



# REVISTA DE GASTROENTEROLOGÍA DE MÉXICO

[www.elsevier.es/rgmx](http://www.elsevier.es/rgmx)



## REVIEW ARTICLE

# New tools for the grading and scoring of gastro-esophageal reflux disease



S. Siboni<sup>a,\*</sup>, M. Sozzi<sup>a</sup>, P. Visaggi<sup>b</sup>, R. De Maron<sup>a</sup>, E. Asti<sup>a</sup>

<sup>a</sup> División de Cirugía General y de Urgencia, IRCCS Policlinico San Donato, Universidad de Milán, Milan, Italy

<sup>b</sup> División de Gastroenterología, Departamento de Investigación Traslacional y Nuevas Tecnologías, Universidad de Pisa, Pisa, Italy

Received 16 April 2025; accepted 14 May 2025

Available online 10 September 2025

## KEYWORDS

Gastro-esophageal  
reflux disease;  
Endoscopy;  
High-resolution  
manometry;  
Reflux monitoring  
study

**Abstract** Gastroesophageal reflux disease (GERD) is a prevalent condition characterized by the retrograde flow of gastric contents into the esophagus, significantly impacting quality of life. Traditional diagnostic approaches often lack precision due to symptom overlap with other conditions. This review introduces the American Foregut Society (AFS) classification, Milan score, pHoenix score, COuGH RefluX score, and Lyon score, five novel tools designed to enhance the objectivity, reproducibility, and clinical relevance of GERD diagnosis.

The AFS classification refines the endoscopic assessment of esophagogastric junction (EGJ) integrity by incorporating measurable parameters (hiatal hernia length, hiatal aperture diameter, and flap valve), overcoming the subjectivity of the Hill classification. The Milan score, derived from high-resolution manometry, integrates four parameters (ineffective esophageal motility, EGJ-contractile integral, EGJ morphology, and straight leg raise response) to quantify anti-reflux barrier (ARB) disruption.

The pHoenix score, developed for prolonged wireless pH monitoring, weights supine AET more heavily, addressing limitations of the DeMeester score and Lyon 2.0 consensus. The COuGH RefluX score, a clinical prediction model for laryngopharyngeal symptoms, uses six parameters (cough, obesity, globus, hiatal hernia, regurgitation, male sex) to stratify GERD likelihood. Finally, the Lyon score integrates endoscopic and pH-impedance data, categorizing patients into phenotypes (from no GERD to severe GERD) and predicting treatment outcomes.

These tools collectively address diagnostic challenges by standardizing assessments and improving patient stratification. By reducing diagnostic ambiguity and guiding personalized therapy, these innovations hold promise for transforming GERD management, particularly in selecting candidates for escalated medical or surgical interventions.

© 2025 Asociación Mexicana de Gastroenterología. Published by Masson Doyma México S.A. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

\* Corresponding author at: División de Cirugía General y de Urgencia, IRCCS Policlinico San Donato, Piazza Malan 1, San Donato Milanese, 20097, Milano, Italy. Tel.: +39 0252774896.

E-mail address: [Stefano.siboni@grupposandonato.it](mailto:Stefano.siboni@grupposandonato.it) (S. Siboni).

## PALABRAS CLAVE

Enfermedad por reflujo gastroesofágico; Endoscopia; Manometría de alta resolución; Estudio de monitoreo de reflujo

## Nuevas herramientas para la clasificación y puntuación de enfermedad por reflujo gastroesofágico

**Resumen** La enfermedad por reflujo gastroesofágico (ERGE) es una condición prevalente caracterizada por el flujo retrógrado de contenido gástrico al esófago, con un impacto significativo en la calidad de vida. Los abordajes diagnósticos tradicionales a menudo carecen de precisión debido al traslape de síntomas con otras condiciones. La presente revisión presenta la clasificación de la *American Foregut Society* (AFS), el puntaje de Milán, el puntaje pHoenix, el puntaje COUGH Reflux y el puntaje de Lyon, cinco nuevas herramientas diseñadas para mejorar la objetividad, reproducibilidad y la relevancia clínica del diagnóstico de ERGE.

La clasificación de la AFS refina la evaluación endoscópica de la integridad de la unión esofagogastrica (UEG) al incorporar parámetros medibles (longitud de hernia hiatal, diámetro de apertura hiatal y válvula "flap"), superando la subjetividad de la clasificación de Hill. El puntaje de Milán, derivado de la manometría de alta resolución (MAR), integra cuatro parámetros (motilidad esofágica ineficaz, integral contráctil de UEG, morfología UEG y respuesta de elevación de pierna recta) para cuantificar la disrupción de la barrera antirreflujo (BAR).

El puntaje de pHoenix, desarrollado para el monitoreo prolongado remoto del pH, mide el tiempo de exposición a ácido (TEA) supino más profundamente, abordando las limitaciones del puntaje de DeMeester y el consenso de Lyon 2.0. El puntaje COUGH Reflux, un modelo de predicción clínica para síntomas laringofaríngeos, utiliza seis parámetros (tos, obesidad, globo, hernia hiatal, regurgitación y sexo masculino) para estratificar la probabilidad de ERGE. Finalmente, el puntaje de Lyon integra datos endoscópicos y de impedancia de pH, categorizando pacientes en fenotipos (de sin ERGE a ERGE grave) y prediciendo desenlaces de tratamientos.

Estas herramientas, de manera colectiva, abordan los desafíos de diagnóstico al estandarizar las evaluaciones y mejorar la estratificación de los pacientes. Al reducir la ambigüedad diagnóstica y guiar la terapia personalizada, estas innovaciones prometen transformar el manejo de ERGE, particularmente al seleccionar candidatos para intervenciones médicas o quirúrgicas de mayor impacto.

© 2025 Asociación Mexicana de Gastroenterología. Publicado por Masson Doyma México S.A. Este es un artículo Open Access bajo la CC BY-NC-ND licencia (<http://creativecommons.org/licencias/by-nc-nd/4.0/>).

## Introduction

Gastroesophageal reflux disease (GERD) is a common condition where gastric contents flow backward into the esophagus, causing symptoms that can significantly affect quality of life.<sup>1</sup> GERD presents with a wide range of symptoms, which are often categorized into two groups: "typical" symptoms, such as heartburn and regurgitation, which have a high likelihood of being linked to GERD, and "atypical symptoms", like cough, asthma or hoarseness, which are less reliably associated with the condition.<sup>2</sup>

For patients experiencing typical GERD symptoms with no alarm symptoms, such as unintended weight loss or dysphagia, the American Gastroenterological Association recommends an 8-week trial of proton pump inhibitor (PPI) therapy, taken once daily before a meal. This recommendation is supported by moderate-level evidence and is considered a strong guideline.<sup>3</sup> Diagnostic workup is typically indicated after the PPI trial, in case of inadequate response, or upfront, if the presenting symptoms are atypical.<sup>4</sup>

The first-line diagnostic test for GERD is typically endoscopy. This procedure can confirm GERD by identifying specific findings, such as Los Angeles grade B, C, or D esophagitis, histologically confirmed Barrett's esophagus, or

a peptic stricture. Additionally, endoscopy provides valuable insights into the extent of esophagogastric junction (EGJ) disruption, assessed using the Hill classification, and can detect a hiatal hernia, a condition where the lower esophageal sphincter (LES) and the crural diaphragm (CD) are separated, which is a major risk factor for GERD.<sup>5,6</sup>

If endoscopy does not provide conclusive evidence of GERD or if surgical intervention is being considered, esophageal function testing becomes necessary. High-resolution manometry (HRM) is a key tool in this phase. While HRM cannot directly diagnose GERD, it plays a critical role, beyond simply guiding catheter placement for reflux testing, by ruling out other conditions that may mimic GERD, such as achalasia, other motility disorders, or behavioral conditions.<sup>7–9</sup> Over the past decade, HRM has been extensively explored to identify variables and provocative maneuvers that can quantify the disruption of the anti-reflux barrier (ARB) and differentiate GERD patients from healthy individuals.<sup>10–15</sup>

For cases in which endoscopy is inconclusive, reflux monitoring is considered the gold standard for diagnosing GERD. Over the years, different criteria have been developed for interpreting 24 h catheter-based reflux monitoring studies. In 1974, Johnson and DeMeester introduced the DeMeester score, which evaluates 6 reflux parameters and remains

widely used by surgeons to select candidates for anti-reflux surgery (ARS).<sup>16</sup> More recently, the Lyon 2.0 consensus proposed updated benchmarks: an acid exposure time (AET) greater than 6% indicates definitive GERD, an AET between 4 and 6% is inconclusive, and an AET below 4% suggests no GERD. For inconclusive cases (AET 4-6%), additional criteria are used to confirm or rule out GERD, including a total of more than 80 reflux episodes per day, a mean nocturnal baseline impedance (MNBI) below 1,500 ohms, or an association between reflux events and symptoms.<sup>4,17</sup>

While these established criteria and tests provide solid confidence in objective GERD diagnosis, several novel tools to assess GERD severity and treatment response have been developed in the past decade. These innovations aim to make GERD diagnosis more precise and less invasive, and to better identify patients who may benefit from treatment escalation. Our review discusses these new tools, focusing on current evidence, practicality, potential impact on clinical practice, and limitations.

