Gender				
Age	**2007**	**2008**	**2009**	**2010**	**Grand Total**	**2007**	**2008**	**2009**	**2010**	**4 year overall***
0-17	$54,432	$58,818	$95,343	$101,793	$310,387	155	167	238	259	819
18-34	$149,952	$202,667	$229,545	$260,202	$842,366	451	554	607	707	2,319
35-49	$397,924	$410,614	$544,541	$549,633	$1,902,713	1,225	1,211	1,474	1,614	5,524
50-64	$537,683	$540,828	$706,873	$802,164	$2,587,549	1,668	1,700	2,045	2,542	7,955
65-79	$68,744	$67,058	$55,455	$48,267	$239,524	492	434	531	749	2,206
80+	$7,246	$17,650	$8,913	$9,857	$43,665	102	133	121	171	527
Total	**$1,215,982**	**$1,297,634**	**$1,640,671**	**$1,771,916**	**$5,926,203**	**4,093**	**4,199**	**5,016**	**6,042**	**19,350**
% Female	**2007**	**2008**	**2009**	**2010**	**Grand Total**	**2007**	**2008**	**2009**	**2010**	**4 year overall***
0-17	59.6%	62.0%	52.3%	51.8%	55.3%	54.3%	59.7%	53.3%	59.3%	58.1%
18-34	62.5%	70.8%	61.5%	60.4%	63.6%	62.4%	67.9%	61.7%	60.8%	63.8%
35-49	63.6%	62.8%	63.3%	62.4%	63.0%	63.1%	62.5%	63.0%	64.2%	63.1%
50-64	60.6%	59.3%	61.1%	61.0%	60.5%	61.2%	60.0%	61.0%	61.2%	61.1%
65-79	46.3%	49.1%	45.2%	47.3%	47.0%	53.4%	58.2%	53.1%	56.3%	55.7%
80+	41.5%	51.1%	41.2%	57.0%	47.9%	44.8%	50.3%	50.6%	52.9%	51.0%
Total	**59.7%**	**60.0%**	**59.3%**	**59.0%**	**59.4%**	**58.2%**	**59.9%**	**58.4%**	**59.6%**	**59.5%**

Figure 4a: PEB Patient Counts, All Upper Endoscopy vs GERD Diagnosis UE, 2007-2010

PEB Patient Counts for UE, 2007-2010

	2007	2008	2009	2010	2011
All Upper Endoscopies (UE)	4662	5569	6010	5998	
All GERD Diagnosis UE	2578	3087	3366	3335	
GERD % of all UE	55.3%	55.4%	56.0%	55.6%	

9.0%/yr avg growth

9.3%/yr avg
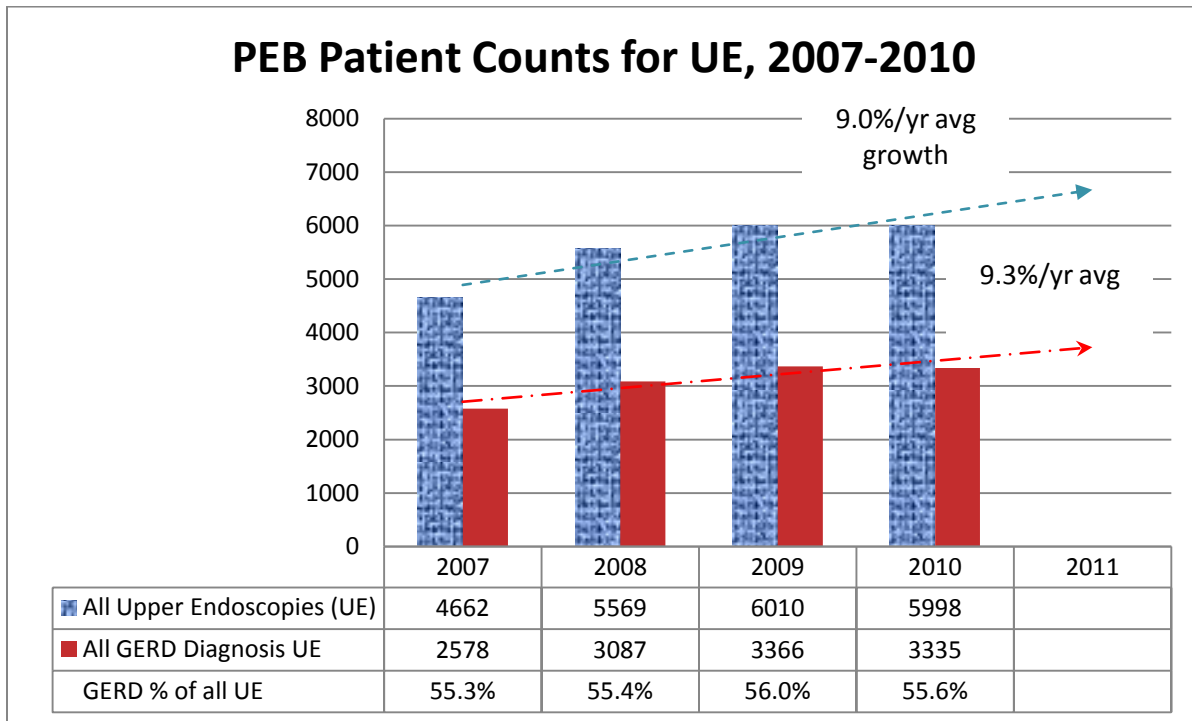
Figure 4b: Medicaid Patient Counts, All Upper Endoscopy vs GERD Diagnosis UE, 2007-2010

Medicaid Patient Counts for UE, 2007-2010

	2007	2008	2009	2010	2011
All Upper Endoscopies (UE)	7794	7899	9457	11481	
All GERD Diagnosis UE	4093	4199	5016	6031	
GERD % of all UE	52.5%	53.2%	53.0%	52.5%	

14.2%/yr avg growth

14.1%/yr avg growth
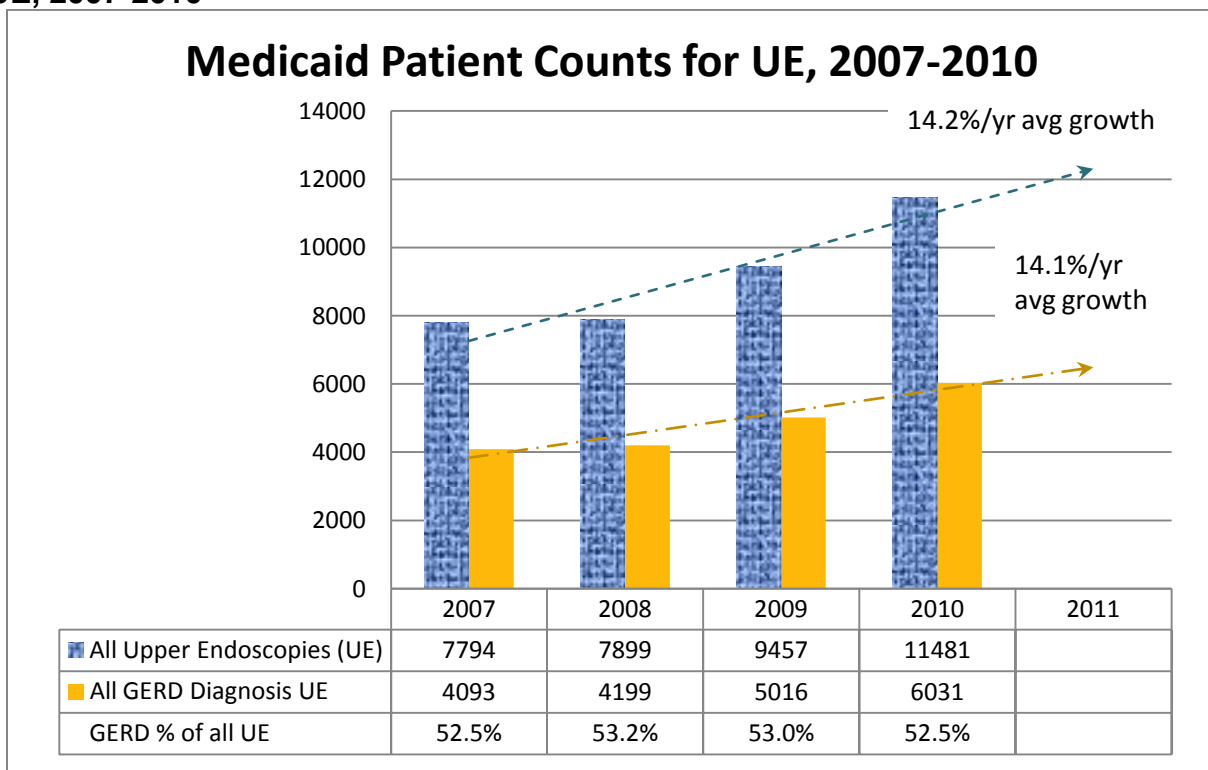
Figure 5a: PEB Payments for All Upper Endoscopy vs GERD Diagnosis UE, 2007-2010

PEB Payments for UE, 2007-2010

	2007	2008	2009	2010	
All Upper Endoscopies (UE)	$2,707,776	$3,568,862	$4,135,016	$4,083,934	
All GERD Dx endoscopies	$1,576,355	$2,058,633	$2,363,815	$2,277,442	
% GERD in all UE	58.2%	57.7%	57.2%	55.8%	

15.5%/yr avg growth

13.9%/yr
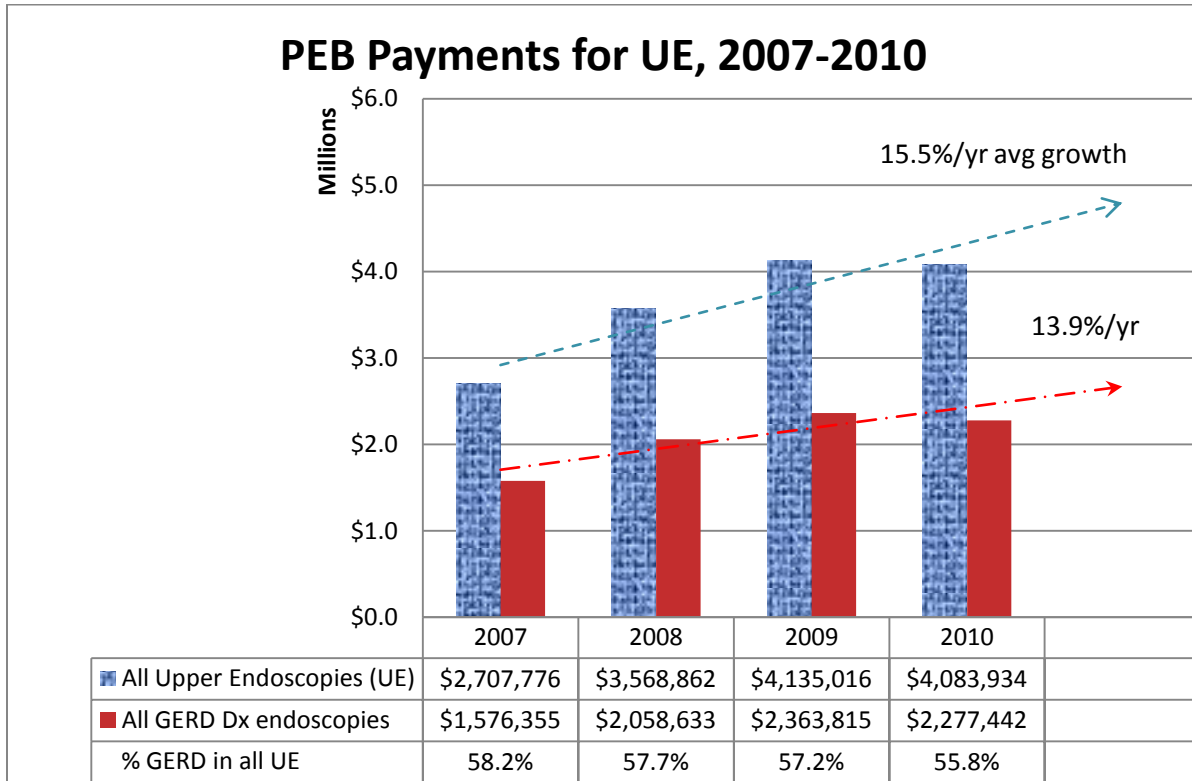
Figure 5b: Medicaid Payments for All Upper Endoscopy vs GERD Diagnosis UE, 2007-2010

Medicaid Payments for UE, 2007-2010

	2007	2008	2009	2010	2011
All Upper Endoscopies (UE)	$2,299,776	$2,361,653	$2,980,410	$3,250,317	
All GERD Diagnosis UE	$1,215,982	$1,297,634	$1,640,671	$1,772,311	
GERD % of all UE	52.9%	54.9%	55.0%	54.5%	

