Hill classification is superior to the axial length of a hiatal hernia for assessment of the mechanical anti-reflux barrier at the gastroesophageal junction

Background and study aims: The pathogenesis of gastroesophageal reflux disease (GERD) is multifactorial, including the mechanical anti-reflux barrier of the gastroesophageal junction. This barrier can be evaluated endoscopically in two ways: by measuring the axial length of any hiatal hernia present or by assessing the gastroesophageal flap valve. The endoscopic measurement of axial length is troublesome because of the physiological dynamics in the area. Grading the gastroesophageal flap valve is easier and has proven reproducible. The aim of the present study was to compare the two endoscopic grading methods with regard to associations with GERD.

Patients and methods: Population-based subjects underwent endoscopic examination assessing the axial length of hiatus hernia, the gastroesophageal flap valve using the Hill classification, esophagitis using the Los Angeles (LA) classification, and columnar metaplasia using the Z-line appearance (ZAP) classification. Biopsies were taken from the squamocolumnar junction to assess the presence of intestinal metaplasia. Symptoms were recorded with the validated Abdominal Symptom Questionnaire. GERD was defined according to the Montreal definition.

Results: In total, 334 subjects were included in the study and underwent endoscopy; 86 subjects suffered from GERD and 211 presented no symptoms or signs of GERD. Based on logistic regression, the estimated area under the curve statistic (AUC) for Hill (0.65 [95% CI 0.59 – 0.72]) was higher than the corresponding estimate for the axial length of a hiatal hernia (0.61 [95% CI 0.54 – 0.68]), although the difference was not statistically significant (P=0.225).

Conclusion: From our data, and in terms of association with GERD, the Hill classification was slightly stronger compared to the axial length of a hiatal hernia, but we could not verify that the Hill classification was superior as a predictor. The Hill classification may replace the axial length of a hiatal hernia in the endoscopic assessment of the mechanical anti-reflux barrier of the gastroesophageal junction.

Introduction

Gastroesophageal reflux (GER) is a common disorder with an approximate prevalence of 10–20% in the Western world [1]. The clinical manifestations of GER (i.e. heartburn and acid regurgitation) are due to reflux of gastric content into the esophagus. The present theory for the pathogenesis of GER is multifactorial involving the lower esophageal sphincter (LES), diaphragmatic crus, esophageal acid clearance, gastric acid secretion, gastric emptying, and intra-abdominal pressure [2]. The Montreal definition of gastroesophageal reflux disease (GERD) states that GERD is present when the reflux of stomach contents causes troublesome symptoms and/or complications [3]. Possible complications include esophagitis and Barrett’s esophagus. In Barrett’s esophagus, the normal squamous epithelium of the distal esophagus has been replaced by a columnar epithelium, giving rise to a columnar lined esophagus (CLE). The definition of Barrett’s esophagus has always been controversial and no uniformly accepted criteria exist. The most widespread definition of Barrett’s esophagus requires histologically verified intestinal metaplasia in the segment of CLE [4].

The gastroesophageal junction (GEJ) is the anatomical area where the distal esophagus joins the proximal stomach. Under normal conditions, it is located at the level of the diaphragmatic crus. The location of the GEJ is however not static, and moves several centimeters during swallowing and breathing [5]. During swallowing, the longitudinal smooth muscle of the esophagus contracts which shortens the esophagus, resulting in a physiological herniation. The GEJ is later re-
turned to its original location by elastic supporting structures, especially by the phrenoesophageal membrane. When the GEJ together with the LES and the gastric cardia are permanently displaced upward into the thoracic cavity through the diaphragmatic hiatus, a hiatal hernia is present [2,5,6]. Studies have shown that a hiatal hernia impairs the LES pressure and the sphincter function of the diaphragm [7–9]. The presence and axial length of a hiatal hernia have also been shown to correlate with the severity of GER [10].