## The American Foregut Society Classification

### Introduction

Endoscopy is the first-line diagnostic test for gastrointestinal symptoms, enabling direct visualization of the upper gastrointestinal tract. It plays a critical role in excluding conditions, such as malignancies, gastric atrophy, and peptic ulcer disease, from the GERD diagnostic pathway.<sup>3,18</sup> The Lyon 2.0 consensus further solidifies endoscopy's crucial role by defining specific findings, such as Los Angeles grade B, C, or D esophagitis, Barrett's esophagus, or peptic stricture, as diagnostic for GERD. Beyond diagnosis, endoscopic evaluation of the EGJ provides essential insights into the anatomic factors contributing to reflux. To address limitations in earlier EGJ classifications, the American Foregut Society (AFS) has introduced a novel classification system for assessing EGJ integrity, offering a more comprehensive and standardized approach to GERD evaluation.<sup>19</sup>

### Background and limitations of previous scores

Since 1996, the Hill classification has been considered the primary tool for endoscopic assessment of the gastroesophageal flap valve, a critical component of the anti-reflux barrier (ARB).<sup>5</sup> Despite its relevance, the Hill classification has not been fully adopted in routine practice due to several limitations. It primarily focuses on the flap valve, giving only minimal attention to hiatal hernia. Additionally, its reliance on subjective assessments rather than measurable parameters, without a standardized endoscopic technique or nomenclature, has reduced its reliability. For these reasons, the distinction between grades I and II, both considered normal, lacks clinical significance, whereas the differentiation between grades II and III shows poor correlation with pathologic reflux rates. These shortcomings have limited the Hill classification's utility in both clinical practice and research, highlighting the need for a more robust system.

### Description of the score

The AFS classification represents a significant improvement by incorporating objective measurements of three key components of the ARB, summarized by the acronym "LDF": L (length) measures hiatal hernia axial length, D (diameter) assesses the hiatal aperture size, using the standard endoscopic diameter of approximately 1 cm as a reference, and F (flap valve) evaluates the presence (F+) or absence (F-) of a functioning gastroesophageal flap valve at the angle of His (Fig. 1).

This system assigns grades from 1 (normal) to 4 (severe anatomic disruption), with the final grade determined by the component exhibiting the most significant abnormality, reflecting the weakest link in the ARB. To ensure consistency, the AFS classification provides a clear methodology, recommending prolonged insufflation (30-45 s) and rotational maneuvers in the retroflexed position, to assess for sliding hiatal herniation. These standardized techniques may reduce inter-observer variability and help to prevent undergrading of EGJ disruption, addressing a key limitation of previous systems.

### Evidence and validation studies




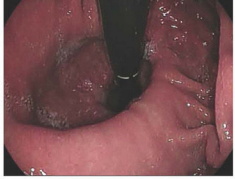
Recent prospective studies have validated the AFS classification using objective physiologic measurements.<sup>12-20</sup> In a study involving 56 patients with suspected GERD, who underwent endoscopy, HRM, and reflux monitoring study, the AFS classification showed a strong correlation with both pathologic reflux and manometric EGJ disruption.<sup>20</sup>

The study found a progressive increase in the prevalence of pathologic AET (>6%) across the AFS grades: 0% in grade I, 5.9% in grade II, 52% in grade III, and 77.8% in grade IV ( $p < 0.001$ ). In contrast, the Hill classification demonstrated poor discriminatory ability in the same population, with similar rates of pathologic reflux in grades II and III (42.1% vs. 37.9%,  $p = 0.411$ ), underscoring its limitations.

In the same study, each component of the AFS classification correlated with specific HRM parameters, reflecting different mechanisms of EGJ disruption: the L component correlated with EGJ morphology and intra-abdominal LES length, the D component showed significant association with EGJ-contraction integral (EGJ-CI) and LES basal pressure, and the F component demonstrated correlation with the Straight Leg Raise (SLR) maneuver. These findings confirm the physiologic relevance of the AFS classification components and its ability to stratify patients according to the severity of EGJ disruption. Another recent paper demonstrated the superiority of the AFS classification to the Hill classification, in terms of inter-observer variability, and confirmed its superior ability to predict AET.<sup>21</sup>

### Clinical Implications

Though still in the early stages of validation, the AFS classification offers a promising advancement in endoscopic EGJ assessment. Its precise, reproducible protocol, with measurable parameters, minimizes subjective interpretation, potentially improving patient selection for further investiga-

|                                | Grade 1   | Grade 2   | Grade 3  | Grade 4   |
|--------------------------------|---|---|--|---|
|                                |  |  |  |  |
|                                | <b>1</b>  | <b>2</b>  | <b>3</b>   | <b>4</b>  |
| <b>Hiatal axial length (L)</b> | None (0 cm)   | None (0 cm)   | 0–2 cm   | >2 cm   |
| <b>Hiatal aperture (D)</b>     | Snug to scope<br>1 cm   | Loose<br>1–2 cm   | Open<br>2–3 cm   | Wide open<br>> 3 cm   |
| <b>Flap valve (F)</b>          | Present, full lip<br>with Omega<br>shape (F+)                                     | Absent, thinning<br>& flattening<br>valve lip (F-)                                | Absent (F-)  | Absent (F-)   |

**Figure 1** The American Foregut Society classification of esophagogastric junction integrity.

tions and reducing unnecessary diagnostic testing in patients with an intact EGJ.

### Limitations and future research

Despite its advancements, the AFS classification has certain limitations that warrant consideration. The detailed assessment methodology may require additional training for endoscopists, and further studies on inter-observer variability are needed to ensure consistency among practitioners. Additionally, the recommended procedures, such as prolonged insufflation and provocative maneuvers, extend the duration of routine endoscopy, which could pose challenges in busy clinical settings. Finally, current validation studies have also been limited to tertiary referral centers for esophageal diseases, potentially limiting applicability to community practice. Finally, the AFS classification does not yet address the evaluation of patients who have undergone antireflux interventions, even though the American Foregut Society has recently proposed a separate comprehensive endoscopic evaluation for post-ARS patients.<sup>22</sup>

Additional studies in different clinical settings are needed to confirm the classification's utility in real-world scenarios outside of tertiary care centers, to assess its long-term predictive value for guiding intervention selection, and to further validate its ability to predict pathologic GERD and stratify patients effectively.

### Conclusion

The AFS classification is a promising advancement in the endoscopic assessment of EGJ integrity, providing a standardized and objective framework for evaluating patients with suspected GERD. Although further validation is needed, particularly regarding its generalizability across different practice settings and prognostic value, the classification

holds substantial potential to improve both clinical practice and research in GERD management.