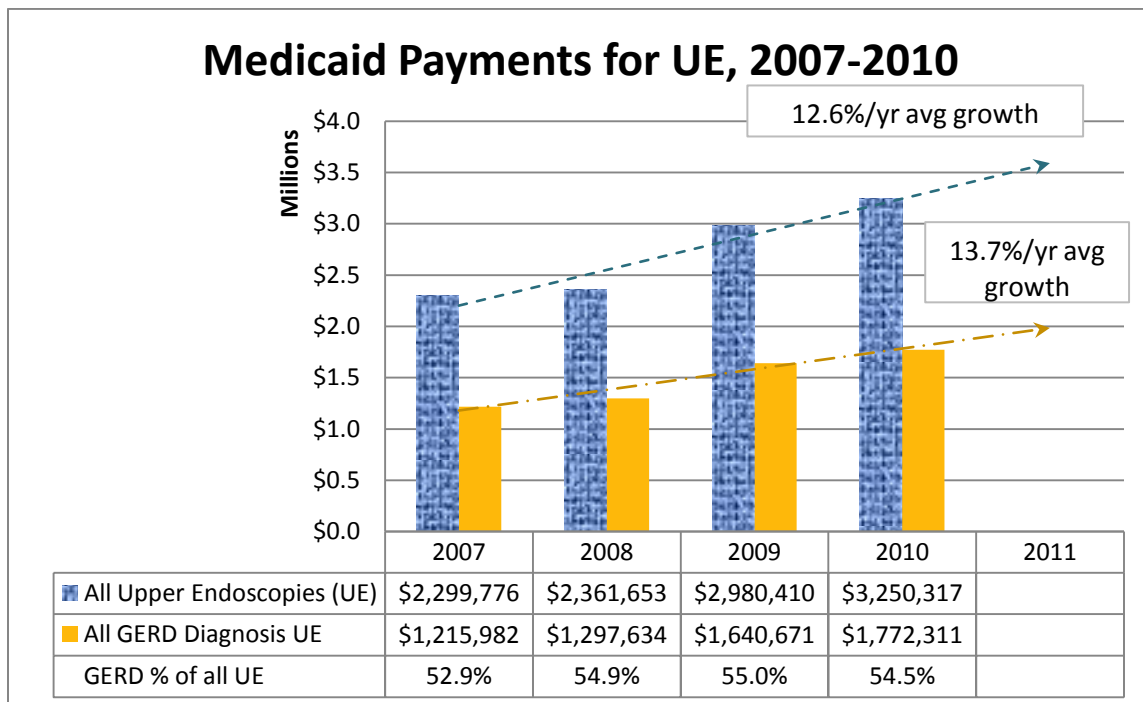12.6%/yr avg growth

13.7%/yr avg growth

Figure 6a: PEB Top 10 Diagnosis codes for GERD UEs, Descending by 4 year overall payments

PEB GERD Diagnoses	Diag Code	2007	2008	2009	2010	Grand Total	% Total
ESOPHAGEAL REFLUX	530.81	$248,766	$387,451	$448,739	$350,709	**$1,435,666**	**17.3%**
BARRETT'S ESOPHAGUS	530.85	$192,269	$224,967	$279,074	$284,288	**$980,598**	**11.8%**
DYSPHAGIA NOS	787.20	$41,876	$180,474	$253,757	$323,591	**$799,698**	**9.7%**
STOMACH FUNCTION DIS NEC	536.8	$157,345	$195,668	$215,724	$190,094	**$758,831**	**9.2%**
REFLUX ESOPHAGITIS	530.11	$117,279	$158,795	$197,673	$162,435	**$636,182**	**7.7%**
ESOPHAGEAL STRICTURE	530.3	$122,047	$157,034	$177,774	$157,093	**$613,948**	**7.4%**
GSTR/DDNTS NOS W/O HMRHG	535.50	$122,096	$140,961	$128,968	$147,132	**$539,157**	**6.5%**
ABDMNAL PAIN EPIGASTRIC	789.06	$91,566	$135,111	$141,850	$146,014	**$514,540**	**6.2%**
OTH SPF GSTRT W/O HMRHG	535.40	$75,011	$90,246	$91,263	$83,551	**$340,071**	**4.1%**
ESOPHAGITIS, UNSPECIFIED	530.10	$45,576	$56,474	$79,365	$73,479	**$254,894**	**3.1%**

Figure 6b: Medicaid Top 10 Diagnosis codes for GERD UEs, Descending by 4 year overall payments

Medicaid GERD Diagnoses	Diag Code	2007	2008	2009	2010	Grand Total	% Total
ESOPHAGEAL REFLUX	530.81	$159,156	$156,727	$219,874	$230,105	**$765,862**	**12.9%**
DYSPHAGIA NOS	787.2	$43,406	$154,363	$209,826	$209,679	**$617,274**	**10.4%**
REFLUX ESOPHAGITIS	530.11	$103,112	$105,167	$157,656	$171,217	**$537,152**	**9.1%**
STRICTURE AND STENOSIS OF ESOPHAGUS	530.3	$100,167	$99,616	$146,421	$165,560	**$511,765**	**8.6%**
ABDMNAL PAIN EPIGASTRIC	789.06	$98,018	$120,847	$131,383	$139,204	**$489,452**	**8.3%**
UNSPECIFIED GASTRITIS AND GASTRODUODENITIS	535.5	$93,573	$99,149	$113,033	$136,981	**$442,737**	**7.5%**
BARRETT'S ESOPHAGUS	530.85	$67,263	$81,999	$89,444	$123,967	**$362,673**	**6.1%**
DYSPEPSIA AND OTHER SPECIFIED DISORDERS OF FUNCTION OF STOMACH	536.8	$79,992	$78,099	$108,509	$92,375	**$358,975**	**6.1%**
OTHER SPECIFIED GASTRITIS	535.4	$73,458	$89,584	$78,411	$98,227	**$339,680**	**5.7%**
ESOPHAGEAL REFLUX	530.1	$33,176	$44,115	$73,905	$79,769	**$230,964**	**3.9%**

Figure 7a: All PEB GERD DX Upper Endoscopies, General Symptoms vs Objective Diagnoses, Highest to Lowest Total Payments 2007-2010

All PEB GERD Dx UE, Diagnosis Category	2007	2008	2009	2010	Grand Total	% of Category Total
General Symptoms Top Diagnoses	**$770,611**	**$972,457**	**$1,084,923**	**$1,159,927**	**$3,987,919**	
DYSPHAGIA NOS	$41,876	$180,474	$253,757	$323,591	$799,698	20.1%
STOMACH FUNCTION DIS NEC	$157,345	$195,668	$215,724	$190,094	$758,831	19.0%
GSTR/DDNTS NOS W/O HMRHG	$122,096	$140,961	$128,968	$147,132	$539,157	13.5%
ABDMNAL PAIN EPIGASTRIC	$91,566	$135,111	$141,850	$146,014	$514,540	12.9%
OTH SPF GSTRT W/O HMRHG	$75,011	$90,246	$91,263	$83,551	$340,071	8.5%
HEARTBURN	$39,776	$47,225	$56,684	$56,694	$200,379	5.0%
ABDMNAL PAIN UNSPCF SITE	$22,771	$39,322	$49,609	$51,567	$163,270	4.1%
DEL - DYSPHAGIA	$151,753				$151,753	3.8%
DYSPHAGIA NEC	$11,387	$25,404	$31,697	$35,500	$103,987	2.6%
CHEST PAIN NOS	$19,996	$27,768	$34,562	$19,423	$101,750	2.6%
DYSPHAGIA,PHARYNGOESOPH	$5,044	$35,897	$19,369	$13,781	$74,090	1.9%
ABDMNAL PAIN GENERALIZED	$8,229	$8,504	$15,525	$27,353	$59,611	1.5%
ESOPHAGEAL DISORDER NOS	$3,387	$9,380	$21,469	$23,535	$57,770	1.4%
ABDMNAL PAIN OTH SPCF ST	$6,501	$10,744	$5,796	$16,045	$39,085	1.0%
OTHER DSRDERS ESOPHAGUS	$8,151	$13,718	$3,674	$9,671	$35,215	0.9%
CHEST PAIN NEC	$2,819	$8,343	$7,573	$8,819	$27,554	0.7%
PERSISTENT VOMITING	$2,657	$841	$4,424	$883	$8,806	0.2%
DYSPHAGIA, ORAL PHASE		$1,303	$245	$3,968	$5,516	0.1%
DYSPHAGIA, OROPHARYNGEAL		$686	$856	$820	$2,362	0.1%
GSTR MCSL HYPRT W/O HMRG		$469	$986	$851	$2,306	0.1%
DYSPHAGIA, PHARYNGEAL		$191	$893		$1,084	0.0%
STOMACH FUNCTION DIS NOS	$246			$636	$882	0.0%
ALCHL GASTRTIS W/O HMRHG		$203	$0		$203	0.0%
Objective Diagnosis	**$805,743**	**$1,086,176**	**$1,278,892**	**$1,117,514**	**$4,288,326**	
ESOPHAGEAL REFLUX	$248,766	$387,451	$448,739	$350,709	$1,435,666	33.5%
BARRETT'S ESOPHAGUS	$192,269	$224,967	$279,074	$284,288	$980,598	22.9%
REFLUX ESOPHAGITIS	$117,279	$158,795	$197,673	$162,435	$636,182	14.8%
ESOPHAGEAL STRICTURE	$122,047	$157,034	$177,774	$157,093	$613,948	14.3%
ESOPHAGITIS, UNSPECIFIED	$45,576	$56,474	$79,365	$73,479	$254,894	5.9%
ACUTE GASTRTIS W/O HMRHG	$44,233	$52,868	$53,420	$54,329	$204,850	4.8%
OTHER ESOPHAGITIS	$19,574	$29,385	$28,447	$22,266	$99,673	2.3%
ULC ESOPHAGUS W/O BLEED	$11,250	$11,915	$11,355	$9,313	$43,833	1.0%
ACUTE ESOPHAGITIS	$3,359	$4,249	$3,028	$3,124	$13,759	0.3%
ULCER ESOPHAGUS W BLEED	$644	$2,202	$16	$478	$3,340	0.1%
DEL - ESOPHAGITIS	$745	$836			$1,582	0.0%
Grand Total	**$1,576,355**	**$2,058,633**	**$2,363,815**	**$2,277,442**	**$8,276,245**	

Figure 7b: All Medicaid GERD DX Upper Endoscopies, General Symptoms vs Objective Diagnoses, Highest to Lowest Total Payments 2007-2010

All Medicaid GERD UE Dx, by Diagnosis Category	2007	2008	2009	2010	Grand Total	% of Cat. Total
General	**$644,969**	**$718,244**	**$851,851**	**$892,558**	**$3,107,622**	
DYSPHAGIA NOS	$43,406	$154,363	$209,826	$209,679	$617,274	19.9%
ABDMNAL PAIN EPIGASTRIC	$98,018	$120,847	$131,383	$139,204	$489,452	15.8%
UNSPECIFIED GASTRITIS	$93,573	$99,149	$113,033	$136,981	$442,737	14.2%
DYSPEPSIA AND OTHER SPECIFIED DISORDERS	$79,992	$78,099	$108,509	$92,375	$358,975	11.6%
OTHER SPECIFIED GASTRITIS	$73,458	$89,584	$78,411	$98,227	$339,680	10.9%
ABDMNAL PAIN UNSPCF SITE	$41,301	$42,737	$49,220	$63,268	$196,526	6.3%
HEARTBURN	$28,742	$30,380	$50,135	$39,365	$148,623	4.8%
DYSPHAGIA NOS	$141,497		$9		$141,507	4.6%
ABDMNAL PAIN GENERALIZED	$11,839	$10,778	$23,422	$20,983	$67,022	2.2%
ABDMNAL PAIN OTHER, MULTI	$6,469	$14,923	$14,789	$15,490	$51,672	1.7%
DYSPHAGIA NEC	$1,259	$16,990	$11,346	$12,417	$42,012	1.4%
UNSPECIFIED DISORDER OF ESOPHAGUS	$1,845	$4,347	$11,780	$20,348	$38,320	1.2%
DYSPHAGIA,PHARYNGOESOPH	$838	$18,313	$10,953	$8,158	$38,262	1.2%
OTHER ESOPHAGEAL	$6,210	$8,814	$4,651	$12,539	$32,214	1.0%
CHEST PAIN NOS	$6,980	$11,787	$7,725	$4,681	$31,172	1.0%
CHEST PAIN NEC	$3,550	$6,200	$9,334	$7,074	$26,158	0.8%
DYSPHAGIA, OROPHARYNGL	$624	$5,452	$6,632	$4,201	$16,909	0.5%
PERSISTENT VOMITING	$3,972	$1,300	$6,592	$4,725	$16,589	0.5%
UNSPECIFIED FUNCTIONAL DISORDER OF STOMACH	$746	$825	$1,916	$1,786	$5,273	0.2%
DYSPHAGIA, PHARYNGEAL		$999	$1,571	$553	$3,123	0.1%
DYSPHAGIA, ORAL	$470	$877	$612	$160	$2,119	0.1%
ALCOHOLIC GASTRITIS		$1,398		$154	$1,552	0.0%
ALCOHOLIC GASTRITIS, WITHOUT HEMORRHAGE	$181	$81		$189	$451	0.0%
Objective	**$571,013**	**$579,390**	**$788,820**	**$879,753**	**$2,818,976**	
ESOPHAGEAL REFLUX	$159,156	$156,727	$219,874	$230,105	$765,862	27.2%
REFLUX ESOPHAGITIS	$103,112	$105,167	$157,656	$171,217	$537,152	19.1%
STRICTURE/ STENOSIS OF ESOPH	$100,167	$99,616	$146,421	$165,560	$511,765	18.2%
BARRETT'S ESOPHAGUS	$67,263	$81,999	$89,444	$123,967	$362,673	12.9%
ESOPHAGITIS, UNSPECIFIED	$33,176	$44,115	$73,905	$79,769	$230,964	8.2%
ACUTE GASTRITIS	$57,380	$49,897	$45,942	$40,985	$194,204	6.9%
OTHER ESOPHAGITIS	$29,871	$25,855	$32,939	$43,455	$132,120	4.7%
ULCER OF ESOPHAGUS	$14,422	$9,756	$17,180	$17,361	$58,719	2.1%
ACUTE ESOPHAGITIS	$3,535	$5,121	$4,823	$6,109	$19,588	0.7%
ULCER OF ESOPHAGUS W/BLEEDING	$2,930	$1,138	$636	$1,205	$5,909	0.2%
ACUTE GASTRITIS				$21	$21	0.0%
Grand Total	**$1,215,982**	**$1,297,634**	**$1,640,671**	**$1,772,311**	**$5,926,598**	