Esophagogastroduodenoscopy (EGD) is the standard investigation method for assessing the upper gastrointestinal tract. The competence of the mechanical anti-reflux barrier can be evaluated endoscopically in two ways; one way is to measure the axial length of any hiatal hernia present (between the hiatus and the GEJ). At endoscopy, the hiatus is represented by the diaphragmatic pinch and the GEJ is defined by the proximal margin of the gastric mucosal folds [11]. Due to the physiologic dynamics in this area, it can be difficult to measure the length of a hiatal hernia [5]. It has also been shown that the interobserver agreement with regard to endoscopic measurement is poor even under ideal conditions [12]. It is not clear at which length a hiatal hernia becomes clinically significant, and since the GEJ is not static, most endoscopists use a 2 cm cutoff [6]. Another way to assess the GEJ is by grading the gastroesophageal flap valve (GEFV) using the Hill classification (Figs. 1–4) [13]. Studies have shown an association between higher Hill grades and the frequency of GERD [8, 14]. Higher Hill grades are also associated with lower LES pressure [8], increased prevalence of hiatal hernia [15], and are able to predict poor response to proton pump inhibitor treatment [16]. The Hill classification has been proven to be reproducible and provides useful information when evaluating patients with suspected GERD who are undergoing endoscopy [8]. Esophagitis can be defined endoscopically and classified according to the Los Angeles (LA) classification [17]. The extent of CLE can be evaluated in a standardized manner endoscopically by the Z-line appearance (ZAP) classification. The ZAP classification has been proven highly reproducible, and is associated both with the prevalence of intestinal metaplasia, as well as with GERD [18–20].

The aim of the present study was to investigate how the two ways of assessing the competence of the mechanical anti-reflux barrier of the GEJ correlated with GERD. We hypothesized that the Hill classification would be superior to measuring the axial length of any hiatal hernia present in terms of association with GERD.

Materials and methods

A thorough description of the setting, population, endoscopy, and symptom evaluation has been given elsewhere [21].

Study population

A population cohort in Östhammar, Sweden, has been studied for over two decades with regard to gastrointestinal symptoms. The study population is representative of the general Swedish population with regard to gender, age, income, and other potential selection factors. In 1988, a questionnaire with regard to abdominal symptoms (ASQ) [22] was sent by mail to the cohort for the first time. In 1989, 1995, and 2011 slightly updated forms of questionnaires were sent to the same population. In 2012, the same population was invited to participate in the present study. To be invited to the EGD, the subjects had to have participated in the 2011 study. Fig. 5 illustrates the study population and the dropouts.

Exclusion criteria for EGD were angina pectoris, myocardial infarction (in the last 6 months), congestive heart failure, severe lung disease, severe liver disease, esophageal varices, treatment with anticoagulants, need for anesthesia for the endoscopy, earlier surgery of the stomach, and those 80 years or above in age. Of the 388 individuals who completed the EGD, 54 were excluded because of missing data, leaving 334 individuals in the present study.

Endoscopy

Each EGD was performed by one of five experienced endoscopists. The endoscopists were scheduled to work at the research facility during different weeks. Research assistants, who were unaware which endoscopist was going to perform the endoscopy, invited the eligible participants to the endoscopy. Before the
study, a consensus meeting led by an external expert (Professor Lars Lundell) reviewed multiple video recordings according to the study protocol. Each endoscopist was monitored on the first day by the project leader (LA) and a part of the procedure (the hiatus and distal esophagus) was video recorded.

A structured endoscopic protocol was used to investigate the esophagus, stomach, and upper duodenum. Biopsies for histopathologic examination were taken according to a strict protocol. The mechanical anti-reflux barrier of the gastroesophageal junction was assessed in two ways:

1. The axial length of any hiatus hernia present: The axial length of the hiatal hernia was defined as the distance between the GEJ and the hiatus. Using the hash marks on the endoscope, the distance between the GEJ (represented by the transition from the gastric folds to tubular esophagus) and the hiatus (represented by the diaphragmatic pinch) was measured at the incisors. The axial length of the hiatal hernia was measured in centimeters on withdrawal of the endoscope [23].

2. The GEFV or Hill classification was graded I–IV according to the Hill classification [13] (Fig. 1–4).

If esophagitis was present, it was graded according to the Los Angeles (LA) classification [17]. If columnar metaplasia was present, it was graded according to the ZAP classification [20]:

- Grade 0: The Z-line is sharp and circular.
- Grade I: The Z-line is irregular and shows tongue-like protrusions and/or islands of columnar epithelium.
- Grade II: Distinct tongues of columnar epithelium < 3 cm can be seen.
- Grade III: Distinct tongues of columnar epithelium, or a cephaled displacement of the Z-line, >3 cm, can be seen.

**Definition of GERD**

The Montreal Definition states that GERD is present when the reflux of stomach contents causes troublesome symptoms and/or complications [3]. It has previously been shown that reflux symptoms present at least weekly impair quality of life [24] and may thus be considered troublesome. The most common complications of GERD are esophagitis and Barrett’s esophagus. Hence, for this study, GERD was defined as present if (1) the subject reported acid regurgitation and/or heartburn on at least a weekly basis, and/or (2) esophagitis or Barrett’s esophagus was present. Barrett’s esophagus was considered present if intestinal metaplasia was found in a segment of ZAP Grade II or III (i.e. distinct, obvious tongues of metaplastic, columnar-appearing epithelium).