### The Milan score

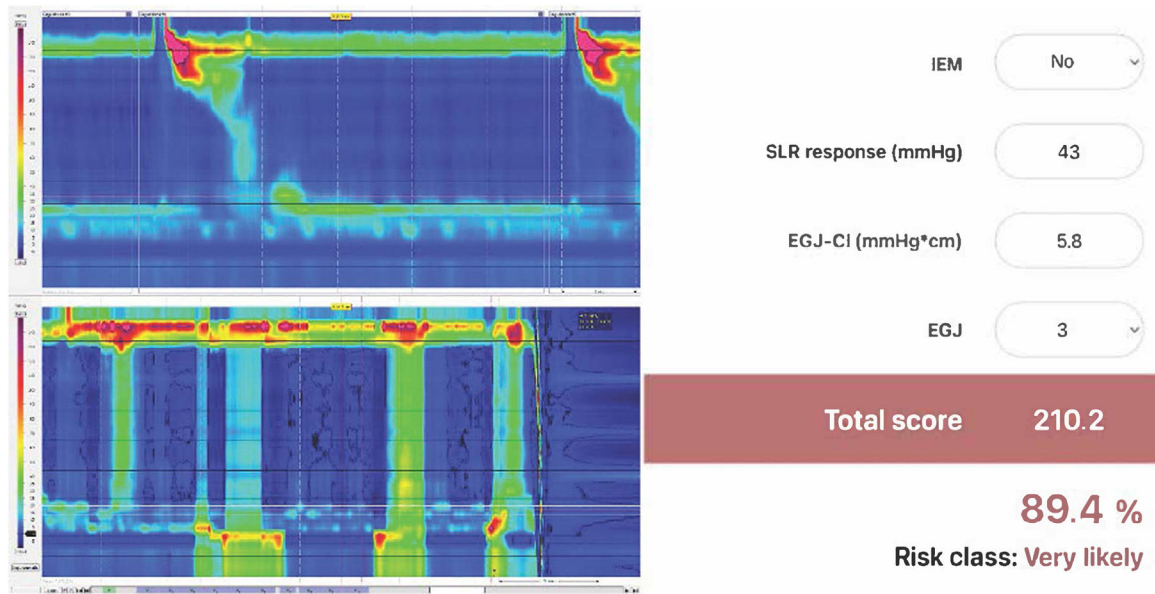
#### Introduction

HRM plays an ancillary role in the diagnostic pathways of GERD, primarily used to exclude major motility disorders that mimic GERD symptoms and to accurately localize the LES for reflux monitoring catheter placement.<sup>8</sup> However, despite its secondary role, consensus papers emphasize HRM's growing importance in evaluating ARB disruption and in the preoperative assessment prior to ARS.<sup>9,23–26</sup>

Unlike traditional gold-standard methods (pH monitoring and endoscopy), which focus on detecting acid reflux or esophageal lesions, HRM offers a unique ability to quantify underlying functional abnormalities. The recent introduction of the Milan score leverages this capability, providing a single parameter to assess ARB disruption and predict objective GERD.

#### Background and limitations of previous scores

The exploration of manometric abnormalities correlating with objective GERD measures in patients with upper gastrointestinal symptoms is well-documented.<sup>27</sup> One of the fathers of modern esophageal surgery, Dr. DeMeester, extensively studied conventional manometry potential. He demonstrated the impact of LES characteristics, including total and intra-abdominal LES length and basal pressure,<sup>28</sup> body contraction,<sup>29</sup> hiatal hernia, and EGJ response to increased abdominal pressure,<sup>30,31</sup> on the pathophysiology of GERD. These factors have been adapted into HRM metrics, including the LES Pressure Integral (LESPI),<sup>32</sup> EGJ-CI,<sup>10</sup> Ineffective Esophageal Motility (IEM),<sup>33</sup> EGJ morphology,<sup>11,34–37</sup>



**Figure 2** Example of a HRM plot and the calculation of the Milan score using the website [www.milanscore.com](http://www.milanscore.com).

thoracoabdominal pressure gradient (TAPG),<sup>38</sup> and the Straight Leg Raise (SLR) maneuver.<sup>12,13</sup>

In 2020, Masuda et al.<sup>14</sup> proposed a manometric index combining EGJ morphology, LESPI, and TAPG. However, its receiver operating characteristics (ROC) analysis yielded an area under the curve (AUC) of 0.615, with a sensitivity of 56% and specificity of 60.7% at the optimal cut-off, indicating limited diagnostic accuracy and highlighting the need for a more robust scoring system.

### Description of the score

The Milan score is designed to quantify ARB disruption and to estimate the risk of objective GERD in patients undergoing HRM and pH-studies for persistent symptoms. It integrates 4 key HRM parameters:

- 1 IEM: defined as > 70% ineffective swallows or  $\geq 50\%$  failed swallows, it is a measure of esophageal clearance time.
- 2 EGJ-CI: calculated during the reference period using the distal contractile integral (DCI) tool placed over the EGJ, and adjusted for respiration and gastric pressure, it quantifies the strength and duration of EGJ contraction, reflecting its competency.
- 3 EGJ morphology: classified as type 1 in cases of superimposed LES and CD, type 2 in cases of LES-CD separation < 3 cm, and type 3 with separation  $\geq 3$  cm. EGJ morphology identifies the presence and size of hiatal hernia.
- 4 SLR response: performed in the supine position as previously described,<sup>13</sup> with double leg raise if intra-abdominal pressure (IAP) augmentation was insufficient with single leg raising. The SLR measures the ability of the EGJ to counteract the IAP increase, a major pathophysiologic factor of GERD.

The Milan score employs a mathematical formula to weigh these parameters, yielding a final score. The

Milan score is computed using the online calculator tool ([www.milanscore.com](http://www.milanscore.com)) (Fig. 2). A value  $\geq 137$  indicates a 50% risk for objective GERD. This structured approach provides a reproducible measure of ARB dysfunction.

### Evidence and validation studies

The Milan score was developed using a multicenter cohort of 295 patients and externally validated in a separate cohort of 233 patients.<sup>39</sup> The ROC analysis showed an AUC of 0.880 in predicting pathologic GERD, with strong discrimination and calibration (corrected Harrell's c-index = 0.90, integrated calibration index 0.07) in the validation cohort. Subsequent studies have explored its applicability across different clinical scenarios. In a paper published by the same multicenter group, the effectiveness of the Milan score was confirmed in patients with type 2 EGJ morphology, a challenging group due to partial LES-CD separation, achieving an AUC of 0.858 in identifying those at risk for objective GERD.<sup>40</sup>

Another study demonstrated its utility in patients with laryngopharyngeal symptoms (LPS), showing a sensitivity of 57.1% and a specificity of 91.3%, suggesting the potential to streamline diagnostic pathways and better select candidates for further testing.<sup>41</sup>

Additionally, a monocentric study of 160 patients demonstrated the Milan score's ability to predict successful outcomes, post-ARS, offering an objective ARB assessment after intervention. Despite limitations, such as its retrospective design, lack of objective outcome measures, and inclusion of varied surgical techniques, this finding underscores the score's promise in surgical contexts.<sup>42</sup>

### Clinical implications

The Milan score's primary strength lies in its ability to quantify the degree of ARB disruption. The comprehensive nature of the score, by integrating multiple HRM parameters, makes

it accessible to most patients undergoing HRM. By providing risk rates for GERD, it can also serve as an upfront test to identify patients who may benefit from further invasive tests or to exclude low-risk individuals from further evaluation.

## Limitations and future research

Despite its advantages, the Milan score has some limitations that should be considered. The score relies on HRM parameters, which requires specialized equipment and expertise that may not be available in all clinical settings.

The SLR maneuver, a cornerstone of the score, poses challenges, as not all patients can perform it correctly, potentially skewing results. The current definition of an effective SLR (a 50% increase in intra-esophageal pressure during the maneuver over baseline) is based on expert opinion and lacks objective validation, warranting further studies. Additionally, the SLR significant weight in the score may be overestimated; larger studies could refine this balance or confirm its predictive power. The weighting of all four components might also benefit from integration with clinical parameters to improve accuracy.

Given the multifactorial nature of GERD and the overlap with other conditions, the Milan score cannot replace a definitive diagnosis but should complement clinical assessment, endoscopy, and reflux monitoring studies. Future research should also evaluate its predictive value for outcomes after medical or surgical treatments to fully validate its clinical utility.

## Conclusion

The Milan score represents a significant advancement in GERD diagnosis by offering a physiologically grounded method to stratify risk, based on HRM parameters. Its comprehensive evaluation of esophageal function and ARB integrity enhances the ability to identify objective GERD, paving the way for personalized diagnostic and therapeutic strategies. While further validation is needed, particularly regarding its applicability in diverse clinical settings and outcomes after treatment, the Milan score has substantial potential to improve GERD management by enabling more personalized diagnostic and therapeutic strategies.