Figure 8a: PEB Payments for General Symptoms vs Objective Diagnosis, Endoscopies with a GERD Diagnosis 2007-2010
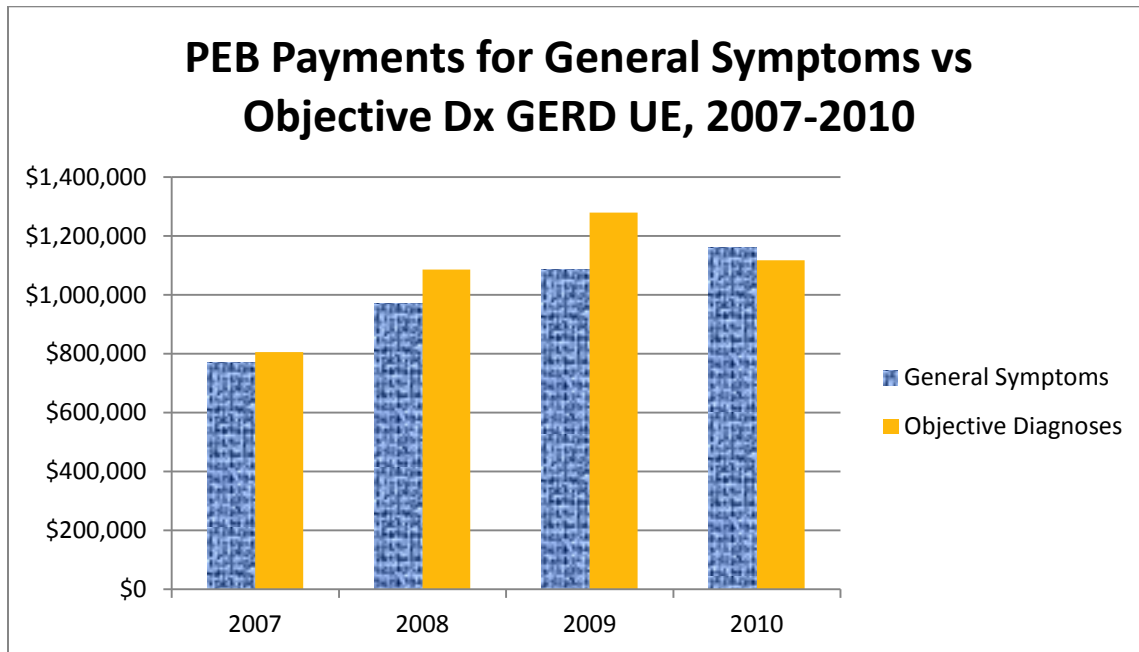

Figure 8b: PEB Patient Counts for General Symptoms vs Objective Diagnosis, Endoscopies with a GERD Diagnosis 2007-2010
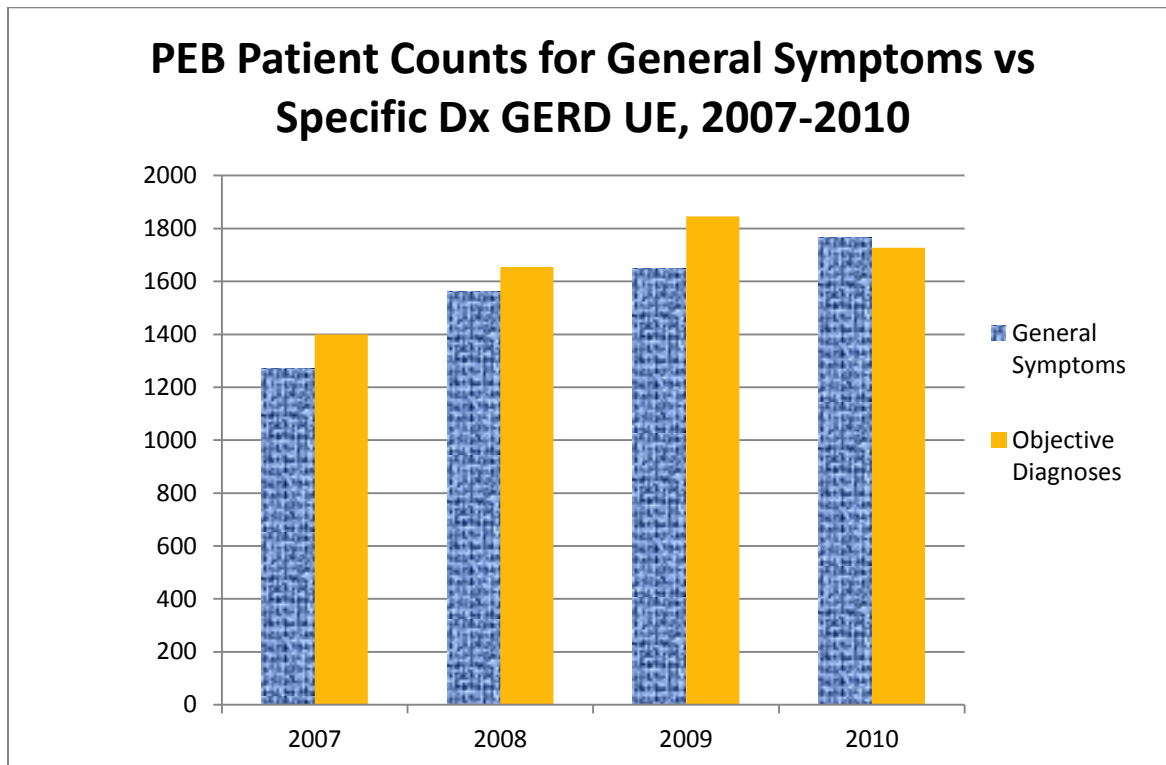
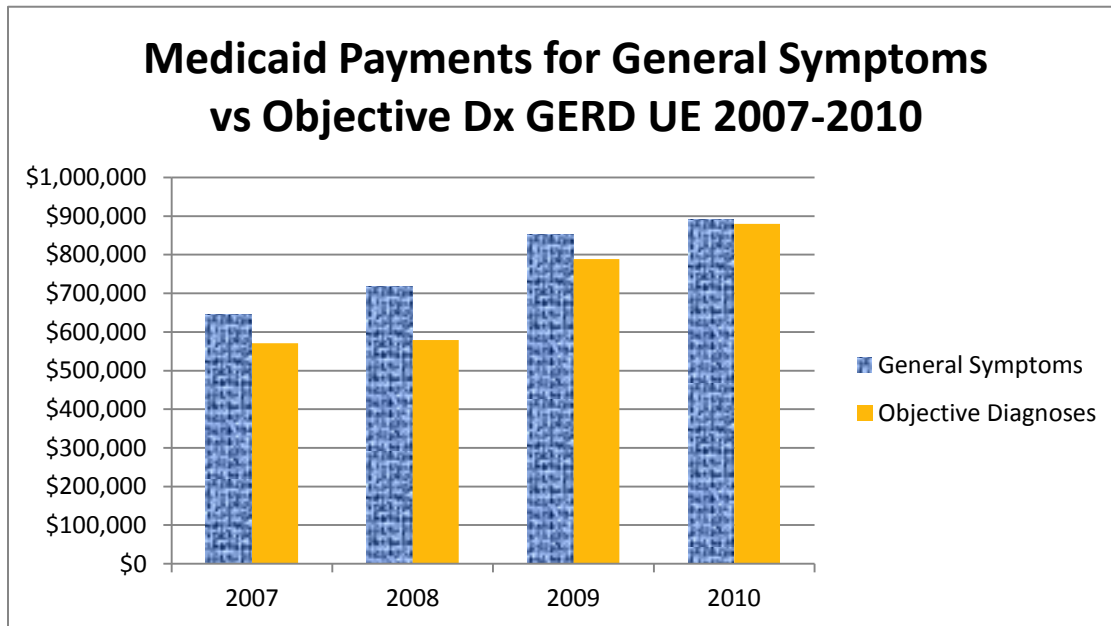**Figure 8c: Medicaid Payments for General Symptoms vs Objective Diagnosis, Endoscopies with a GERD Diagnosis 2007-2010**

Medicaid Payments for General Symptoms vs Objective Dx GERD UE 2007-2010

Figure 8d: Medicaid Patient Counts for General Symptoms vs Objective Diagnosis, Endoscopies with a GERD Diagnosis 2007-2010

Medicaid Patient Counts for General Symptoms vs Objective Dx GERD UE 2007-2010
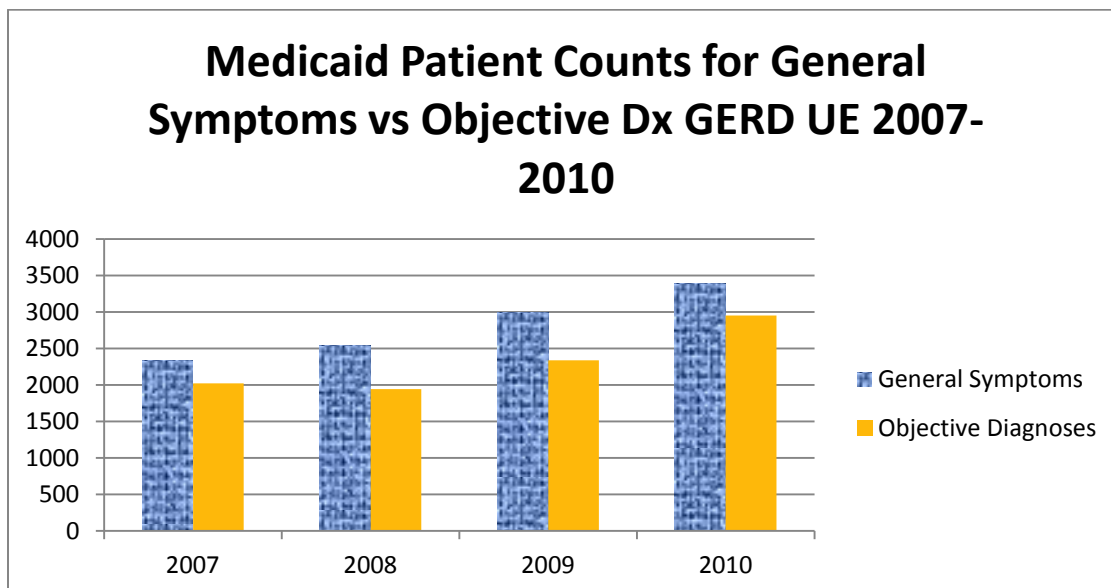
Figure 9a: Repeated procedures: PEB GERD Diagnosis Upper Endoscopy

- Maximum UE procedures for one patient = 16
- 1496 (of total 10305) patients had a repeat endoscopy with a GERD Diagnosis in 4 years (14.5%)
- Of those who had repeats, 1156 of 1496 had only one (77%)
- In those patients who had repeated GERD Diagnosis UE, each had an average of 1.73 UE per person
- Repeats averaged 397 days between.

PEB Patients w/ Repeated Endoscopies w/ UE Diagnoses

Procedures	Patients
16	1
12	1
10	2
9	1
8	4
7	8
6	9
5	19
4	62
3	231
2	1156

Figure 9b: Repeated procedures: Medicaid GERD Diagnosis Upper Endoscopy

- Maximum UE procedures for one patient = 17
- 538 (of total 19339) patients had a repeat endoscopy with a GERD Diagnosis in 4 years (4.5%)
- Of those who had repeats, 350 of 538 had only one (65%)
- In those patients who had repeated GERD Diagnosis UE, each had an average of 1.9 UE per person
- Repeats averaged 190 days between.

Medicaid Patients with Repeated Endoscopies w/ UE Diagnoses

Procedures	Patients
17	1
16	1
15	1
14	4
10	8
9	4
8	6
7	13
6	15
5	10
4	42
3	83

| | 2 | 350 | |

Related Medical Codes		
	Diagnosis Codes likely to indicate GERD	Diagnosis codes Symptom Classification
530.1	Esophagitis	Objective Diagnosis
530.11	Reflux esophagitis	Objective Diagnosis
530.12	Acute esophagitis	Objective Diagnosis
530.19	Other esophagitis	Objective Diagnosis
530.2	Ulcer of esophagus	Objective Diagnosis
530.21	Ulcer of esophagus with bleeding	Objective Diagnosis
530.3	Stricture and stenosis of esophagus	Objective Diagnosis
530.81	Esophageal reflux	Objective Diagnosis
530.85	Barrett's esophagus	Objective Diagnosis
530.89	Other	General Symptoms
530.9	Unspecified disorder of esophagus	General Symptoms
535	Acute gastritis	Objective Diagnosis
535.0	Acute gastritis, without mention of hemorrhage	Objective Diagnosis
535.2	Alcoholic gastritis	General Symptoms
535.3	Alcoholic gastritis, without mention of hemorrhage	General Symptoms
535.4	Other specified gastritis	General Symptoms
535.5	Unspecified gastritis and gastroduodenitis	General Symptoms
536.2	Persistent vomiting	General Symptoms
536.8	Dyspepsia and other specified disorders of function of stomach	General Symptoms
536.9	Unspecified functional disorder of stomach	General Symptoms
786.5	CHEST PAIN NOS	General Symptoms
786.59	CHEST PAIN NEC	General Symptoms
787.1	HEARTBURN	General Symptoms
787.2	DYSPHAGIA NOS	General Symptoms
787.21	DYSPHAGIA, ORAL	General Symptoms
787.22	DYSPHAGIA, OROPHARYNGEAL	General Symptoms
787.23	DYSPHAGIA, PHARYNGEAL	General Symptoms
787.24	DYSPHAGIA,PHARYNGOESOPH	General Symptoms
787.29	DYSPHAGIA NEC	General Symptoms
789	ABDMNAL PAIN UNSPCF SITE	General Symptoms
789.06	ABDMNAL PAIN EPIGASTRIC	General Symptoms
789.07	ABDMNAL PAIN GENERALIZED	General Symptoms
789.09	ABDMNAL PAIN OTHER, MULTI	General Symptoms

Methods

A systematic review using best evidence methodology was used to search and summarize evidence for Key Questions #1 through #5 as outlined below:

- Completed search of the Medicaid Evidence-based Decisions (MED) Project primary evidence sources.

- Existing high quality systematic reviews (SRs) and technology assessments (TAs) summarized for each Key Question.