**Histology**

For the present study, biopsies from the squamocolumnar junction were investigated. To eliminate any doubt concerning the origin of the intestinal metaplasia, the intention was to obtain biopsies from the Z-line containing both squamous and columnar epithelium. The biopsy specimens were fixed in 4% phosphate-buffered formaldehyde then processed and embedded in paraffin.
GER, gastroesophageal reflux; GERD, gastroesophageal reflux disease; ZAP, Z-line appearance.

All values except mean are given as n and (%). GERD was considered to be present if (1) the subject reported acid regurgitation and/or heartburn on a weekly basis, and/or (2) esophagitis or Barrett’s esophagus was present. Barrett’s esophagus was considered present if intestinal metaplasia was found in a segment of ZAP Grade II or III (i.e., distinct, obvious tongues of metaplastic, columnar-appearing epithelium). GER was considered present if the subject reported acid regurgitation and/or heartburn less frequently than on a weekly basis, without signs of esophagitis or Barrett’s esophagus.

**Statistics**

Using logistic regression models, we compared the predictive power of the two endoscopic measurements (hiatal hernia length and Hill classification) on GERD. The evaluation of hiatal hernia was based on hiatal hernia as a continuous variable and as factor variables categorized as (0, 1, 2, 3, 4, 5, ≥6 cm), (≥2 cm, >2 cm) and (≥3 cm, >3 cm). The Hill classification was evaluated as a continuous variable (based on category scores 1–4 representing grades I–IV) and as factor variables: grades (I, II, III, IV) and (≤II, ≥III). The evaluation of the Hill classification and hiatal hernia as a predictor of GERD was performed in two steps.

1. The Akaike’s information criterion (AIC) and Bayesian information criterion (BIC) were used to find the best candidate variable type for each of Hill and hiatal hernia (from the different types of variables listed above), given no other variables in the model. AIC and BIC penalize for the addition of parameters and thus we avoid that the final candidate type of variable automatically is the one with the most parameters (e.g., hiatal hernia categorized in 1 to 6 cm vs hiatal hernia dichotomized). For each endoscopic measurement, a likelihood ratio chi-squared test was performed to test if a model containing an endoscopic measurement statistically significantly contributed to the model.

2. For each of the best candidate variable types of the Hill classification and hiatal hernia, the predictive power was evaluated by constructing receiver–operating characteristic (ROC) curves and calculating the area under curve (AUC) statistic with 95% confidence intervals. To assess how the results would generalize to an independent data set, 10-fold data cross-validations (removing 10 fold for the testing set and modeling the remaining training set) of the AUC estimates were performed. Bias corrected and accelerated (BCa) confidence intervals were estimated by bootstrapping the data and repeating the cross-validation 1000 times. All two-sided P values <0.05 were considered statistically significant.

Data analyses were performed with Stata/IC software (Stata Statistical Software: Release 13, StataCorp LP, College Station, TX, United States). Cross-validation was performed in the statistical software package R (version 3.2.2, R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria).

**Ethics**

The 1989 study was approved by the Ethical Review Board of the Medical Faculty of Uppsala University (Dnr. 1989/220). Approval for the 2011–2012 study was obtained from the Ethics Committee of Uppsala University (Dnr. 2010/443), and all participants gave their informed consent.

**Results**

A total of 334 subjects were included in the study. Out of these, 86 subjects were found to have GERD according to the Montreal definition and 211 did not have GERD. Thirty-seven subjects reported GER less frequently than on a weekly basis, and none of these 37 were found to have esophagitis or Barrett’s esophagus. Since these 37 subjects reported symptoms suggestive of GER, but did not fulfill the criteria for GERD, they were excluded from the final analysis so as to define the groups of subjects with and without GERD more clearly (Table 1). The subjects with GERD had a mean Hill-grade of 2.7 and a mean hiatal hernia length of

---

**Table 1** Demographic data and endoscopic findings in the subjects with and without GERD, and in the subjects with more infrequent symptoms of GER.
1.9 cm, while the subjects without GERD had a mean Hill-grade of 2.1 and a mean hiatal hernia length of 1.3 cm. The 37 subjects with less frequent symptoms had values in between those with and without GERD, with a mean Hill-grade of 2.2 and a mean hiatal hernia length of 1.6 cm.

The concordance between the two endoscopic classification systems is presented in Table 2 (Kendall Tau-β correlation coefficient: 0.38).