## The Phoenix score

### Introduction

Until the DeMeester Score (DMS) was introduced in 1974, GERD diagnosis had relied heavily on symptom assessment, a method often inaccurate due to overlap between patients and healthy individuals.<sup>16,43</sup> The DMS was the first attempt to provide an objective measure of esophageal AET, establishing normative values for GERD diagnosis. More recently, the Lyon 2.0 consensus defined objective GERD based on AET and provided criteria for prolonged wireless reflux monitoring.<sup>4</sup> However, both methods have limitations, prompting the development of the pPhoenix score to address these gaps and improve diagnostic precision.<sup>44</sup>

## Background and limitations of previous scores

Over the years, the DMS has become a cornerstone in GERD diagnosis, considered the gold standard by the surgical community for selecting patients for ARS. The DMS is a composite score that integrates six parameters: recumbent, upright, and total reflux time, number of total episodes, number of episodes over 5 minutes, and longest episode, weighting each one based on standard deviations from healthy controls. Parameters with greater variability (e.g., number of episodes, upright reflux) contribute less to the score, whereas parameters that are seldom found in the controls (e.g., long episodes or recumbent reflux) carry more weight. This approach revolutionized GERD diagnosis by introducing objective AET metrics and has been correlated with endoscopic findings like esophagitis and Barrett's esophagus.<sup>45–48</sup>

However, the DMS presents certain limitations: its development cohort was small, it requires patients to accurately report supine and meal periods, and it was validated only for 24 h catheter-based pH studies, making it susceptible to day-to-day variability and lacking thresholds for prolonged reflux monitoring.

In contrast, the Lyon 2.0 consensus defines pathologic GERD solely by AET (> 6%), introducing a borderline range (4–6%) that requires additional criteria and establishing specific indications for definitive diagnosis in patients undergoing wireless 96 h reflux monitoring tests. The Lyon 2.0 does not take into account other pH parameters that might affect the severity of GERD, in particular supine reflux; however, its 6% threshold seems to effectively identify true GERD with excellent specificity. To address these gaps, specifically the DMS lack of thresholds for prolonged monitoring and the Lyon 2.0 absence of weighting for positional reflux, Latorre-Rodriguez et al. developed the pPhoenix score.

## Description of the score

The pPhoenix score is a composite measure derived from a cohort of patients with AET 2–6% undergoing 48 h wireless pH monitoring that uses the following formula to weigh supine and upright reflux:  $p\text{Phoenix} = (\% \text{ upright AET} \times 0.991) + (\% \text{ supine AET} \times 1.286)$ .

The score sets cut-offs at 8.45 (upper) and 7.06 (lower), defining a grey area for patients with scores in between. Its innovation lies in assigning greater weight to supine AET, reflecting its stronger association with GERD complications compared with upright reflux.

This positional weighting distinguishes the pPhoenix score from Lyon 2.0, which treats all reflux equally, based on total AET.

## Evidence and validation studies

The pPhoenix score's initial validation showed promising results, with an AUC of 0.957 in the identification of pathologic GERD, defined as pathologic DMS.<sup>44</sup>

Notably, it reduced the proportion of patients classified as borderline per Lyon 2.0 criteria from 77.2 to 13.2% ( $p < 0.001$ ), offering clearer diagnoses for patients with inconclusive total AET.

The score demonstrated good sensitivity and specificity across thresholds, with internal validation via bootstrapping, confirming its robustness.

## Clinical implications

The pHoenix score offers an objective measure of reflux that prioritizes supine AET, which is more closely linked to GERD complications. Unlike the DMS, it defines a grey area, which is crucial, given the progressive nature of GERD, and establishes clear positivity criteria for prolonged wireless reflux monitoring. Compared with Lyon 2.0 criteria, it could be helpful in the identification of pathologic GERD in patients with borderline AET (4-6%).

## Limitations and future research

Despite its potential, the pHoenix score has some limitations. Its initial validation was conducted at a single center, limiting generalizability across diverse populations. Multi-center studies are needed to confirm its performance in varied clinical settings and larger cohorts, which would also help refine diagnostic thresholds.

The original study's focus on patients with AET between 2 and 6% may have introduced selection bias, potentially excluding those with clearly normal or pathologic reflux patterns. Additionally, the accuracy of self-reported supine periods during 48 h monitoring, despite the use of electronic diaries, could affect reliability, as inconsistencies in reporting may skew positional data.

Further research should include head-to-head comparisons with other GERD diagnostic methods, such as impedance-pH monitoring and symptom assessment tools, to clarify the pHoenix score's advantages. Prospective studies evaluating patient outcomes after medical or surgical therapy are also essential to validate its clinical utility. Finally, since the score was developed using wireless reflux monitoring, its applicability to 24 h catheter-based pH studies requires validation.

## Conclusion

The pHoenix score represents an advancement in GERD diagnosis by incorporating positional reflux patterns into a weighted composite measure. Its emphasis on supine AET and the proposal of diagnostic criteria for prolonged wireless monitoring addresses key limitations of the DMS and Lyon 2.0. While further validation across diverse settings is needed, the pHoenix score has the potential to enhance GERD diagnosis.

## The COuGH RefluX score

### Introduction

Laryngopharyngeal symptoms, including chronic cough, globus sensation, and hoarseness, have an increasing impact on quality of life and social performance, leading to a significant increase in gastroenterology and ENT clinic consultations.<sup>49</sup> Determining which of these patients are

**Table 1** Calculation of the COuGH RefluX score.

| Variables                        | Points | Results          |
|----------------------------------|--------|------------------|
| <i>Cough</i>                     | 1.5    | Score $\leq 2.5$ |
| <i>Obesity/overweight</i>        |        | Unlikely GERD    |
| BMI 25-30                        | 1.5    | Score 3.0-4.5    |
| BMI $\geq 30$                    | 2.0    | Inconclusive     |
| <i>Globus</i>                    | -1.0   | GERD             |
| <i>Hiatal hernia</i> $\geq 1$ cm | 1      | Score $\geq 5$   |
| <i>Regurgitation</i>             | 1.5    | Likely           |
| <i>Male sex</i>                  | 1.5    | GERD             |

BMI: body mass index; GERD: gastroesophageal reflux disease.

affected by true GERD remains difficult, as the LPS spectrum includes a wide range of non-specific clinical manifestations that can be caused by multiple conditions. This diagnostic uncertainty often leads to empiric PPI therapy, which may be ineffective when GERD is not the underlying cause.<sup>50</sup> In order to avoid PPI over-prescription, upfront esophageal physiologic tests should be performed,<sup>4,51</sup> but they are not always available in all clinical settings. The COuGH RefluX score emerged as a novel clinical prediction tool designed specifically to address this diagnostic gap by stratifying patients with laryngopharyngeal symptoms based on their likelihood of having GERD.<sup>52</sup>

## Background and limitations of previous scores

Traditionally, GERD diagnosis has always been preceded by empiric PPI trials that, while convenient, led to diagnostic uncertainty. The PPI trial lacks specificity and may lead to false positive diagnoses due to the placebo effect or symptomatic improvement unrelated to acid suppression.<sup>53</sup> This approach can result in unnecessary long-term medication use with associated costs and potential side effects.<sup>54</sup>

Although alternative tests and parameters have been proposed,<sup>55,56</sup> definitive diagnosis relies on endoscopy and reflux monitoring study, preferably using prolonged wireless monitoring techniques.<sup>57</sup> Endoscopy has limited sensitivity in patients with LPS, as up to 70% of LPS patients have non-erosive reflux disease.<sup>58</sup>

Ambulatory reflux monitoring study is invasive, costly, often unavailable in primary care settings, and may not be well-tolerated by patients.

Prior to the COuGH RefluX score, no validated clinical prediction tool existed specifically for identifying GERD in patients presenting with LPS, as questionnaires and scoring systems did not demonstrate sufficient diagnostic accuracy.<sup>59,60</sup>

### Description of the score

The COuGH RefluX score is a clinical prediction model that incorporates 6 readily assessable parameters: Cough, Overweight/Obesity, Globus sensation, Hiatal hernia, Regurgitation, and male sex.<sup>52</sup> Each parameter is assigned a specific point value based on its relative predictive strength for GERD (Table 1).

The score calculation is simple, requiring only basic clinical information that can be obtained during routine patient evaluation without specialized testing, with the pos-

sible exception of hiatal hernia detection, which typically requires endoscopy. The total score ranges from 0 to 6 with a lower threshold of 2.5 and an upper threshold of 5.0 to predict proven GERD.