- If there were two or more comparable SRs or TAs identified and one was more recent, of better quality, or more comprehensive, then the other review(s) were excluded.

- Additional search of the MEDLINE® database completed to identify subsequently published studies. Individual studies published after the search dates of the last high quality review were appraised and synthesized with the results of the high quality SRs and TAs.

- If there were no high quality reviews identified, a search, appraisal, and summary of primary individual studies was completed for the last 10 years (January 2002 to January 2012).

Evidence

Search strategy

For this WA Health Technology Assessment (WA HTA) report, a search was conducted to identify published SRs and individual studies (from January 2002 to January 2012) in the MEDLINE® database. The MEDLINE® search strategy is provided in Appendix A. A list of excluded studies with reasons for exclusion is provided in Appendix B. An additional search using the MED Project primary sources was completed to identify SRs and TAs. The primary sources searched included: Cochrane Library (Wiley Interscience), UK National Institute for Health and Clinical Excellence (NICE), Blue Cross/Blue Shield Health Technology Assessment (HTA) program, Veterans Administration TA program, BMJ Clinical Evidence, the Canadian Agency for Drugs and Technologies in Health (CADTH), and the Agency for Health Research and Quality (AHRQ).

Inclusion Criteria
- Articles were included if they were peer reviewed English-language publications.

- Included abstractable information about adults presenting with initial complaints of upper gastrointestinal (GI) discomfort, dyspepsia or GERD.

- Study Design:
  - For Key Questions #1, #3 to #5, SRs, TAs, meta-analyses, randomized controlled trials (RCTs), and controlled clinical trials or comparative observational studies;
  - For Key Question #2, SRs, TAs, meta-analyses, RCTs, cohorts, and case series; and

     o For Key Question #6, economic evaluations, cost-effectiveness analyses, and economic simulation models.

Exclusion Criteria

- Long-term treatment for GERD

- Confirmed Barrett's Esophagus diagnosis

- Wireless capsule endoscopy

- Previous GI and anti-reflux surgeries

- Exclusively Asian populations

Quality Assessment - Evidence

The methodological quality of the included studies was assessed using standard instruments developed and adapted by the Center for Evidence-based Policy and the MED Project that are modifications of the systems used by NICE and the Scottish Intercollegiate Guidelines Network (SIGN) (NICE 2009; SIGN 2009). All studies were assessed by two independent and experienced raters. In cases where there was not agreement about the quality of the study or guideline the disagreement was resolved by conference or the use of a third rater. Quality assessment checklists are provided in Appendix G.

The overall strength of evidence was rated using a modified version of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system (Guyatt 2008). Each study was assigned a rating of good, fair, poor, based on its adherence to recommended methods and potential for biases. In brief, good quality SRs included a clearly focused question, a literature search that was sufficiently rigorous to identify all relevant studies, criteria used to select studies for inclusion (e.g., RCTs) and assess study quality, and assessments of heterogeneity to determine if a meta-analysis would be appropriate. Good quality RCTs clearly described the population, setting, intervention and comparison groups; randomly allocated patients to study groups; concealed allocation; had low dropout rates; and reported intention-to-treat analyses. Good quality SRs and RCTs also had low potential for bias from conflicts of interest and funding source. Fair quality SRs and RCTs had incomplete information about methods that might mask important limitations. Poor quality SRs and RCTs had clear flaws that could introduce significant bias.

A summary judgment for the overall quality of evidence was assigned to each Key Question and outcome (Guyatt 2008). The GRADE system defines the quality of a body of evidence for an outcome in the following manner:

- High: Further research is *very unlikely* to change our confidence in the estimate of effect. Typical sets of studies would be large RCTs without serious limitations.

- Moderate: Further research is *likely* to have an important impact on our confidence in the estimate of effect and may change the estimate. Typical sets of studies would be RCTs with some limitations or well-performed observational studies with additional strengths that guard against potential bias and have large estimates of effects.

- Low[1]: Further research is *very likely* to have an important impact on our confidence in the estimate of effect and is likely to change the estimate. Typical sets of studies would be RCTs with very serious limitations or observational studies without special strengths.

Quality Assessment – Economic studies

The methodological quality of the studies was assessed using a standard instrument developed and adapted by the Center for Evidence-based Policy and the MED Project that is based on modifications of the BMJ (Drummond 1996), the Consensus on Health Economic Criteria list (Evers 2005), and the NICE economic evaluation checklist (NICE 2009). In brief, good quality economic evaluations include a well described research question with economic importance and detailed methods to estimate the effectiveness and costs of the intervention. A sensitivity analysis is provided for all important variables and the choice and values of variables are justified. Good quality economic evaluations also have low potential for bias from conflicts of interest and funding sources. Fair quality economic evaluations have incomplete information about methods to estimate the effectiveness and costs of the intervention. The sensitivity analysis may not consider one or more important variables, and the choice and values of variables are not completely justified. All of these factors might mask important study limitations. Poor quality economic evaluations have clear flaws that could introduce significant bias. These could include significant conflict of interest, lack of sensitivity analysis, or lack of justification for choice of values and variables. All studies were assessed by two independent and experienced raters. In cases where there was not agreement about the quality of the study the disagreement was resolved by conference or the use of a third rater. The economic evaluation checklist is provided in Appendix G.

Guidelines

Search Strategy

A search for relevant clinical practice guidelines (CPGs) was conducted, using the following sources: the National Guidelines Clearinghouse database, the Institute for Clinical Systems Improvement (ICSI), SIGN, NICE, the Veterans Administration/Department of Defense (VA/DOD) guidelines, US Preventive Services Task Force (USPSTF), Australian National Health and Medical Research Council, New Zealand Guidelines Group, and the Center for Disease Control and Prevention (CDC). Included guidelines were limited to those published after 2006.

Quality Assessment

The methodological quality of the guidelines was assessed using an instrument (Appendix G) adapted from the Appraisal of Guidelines Research and Evaluation (AGREE) Collaboration (AGREE Next Steps Consortium 2009). The guidelines were rated by two individuals. A third rater was used to obtain consensus if there were disagreements. Each guideline was assigned a rating of good, fair, poor based on its adherence to recommended methods and potential for biases. A guideline rated as good quality fulfilled all or most of the criteria. A fair quality guideline fulfilled some of the criteria and those criteria not fulfilled were thought unlikely to

---

[1] The MED Project collapses the low and very low GRADE categories because they usually have the same policy implications.

alter the recommendations. If no or few of the criteria were met, the guideline was rated as poor quality.

Policies
At the direction of the WA HTA program, select payer policies were searched and summarized. Aetna, Regence Blue Cross Blue Shield, GroupHealth, and Medicare National and Local Coverage Determinations (NCD and LCD, respectively), were searched using the payers' respective websites.

Findings

The MED Project primary sources identified two SRs and TAs and nine economic studies. For the Key Questions, the MEDLINE® search retrieved 1,316 citations, of which one review and seven articles representing seven studies were included.

KQ1. What is the evidence of effectiveness for early treatment strategies that include upper endoscopy compared with empiric medical management?

One good quality systematic review (Delaney 2005) and one fair quality prospective cohort study (Madan 2005) were identified for this Key Question.

Systematic reviews and technology assessments
One good quality systematic review (Delaney 2005) was identified in the search of core sources. Delaney and colleagues (2005) completed a Cochrane analysis of management strategies, including initial investigation and empirical treatments, for people with dyspepsia. Participants were patients presenting either to primary care or to an endoscopy unit, without prior endoscopic findings. The authors use a broadly inclusive definition of dyspepsia that includes "heartburn, epigastric pain, bloating, discomfort, early satiety, water brash or loss of appetite" (p. 4), noting that "[w]e are considering uninvestigated patients at the community level where there is considerable overlap between epigastric and heartburn symptoms" (p. 3).

Delaney and colleagues (2005) identified 25 randomized controlled trials that met inclusion criteria. They performed systematic reviews of five different comparisons, two of which are relevant to this report ("initial endoscopy versus acid suppression" and "*H. pylori* test and treat versus endoscopy").

Initial endoscopy versus acid suppression
Delaney and colleagues (2005) included five studies that compared prompt endoscopy against initial pharmacological therapy, and data from four of the trials (N=1125) was able to be pooled. One study (Goodson 1989) identified by Delaney and colleagues (2005) looked at "early investigation" with a barium meal study versus acid suppression, and is not considered in their report. One of the five endoscopy studies (Laheij 1998) was excluded from meta-analysis because it reported only symptom-free days rather than a global measure of improvement. The meta-analysis of the remaining four trials showed no statistically significant difference in relative risk (RR) (RR 0.89, 95% CI 0.77 to 1.02) for symptom improvement between the two interventions.

H. pylori test and treat versus endoscopy
Delaney and colleagues (2005) included five trials that compared a test for *H. pylori* with subsequent eradication therapy, as indicated, against prompt endoscopy. Three studies included patients who had been referred to an endoscopy center by GP physicians (Heaney 1999; Lassen 1998; McColl 2002), while a fourth randomized patients in the primary care setting (Arents 2003). An unpublished study (Duggan 1999) compared test and treat with both prompt endoscopy and empirical PPI in patients presenting to primary care. Finally, the authors

included individual patient data from a small primary care study (Myres 2002) that failed to adequately recruit and was terminated. Individual data was obtained from all studies except Heaney (1999) (p. 10). All studies used a symptom score to assess outcome, but four different tools were used and these were therefore dichotomized in the meta-analysis into "improved" versus "not improved."

Pooling data at the trial level, Delaney and colleagues (2005) found no statistically significant difference in outcomes between the two strategies (RR 0.95, 95% CI 0.79 to 1.15). There was high heterogeneity, due to the inclusion of the Duggan (1999) study, which showed a significant benefit to endoscopy-based management. This was addressed with the individual patient data (IPD) meta-analysis of the 1,924 participants, which eliminated the heterogeneity and showed a small but statistically significant benefit of endoscopy-based management (Peto OR 0.75, 95% CI 0.58 to 0.96; RR 0.95, 95% CI 0.92 to 0.99).

RCTs
No randomized controlled trials were identified in our search beyond those used in the above meta-analyses.

Other study designs (e.g. cohort studies)
Our search identified one fair quality prospective cohort study (Madan 2005) addressing the sensitivity and specificity of six different modalities for diagnosis of GERD. Madan and colleagues (2005) performed a series of six diagnostic tests on 70 patients between the ages of 18 and 80 presenting with heartburn and/or reflux for at least two days per week for six weeks. Following clinical evaluation and determination of eligibility for the study, patients ceased all use of proton pump inhibitors (PPIs) and H2-blockers for one week. Patients then underwent upper endoscopy with biopsy and histological evaluation. Upon completion of endoscopy, the subjects took omeprazole 40mg every morning and 20mg every evening for one week. A symptom score reduction of at least 50% was considered a "positive" omeprazole challenge test (OCT). Patients again stopped acid-suppressing medication for one week before undergoing the final two tests, GER scintigraphy and 24-hour esophageal pH monitoring.[2] The authors note that "this is probably the first study that has compared all available diagnostic modalities in a large cohort of patients under similar conditions," noting that the most recent such trial prior to theirs was in 1978 (Madan 2005, p. 36).

The Madan study (2005) considered a concordance of three out of the six tests to be a "true positive" case of GERD. The authors then calculated sensitivity and specificity for each individual test against this gold standard. Parameters for the studies most relevant to this report are given in Table 1 below.

---

[2] Gastrointestinal scintigraphy is a procedure whereby radioopaque material (technetium-99m sulfur colloid) is orally administered and the patient is placed supine under a gamma camera. Serial images are obtained to determine the number of reflux events, duration, and proximal extent of the reflux. Esophageal pH monitoring involves the placement of a thin plastic catheter nasogastrically to position a sensor just above the lower esophageal sphincter. The sensor records each reflux of acid, which is registered by a recorder attached to the nasal end of the catheter.

Table 1. Accuracy of diagnostic tests when gold standard taken as concordance of three or more tests (adapted from Madan 2005)

Test	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)	Positive likelihood ratio	Negative likelihood ratio
Omeprazole challenge test	84.4	56	77.5	66.6	1.9	0.29
Endoscopy	64.4	84	87.8	56.7	4	0.43
Histology	82.2	60	78.7	65.1	2.06	0.3
pH monitoring	77.7	92	94.5	69.6	9.75	0.24

As a single test, 24-hour pH monitoring had the highest level of diagnostic accuracy (true positive plus true negative tests). This also held true when only endoscopy-negative reflux disease (ENRD) patients were considered. The authors note, however, that the sensitivity of pH monitoring is lower than typically desired for a gold-standard test, as almost a quarter of patients with GERD will have normal pH readings. The authors also note that pH monitoring is not as widely available or inexpensive as some of the other tests. A sequential combination of OCT, endoscopy and histology was found to have a sensitivity of 100%. The authors conclude therefore that serial use of these three tests (OCT followed by endoscopy for those who test negative, with histology performed on those without visible endoscopic evidence of GERD) should be the standard diagnostic workup, and that "pH monitoring is required only to confirm or exclude GERD in doubtful cases" (Madan 2005, p. 35).

Overall summary, quality and limitations of the evidence

Systematic Review
One good quality systematic review (Delaney 2005) included two separate meta-analyses, one of early endoscopy versus empiric PPI and one of early endoscopy versus test-and-treat for *H. pylori.* The first meta-analysis found no difference in symptomatic cure at 12 months between endoscopy and PPI arms. The second meta-analysis was first done by pooling trial-level data and found no difference in effect but a high degree of heterogeneity. When an alternate analysis using individual patient data was done, heterogeneity was eliminated and a small but statistically significant benefit to upper endoscopy emerged (Peto OR 0.75, 95% CI 0.58 to 0.96; RR 0.95, 95% CI 0.92 to 0.99).

