Age and alarm symptoms do not predict endoscopic findings among patients with dyspepsia: a multicentre database study


Abstract

Introduction—Symptoms of dyspepsia are common but most patients do not have major upper gastrointestinal pathology. Endoscopy is recommended for dyspeptic patients over the age of 45, or those with certain “alarm” symptoms. We have evaluated the effectiveness of age and “alarm” symptoms for predicting major endoscopic findings in six practising endoscopy centres.

Methods—Clinical variables of consecutive patients with dyspepsia symptoms undergoing upper endoscopy examinations were recorded using a common endoscopy database. Patients who had no previous upper endoscopy or barium radiography were included. Stepwise multivariate logistic regression was used to identify predictors of endoscopic findings. The accuracy of these for predicting endoscopic findings was evaluated with receiver operating characteristic analysis. The sensitivity and specificity of age thresholds from 30 to 70 years were evaluated.

Results—Major pathology (tumour, ulcer, or stricture) was found at endoscopy in 787/3815 (21%) patients with dyspepsia. Age, male sex, bleeding, and anaemia were found to be significant but weak independent predictors of endoscopic findings. A multivariate prediction rule based on these factors had poor predictive accuracy (c statistic=0.62). Using a simplified prediction rule of age ≥45 years or the presence of any “alarm” symptom, sensitivity was 87% and specificity was 26%. Increasing or decreasing the age cut off did not significantly improve the predictive accuracy.

Conclusions—Age and the presence of “alarm” symptoms are not effective predictors of endoscopic findings among patients with dyspepsia. Better clinical prediction strategies are needed to identify patients with significant upper gastrointestinal pathology.

(Gut 2001;49:29–34)

Keywords: dyspepsia; endoscopy; age

Symptoms of recurrent upper abdominal pain or dyspepsia are experienced by approximately 25–40% of the general population. These symptoms accounted for more than 11 million office visits annually to physicians. Multiple diagnostic tests are available for evaluating dyspepsia, including therapeutic trials, testing for Helicobacter pylori, upper gastrointestinal radiography, and endoscopy. Computerised decision analysis studies have yielded inconsistent conclusions on the optimal diagnostic and therapeutic strategy due in large part to the different assumptions underlying these models.

A randomised controlled clinical trial found that initial endoscopy followed by directed therapy is associated with lower costs and fewer days out of work than empiric therapy with histamine H2 receptor antagonists.

Endoscopy is the most accurate method of diagnosis of most conditions associated with dyspepsia, including gastric cancer, peptic ulcer disease, oesophagitis and gastro-duodenitis. However, endoscopy involves some discomfort, significant social inconvenience, and cost. Attempts to identify those patients most likely to benefit from endoscopy have met with variable success. Clinical parameters such as dyspepsia subtypes (ulcer-like, reflux-like, and motility-like) have been shown not to predict pathological conditions. Other factors including age and “alarm symptoms,” such as weight loss, recurrent vomiting, dysphagia, bleeding, or anaemia have been shown to be predictive in some studies but not in others. The American Gastroenterological Association currently recommends endoscopy in all patients over the age of 45 and those with alarm symptoms.

We have used an endoscopic database system of upper endoscopy examinations performed in routine clinical practice at six different medical centres to identify risk factors that predict major endoscopic findings. We have attempted to validate the ability of these risk factors to identify high risk patients and therefore to help refine indications for endoscopy procedures.

Methods

The study was considered exempt by the Institutional Review Board of the Medical University of South Carolina. Patient information on all upper endoscopies was routinely collected using a common electronic endoscopic database (GI-Trac2.0, ASD Medtrac) at each of six medical centres over an average of five years. These data were used to derive a prediction rule of age and other clinical predictors.