## Evidence and validation studies

The original paper developed and validated the COuGH RefluX score in 856 patients (304 training, 552 validation). The training cohort established a model with an AUC of 0.68. A threshold of 2.5 was 82% sensitive, and 5 was 79% specific for proven GERD. In the validation cohort (AUC 0.67), sensitivity was 79% at 2.5, and specificity was 81% at 5. A lower threshold of 1.5 increased sensitivity to 93%, enhancing the negative predictive value, whereas excluding hiatal hernia slightly reduced performance (AUC 0.63).

A subsequent study evaluated the practical application of the COuGH RefluX score in 232 patients, classifying them as unlikely (126 patients), inconclusive (74 patients), or likely (32 patients) GERD. The authors found proven GERD rates increasing progressively (8%, 36.5%, 65.5%;  $p < 0.001$ ). After excluding erosive esophagitis and Barrett's esophagus, 196 patients showed increasing PPI response rates across groups (12.4%, 45.0%, 73.3%;  $p < 0.001$ ). Multivariate analysis identified COuGH RefluX scores and lower MNBI as independent predictors of PPI response (scores  $\geq 5.0$ , OR = 15.772; MNBI, OR = 0.915;  $p < 0.001$ ).<sup>61</sup>

## Clinical implications

The COuGH RefluX score provides a standardized risk stratification for GERD in patients with LPS. Patients in the "unlikely GERD" category showed clinical characteristics similar to the healthy population, with a low rate of pathologic GERD, low likelihood of PPI response, and higher MNBI. The application of this tool has the potential to avoid unnecessary PPI therapy, reducing inappropriate prescribing.

In the "likely GERD" category, the score provides support for initiating PPI therapy with increased confidence in cases of initial medical evaluation, or to consider invasive physiology tests in cases of chronic LPS. The documented correlation between higher scores and better PPI response allows for more personalized treatment planning and setting of realistic expectations with patients.

Additionally, the COuGH RefluX score may be a practical tool in the hands of specialists in many areas, helping to create a common ground, in particular among gastroenterologists, otolaryngologists, pulmonologists, surgeons, and primary care physicians, who frequently co-manage these patients.

## Limitations and future research

Despite its promising utility, several limitations of the COuGH RefluX score must be acknowledged. The score relies partly on subjective symptom reporting, which introduces potential variability and reporting bias. Standardized symptom assessment tools could enhance the reliability of this component.

The detection of hiatal hernia, one of the score's parameters, requires endoscopy, which may not be readily available in all clinical settings. This could limit its applicability in primary care environments without easy access to endoscopic services. The scoring tool was shown to have moderate efficacy, even without hiatal hernia, but it is still an important component of the score. Moreover, the small cohort of patients with isolated laryngeal symptoms was underpowered for subgroup modeling. However, this cohort of patients is not a rare encounter in clinical practice and deserves specific validation.

Additionally, the COuGH RefluX score does not replace the need for clinical judgment or physiologic tests. The moderate AUC values indicate that while the score is a valuable and practical predictive tool, it is not a perfect discriminator and should be interpreted within the broader clinical context.

## The Lyon score

### Introduction

In 2024, the Lyon Consensus 2.0 updated the modern definition of reflux disease and introduced the concept of actionable GERD, which refers to settings where a high confidence in GERD diagnosis is essential, including long-term acid suppression following a successful PPI-trial, patients with GERD symptoms refractory to PPIs or with atypical symptoms, and patients requiring escalation of medical management or invasive treatment. Endoscopic and ambulatory reflux monitoring criteria were defined as conclusive, supportive, borderline, or non-supportive to diagnose objective GERD. Moreover, the Lyon 2.0 introduced adjunctive parameters, such as the number of reflux episodes (REs), MNBI, or reflux symptom association, to define or refute GERD in cases of borderline AET (4-6%), describe the thresholds to be used with prolonged wireless monitoring and with on-PPI testing, and added Los Angeles grade B to the definition of objective GERD.<sup>62</sup>

The Lyon score was subsequently developed to integrate findings from different diagnostic modalities (impedance pH study [MII-pH] and endoscopy), based on the Lyon 2.0 framework. This novel scoring system aims to translate the concept and evidence behind the Lyon 2.0 into clinical practice, condensing multiple testing metrics into a simple tool.<sup>63</sup>

### Background and limitations of previous scores

Conventional approaches to GERD diagnosis have relied heavily on single parameters or limited composite metrics, each with significant limitations. The DeMeester score, although widely used particularly by the surgical community, focuses primarily on acid exposure parameters, without considering endoscopic data or impedance parameters.<sup>16</sup> Similarly, isolated assessment of AET may result in a substantial number of inconclusive cases (AET 4-6%) and a greater dependence on the day-to-day variability of catheter-based reflux monitoring.

The limitations of these approaches in assessing mucosal integrity, reflux episode frequency, and multiple diagnostic

**Table 2** Calculation and phenotypes of the Lyon score.

| Esophagitis  |         |                              | AET   |     | Reflux episodes         |     | MNBI            |     | Reflux-symptom association |     | Hiatus hernia |    |
|--|---------|------------------------------|-------|-----|-------------------------|-----|-----------------|-----|----------------------------|-----|---------------|----|
| Normal   | None    | 0                            | <4.0% | 0   | <40                     | 0   | >2500           | 0   | No                         | 0   | No            | 0  |
| Abnormal   | LA-A    | 2                            | 4-6%  | 3.5 | >40                     | 2.5 | 1500-2499       | 1.5 | Yes                        | 1.5 | Yes           | 1  |
|  |         |                              |       |     |                         |     | <1500           | 2.0 |                            |     |               |    |
| Conclusive   | LA-B    | 4.5                          | >6.0% | 4.5 | NA                      |     | NA              |     | NA                         |     | NA            |    |
| Severe   | LA-C    | 5.5                          | >10%  | 5   |                         | NA  |                 | NA  |                            | NA  |               | NA |
|  | LA-D    | 6.5                          |       |     |                         |     |                 |     |                            |     |               |    |
| Phenotype  | No GERD | Isolated permeability defect |       |     | Reflux hypersensitivity |     | Borderline GERD |     | Conclusive GERD            |     | Severe GERD   |    |
| Lyon score   | 0-0.5   | 0.5-2.0                      |       |     | 1.0-3.0                 |     | ≥ 3.0           |     | ≥ 5.0                      |     | ≥ 10.0        |    |
| AET: acid exposure time; GERD: gastroesophageal reflux disease; LA: Los Angeles classification; MNBI: mean nocturnal baseline impedance; NA: not applicable. |         |                              |       |     |                         |     |                 |     |                            |     |               |    |

modalities result in diagnostic uncertainty in response to antireflux therapy.

The updated Lyon Consensus 2.0 addresses many of these limitations by establishing a comprehensive framework for GERD diagnosis that incorporates multiple diagnostic parameters. The Lyon score attempts to fill the gap between sophisticated diagnostic criteria and a practical tool for clinical use.

### Description of the score

The Lyon score integrates multiple parameters from endoscopy and pH-impedance monitoring into a single, comprehensive scoring system.

The Lyon score includes two endoscopic and four MII-pH parameters:

- 1 Esophagitis, graded according to the Los Angeles classification as follows: grade A, one or more mucosal erosions < 5 mm; grade B, erosions > 5 mm; grade C, mucosal breaks covering less than 75% of the esophageal circumference; and grade D, ≥ 75% of the esophagus.<sup>64</sup>
- 2 Hiatal hernia, detected on endoscopy or HRM.
- 3 AET, weighted based on established thresholds from the Lyon Consensus 2.0, with higher scores for increased acid exposure.
- 4 MNBI, a measure of mucosal integrity, with lower values indicating compromised esophageal mucosa, calculated manually as the mean of three 10 min night measurements.<sup>65</sup>
- 5 Number of Reflux Episodes, quantifying the frequency of reflux events, identified by software and then manually verified according to the Wingate consensus.<sup>66</sup>
- 6 Reflux-symptom Association, defined as either symptom index > 50% or symptom association probability > 95%.<sup>67</sup>

Each parameter is assigned a specific weight based on its relative importance in GERD diagnosis and correlation with

treatment outcomes, and the weighted values are added together to create a composite score (Table 2).