A limitation of the first meta-analysis performed by Delaney and collegeaues (2005) is that, in one of the included studies (Bytzer 1994), patients were recruited after having been referred by a GP physician to an endoscopy center. This would bias the results in favor of endoscopy, since patients who are referred out of a GP's office have a higher likelihood of pathology. This same study also failed to treat the patients found to have peptic ulcer disease on endoscopy with *H.*

pylori eradication therapy, as was done in the other three studies. This would bias the results in the opposite direction for this study, since patients undergoing endoscopy did not receive the same treatment as endoscopy patients in other trials. Finally, the authors note that two trials (Duggan 1999; Lewin 1999a) were not yet peer-reviewed at the time of their inclusion into this review and should be taken with caution.

There are several possible explanations for the finding that endoscopy was more effective than a test-and-treat strategy in the IPD meta-analysis. Unblinded studies are known to overestimate effect. Three of the five included studies allowed patients with positive endoscopic findings to be treated with *H. pylori* eradication therapy. Endoscopy patients used more PPIs than test-and-treat patients, and this may account for the improvement in symptoms. Finally, this meta-analysis may have simply amplified a bias present in the smaller studies.

Cohort Study
A single fair quality prospective cohort study (Madan 2005) of 70 patients found that 24-hour pH monitoring is the most accurate single diagnostic test for GERD, when a concordance of three separate tests is taken as the gold standard. However, the authors note that there are barriers to its widespread use including invasiveness, cost, and availability. When making clinical decisions, the most "accurate" test may not be the best choice, depending upon the goals of physician and patient. In the case of GERD, sensitivity may be more important than overall accuracy, since the goal of the work-up is to minimize false negatives and treat all cases of GERD appropriately. The authors conclude that a serial application of an omeprazole challenge test, endoscopy, and finally histopathology achieves a sensitivity of 100% for GERD diagnosis.

There are several limitations to this cohort study. It should be noted that, although histopathology was treated as a separate test in the calculations of this study, in the real clinical setting it will only be done in the setting of and in conjunction with endoscopy and never entirely on its own. Second, the authors did not perform any testing for *H. pylori* in their study population. It is likely that some of the subjects would have tested positive and may have responded to eradication therapy, eliminating the need for subsequent endoscopic investigation. The study population is probably not representative of the typical US population presenting for initial evaluation in primary care, as evidenced by the high prevalence of Barrett's esophagus in the study population (4 of 70). Finally, this study did not report confidence intervals for its measures of test performance.

Overall, the evidence does not point to a clinically relevant benefit of prompt upper endoscopy over test and treat strategies or empiric PPI therapy for uninvestigated GERD symptoms in the primary care setting.

 Overall strength of evidence: High

KQ2. Are there clinical signs and symptoms useful to identify patients for whom early endoscopy is effective to improve health outcomes and/or disease management?

One good quality meta-analysis (Vakil 2006), one good-quality prospective cohort study (Marmo 2005), and three fair quality prospective cohort studies (Bowrey 2005, Rossi 2002, Veldhuyzen van Zanten 2006) were identified for this Key Question.

Systematic reviews and technology assessments
The search strategy identified one good quality meta-analysis of 17 prospective cohort studies (Vakil 2006) relevant to this Key Question. Vakil and colleagues (2006) analyzed data from 17 studies to determine the diagnostic performance of alarm symptoms (e.g., weight loss, dysphagia, and anemia), computer models based on symptom questionnaires, and professional clinical opinion in the prediction of gastrointestinal malignancy. Studies were included if their populations were greater than or equal to 16 years of age; if they evaluated at least 100 patients; if they collected data prospectively; if dyspepsia and alarm symptoms were recorded along with endoscopic diagnosis; and if the entire study diagnosed at least one GI malignancy.

A total of 57,363 patients were included in the meta-analysis of 17 studies. A total of 458 (0.8%) of the patients had cancer. The performance statistics for the investigated predictors of GI malignancy are given in Table 2. Overall, the sensitivity of "alarm symptoms" varied from 0% to 83% with considerable heterogeneity between studies [note: studies with sensitivity equal to 0% were excluded from pooled analyses and are not reflected in Table 2]. Anemia was the least sensitive indicator, with a pooled sensitivity of 13% with no significant heterogeneity. However, anemia was highly specific (Sp=95%, 95% CI 92% to 97%). Clinical opinion also had a low sensitivity (Sn=29%, 95% 10% to 88%), but a high specificity (Sp=97%, 95% CI 96% to 99%).

The most sensitive predictor was computer modeling, with a sensitivity of 96% (95% CI 92% to 100%) with no significant heterogeneity. Specificity of computer modeling was low at 34% (95% CI 27% to 44%). The presence of one or more alarm symptom, which is a criterion used in many treatment guidelines, was only 67% sensitive for GI malignancy (95% CI 54% to 83%) with heterogeneity; and 66% specific (95% CI 55%, 79%) with a high degree of heterogeneity. Vakil and colleagues (2006) conclude that "alarm features, clinical diagnoses, and computer models are relatively inaccurate predictors of an underlying malignancy" (p. 396).

RCTs
No randomized controlled trials were identified in our search.

Table 2. Performance of Predictors for GI Malignancy (adapted from Vakil 2006)

	N; number of studies	Sensitivity (95% CI)	Specificity (95% CI)	+ LR for malignancy (95% CI)	Heterogeneity present?
Clinical opinion	n=3159; 3 studies	29% (10%, 88%)	97% (96%, 99%)	Not Reported	Not reported
Computer models based on symptom questionnaires	n=8043; 5 studies	96% (92%, 100%)	34% (27%, 44%)	1.49 (1.33, 1.67)	No, $Q=4.4_4$ and P=0.36 for sens; Yes, $Q=170_4$ and P<0.0001 for spec
Weight loss	n=48499; 8 studies	49% (37%, 65%)	84% (81%, 87%)	Not reported	Yes, $Q=34.0_7$ and P<0.001 for sens; $Q=143_7$ and P<0.001 for spec
Dysphagia	n=9646; 5 studies	39% (23%, 66%)	85% (78%, 92%)	Not reported	Yes, $Q=19.5_4$ and P<0.001 for sens; $Q=852_4$ and P<0.001 for spec
Anemia	n=42,327; 4 studies	13% (8%, 20%)	95% (92%, 97%)	Not reported	No for sens; Yes for spec with $Q=77_2$ and P<0.001
≥1 alarm feature	n=46,161; 7 studies	67% (54%, 83%)	66% (55%, 79%)	2.74 (1.47, 5.24)	Yes, $Q=7.6_4$ and P=11 for sens; $Q=1779_4$ and P<0.001 for spec.
All approaches	n=57,363; 17 studies	Pooled DOR 7.49 (4.37, 12.8)		Not reported	Yes; $\chi^2=44_{16}$, P<0.001

Other study designs (e.g. cohort studies)
The search identified one good and three fair quality prospective case series (Marmo 2005; Bowrey 2006, Rossi 2002, Veldhuyzen van Zanten 2006) of relevance to this Key Question. Rossi and colleagues (2002) assessed 1,777 consecutive patients presenting to an "open-access" upper endoscopy facility in Northern Italy for the presence of American Society for Gastrointestinal Endoscopy (ASGE)-approved indications, based on the current ASGE guideline at the time. The authors then correlated the presence of these indications with "clinically relevant" endoscopic findings. Those findings deemed to be clinically relevant are listed below:

- Erosive gastritis, esophagitis, or duodenitis;
- Duodenal ulcer;
- Barrett's esophagus;

- Gastric ulcer;
- Gastric neoplasms;
- Esophageal varices;
- Esophageal stenosis;
- Esophageal neoplasms; and
- Gastric varices.

Endoscopy was appropriate by ASGE guidelines in 84.4% of cases. The most common indication was dyspepsia (53.5%; 27% of patients with dyspepsia were greater than 45 years of age). Diagnosis of clinically relevant disease occurred in 47.4% of patients with an ASGE-approved indication, but only 28.8% in the absence of such an indication (OR 2.23, 95% CI 1.55 to 3.22, p < 0.01). The overall pre-test probability of making relevant findings on endoscopy was 45%. This was marginally improved to 47% with the presence of ASGE-approved criteria, and markedly decreased to 29% in the absence of ASGE criteria (Rossi 2002, p. 717). Only 6.5% of patients with ASGE criteria had normal endoscopy, while 11.9% of those without criteria were normal. Gastric or esophageal cancer was discovered in 32/788 patients with criteria (4%), and 6/198 patients with no criteria (3%).

Alarm symptoms prompted endoscopy in 13% of patients. Specific alarm symptoms addressed in the Vakil meta-analysis (2006) were also measured in the Rossi study (2002). Dysphagia had a 1% sensitivity and a 99% specificity for relevant disease (95% CI not reported); positive likelihood ratio (LR) was 1.24, and negative LR was 0.99. Anemia had a 3% sensitivity and 97% specificity (95% CI not reported); positive LR was 0.84 and negative LR, 1.01.

The authors conclude that appropriateness criteria proposed by the ASGE are useful for certain activities, such as monitoring quality assurance programs in open-access endoscopy. However, they should not be used to select which patients undergo endoscopy as this practice is "neither reliable nor safe and carries the risk of missing clinically relevant diagnoses" (Rossi 2002, p. 719).

In a good-quality prospective study, Marmo and colleagues (2005) collected data on all patients presenting for endoscopy at one of four community-based hospitals over a period of two years, excluding those with alarm symptoms or who regularly used NSAIDs (final included n=5,224). Twenty-two patients without alarm symptoms were found to have cancer (0.4%) whereas cancer was found in 36 of 464 patients with alarm symptoms (7.7%). Alarm symptoms were considered to be unexplained weight loss, recurrent vomiting, dysphagia, hematemesis or melena, anemia, and palpable mass. Of the patients who had none of these, mean age among those with malignancy was significantly higher than in those without (66.2 ± 14.7 vs 47.9 ± 15.8, p<0.0001), and mean age of males with malignancy was lower than for females (63.6 ± 14.1 for males and 72.7 ± 15.4 for females, p=0.08).

In a fair-quality prospective cohort study, Bowrey and colleagues (2006) evaluated cancer dectection rates in a population of patients referred by their primary care physician for endoscopy. In this "open-access" system, any patient over 35 with dyspepsia or dysphagia was

eligible for referral. Of the 4,018 patients completing endoscopy, 138 (3%) were found to have malignancy. Of those with malignancy, 85% presented with alarm symptoms, but 15% (n=19) had no alarm symptoms. Those without any alarm symptoms tended to have a more favorable tumor grade than those with alarm symptoms (11% stage IV vs 47% stage IV), and a longer median survival time (39 months vs 4 months for patients with epigastric mass or 10 months for patients with unexplained weight loss). The authors conclude that open-access endoscopy confers a survival advantage, but fail to take into account the high risk of lead-time bias; that is, the possibility that patients without alarm symptoms whose cancers were diagnosed early would have died at the same point in time if their cancers were diagnosed later, i.e., after the appearance of alarm symptoms.

A fair-quality prospective study by Veldhuyzen van Zanten and colleagues (2005) used data from the Canadian Adult Dyspepsia Empirical Therapy Prompt Endoscopy (CADET-PE) trial to assess the prevalence of Barrett's esophagus in adults (>18) with dyspepsia and to identify potential risk factors for its presence. Barrett's esophagus was identified in 15 of 379 (4%) patients over the age of 50, and in 10 of 661 (1.5%) patients age 50 and younger, a statistically significant difference (p = 0.013). Among patients with identified Barrett's esophagus, 64% had reported heartburn or acid regurgitation as their dominant symptom, whereas this was the dominant symptom in only 37% of patients without BE (p = 0.0062). The BE patients had an average length of symptoms of 10 years, but 44% had symptoms < 5 years and 16% less than 1 year (p. 597).

Overall summary, quality and limitations of the evidence
One good quality meta-analysis (Vakil 2006) of a large number of patients found that alarm symptoms, clinical opinion, and computer modeling programs based on symptom questionnaires were all unreliable predictors of gastrointestinal malignancy. Sensitivity ranged from 0% to 83% while specificity varied from 40% to 98%. A fair quality prospective cohort study (Rossi 2002) determined that ASGE guideline criteria were poorly correlated with clinically relevant endoscopic findings, although their presence does marginally increase the pre-test probability of endoscopy (from 45% to 47%) and their absence lowers it (from 45% to 29%). A good quality prospective study (Marmo 2005) found a much higher prevalence of malignancy in patients with alarm symptoms than in those without (7.7% vs 0.4%). A fair-quality prospective study found that 15% of patients with malignancy presented with no alarm symptoms; however, due to the high risk of lead-time bias it is not clear that there is a survival advantage to early endoscopy for these patients.

The meta-analysis by Vakil and colleagues (2006) was of good quality and has relatively few limitations. It did include one study from China, where there is a higher prevalence of gastric malignancy, which would have tended to overestimate the positive predictive value of alarm symptoms. The authors report that exclusion of this China study did not alter the conclusions. Some of the studies took place in secondary care, and thus a referral bias may be present. The direction of such a bias on the performance of alarm symptoms as harbingers of malignancy cannot be predicted.

One prospective cohort study (Rossi 2002) looked at consecutive patients presenting to an open-access endoscopy center. Seventy percent were outpatient, while 30% were hospital inpatients at the time of the study. The impact of including these two disparate populations is not assessed. The study was published in 2002, and ASGE guidelines have since been updated, and so conclusions must be taken with caution.

If alarm features and the symptoms included in specialty guidelines are not reliable, how should clinicians choose patients for endoscopy? Vakil and colleagues (2006) suggest that, in the absence of compelling predictors, the concept of "alarm symptoms" should not be abandoned at this time. They suggest age greater than 55 as "the most logical alternative strategy… because the incidence of upper GI malignancy is negligible in Western populations at younger ages and only rises in prevalence above the age of 55 years" (Vakil 2006, p. 398). The use of age as a factor in pre-test probability of malignancy is supported by the prospective cohort studies considered here. Vakil and colleagues recommend further research into alarm symptoms and physical exam findings, as diagnostic accuracy may improve when certain features occur in concert. They also note that the alarm symptoms themselves are not well-defined: how much weight loss qualifies as alarming? Is dysphagia concerning if it is not progressive? Future research should endeavor to define alarm symptoms more precisely.