Abbreviations used in this paper: ASA, American Society of Anesthesia; ROC, receiver operating characteristic.
years. The centres included: the University of California, Davis; Tufts New England Medical Center; University of Kentucky, Lexington; Wilford Hall-United States Air Force Medical Center; Durham Veterans Affairs Medical Center; and the Medical University of South Carolina. Four of these centres are academic referral centres but also serve as primary care centres for their region. Two of these centres are military or veterans administration (open only to US military veterans) and serve as the primary care centres for veterans in their region. All centres had open access endoscopy centres to which primary care providers could directly refer without consultation with a gastroenterologist. The endoscopist entered the procedure data, including presenting symptoms, pre-procedure diagnosis, recent tests, and diagnostic findings, immediately following endoscopy. All physicians were trained in the use of the database system and in the definitions of the clinical variables. All data, excluding patient demographics, were downloaded to the Clinical Innovation Group of the Foundation for Research Development at the Medical University of South Carolina for analysis. Data quality assurance was assessed by random review of 5% of the procedures and compared with clinical variables in the patient’s medical chart. Only variables with >80% correlation between the chart and database were used in this analysis. The dataset of upper endoscopy procedures contained 162 variable fields and 29,596 visit records for 20,659 patients. Critical variables including age, sex, American Society of Anesthesia (ASA) score, procedure indication, endoscopic findings, and complications were required fields for all patients. Up to three indications for each procedure were recorded for each patient. Only patients who had dyspepsia as an indication were considered. Follow up patient visit records were excluded from the analysis. Patients were included if they had dyspepsia symptoms, complete diagnostic examination of the oesophagus, stomach and duodenum, no proven pre-procedure upper gastrointestinal diagnosis; no previous variceal treatment, percutaneous endoscopic gastrostomy, dilation, stenting, tumour ablation, or foreign body removal; and no recent endoscopy or barium meal. Using these criteria, 3815 (18%) patients were eligible for analysis. Outcome was classified as patients having either a major final diagnosis or otherwise. Major was defined as any ulcer, tumour, or stricture found in the upper region of the gastrointestinal tract. All statistical analysis procedures were performed using the SAS 6.12 statistical package (SAS Institute, Cary, North Carolina, USA).

Table 1  Major endoscopic findings among 3815 patients undergoing upper endoscopy

<table>
<thead>
<tr>
<th>Endoscopic finding</th>
<th>Frequency (No (%))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ulcer</td>
<td>692/3815 (18%)</td>
</tr>
<tr>
<td>Cancer</td>
<td>81/3815 (2%)</td>
</tr>
<tr>
<td>Stricture</td>
<td>14/3815 (0.3%)</td>
</tr>
</tbody>
</table>

Stepwise multiple logistic regression analysis was used to identify significant predictor variables of a major final diagnosis. The prediction model was built using SAS stepwise logistic regression analysis on the exploratory sample population with an entry criterion of p<0.05. The stepwise procedure added the independent variables to the model one at a time. In the final model variables were removed if the retention criterion of p<0.05 was not met. The study population was randomly divided into an exploration group (n=1908) and a validation group (n=1907). Thirteen predictor variables (age, sex, ASA classification of comorbidity, bleeding, vomiting, heartburn, dysphagia, weight loss, early satiety, chest pain, odynophagia, anaemia, and feeding problems) were included in the model building process. Interaction among the predictor variables was analysed using the Breslow day test for homogeneity. Once the model was established using the exploratory group, the parameter estimates were applied to the validation group to test the predictive accuracy of the model. The predictive value of the model was assessed with a receiver operating characteristic (ROC) curve. The curve represents the relationship between sensitivity and specificity for the prediction of a major final diagnosis.

Because prediction rules based on logistic regression are often too complex for routine clinical use, we also developed a simplified clinical prediction rule. The variables specified in the logistic regression model were evaluated using a univariate logistic analysis to justify inclusion in the simplified model. The simplified rule corresponds to current consensus recommendations on endoscopic evaluation of
Table 2 Multivariate risk factors for predicting major upper gastrointestinal pathology

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter estimate</th>
<th>Odds ratio</th>
<th>OR CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.9493</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Age (≥45)</td>
<td>0.5401</td>
<td>1.716</td>
<td>1.331–2.221</td>
</tr>
<tr>
<td>Sex (male)</td>
<td>0.3360</td>
<td>1.399</td>
<td>1.080–1.801</td>
</tr>
<tr>
<td>Anaemia</td>
<td>0.8218</td>
<td>2.274</td>
<td>1.249–4.042</td>
</tr>
<tr>
<td>Bleed</td>
<td>1.0642</td>
<td>2.899</td>
<td>1.889–4.419</td>
</tr>
</tbody>
</table>

Model: probability of a major endoscopic finding = −1.9493 + (0.5401×age) + (0.3360×sex) + (0.8218×anaemia) + (1.0642×bleed), where age, sex (f/m), anaemia (n/y), and bleed (n/y) are coded 0 or 1.

OR, odds ratio; CI, confidence interval.

Results

Data were collected on 20 659 patients undergoing routine upper endoscopy from a total of six centres (four academic, one veterans administration, and one armed forces centre) using the GI-TRAC database over an average of five years. A total of 3815 (18%) patients had dyspepsia as an indication for the procedure and met the inclusion criteria. The mean age of the study sample was 47 (SD 18) years. There were 1994 females (52%) and 1821 males (48%). In the entire population (3815), bleeding was seen in 214 (6%), anaemia in 139 (4%), and dysphagia in 128 (3%). A total of 787 (21%) patients were classified as having a major endoscopic finding (tumour, ulcer, stricture), including 81 (2%) with cancer. The major findings at endoscopy among the 3815 patients are shown in table 1.