Moreover, the Lyon score provides clear thresholds for categorizing patients into diagnostic phenotypes: functional heartburn, permeability defect, reflux hypersensitivity, inconclusive GERD, conclusive GERD, and severe GERD.

### Evidence and validation studies

In the validation study published by Gyawali et al. in 2024, the development cohort consisted of 281 patients with GERD symptoms, 263 of whom were on PPI therapy and 18 who underwent ARS. The median Lyon score was 6.5 (IQR 2.5-10.5). Good outcome was defined as 50% symptom improvement using a visual analog scale (VAS) ranging from 0 (no symptoms) to 100 (severe symptoms).<sup>68</sup> Response rates rose with GERD evidence: from 7.1% in the functional heartburn phenotype, to 37.5% in borderline GERD, to ≥ 76.6% in conclusive GERD ( $p < 0.001$ ). The AUC for predicting good outcomes was 0.819. The Lyon score outperformed the DeMeester score (AUC 0.779,  $p = 0.019$ ) in predicting outcomes after GERD treatment, with an optimal threshold of 6.25 (81.2% sensitivity, 73.4% specificity).

Two different validation cohorts were evaluated. The European (215 patients from two centers in Italy) was heartburn-dominant, while the Asian (258 patients from Taiwan) was regurgitation predominant and consisted of previously untreated patients. Median Lyon scores were 2.5 in the European cohort and 4.3 in the Asian cohort, with AUCs of 0.908 and 0.637, respectively ( $p < 0.001$ ), in predicting treatment response. The Lyon score consistently outperformed individual components and the DeMeester score, especially in the regurgitation-heavy Asian cohort. At a 2.75 threshold (95% sensitivity), it identified symptom response in 85.6% of the European cohort and in 77.9% of the Asian cohort; at the 11.5 threshold (95% specificity), only 14.4% of patients in the European cohort and 12.3% in the Asian cohort had no symptom response.

## Clinical implications

The Lyon score has the potential to offer significant clinical advantages for the diagnosis of GERD. By integrating and weighting metrics from endoscopy and MII-pH, it provides a comprehensive tool to stratify patients with GERD symptoms into distinct phenotypes, ranging from functional heartburn to severe GERD. The Lyon score's ability to predict symptom response to antireflux therapy, ranging from 7.1 in functional heartburn phenotype to  $\geq 75\%$  in conclusive GERD, may reduce unnecessary treatments and improve overall therapeutic outcomes, enabling tailored treatment decisions. Patients with low scores might benefit from conservative approaches, such as lifestyle changes or minimal medical therapy, whereas those with higher scores could be prioritized for more invasive options, such as ARS. The score's reliance on widely available diagnostic tests and potential for automation further enhance its practicality, offering a reproducible, readily available, user-friendly tool. If validated prospectively, the Lyon score could make the GERD diagnosis more efficient, offering a better selection of good candidates for GERD treatment escalation.

## Limitations and future research

Despite its promising performance, some limitations of the Lyon score must be acknowledged. The retrospective design of the developmental studies introduces potential biases related to data collection and patient selection. Prospective validation in diverse clinical settings remains necessary to confirm the score's robustness and generalizability.

The abovementioned study focused on patients with typical GERD symptoms off antisecretory therapy, therefore validation in atypical presentations and on-PPI cohorts is warranted. Variability in symptom profiles across developmental and validation cohorts, such as heartburn predominance in Europeans versus regurgitation in Asians, suggests potential ethnic or geographic influences on performance. Additionally, the score's reliance on pH-impedance metrics raises questions about its applicability in settings where wireless pH monitoring is preferred. Future research should prioritize prospective, multicenter trials to externally validate the score across diverse populations and symptom phenotypes. Studies targeting specific therapeutic cohorts, medical vs. surgical, integration with other scores (e.g., Milan score or COuGH RefluX score) may provide a more robust validation and enhance their clinical utility.

## Discussion

The diagnosis and management of GERD have evolved significantly over the past few decades, driven by advancements in diagnostic technologies and a deeper understanding of its pathophysiology. This review highlights several novel tools the AFS Classification, Milan score, pHoenix score, COuGH RefluX score, and Lyon score, novel tools that aim to enhance the precision, objectivity, and clinical relevance of GERD assessment. These tools address longstanding challenges in

GERD diagnosis, such as subjectivity of earlier methods, limitations of single-parameter assessments, and the need for treatment stratification. A summary of the modality and performance of the discussed scores is shown in [Table 3](#).

All the scores discussed above share the common intent of providing a more objective evaluation. Historically, GERD diagnosis relied heavily on symptom reporting and empiric PPI trials, but this approach has failed, given the significant symptom overlap with other conditions and placebo effects. The introduction of the DMS in 1974 marked a pivotal shift toward objective measurement by quantifying AET and other reflux parameters.<sup>16</sup> Several decades later, the Lyon Consensus established AET benchmarks and incorporated adjunctive metrics, such as the MNBI and reflux-symptom association.<sup>4</sup> However, even if its threshold of AET (6%) correctly identifies true GERD, it did not take into account other reflux monitoring parameters (e.g., supine reflux) and it lacked a simple and practical tool to convert its concepts into clinical practice. To overcome these shortcomings, the pHoenix and the Lyon scores have been recently introduced.

Because clinical assessment, endoscopy, and HRM are an integral part of the GERD work-up and traditional metrics suffered from subjectivity and a lack of diagnostic accuracy, different research groups proposed the COuGH RefluX Score, the AFS Classification, and the Milan score, with the aim to objectively stratify LPS for GERD likelihood, standardize the endoscopic assessment of the EGJ, and quantify ARB disruption.<sup>19,39</sup>

Together, these tools reflect a paradigm shift toward precision medicine in GERD, aiming to reduce diagnostic ambiguity and optimize therapeutic decision-making.

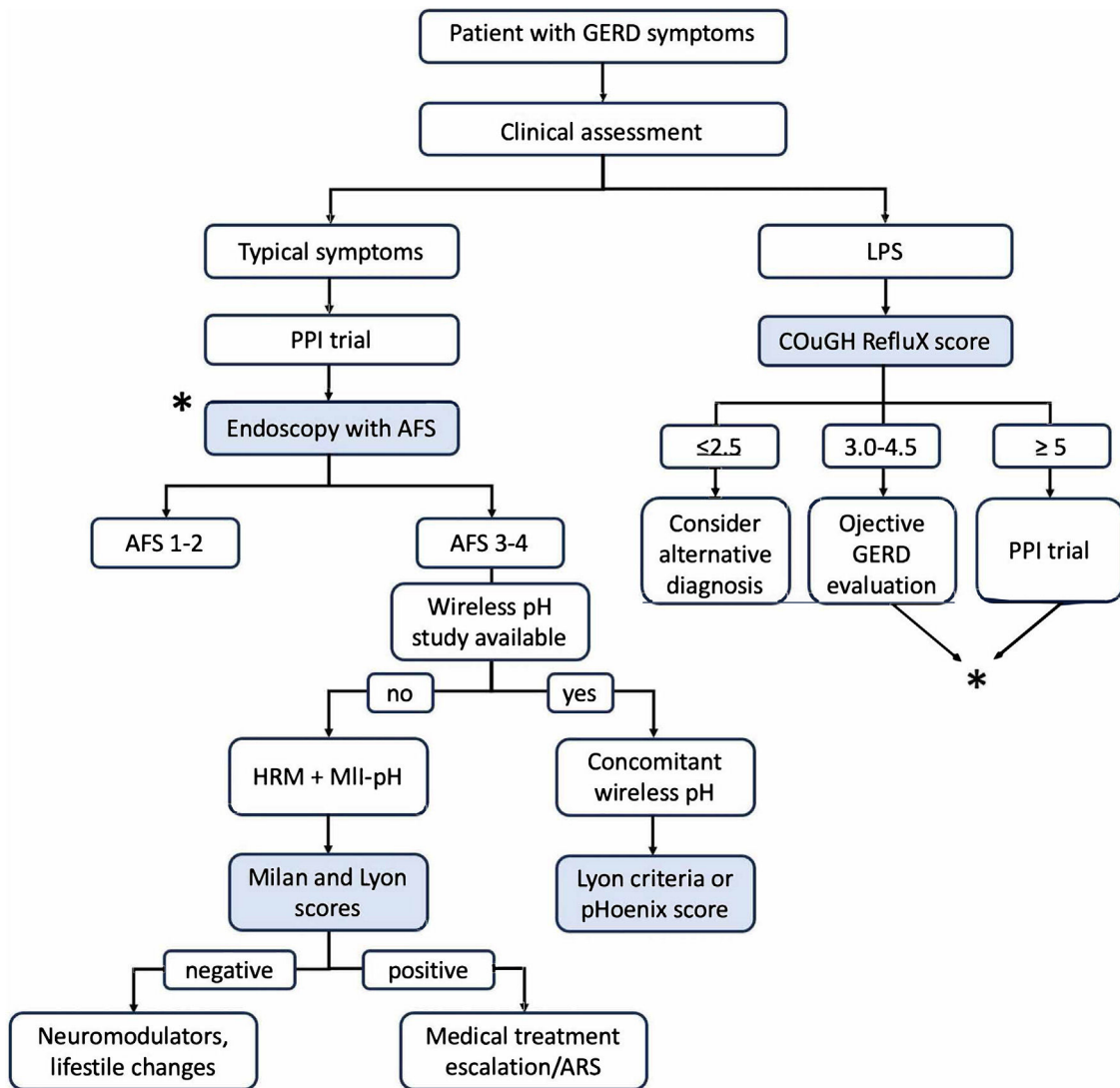
Each tool brings unique strengths to the diagnostic landscape. The AFS classification improves upon the Hill classification, by introducing measurable parameters (length, diameter, and flap valve integrity) and a standardized endoscopic protocol, reducing subjectivity, and enhancing reproducibility. Its correlation with physiologic measures, such as the AET and HRM parameters, underscores its potential to identify patients with significant EGJ disruption, who may benefit from escalated therapy. Similarly, the Milan score enhances the HRM capability to assess functional ARB defects, offering a single, predictive metric with excellent diagnostic accuracy for objective GERD. Its ability to guide preoperative assessment for ARS highlights its clinical utility in complex cases.