 Overall strength of evidence: Moderate

KQ3. For what diagnoses and within what time frames, is repeat endoscopy indicated versus other tests or no follow-up tests for surveillance of disease progression and/or treatment response? Does repeat endoscopy change treatment and outcome?

One good quality prospective cohort study (Westbrook 2005) was identified for this Key Question.

Systematic reviews and technology assessments
The search did not identify any systematic reviews or technology assessments relevant to this Key Question.

RCTs
The search did not identify any randomized controlled trials relevant to this Key Question.

Other study designs (e.g. cohort studies)
The study identified one good quality prospective cohort study (Westbrook 2005) relevant to this Key Question. Westbrook and colleagues (2005) investigated 302 patients who had undergone endoscopy following gastroenterology referral for dyspeptic symptoms at one of two Australian teaching hospitals 18 months prior. Patients were interviewed at study onset (FU1) and then again 8 to 9 years post-index endoscopy (FU2) to determine persistence of symptoms, use of acid-suppressing medications, and use of repeat endoscopy.

In all, 34% (95% CI 29.0% to 39.8%) of patients were asymptomatic at FU1 and 41% (95% CI 35.6% to 46.6%) were symptom-free at FU2, indicating that the majority of these patients had ongoing symptoms. The authors found no association between endoscopic diagnosis (normal, normal with reflux symptoms, peptic ulcer disease, esophagitis only, or esophagitis with peptic ulcer disease) and symptomatic outcome at either FU1 or FU2 (χ^2=14.7, df=12, P > 0.05). At FU2, 31% of patients (n=92) had undergone repeat endoscopy, including 50% of patients whose initial diagnosis was "esophagitis only." There was no association between repeat endoscopy and symptomatic outcome (χ^2=0.6, df=1, P > 0.05).

Overall summary, quality and limitations of the evidence
Only one study, a prospective cohort study of good quality (Westbrook 2005), addressed the question of repeat endoscopy in patients who initially presented with dyspeptic symptoms and had non-malignant endoscopic findings. About a third of these patients underwent a subsequent endoscopy within nine years of the index study. The results of these later endoscopies are not known; however, patients who had further endoscopy were neither more nor less likely than other patients to be symptomatic at FU2 (χ^2=0.6, df=1, P > 0.05).

Data was collected by one investigator via structured telephone interview, lowering the risk of reporting bias. Recall bias is a possibility, but the authors did query patients at FU1 and FU2 regarding the symptoms that had prompted their initial presentations, and found good concordance with the medical record (Westbrook 2005, p. 620). This would indicate that the patients' recollections of their illnesses were reliable. The fact that all patients were at one of two Australian hospitals may limit generalizability to other populations. The lack of information on the diagnoses made at subsequent endoscopies is a major limitation of this study.

Overall, evidence is limited on the question of repeat endoscopy for any patients with initial dyspepsia who have non-malignant findings on their index endoscopy.

  Overall strength of evidence: Low

KQ4. What are the potential harms of performing upper endoscopy in the diagnostic or treatment planning workup of adults with upper GI symptoms? What is the incidence of these harms? Include consideration of progression of treatment in unnecessary or inappropriate ways.

One decision analysis (Spiegel 2002) and one good quality meta-analysis (Ford 2005) were identified for this Key Question.

Systematic reviews and technology assessments
There is very little mention of harms of endoscopy in the systematic reviews, meta-analyses, and economic evaluations identified by this search. The decision analysis by Spiegel and colleagues (2002) indicates that "the most common complications of endoscopy are cardiorespiratory and generally require only additional observation. Our model assumed a 0.02% probability of severe endoscopic complications requiring hospitalization and surgery. The

costs of severe endoscopic complications were modeled after the surgical repair of a perforation" (p. 1275).

One meta-analysis of five trials comprising 1924 patients stated "there were no adverse events reported as a direct result of endoscopy" (Ford 2005, p. 1842).

RCTs
The search did not identify any randomized controlled trials relevant to this Key Question.

Other study designs (e.g. cohort studies)
The search did not identify any other comparative studies relevant to this Key Question.

Overall summary, quality and limitations of the evidence
Because so little information could be gathered from the initial search, we conducted a separate MEDLINE® search for complications of endoscopy (see Appendix A). Unfortunately there were no studies that could be included. All but one of the systematic reviews, meta-analyses, and economic evaluations neglected to factor harms of endoscopy into their reports.

According to the authors of one economic evaluation, most harms of endoscopy are cardiorespiratory in nature; that is, related to the procedure sedation rather than the endoscope itself. These authors used a 0.02% incidence of severe harms and modeled their economic assumptions on the surgical repair of perforation.

Our search identified no data on harms associated with empiric acid-suppression or *H. pylori* test-and-treat treatment strategies.

Overall strength of evidence: Insufficient

KQ5. What is the evidence that upper endoscopy has differential efficacy or safety issues in sub populations? Including consideration of:
- **Gender**
- **Age**
- **Psychological or psychosocial co-morbidities**
- **Other patient characteristics or evidence based patient selection criteria, especially comorbidities of diabetes, high BMI, and chronic ingestion of alcohol**
- **Provider type, setting or other provider characteristics**
- **Payer / beneficiary type: including worker's compensation, Medicaid, state employees?**

One good qualtiy meta-analysis (Ford 2005), two good-quality economic evaluations (Barton 2008; Makris 2003) , three prospective cohort studies (Marmo 2005, good quality; Bowrey 2006, fair quality; Veldhuyzen van Zanten 2006, fair quality), and one poor quality retrospective cohort study (Connor 2004) were identified for this Key Question.

Systematic reviews and technology assessments

One meta-analysis (Ford 2005) included prespecified subgroup analyses. Ford and colleagues (2005), in their individual patient data analysis comparing prompt endoscopy with *H. pylori* test-and-treat, considered gender, age (less than 50 years or 50 years and older), predominant symptom at trial entry (epigastric pain or heartburn), and initial *H. pylori* status. In these analyses, the only difference observed was a small but statistically significant effect in favor of endoscopy for patients aged 50 years and older (RR=0.90, 95% CI 0.82 to 1.00, p < 0.05). The relative risks (95% CI) of all subgroup analyses are as follows:

Females:	0.95 (0.90 to 1.00)
Males:	0.97 (0.0.91 to 1.04)
50 years and older:	0.90 (0.82 to 1.00, P < 0.05)
Less than 50 years:	0.97 (0.93 to 1.01)
Predominant heartburn:	0.99 (0.93 to 1.04)
Predominant epigastric pain:	0.96 (0.89 to 1.04)
H. pylori positive:	0.93 (0.86 to 1.01)
H. pylori negative:	0.97 (0.93 to 1.02)

Two economic evaluations (Barton 2008; Makris 2003) considered subgroups to some extent. Barton and colleagues (2008) executed their simulation model in both a 30 year old and a 60 year old hypothetical population. In both populations, endoscopy with biopsy for *H. pylori* and eradication as indicated (contributing 4.3496 QALYs in 30 year olds and 4.3860 QALYs in 60 year olds) was more effective than endoscopy without biopsy (contributing 4.3387 QALYs in 30 year olds and 4.3712 QALYs in 60 year olds). Also in both populations, a trial of PPI followed by endoscopy with biopsy was the most effective strategy (contributing 4.3541 QALYs in 30 year olds and 4.3942 QALYs in 60 year olds).

Makris and colleagues (2003) also divided their hypothetical populations into age cohorts, ≤45 years old (Group A), and 45 years old and older (Group B). Unfortunately, effectiveness data were not reported for Group B and so comparison cannot be done.

RCTs
The search did not identify any randomized controlled trials relevant to this Key Question.

Other study designs (e.g. cohort studies)
This search identified one good quality prospective cohort study (Marmo 2005), two fair quality prospective cohort studies (Bowrey 2006 and Veldhuyzen van Zanten 2006), and one poor quality retrospective cohort study (Connor 2004) relevant to this Key Question.

Marmo and colleagues (2005) published a good-quality prospective cohort study of 5,224 adults with uncomplicated dyspepsia, excluding patients with NSAID use or alarm symptoms. Malignancy was discovered in 22 patients (0.4%). Eighteen of these patients (81.8%) were >45 years of age, and sixteen (72.7%) were male. The mean age of males with malignancy was 63.6 ± 14.1 years, which was significantly lower than the mean age of females with malignancy (72.7 ± 15.4 years; p < 0.05). Overall, the mean age of patients with malignancy was significantly higher than the mean age of patients without malignancy (66.2 ± 14.7 years versus 47.9 ± 15.8

years; p < 0.0001). The authors created a prediction tool combining age and gender, with a 35 year cutoff for males and a 57 year cutoff for females, and tested it in a second "split sample" of 3,684 patients. In the split sample, the shift in age cutoff to 35 years for males and 56 years for females was demonstrated to increase diagnostic yield by 12% (i.e., 12% more patients with cancer are included in these selection criteria, without increasing the total number of endoscopies performed). Marmo and colleagues conclude that "Only the combination of age and gender was able to predict upper GI malignancy in patients with uncomplicated dyspepsia. All the other variables considered as possible risk factors [*H. pylori* infection, comorbidities, coprescriptions] had no predictive value" (p. 787).

A fair-quality prospective study by Bowrey and colleagues (2005) was primarily aimed at assessing the value of alarm symptoms for predicting malignancy, but did also touch on the issues of age and gender. All patients age 35 years and older presenting for endoscopy (n=4,018) were included. Nineteen of 123 patients diagnosed with malignancy (15.4%) had no alarm symptoms and would be comparable to the "uncomplicated dyspepsia" patients in the Marmo 2005 study; 13 of those were male and 6 were female (p=0.61). Five of those nineteen patients with malignancies and no alarm symptoms were younger than 55 years of age (25%). Overall, the prevalence of malignancy in the cohort increased along with age, with only a 0.3% prevalence among those < 41 years of age, and a 16% prevalence among patients >80 years old.

A fair-quality prospective study by Veldhuyzen van Zanten and colleagues (2006) considered the outcome of histologically diagnosed Barrett's esophagus (BE) in a population undergoing prompt endoscopy as part of the CADET-PE trial. The prevalence of BE was 2.4% overall. The prevalence was significantly higher in patients over 50 years of age (4% versus 1.5%, p = 0.013). The prevalence among males was higher than among females, but this difference fell shy of statistical significance (2.4% of males versus 2% of females, p = 0.068).

Connor and colleagues (2004) performed a retrospective chart review of 264 patients in the Kansas City Veterans Affairs Medical Center Gastroenterology department who had undergone upper endoscopy for dyspeptic symptoms. The definition of dyspepsia included upper abdominal pain or discomfort, with or without heartburn, reflux, nausea, or vomiting. Patients with alarm symptoms were excluded. The population was 95% male, 73% Caucasian, with a mean age of 57 years. The incidence of Barrett's esophagus was 6.1%, and erosive esophagitis was present in 23.8%. These two endoscopic findings were tested for correlation with patient demographics (e.g., age, gender, race), NSAID use, or presence of hiatal hernia. The only significant correlation was with hiatal hernia (Fisher's exact test, p = 0.0032). There were no significant associations between demographics and endoscopic diagnosis.

Overall summary, quality and limitations of the evidence
The search uncovered no evidence related to most of the subpopulations named in the Key Question. Gender was evaluated in a good quality meta-analysis (Ford 2005) and a poor quality retrospective cohort study (Connor 2004) and no differential effectiveness was found. Age was the only factor associated with differential effectiveness in the meta-analysis. Age was identified as an independent risk factor for malignancy in three prospective cohort studies, two

of fair quality (Bowrey 2006, Veldhuyzen van Zanten 2006) and one of good quality (Marmo 2005).

Ford and colleagues (2005) performed subgroup analyses based on age, gender, predominant symptom, and presence of *H. pylori*. There was a small but statistically significant benefit of endoscopy in patients 50 years of age and older (RR=0.90, 95% CI 0.82 to 1.00, p < 0.05); no other associations were found. In a good quality economic evaluation simulation model, relative effectiveness of interventions was the same in hypothetical 30 year olds as in hypothetical 60 year olds (Barton 2008). A poor quality retrospective chart review (Connor 2004) of VA patients failed to find any correlation between significant endoscopic findings (Barrett's esophagus and/or erosive esophagitis) and age, gender, race, or NSAID use. Two fair-quality prospective cohort studies (Bowrey 2006, Veldhuyzen van Zanten 2006) affirm the increasing risk of malignancy or Barrett's esophagus with advancing age. A good quality prospective cohort study (Marmo 2005) proposes a predictive tool using a combination of age and gender (>35 years for males, >56 years for females).

Overall strength of evidence: Moderate (Age), Insufficent (all others)

KQ6. What is the evidence of cost and cost-effectiveness of endoscopy compared to other treatment strategies when used in diagnostic or treatment planning workups of adults with upper GI symptoms?

Economic modeling studies and cost-effectiveness analyses
Our search identified ten studies that met inclusion criteria, six economic model evaluations (Barkun 2010; Barton 2008; García-Altés 2005; Makris 2003; Spiegel 2002; You 2006) and four RCTs (Duggan 2008; Ginannini 2008; Kjeldsen 2007; Klok 2005) that included cost-effectiveness analyses. Since procedural and pharmaceutical costs vary across health care systems, cost-effectiveness evaluations done outside the US should be taken with caution. Only two included studies, a second-order simulation model and a decision analysis, utilized US data (Barton 2008; Spiegel 2002). All studies investigated adult patients who were presenting for initial evaluation and management of upper gastrointestinal symptoms.