Of the different candidate variables representing hiatal hernia from logistic regression, the candidate with hiatal hernia as a continuous variable resulted in the best prediction of GERD with AIC: 351 (likelihood ratio chi-squared test: \( P = 0.0013 \)). Correspondingly, the Hill classification showed the best prediction of GERD when the model was based on category scores (1 – 4) as a continuous variable with AIC: 342 (likelihood ratio chi-squared test: \( P < 0.0001 \)) (Table 3). Including this best-fitting Hill classification variable given that the corresponding hiatal hernia variable was also in the model improved the fit of the model (likelihood ratio chi-squared test: \( P = 0.0008 \) for including Hill as a continuous variable based on category scores). Conversely, including the best-fitting variable of hiatal hernia given that the corresponding Hill classification was also in the model did not improve the fit of the model (likelihood ratio chi-squared test: \( P = 0.1393 \)).

The predictive powers of the best-fitting model for each classification of hiatal hernia and Hill are illustrated in Fig. 6 by ROC curves. The area under the ROC curve was 0.61 (95%CI 0.54–0.68) for hiatal hernia and 0.65 (95%CI 0.59–0.72) for the Hill classification. The difference between the two AUC estimates was not statistically significant (\( P = 0.225 \)). The corresponding 10-fold cross-validated estimates were 0.58 (95%CI 0.51–0.65) for hiatal hernia and 0.62 (95%CI 0.53–0.68) for Hill.

### Table 2

Concordance between the Hill classification and the axial length of hiatal hernia.

<table>
<thead>
<tr>
<th>cm</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>43</td>
<td>40</td>
<td>15</td>
<td>3</td>
<td>101</td>
</tr>
<tr>
<td>1</td>
<td>30</td>
<td>33</td>
<td>21</td>
<td>3</td>
<td>87</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>16</td>
<td>30</td>
<td>18</td>
<td>78</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>8</td>
<td>17</td>
<td>8</td>
<td>40</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>12</td>
<td>19</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>96</td>
<td>100</td>
<td>89</td>
<td>49</td>
<td>334</td>
</tr>
</tbody>
</table>

### Table 3

Logistic regression with GERD as dependent variable and different ways of looking at Hill grade (I – IV) and hiatal hernia length (cm), as independent variables. The group of hiatal hernia ≥ 6 cm consisted of two 6-cm hiatal hernias, one 7-cm hiatal hernia, and one 8-cm hiatal hernia.

<table>
<thead>
<tr>
<th>hiatal hernia</th>
<th>( P ) value</th>
<th>AIC-value</th>
<th>BIC-value</th>
<th>OR</th>
<th>95%CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>continuous</td>
<td>0.0013</td>
<td>351.1</td>
<td>358.5</td>
<td>1.34</td>
<td>1.12–11.61</td>
</tr>
<tr>
<td>ordinal</td>
<td>0.0771</td>
<td>359.7</td>
<td>385.5</td>
<td>1.10</td>
<td>0.53–2.30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>hiatal hernia</th>
<th>( P ) value</th>
<th>AIC-value</th>
<th>BIC-value</th>
<th>OR</th>
<th>95%CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥ 2 cm (dichotomous)</td>
<td>0.0035</td>
<td>352.9</td>
<td>360.3</td>
<td>2.12</td>
<td>1.28–3.53</td>
</tr>
<tr>
<td>≥ 3 cm (dichotomous)</td>
<td>0.0055</td>
<td>353.7</td>
<td>361.1</td>
<td>2.34</td>
<td>1.29–4.22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hill (continuous)</th>
<th>( P ) value</th>
<th>AIC-value</th>
<th>BIC-value</th>
<th>OR</th>
<th>95%CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 0.0001</td>
<td>342.0</td>
<td>349.4</td>
<td>1.75</td>
<td>1.36–2.27</td>
<td></td>
</tr>
<tr>
<td>(ordinal)</td>
<td>0.0001</td>
<td>344.5</td>
<td>359.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hill ≥ III (dichotomous)</th>
<th>( P ) value</th>
<th>AIC-value</th>
<th>BIC-value</th>
<th>OR</th>
<th>95%CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 0.0001</td>
<td>343.5</td>
<td>350.8</td>
<td>3.02</td>
<td>1.80–5.06</td>
<td></td>
</tr>
</tbody>
</table>

AIC, Akaike’s information criterion; BIC, Bayesian information criterion; GERD, gastroesophageal reflux disease.