The pHoenix score addresses limitations in the DeMeester score and Lyon 2.0 by emphasizing supine AET, a known risk factor for GERD complications, and establishing thresholds for prolonged wireless monitoring. Its ability to reclassify borderline AET cases (4–6%) into clearer diagnostic categories could streamline management decisions. The COuGH RefluX score fills a critical gap in evaluating LPS, a notoriously challenging symptom cluster, by providing a simple, clinic-based tool with high sensitivity and specificity at defined thresholds. Finally, the Lyon score stands out for its integrative approach, combining endoscopic and pH-impedance data to stratify patients into actionable phenotypes, with demonstrated predictive power for treatment outcomes (AUC 0.819).

**Table 3** Summary of the setting, modality, and performance of the novel tools.

| Tool               | Setting   | Application  | Modality                         | Threshold                                       | Performance   |
|--------------------|---|--|----------------------------------|---|---|
| AFS Classification | Suspected GERD  | Endoscopic assessment of EGJ integrity                                     | Endoscopy                        | Pathologic: AFS type III-IV                     | Correlation with pathologic AET (>6%):<br>0% (Grade I)<br>5.9% (Grade II)<br>52% (Grade III)<br>77.8% (Grade IV)<br>AUC 0.880 |
| Milan Score        | Typical and atypical GERD symptoms<br>Post-operative assessment | Assessment of ARB disruption   | HRM                              | 137   | Sensitivity 80.4%   |
| pHoenix Score      | GERD symptoms   | Diagnosis of GERD in borderline AET cases (2-6%) with prolonged monitoring | Wireless pH monitoring           | Lower cut-off: 7.06                             | Specificity 85.2%<br>AUC 0.957 in predicting pathologic DeMeester   |
| CouGH RefluX Score | Atypical GERD symptoms (LPS)                                    | Risk stratification for GERD in patients with laryngopharyngeal symptoms   | Questionnaire                    | Higher cut-off: 8.45<br>Low: $\leq 2.5$         | AUC 0.67  |
| Lyon Score         | Typical GERD symptoms   | Diagnosis and phenotyping of GERD, predicting treatment outcomes           | Multi-modal (Endoscopy + MII-pH) | Intermediate: 3 - 4.5<br>High: $\geq 5$<br>6.25 | Sensitivity 79%<br>Specificity 81%<br>AUC 0.819   |
|                    |   |  |                                  |   | Sensitivity 81.2%<br>Specificity 73.4%  |

AFS: American Foregut Society; ARB: anti-reflux barrier; AUC: area under the curve; EGJ: esophagogastric junction; GERD: gastroesophageal reflux disease; HRM: high-resolution manometry; LPS: laryngopharyngeal symptoms; MII-pH: multichannel intraluminal impedance-pH.



**Figure 3** Integration proposal of the novel tools into the diagnostic pathway for GERD.

AFS: American Foregut Society; ARS: anti-reflux surgery; GERD: gastroesophageal reflux disease; HRM: high-resolution manometry; LPS: laryngopharyngeal symptoms; MII-pH: multichannel intraluminal impedance-pH; PPI: proton pump inhibitor.

Despite their promise, these tools face certain limitations that temper their immediate adoption into routine practice, such as their validation in tertiary care centers, a fact that raises questions about applicability in community practice.

Another common thread of these tools is their developmental stage. Most of them have been validated in controlled or specialized settings, with retrospective and monocentric populations, and their performance in diverse populations remains underexplored. Additionally, the multifactorial nature of GERD implies that no single tool can fully capture its complexity, underscoring the need for complementary use with clinical judgment and established diagnostics.

Despite these limitations, the potential impact of these tools on clinical practice is substantial, by providing a more sophisticated and tailored approach to GERD diagnosis. Some diagnostic tests are not readily available in several countries, thus making clinical evaluation and first-line

endoscopic assessment crucial for patient stratification.<sup>69</sup> Under this light, the COuGH RefluX score and the AFS classification during endoscopy might help to select patients presenting with a high risk of GERD and profound EGJ disruption for confirmatory pathophysiologic tests and to reduce unnecessary testing in those with low risk or intact barriers. The COuGH RefluX score stands out as particularly suitable for resource-limited settings due to its reliance on clinical parameters that can be assessed during routine patient evaluations, with the exception of hiatal hernia detection, which may require endoscopy. In primary care or underserved areas where endoscopy is not readily available, the score's performance without hiatal hernia remains moderate, suggesting potential utility as a triage tool. However, its dependence on subjective symptom reporting may introduce variability, necessitating standardized patient questionnaires to enhance reliability. In contrast, the AFS classification offers a standardized endoscopic assessment of EGJ integrity but poses challenges

in general practice due to the need for trained specialists. To improve accessibility, simplified training protocols or telemedicine-guided endoscopic assessments could be explored.

It is well known that PPI response and good outcomes after ARS rely heavily on objective GERD diagnosis.<sup>70,71</sup> The pHOenix score's focus on supine reflux could refine treatment escalation, particularly in patients with nocturnal symptoms or borderline AET, whereas the Lyon score's comprehensive nature and phenotypes might offer objective evidence to select patients for surgery (conclusive and severe GERD) or less invasive therapeutic options (neuro-modulators, PPI).

Finally, the Milan score's capability to quantify ARB disruption and identify cases of severe GERD make it a particularly appropriate and useful test before ARS.

Future research should focus on multicenter, prospective validations to establish the robustness of these tools across diverse populations and clinical scenarios. Integrating clinical parameters (e.g., symptom severity, BMI) with physiologic metrics could enhance diagnostic accuracy, as seen in the COuGH RefluX score's use of obesity and sex. Comparative studies challenging these tools against each other, or using them in combination, will clarify their relative advantages and combined performance.

Based on the performance and features of each tool, we propose a diagnostic algorithm to integrate these new scores into the diagnostic pathway for GERD, as shown in Fig. 3. Long-term outcome studies are critical for validating these scores as useful tools to predict treatment response to medical therapy escalation or ARS.

## Conclusions

The AFS classification, Milan score, pHOenix score, COuGH RefluX score, and Lyon score represent significant advancements in GERD diagnosis, offering objective, standardized tools to enhance precision and personalize treatment. The AFS classification and COuGH RefluX score streamline endoscopic and clinical assessments, respectively, aiding patient triage in primary care. The Milan score and pHOenix score refine pre-operative and borderline AET evaluations, and the Lyon score integrates multimodal data to predict treatment outcomes. The proposed diagnostic pathway (Fig. 3) integrates these tools to guide clinicians from initial symptom assessment to confirmatory testing and treatment escalation, ensuring efficient and tailored GERD management. As validation progresses and practical barriers are addressed, these innovations have the potential to transform GERD management, improving diagnostic accuracy, reducing overtreatment, and optimizing patient outcomes.