The largest evaluation was a good quality economic evaluation (Barkun 2010) performed with an aggregation of four complementary Canadian Adult Dyspepsia Empiric Treatment (CADET) studies. This design was able to aggregate data from 2,236 individual patients in Canada who had enrolled in one of the four studies. Patients were adults presenting to their primary care clinician with at least three months of uninvestigated upper GI symptoms, who did not have alarm symptoms (e.g., unintentional weight loss, vomiting, dysphagia, hematemesis, melena, fever, jaundice, or anemia) and were not regular users of nonsteroidal anti-inflammatory medications. Interventions compared included acid-suppressing medication (omeprazole or ranitidine) based on the Canadian Dyspepsia Working Group clinical management tool (CanDys

omeprazole and CanDys ranitidine);[3] empirical omeprazole or ranitidine; endoscopy followed by omeprazole; and endoscopy followed by ranitidine. Costs were expressed in 2007 Canadian dollars. Incremental cost-effectiveness analyses demonstrated that no single approach was the dominant best choice based on cost-effectiveness. However, at a clinically relevant willingness-to-pay threshold of CAN $30,000-CAN$70,000 per quality-adjusted life-year (QALY), the CanDys Omeprazole arm was the most cost-effective.

A fair quality modeled decision analysis from Spain (García-Altés 2005) identified a locally-validated scoring instrument followed by endoscopy as more cost-effective than prompt endoscopy, *H. pylori* test and scope, *H. pylori* test and treat, and empiric proton-pump inhibitor (PPI) strategies .

Four other modeling studies included testing for *H. pylori* in their analysis, and found it to be the most cost-effective intervention when compared to empiric initial treatment and/or prompt endoscopy. One US study (Barton 2008), a good quality simulation model, found that, among 60 year olds, test-and-treat was the most cost-effective intervention with an incremental cost-effectiveness ratio of US $6,740 per QALY. A very similar study from Canada also found the test-and-treat approach to be more cost-effective, at CAN $2,970 per QALY, compared with empirical acid-suppression, a barium test, endoscopy, or empirical eradication of *H. pylori* (Makris 2003). A third good quality modeled study from China (You 2006) found that the test-and-treat approach was more cost-effective than empiric PPI or prompt endoscopy, but noted that the cost-effectiveness of test-and-treat was sensitive to the prevalence of *H. pylori* in the population.

A good quality decision analysis using US data (Spiegel 2002) found that adding a 6-week trial of PPI, either before test-and-treat or interposed between test-and-treat and endoscopy, increased cost-effectiveness over a strategy of going directly from test-and-treat to endoscopy in a population of patients greater than 45 years old. Acid reflux and regurgitation were excluded as dominant symptoms and no alarm symptoms were present as inclusion criteria for this population. Their conclusion about preferred strategies remained even after sensitivity analysis.

The test-and-treat approach was favored in a fair quality RCT from the UK (Duggan 2008) when compared with prompt endoscopy or empirical PPI. There was not a substantial difference in the effectiveness of these alternative strategies. However, 39% of patients assigned to the empiric PPI arm of the trial subsequently underwent upper endoscopy after an initial treatment failure. This resulted in higher spending amongst the empiric PPI arm and the lowest overall

---

[3] The Canadian Dyspepsia Working Group's "CanDys Clinical Management Tool" recommends stratifying patients into two groups: those in whom symptoms of heartburn or reflux are dominant, and those in whom heartburn and/or reflux are not dominant. In the CanDys omeprazole arm, heartburn-predominant patients were treated empirically with omeprazole 20mg daily for 8 weeks. Other patients were tested using the urea breath test for H. pylori. If positive, they received one week of eradication triple therapy; if negative, they received omeprazole 20mg daily for four weeks. In the CanDys ranitidine arm, patients were started on 150mg ranitidine twice daily, with step-up to omeprazole for heartburn-predominant patients with persistent symptoms after four to eight weeks of ranitidine.

number of endoscopies in the test-and-treat arm. Finally, one poor quality RCT from the Netherlands found test-and-treat to be slightly more effective and less costly than prompt endoscopy, with an incremental cost-effectiveness ratio of €47,412 per QALY (Klok 2005).

In a good quality simulation model using a hypothetical population of US 30 year olds (Barton 2008), empiric PPI was the most cost-effective option with an incremental cost-effectiveness ratio (ICER) of US$9,740 per QALY when compared to a baseline strategy of antacid (agent unspecified) alone with no further interventions.[4] A poor quality Italian RCT (Giannini 2008) favored empiric PPI over endoscopy because it was found to be equally effective, but less expensive than endoscopy. However, this study was at high risk of bias in favor of PPI because the patients with positive endoscopy received identical treatment to those in the empiric PPI arm (40mg omeprazole), while those with negative endoscopy received 20mg omeprazole. Finally, a poor quality cost-effectiveness analysis of RCT data from Denmark (Kjeldsen 2007) showed that while endoscopy was slightly more effective, it was much more costly, with an ICER €13,905 per cure (defined as symptom-free at 12 months). This result was sensitive to age, with ICER increasing for patients less than 45 years of age. An important limitation of these two low quality studies is that they did not include *H. pylori* testing or treatment in their comparisons.

Finally, the individual patient data (IPD) meta-analysis conducted by Ford and colleagues (2005) using data from the Cochrane review included a cost-effectiveness analysis reporting costs and cost-effectiveness as weighted-mean difference (WMD). The authors found that endoscopy cost more than the test-and-treat strategy (WMD US$389, 95% CI $276 to $502), and this was due to the cost of the investigation. Prompt endoscopy was not a cost-effective alternative until the willingness-to-pay threshold was raised to $180,000 per symptom-free patient.

Table 3. Cost-Effectiveness Studies by Quality and Intervention Favored

Favored Intervention	Empiric PPI	*H. pylori* Test & Treat	Screening Questionnaire
Good Quality	• Barton 2008, US (*preferred strategy for hypothetical 30yo pop.*)	• Barton 2008, US (*preferred strategy for hypothetical 60yo pop.*) • Makris 2003, Canada (*preferred strategy for both hypothetical 18-45yo and ≥45yo pops.*) • You 2006, Hong Kong (*hypothetical ≥18yo pop.*) • Barkun 2010, Canada (*individual data from 2,236 Canadians ≥18yo*) • Spiegel 2002 (*T&T→PPI→EGD is the preferred strategy in US patients < 45 yo*) • Ford 2005 (IPD *meta-analysis of Cochrane data*)	
Fair Quality		• Duggan 2008, UK (*762 adults ≥18yo*	• Garcia-Altes

---

[4] The authors explain that "[antacid alone] is not a realistic strategy but was used as a baseline comparison so that information from placebo-controlled trials could be used in the model" (p. 47).

Favored Intervention	Empiric PPI	*H. pylori* Test & Treat	Screening Questionnaire
		presenting to primary care with dyspepsia)	2005, Spain (*hypothetical ≥18yo pop.*)
Poor Quality	• Giannini 2008, Italy (*612 adults ≥18yo presenting to GI centers with ≥3mo of symptoms*) • Kjeldsen 2007, Denmark (*368 adults ≥18yo presenting to primary care with dyspepsia*) **Note:** Neither study included a comparison with *H. pylori* test-and-treat	• Klok 2005, Netherlands (*281 adults ≥18yo presenting to primary care with dyspepsia*)	

Overall summary, quality and limitations of the evidence

With the exception of empiric therapy for US 30 year olds, all five good quality studies, one of two fair quality studies, and one of three poor quality studies favored *H. pylori* test-and-treat as the most cost-effective strategy for adults with uninvestigated symptoms of dyspepsia and/or GERD.

Only two studies, a second-order simulation model and a decision analysis both of good quality, evaluated the cost-effectiveness of different management strategies for new upper gastrointestinal symptoms in a US population (Barton 2008; Spiegel 2002). In the Barton study (2008), empiric PPI was the strategy of choice for 30 year old patients, and test-and-treat for *H. pylori* was the most cost-effective intervention for 60 year olds. Spiegel and colleagues (2002) looked only at patients less than 45 years of age, and determined that adding a 6-week trial of PPI to the test-and-treat strategy improved its cost-effectiveness. A good quality economic evaluation of Canadian individual patient data concluded that no one strategy was the most clearly cost-effective, but at a clinically relevant willingness-to-pay threshold of CAN$30,000-70,000 per QALY, omeprazole treatment based on the CanDys protocol (which incorporates test-and-treat for those without heartburn or reflux as the predominant symptom) was the most cost-effective (Barkun 2010). Two other good quality models (Makris 2003; You 2006) also favored the test-and-treat approach, along with one fair and one poor quality RCT (Duggan 2008; Klok 2005).

One fair quality decision analysis favored a screening questionnaire followed by prompt endoscopy for high-risk patients (García-Altés 2005). Two poor quality RCTs found empiric PPI to be the most cost-effective alternative, but did not include comparison to *H. pylori* testing and treatment (Giannini 2008; Kjeldsen 2007). There were no economic studies that found prompt endoscopy to be the most cost-effective intervention.

Overall strength of evidence: Moderate

Guidelines

Three guidelines (American Gastroenterological Association (AGA) 2008; American Society for Gastrointestinal Endoscopy (ASGE) 2007a; ASGE 2007b) were identified in relation to the role of endoscopy for the diagnosis and early management of dyspepsia and GERD. An additional guideline specifically related to modifications in endoscopic practice for the elderly was identified from the ASGE (2006). Of the guidelines identified, one guideline was rated good quality (AGA 2008), two were rated fair quality (ASGE 2007a; ASGE 2007b), and one was rated poor quality (ASGE 2006).

Role of endoscopy in the diagnosis and management of GERD
One good quality (AGA 2008) and one fair quality (ASGE 2007b) guidelines were identified that discussed the role of endoscopy in the diagnosis and management of GERD. The AGA (2008) guideline recommends endoscopy "with biopsy for patients with an esophageal GERD syndrome with troublesome dysphagia" (p. 1385) and to evaluate patients who have not responded to an empirical trial of twice-daily PPI therapy and who have suspected esophageal GERD symptoms. The AGA (2008) guideline recommends against routine endoscopy for patients with "erosive or nonerosive reflux disease to assess for disease progression" (p. 1387) and finds insufficient evidence for routine upper endoscopy for chronic GERD symptoms to "diminish the risk of death from esophageal cancer" (p. 1389) or for screening of "Barrett's esophagus and dysplasia in adults 50 years or older with greater than 5 to 10 years of heartburn to reduce mortality from esophageal adenocarcinoma" (p. 1389).

The fair quality ASGE (2007a) guideline presents complementary recommendations to the AGA (2008) guideline. The ASGE (2007a) recommends that GERD can be diagnosed based on typical symptoms without the need for endoscopy. For patients who have alarm symptoms (e.g., GERD symptoms that are persistent or progressive despite appropriate medical therapy; dysphagia or odynophagia; involuntary weight loss greater than five percent; evidence of GI bleeding or anemia; finding a mass, stricture, or ulcer on imaging studies; or persistent vomiting) endoscopy is recommended. Additionally the ASGE (2007a) recommends the use of endoscopy for the evaluation of patients with suspected extra-esophageal manifestations of GERD, with recurrent symptoms after endoscopic or surgical antireflux procedures, and for the screening for Barrett's esophagus in selected patients as clinically indicated.

Role of endoscopy in the diagnosis and management of dyspepsia
One guideline of fair quality (ASGE 2007b) was identified that discussed the role of endoscopy in the diagnosis and management of dyspepsia. The ASGE (2007b) acknowledges that "given the large number of patients with dyspepsia, it is not practical to perform endoscopy in all patients with dyspepsia" (p. 1071). Characteristics, such as age and alarm features, are recommended to be used to help determine whether a patient with dyspepsia should undergo initial endoscopy. The ASGE (2007b) suggests that patients between the age 45 to 55 years who have new onset dyspepsia and those who have alarm features (e.g., family history of upper-GI

malignancy; unintended weight loss; GI bleeding or iron deficiency anemia; progressive dysphagia; odynophagia; persistent vomiting; palpable mass or lymphadenopathy; jaundice) should undergo an endoscopy. In addition, the ASGE (2007b) guideline recommends endoscopy be considered for patients without alarm features for whom there is clinical suspicion of malignancy. For patients that are younger than 50 years old and are *H. pylori* negative, the ASGE (2007b) guideline recommends an initial endoscopy or a short trial of PPI acid suppression. The ASGE (2007b) also gives a weak recommendation for endoscopy in "patients with dyspepsia who do not respond to empiric PPI therapy or have recurrent symptoms after an adequate trial" (p. 1074).

Modification of Endoscopy for Elderly Population
One guideline of poor quality (ASGE 2006) was identified that specified modification in endoscopic practice for a specific population. The ASGE (2006) recommends that endoscopy only be performed when the results will influence clinical management or outcome. The guideline states that "advanced age is not a contraindication to endoscopy" (p. 566) and that "preparation for endoscopy in geriatric or aged populations differs little from that for younger adults" (p. 567). Additionally the ASGE (2006) guideline suggests that intensified monitoring during the endoscopic procedure is appropriate for many elderly patients.

Summary of Guidelines and Quality Assessment
The search identified four guidelines of which one was rated as good quality, two were rated fair quality, and one was rated as poor quality. A summary and quality rating for each guideline is provided in Table 4. In general, the guidelines are in accord with the clinical evidence. Because of the poor reliability of alarm symptoms for predicting GI malignancy, guidelines should and do leave some room for clinical judgement. It is reasonable based on evidence to perform endoscopy in patients who have not responded to medical management. The AGA (2008) guideline also recommends against routine surveillance endoscopy for patients with reflux disease to assess for disease progression or resolution, which is supported by the evidence. The guidelines do not, however, take into account the increasing prevalence of gastrointestinal malignancy after age 55. As noted under Key Question #5, age greater than 55 may be a reasonable criterion to apply when choosing whom to investigate.