The effect of increasing or decreasing the age cut off values as a single predictor was evaluated to determine the “optimal” age threshold for predicting endoscopic findings. Patients were grouped by age from 30 to 70 years in increments of five years. Figure 1 illustrates the proportion of major endoscopic findings in each age group and according to the presence of any “alarm” symptoms as predictors of major endoscopic findings (cancer, ulcer, or stricture): 2x2 table of the number of patients in each category.

Table 3 Accuracy of using age (≥45 years) or the presence of any “alarm” symptoms as predictors of major endoscopic findings (cancer, ulcer, or stricture): sensitivity, specificity, and predictive values [95% confidence intervals]

<table>
<thead>
<tr>
<th>Major endoscopic finding</th>
<th>No major endoscopic finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age ≥45 or alarm symptoms</td>
<td>685</td>
</tr>
<tr>
<td>Age ≤45 and no alarm symptoms</td>
<td>102</td>
</tr>
</tbody>
</table>

Table 4 Accuracy of using age (≥45 years) or the presence of any “alarm” symptoms as predictors of major endoscopic findings (cancer, ulcer, or stricture): sensitivity, specificity, and predictive values [95% confidence intervals]

| Positive predictive value | 23% [22–25] |
| Negative predictive value | 88% [86–90] |
| Sensitivity | 87% [85–89] |
| Specificity | 26% [25–28] |

Figure 2 Receiver operating characteristic (ROC) curve showing the relationship between sensitivity and specificity of the prediction rule for major upper gastrointestinal pathology. The diagonal line represents a model which has zero predictive value. C statistic (area under the ROC curve)=0.62.
Discussion
Using a large multi-institution database to collect routine endoscopic data, we have found that age and the presence of "high risk" symptoms are poor predictors of the presence of major endoscopic findings in the upper gastrointestinal tract. The use of these variables to identify patients who do not require upper endoscopy would likely result in high false negative rates (that is, significant under diagnosis or delayed diagnosis) of important conditions such as ulcer disease, cancer, and strictures of the upper gastrointestinal tract. These findings demonstrate the need for better clinical predictors of upper gastrointestinal pathology. In the absence of better clinical predictors, the study also demonstrates the need for less invasive and thus more widely applicable endoscopy.

The high cost of endoscopy and high prevalence of dyspepsia symptoms has led to extensive studies of how to best apply endoscopy. Cohort studies of endoscopy in referral and general practice populations have shown a high prevalence of relevant pathology among patients with dyspeptic symptoms. Adang and colleagues studied 2900 consecutive patients in a referral practice and found that 21% of dyspeptic patients aged 45 years or less and 25% of those over 45 years had significant pathology identified by upper endoscopy. In another cohort study of 2253 dyspeptic patients, Mansi and colleagues found a high prevalence (approximately 70%) of major and minor pathology. A high but variable prevalence of major pathology (20–50%), including 2% with carcinoma, has also been observed in three cohort studies in a general practice setting.

Attempts to use clinical variables such as age and certain "alarm" symptoms such as weight loss, bleeding, dysphagia, anaemia, or recurrent vomiting to predict pathology have met with variable success. Talley et al have developed a scheme for classifying dyspepsia into ulcer-like, dysmotility-like, reflux-like, and unspecified. Population based studies have shown that this classification is a poor predictor of anatomical pathology. Using a population based survey in Norway, Johnsen and colleagues compared endoscopic findings in 309 patients with and 310 patients without dyspepsia. Almost one third of inflammatory conditions and 50% of ulcers were identified in asymptomatic patients, supporting the argument that symptoms are poor markers of anatomical pathology.

The American Gastroenterological Association recommendation and common clinical practice is to perform endoscopy on patients with dyspepsia and "alarm" symptoms or over the age of 45. Younger patients without alarm symptoms can be treated empirically, with endoscopy reserved for when symptoms fail to resolve. The ability of age to discriminate between patients with and without upper gastrointestinal tract pathology is controversial. Christie and colleagues identified all gastric cancer cases within a defined region of England, and retrospectively assessed alarm symptoms from chart review. Gastric cancer was identified in 319 patients of whom 25 were less than 55 years of age. Twenty four of 25 had alarm symptoms. In a similar study by Gillen and McColl performed in Scotland, 169 patients less than 55 years had upper gastrointestinal tract cancer and 164/169 (97%) had alarm symptoms. These authors argue that endoscopy is not necessary in young patients without alarm symptoms due to the low prevalence of cancer in this group. However, these retrospective studies have the potential for recall bias that may cause overestimation of the prevalence of "alarm" symptoms prior to endoscopy. Furthermore, these studies only focused on cancer. The value of endoscopic detection of ulcers, strictures, or other minor conditions was not assessed.