---

**Fig. 6** Receiver-operating characteristic (ROC) curves for each of the hiatal hernia and Hill classifications that resulted in the best prediction of GERD (hiatal hernia as a continuous variable and Hill as a continuous variable based on category scores).
Analysis of the endoscopists revealed that one endoscopist (Group A) did 125 endoscopies, another endoscopist (Group B) did 114 endoscopies, and the remaining 95 endoscopies were done by three endoscopists (Group C). The three groups did not differ with regard to demographics or prevalence of GERD. One of the endoscopists rated both the Hill grade and the axial length of a hiatal hernia significantly higher (Group B) than the other two groups. Excluding the endoscopies done by this endoscopist did not change the results with regard to the association between GERD and the endoscopic classification systems among the remaining endoscopists. The subjects in Group B did, however, only reach a statistically significant association between GERD and the Hill classification, and not with the axial length of a hiatal hernia. The biopsies from the squamocolumnar junction contained both squamous and columnar epithelium in 81% of the subjects.

Discussion

The present study compares two different ways to assess the anti-reflux barrier of the GEJ (axial length of a hiatal hernia and GEVF graded by the Hill classification) endoscopically and the association between these two assessments and GERD. Of these two techniques, the most widely used is to assess the presence and axial length of any hiatal hernia. It has, however, been shown that the interobserver agreement with regard to endoscopic length measurement is poor even under ideal conditions. As a result of this, the concept of hiatal hernia length also suffers from this inherent weakness. Furthermore, there is no consensus at which length the physiological movement of the GEJ becomes a hiatal hernia.

An important strength of this study is that the same study population and sample (with the same abdominal symptom questionnaire (ASQ) and esophagagogastroduodenoscopy (EGD)) were used for both of the endoscopic grading methods. Because of this, we did not adjust for other variables in the logistic regression models (i.e. the need to adjust for confounding due to heterogeneous populations was taken care of by the design). Confounding could still be an issue if the causal mechanisms between the two methods and GERD differed. However, the axial length of a hiatal hernia and the Hill classification evaluate the same mechanism but in different ways and what affects the hiatal hernia length will probably also affect the Hill grade. Another strength was that the endoscopies were done by a rather large group of endoscopists. Even though one of the endoscopists rated the Hill grade and the axial hiatal hernia length differently, this did not affect our main finding. The differences between endoscopists are a reflection of the difficulty in the rating of hiatal hernia and Hill, something that affects the predictive power of the endoscopic measurements. However, the comparison between hiatal hernia and Hill as predictors should not be biased, as GERD was not associated with endoscopist.

A weakness of the present study is the method used to define GERD; including a more objective way of measuring GERD such as 24-h pH measurement might have been useful as a complement to the anamnestic information used. The axial hiatal hernia length and the Hill grade were tested in different ways to determine the strongest predictive power for GERD. Comparing the association between GERD and the two endoscopic classification systems based on logistic regression, the Hill classification showed a slightly stronger association with GERD in our data with AIC- and BIC-values generally lower than those for the axial length of a hiatal hernia (Table 3). The estimated AUC for Hill (0.65 [95% CI 0.59 – 0.72]) was also higher than the AUC for hiatal hernia (0.61 [95% CI 0.54 – 0.68]). However, this study could not statistically significantly verify that Hill was superior as a predictor of GERD compared to hiatal hernia. Since the Hill classification has been shown to be highly reproducible, and since it is a known fact that endoscopic length measurement is difficult even under ideal conditions, it seems reasonable to consider using the Hill classification instead of hiatal hernia length in assessment of the anti-reflux barrier.

Competing interests: None

Institutions

1 Department of Surgical and Perioperative Sciences, Surgery, Umeå University Hospital, Umeå, Sweden
2 Department of Radiation Science, Oncology, Umeå University, Umeå, Sweden
3 Division of Family Medicine, Karolinska Institutet, Huddinge, Sweden
4 Stress Research Institute, Stockholm University, Stockholm, Sweden
5 Uppsala University, Uppsala, Sweden
6 Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden
7 Faculty of Medicine, University of Newcastle, Newcastle, Australia
8 Institute of Pathology, Klinikum Bayreuth, Bayreuth, Germany

Acknowledgments

The present study was supported by Olympus Solna Sverige AB, who supplied equipment for realization of the study. The authors have no other affiliation with this company.

References

8 Kim GH, Kang DH, Song GA et al. Gastroesophageal flap valve is associated with gastroesophageal and gastropharyngeal reflux. J Gastroenterol 2006; 41: 654–661
12 Guda N, Partington S, Vakil NB. Inter- and intra-observer variability in the measurement of length at endoscopy: Implications for the measurement of Barrett’s esophagus. Gastrointest Endosc 2004; 59: 655–658