Table 4. Role of Endoscopy in the Diagnosis and Management of GERD and Dyspepsia

Guideline	Recommended Use of Endoscopy	Not Recommended / Insufficient Evidence	Quality
AGA (2008) [GERD]	• Endoscopy with biopsy for patients with an esophageal GERD syndrome with troublesome dysphagia • Evaluation of patients who have not responded to an empirical trial of twice-daily PPI therapy and who have suspected esophageal GERD symptoms	• Routine endoscopy for patients with erosive or nonerosive reflux disease to assess for disease progression (Recommends Against) • Routine upper endoscopy for chronic GERD symptoms to diminish the risk of death from esophageal cancer (Insufficient Evidence) • Screening of "Barrett's esophagus	Good

Guideline	Recommended Use of Endoscopy	Not Recommended / Insufficient Evidence	Quality
		and dysplasia in adults 50 years or older with greater than 5 to 10 years of heartburn to reduce mortality from esophageal adenocarcinoma (Insufficient Evidence)	
ASGE (2007a) [GERD]	• Patients who have alarm symptoms • Evaluation of patients with suspected extra-esophageal manifestations of GERD • Evaluation of patients with recurrent symptoms after endoscopic or surgical antireflux procedures • Screening for Barrett's Esophagus in selected patients as clinically indicated	• GERD can be diagnosed based on typical symptoms without the need for endoscopy	Fair
ASGE (2007b) [Dyspepsia]	• Patients between 45 to 55 years with new onset dyspepsia • Patients with alarm features • Patients without alarm features for whom there is clinical suspicion of malignancy • Patients younger than 50 years and who are *H pylori* negative, endoscopy or short trial of PPI acid suppression • Patients with dyspepsia who do not respond to empiric PPI therapy or have recurrent symptoms after an adequate trial	n/a	Fair
ASGE (2006) [Considerations for older population]	• If results will influence clinical management or outcomes • Intensified monitoring may be appropriate for many elderly patients	n/a	Poor

Policy Considerations

At the direction of WA HTA, this review searched for Medicare, Aetna, Regence BCBS, and GroupHealth policies addressing coverage of upper endoscopy for patients with symptoms of GERD. The policies identified are summarized below, with further detail and direct web links to each policy provided in Appendix H.

Medicare

A Medicare NCD for "endoscopy" allows coverage of "endoscopic procedures when reasonable and necessary for the individual patient" (Centers for Medicare and Medicaid Services (CMS) 2012). Medicare contractor LCDs may further define criteria of reasonable and necessary. However, there are no relevant LCDs applicable to Washington or the Northwest Region (CMS Region X).

Aetna

Aetna has issued a clinical policy bulletin (CPB) addressing upper gastrointestinal endoscopy (Aetna 2011). The CPB outlines detailed clinical indications for the use of upper endoscopy in the following categories: high-risk screening, diagnostic, therapeutic, and sequential or periodic surveillance. The policy excludes coverage of upper endoscopy as experimental and investigational in several explicit circumstances. Table 5 highlights clinical indications set forth by the policy that are relevant to this review. Please see Appendix Hfor complete copy of the policy and a full list of indications allowing coverage of upper endoscopy.

Table 5. Aetna Policy Outlining Clinical Indications for Upper Endoscopy

Type of use for upper endoscopy	Clinical indications considered medically necessary

Diagnostic	• Evaluation of upper abdominal symptoms that persist despite an appropriate trial of therapy
	
	• Evaluation of upper abdominal symptoms associated with other symptoms or signs suggesting serious organic disease (e.g., anorexia and weight loss) or in persons over 45 years of age
	
	• Evaluation of dysphagia or odynophagia
	
	• Evaluation of esophageal reflux symptoms that are persistent or recurrent despite appropriate therapy
	
	• Evaluation of persistent vomiting of unknown cause
	
	• Evaluation of GI bleeding
	
	    o For persons with active or recent bleeding
	
	    o For presumed chronic blood loss and for iron deficiency anemia when the clinical situation suggests an upper GI source
	
	• Evaluation of dyspepsia when any of the following is present
	
	    o Chronic GI bleeding
	
	    o Epigastric mass
	
	    o Iron deficiency anemia
	
	    o Persistent vomiting
	
	    o Progressive difficulty swallowing
	
	    o Progressive unintentional weight loss

Type of use for upper endoscopy	Clinical indications considered medically necessary
	o Suspicious barium meal (upper GI series) (Note: the above indications are excerpted from the full Aetna policy, provided in Appendix H.)
High-risk screening	• Persons with chronic (5 years or more) gastroesophageal reflux disease (GERD) at risk for Barrett's esophagus. (<u>Note</u>: After a negative screening EGD, further screening EGD is not indicated). • Persons with symptomatic pernicious anemia (e.g., anemia, fatigue, pallor, Red tongue, shortness of breath, as well as tingling and numbness in the hands and feet) to identify prevalent lesions (e.g., carcinoid tumors, gastric cancer). • Persons with cirrhosis and portal hypertension but no prior variceal hemorrhage, especially those with platelet counts less than 140,000/mm^3, or Child's class B or C disease.
Surveillance	• Surveillance of persons with Barrett's esophagus (BE) without dysplasia. For persons with established BE of any length and with no dysplasia, after 2 consecutive examinations within 1 year, an acceptable interval for additional surveillance is every 3 years. • Surveillance of persons with BE and low-grade dysplasia at 6 months. If low-grade dysplasia is confirmed, then surveillance at 12 months and yearly thereafter as long as dysplasia persists. • Surveillance of persons with BE and high-grade dysplasia every 3 months for at least 1 year. After 1 year of no cancer detection, the interval of surveillance may be lengthened if there are no dysplastic changes on 2 subsequent endoscopies performed at 3-month intervals. • Surveillance of persons with a severe caustic esophageal injury every 1 to 3 years beginning 15 to 20 years after the injury. • Surveillance of persons with tylosis every 1 to 3 years beginning at 30 years of age. • Surveillance of recurrence of adenomatous polyps in synchronous and metachronous sites at 3- to 5-year intervals. • Surveillance of persons with familial adenomatous polyposis starting around the time of colectomy or after age of 30 years. • Surveillance of persons with hereditary non-polyposis colorectal cancer.
Indications excluded from coverage	• EGD for routine screening • Evaluation of symptoms that are considered functional in origin. (There are exceptions in which an EGD may be done once to rule out organic disease, especially if symptoms are unresponsive to therapy)

Type of use for upper endoscopy	Clinical indications considered medically necessary
	• Evaluation of metastatic adenocarcinoma of unknown primary site when the results will not alter management
	• Repeat EGD for persons with a prior normal EGD if symptoms remain unchanged
	• Routine evaluation of abdominal pain in children (i.e., without other signs and symptoms suggestive of serious organic disease)
	• Evaluation of radiographical findings of:
	○ Asymptomatic or uncomplicated sliding hiatal hernia
	○ Deformed duodenal bulb when symptoms are absent or respond adequately to ulcer therapy
	○ Uncomplicated duodenal ulcer that has responded to therapy
	• Surveillance for malignancy in persons with gastric atrophy, pernicious anemia, or prior gastric operations for benign disease (e.g., partial gastrectomy for peptic ulcer disease)
	• Surveillance of healed benign disease (e.g., esophagitis or duodenal/gastric ulcer)
	• Surveillance during repeated dilations of benign strictures unless there is a change in status
	• Surveillance of persons with achalasia
	• Surveillance of persons with previous aerodigestive squamous cell cancer
	• Surveillance of persons with gastric intestinal metaplasia
	• Surveillance of persons following adequate sampling or removal of non-dysplastic gastric polyps

GroupHealth
No policies identified addressing upper endoscopy for people with symptoms of GERD.

Regence BCBS Washington
No policies identified addressing upper endoscopy for people with symptoms of GERD.

Overall Summary

Evidence
There are a variety of options for initiating workup and treatment of patients presenting with uninvestigated dyspepsia and/or GERD symptoms. A good quality systematic review (Delaney 2005) and a fair quality prospective cohort study (Madan 2005) show that non-invasive strategies, such as an empiric trial of PPI or *H. pylori* test and treat, are equally as effective as

prompt endoscopy for achieving symptom improvement. A 24-hour esophageal pH study might be the gold standard for GERD diagnosis according to the Madan study (2005), but its clinical usefulness is limited by invasiveness, cost, and availability.

There is wide acceptance of the use of "alarm symptoms" such as anemia, dysphagia, and unintentional weight loss to determine patients' need for prompt endoscopy. Patients with these symptoms are excluded from trials of endoscopy for GERD or dyspepsia, on the grounds that they represent a population with a higher-than-normal risk of malignancy. Clinical guidelines invoke these alarm features as indications to bypass empiric treatment or non-invasive testing and move straight to endoscopy. One good quality meta-analysis (Vakil 2006) and one fair quality prospective case series (Rossi 2002) both agreed that alarm symptoms, as well as clnical opinion, are poor predictors of gastrointestinal malignancy. However, at this time there is no compelling evidence as to what should replace them. Vakil and colleagues (2006) do point out the very low incidence of GI malignancy in Western populations below the age of 55, and note that an age cutoff may be appropriate in formulating a strategy for use of endoscopy.

One meta-analysis that included prespecified subgroup analyses did show that there was a small but statistically significant effect in favor of endscopy for patients aged 50 years and older. Other subgroup analyses based on gender, predominant symptom, and presence of *H. pylori* showed no difference in the effectiveness of endoscopy between groups. A poor quality retrospective cohort study failed to demonstrate any significant associations between meaningful endoscopic findings and patient demographics (e.g., age, race, or gender) or NSAID use.

Patients with findings of malignancy or other serious pathology on endoscopy will be followed up appropriately. But for those whose endoscopic diagnosis was nothing more serious than esophagitis and/or peptic ulcer disease, is there an indication to perform a follow-up endoscopy? One good quality prospective cohort study (Westbrook 2005) followed patients who had presented initially with dyspepsia for eight to nine years after their index endoscopy. More than half of patients had persistent symptoms, and those who had undergone repeat endoscopy (31%) were neither more nor less likely to be symptomatic than those who had not. The study did not, however, assess the findings of these subsequent endoscopies.

There is very little recent data on the harms of upper endoscopy when performed for dyspepsia and/or GERD. The author of one economic evaluation noted that complications are commonly cardiorespiratory (related to sedation), and for purposes of the model used an incidence of severe harms of 0.02%. We found no studies reporting harms associated with empiric acid-suppressing medication or *H. pylori* test-and-treat.

There have been several studies of varying quality that have attempted to determine the most cost-effective means of managing the uninvestigated patient with dyspepsia. Five good quality economic evaluations, along with one of fair quality and one poor quality study, have identified an *H. pylori* test-and-treat strategy as the most cost-effective option. The one exception is a US study that looked at a hypothetical population of 30 year olds and preferred empiric PPI for this younger age group. Two poor quality RCTs also recommended empiric PPI as a more cost-

effective choice than endoscopy. There were no studies that demonstrated prompt endoscopy to be the most cost-effective option.

Guidelines

Four guidelines (one good, two fair, and one poor quality) discuss the role endoscopy in the diagnosis and early management of dyspepsia and GERD. One guideline (AGA 2008) recommends endoscopy to evaluate patients who have not responded to PPI therapy and have suspected GERD symptoms. A fair quality guideline (ASGE 2007a) recommends the use of endoscopy for the screening of Barrett's esophagus as clinical indicated, in patients with recurrent symptoms after endoscopic or surgical antireflux procedures, and/or patients with suspected extra-esophageal manifestations of GERD. One poor quality guideline (ASGE 2006) recommends endoscopy only be conducted in an elderly population when the results will influence clinical management or outcomes, that endoscopy preparation does not differ for geriatric populations, and that intensified monitoring may be appropriate for many elderly patients. The AGA (2008) recommends against routine endoscopy for patient with GERD for assessment of disease progression, and finds insufficient evidence for routine upper endoscopy to reduce mortality from esophageal cancer.

One fair quality guideline (ASGE 2007b) specifically recommends endoscopy in patients between the age 45 to 55 years who have new onset dyspepsia and those who have alarm features should undergo an endoscopy, or in patients without alarm features for whom there is clinical suspicion of malignancy. The ASGE (2007b) guideline recommends either endoscopy or a short trial of PPI acid suppression for patients who are younger than 50 years old and are *H. pylori* negative.

Policies
This review identified two payers, Medicare and Aetna, with policies addressing coverage of upper endoscopy for patients with symptoms of GERD. Medicare has issued a general NCD for "endoscopy" allowing coverage of "endoscopic procedures when reasonable and necessary for the individual patient" (CMS 2012). There are no LCDs applicable to Washington or the Northwest Region that further define criteria constituting "reasonable and necessary" use of the procedure. Among private payers, Aetna has issued a policy setting forth detailed clinical indications for the use of upper endoscopy in the following categories: high risk screening, diagnostic, therapeutic, and sequential or periodic surveillance. The policy excludes coverage of upper endoscopy as experimental and investigational in several explicit circumstances.

Discussion and Limitations

Upper endoscopy for diagnosis of GERD and other upper gastrointestinal symptoms is a thorny topic, subject to many sources of imprecision and potential bias. First, there is the problem of defining which symptoms are indicative of gastro-esophageal reflux disease, and which are dyspepsia. In a primary care office setting, patients are rarely clear-cut members of one category or the other. Second, there is the question of practice setting. Some studies look only at patients primary care, while others include patients in a specialty referral setting such as an

endoscopy center. Depending upon the health care system, patients may or may not be able to self-refer into these specialty centers. Therefore it becomes unclear whether patients in a primary care setting and those in a specialty center are in fact comparable populations.

There is not a consensus on how outcomes should be measured in patients who are treated for dyspepsia or GERD. Several symptom scoring tools exist, some of which are validated. When data are pooled into meta-analyses, these symptom scores are necessarily dichotomized into "cured" versus "not cured," or "improved" versus "not improved." A cost-utility analysis that converts these symptom scores into QALYs is one further step removed from the actual patient experience.

Economic modeling studies and cost-effectiveness analyses in this report came to a consensus around one type of intervention as being generally the most cost-effective (test-and-treat), and initial endoscopy as being less cost-effective. Using incremental cost-effectiveness ratios (ICERs) allows for some degree of comparison across multiple nations whose health care costs may be defined in radically different ways.