Unlike other relevant studies, we have included cancer, ulcers, and strictures in the list of major pathological conditions. Endoscopy is clearly indicated for the detection of strictures because of the ability to deliver therapy such as dilation. Endoscopy is also the only test capable of distinguishing benign from malignant gastroduodenal ulcers, and of assessing the risk of bleeding from ulcers. Endoscopy also may improve quality of life and reduce costs of unnecessary empiric therapy, independent of its role in cancer and ulcer detection. One cohort study and three randomised controlled trials have evaluated the impact of initial endoscopy versus empiric therapy on cost and quality of life. In a study of 196 dyspeptic patients (mean age 43 years) before and after endoscopy, Wiklund and colleagues found a significant increase in quality of life, physical activity, and sleep scores in response to endoscopy. In a study of 414 young (median age 44 years) dyspeptic patients who were randomised to empiric therapy with anti-histamine H2 antagonists versus initial endoscopy, Bytzer and colleagues found higher costs for empiric therapy but no difference in quality of life. Higher costs were mostly due to increased time away from work and pharmaceutical costs. Another trial by Laheij and colleagues randomised 80 patients to empiric therapy with proton pump inhibitors versus initial endoscopy directed therapy. In contrast with the Bytzer study, they found a higher cost in the endoscopic group. This study only evaluated the cost during the initial year, and did not account for the ongoing high cost of proton pump inhibitor therapy. In addition, the diagnosis of at least two upper gastrointestinal malignancies was delayed in the empiric therapy group. A third randomised study of H pylori negative patients by Asante and colleagues found no differences in short term costs and quality of life for the empiric therapy versus initial endoscopy. This study did not evaluate the cost of long term proton pump therapy, the rate of subsequent endoscopy for failed empiric therapy, or the rate of missed cancers in the empiric group.

Decision analysis and cost effectiveness analyses have also been used to evaluate endoscopic diagnosis versus empiric therapy for dyspepsia. Silverstein and colleagues found
that the cost of initial endoscopy versus empiric therapy was a "toss up." The costs were influenced by the choice of H2 antagonists versus proton pump inhibitors, and by initial testing and treatment of H pylori. Initial endoscopy was less expensive if proton pump inhibitors were used, or if endoscopy was used to confirm ulcer disease prior to H pylori eradication.

Endoscopy based studies at referral centres are potentially biased by pre-selection of higher risk patients, such as those who have already failed empiric therapy with antisecretory medications. It is unlikely that our results are significantly affected by referral bias as the prevalence of major endoscopic findings in our study (787/3815 or 21%) is similar to that in the studies of unselected dyspeptic patients in general practice settings. Studies of clinical predictors of endoscopic findings in smaller primary care settings have also shown that symptoms are a poor predictor of anatomical pathology. We were unable to specifically determine if antacid therapy or non-steroidal anti-inflammatory medications affect the accuracy of our prediction rule.

Collection of clinical outcome data provides important evidence of the effectiveness and accuracy of decisions made in "real" clinical practice which cannot be generalised from highly controlled clinical trials. A potential limitation of studies performed in a clinical practice setting is the completeness and accuracy of the data collected. We have attempted to maintain data accuracy by requiring that each endoscopist enter a limited number of mandatory "critical" variables such as demographic information, indications, comorbidity, and endoscopic findings. We have also performed standardised data quality assessment procedures to ensure that these critical variables are accurate.

While it may be optimistic to expect age and "alarm" symptoms to predict endoscopic findings, these are still commonly used in clinical practice to decide which patients require endoscopy. Any imperfect prediction rule will require some sacrifice of sensitivity to gain specificity and vice versa. This is minimal in highly accurate prediction rules. The current recommendations of the American Gastroenterological Association for endoscopy are based on the presence of a predetermined set of "alarm" symptoms (anaemia, dysphagia, bleeding, etc.), although there are limited empirical data that these are the most important clinical predictors. Our study used multivariate analysis to determine which demographic features and symptoms were associated with major endoscopic findings among patients with dyspepsia. Although the results support the argument that age, anaemia, and bleeding symptoms are independent predictors of endoscopic findings, the predictive accuracy was very low, and there was no evidence that age >45 years is better than either higher or lower age cut-offs.

In summary, we have found that specific age cut-offs and "alarm" symptoms are inaccurate and should not be used for deciding who to select for endoscopy among patients with dyspepsia. To improve the identification of important upper gastrointestinal pathology, more effective strategies are needed among patients with dyspepsia. These include better clinical classification schemes, such as those already in development, larger prospective endoscopy and clinical database systems, less invasive and costly endoscopy, and standardisation of clinical and endoscopic terminology.

Dr Wallace was supported by the American Digestive Health Foundation, Olympus Advanced Endoscopy Training Award, and the Wilson Cook Endoscopic Research Career Development Award.