## Funding

No funding of any kind was received for this article.

## Conflicts of interest

The authors declare no conflicts of interest.

## References

- Kahrilas PJ, Jonsson A, Denison H, et al. Impact of regurgitation on health-related quality of life in gastro-oesophageal reflux disease before and after short-term potent acid suppression therapy. *Gut*. 2014;63:720–6, <http://dx.doi.org/10.1136/gutjnl-2013-304883>.
- Calabrese F, Pasta A, Bodini G, et al. Applying Lyon consensus criteria in the work-up of patients with extra-oesophageal symptoms — a multicentre retrospective study. *Aliment Pharmacol Ther*. 2024;59:1134–43, <http://dx.doi.org/10.1111/apt.17934>.
- Katz PO, Dunbar KB, Schnoll-Sussman FH, et al. ACG clinical guideline for the diagnosis and management of gastroesophageal reflux disease. *Am J Gastroenterol*. 2022;117:27–56, <http://dx.doi.org/10.14309/ajg.0000000000001538>.
- Gyawali CP, Yadlapati R, Fass R, et al. Updates to the modern diagnosis of GERD: Lyon consensus 2.0. *Gut*. 2024;73:361–71, <http://dx.doi.org/10.1136/gutjnl-2023-330616>.
- Hill LD, Kozarek RA, Kraemer SJ, et al. The gastroesophageal flap valve: in vitro and in vivo observations. *Gastrointest Endosc*. 1996;44:541–7, [http://dx.doi.org/10.1016/S0016-5107\(96\)70006-8](http://dx.doi.org/10.1016/S0016-5107(96)70006-8).
- Fass R. Epidemiology and pathophysiology of symptomatic gastroesophageal reflux disease. *Am J Gastroenterol*. 2003;98:S2–7, [http://dx.doi.org/10.1016/S0002-9270\(03\)00009-1](http://dx.doi.org/10.1016/S0002-9270(03)00009-1).
- Patel A, Posner S, Gyawali CP. Esophageal high-resolution manometry in gastroesophageal reflux disease. *JAMA*. 2018;320:1279–80, <http://dx.doi.org/10.1001/jama.2018.8694>.
- Savarino E, de Bortoli N, Bellini M, et al. Practice guidelines on the use of esophageal manometry — a GISMA-SIGE-AIGO medical position statement. *Dig Liver Dis*. 2016;48:1124–35, <http://dx.doi.org/10.1016/j.dld.2016.06.021>.
- Gyawali CP, Roman S, Bredenoord AJ, et al. Classification of esophageal motor findings in gastro-esophageal reflux disease: conclusions from an international consensus group. *Neurogastroenterol Motil*. 2017;29:e13104, <http://dx.doi.org/10.1111/nmo.13104>.
- Nicodème F, Pipa-Muniz M, Khanna K, et al. Quantifying esophagogastric junction contractility with a novel HRM topographic metric, the EGJ-Contractile Integral: normative values and preliminary evaluation in PPI non-responders. *Neurogastroenterol Motil*. 2014;26:353–60, <http://dx.doi.org/10.1111/nmo.12267>.
- Tolone S, De Bortoli N, Marabotto E, et al. Esophagogastric junction contractility for clinical assessment in patients with GERD: a real added value? *Neurogastroenterol Motil*. 2015;27:1423–31, <http://dx.doi.org/10.1111/nmo.12638>.
- Rogers BD, Rengarajan A, Ali IA, et al. Straight leg raise metrics on high-resolution manometry associate with esophageal reflux burden. *Neurogastroenterol Motil*. 2020;32:e13929, <http://dx.doi.org/10.1111/nmo.13929>.
- Siboni S, Kristo I, Rogers BD, et al. Improving the diagnostic yield of high-resolution esophageal manometry for GERD: the “Straight Leg-Raise” international study. *Clin Gastroenterol Hepatol*. 2023;21:1761–70.e1, <http://dx.doi.org/10.1016/j.cgh.2022.10.008>.

14. Masuda T, Mittal SK, Kovacs B, et al. Simple manometric index for comprehensive esophagogastric junction barrier competency against gastroesophageal reflux. *J Am Coll Surg*. 2020;230:744–55.e3, <http://dx.doi.org/10.1016/j.jamcollsurg.2020.01.034>.
15. Akimoto S, Singhal S, Masuda T, et al. Esophagogastric junction morphology and distal esophageal acid exposure. *Dig Dis Sci*. 2016;61:3537–44, <http://dx.doi.org/10.1007/s10620-016-4331-y>.
16. Johnson LF, DeMeester TR. Twenty-four-hour pH monitoring of the distal esophagus. A quantitative measure of gastroesophageal reflux. *Am J Gastroenterol*. 1974;62:325–32. PMID: 4432845.
17. Visaggi P, Mariani L, Svizzero FB, et al. Clinical use of mean nocturnal baseline impedance and post-reflux swallow-induced peristaltic wave index for the diagnosis of gastro-esophageal reflux disease. *Esophagus*. 2022;19:525–34, <http://dx.doi.org/10.1007/s10388-022-00933-6>.
18. Kahrilas PJ, Shaheen NJ, Vaezi MF, et al. American Gastroenterological Association Medical Position Statement on the management of gastroesophageal reflux disease. *Gastroenterology*. 2008;135:1383–91.e5, <http://dx.doi.org/10.1053/j.gastro.2008.08.045>.
19. Nguyen NT, Thosani NC, Canto MI, et al. The American Foregut Society white paper on the endoscopic classification of esophagogastric junction integrity. *Foregut*. 2022;2:339–48, <http://dx.doi.org/10.1177/26345161221126961>.
20. Siboni S, Sozzi M, Kersik A, et al. Does the American Foregut Society endoscopic classification of esophagogastric junction integrity predict acid reflux and esophagogastric junction disruption? *Foregut*. 2024, <http://dx.doi.org/10.1177/26345161241230940>.
21. Swei E, Almario JA, Abou-Diwan E, et al. American Foregut classification is superior to Hill classification in evaluation of the gastroesophageal junction. *Foregut*. 2024;4:404–10, <http://dx.doi.org/10.1177/26345161241293522>.
22. Nguyen NT, Bell R, Abu-Dayyeh B, et al. American Foregut Society white paper on the endoscopic assessment of the gastroesophageal valve after anti-reflux surgery. *Foregut*. 2024, <http://dx.doi.org/10.1177/26345161241300480>.
23. Pauwels A, Boecxstaens V, Andrews CN, et al. How to select patients for antireflux surgery? The ICARUS guidelines (international consensus regarding preoperative examinations and clinical characteristics assessment to select adult patients for antireflux surgery). *Gut*. 2019;68:1928–41, <http://dx.doi.org/10.1136/gutjnl-2019-318260>.
24. Jobe BA, Richter JE, Hoppo T, et al. Preoperative diagnostic workup before antireflux surgery: an evidence and experience-based consensus of the Esophageal Diagnostic Advisory Panel. *J Am Coll Surg*. 2013;217:586–97, <http://dx.doi.org/10.1016/j.jamcollsurg.2013.05.023>.
25. Visaggi P, Ghisa M, Barberio B, et al. Gastro-esophageal diagnostic workup before bariatric surgery or endoscopic treatment for obesity: position statement of the International Society of Diseases of the Esophagus. *Dis Esophagus*. 2024;37:doae006, <http://dx.doi.org/10.1093/dote/doae006>.
26. Salvador R, Pandolfino JE, Costantini M, et al. The role of high-resolution manometry before and following antireflux surgery: the Padova consensus. *Ann Surg*. 2025;281:124–35, <http://dx.doi.org/10.1097/SLA.0000000000006297>.
27. Siboni S, Bonavina L, Rogers BD, et al. Effect of increased intra-abdominal pressure on the esophagogastric junction: a systematic review. *J Clin Gastroenterol*. 2022;56:821–30, <http://dx.doi.org/10.1097/MCG.0000000000001756>.
28. DeMeester TR, Wernly JA, Bryant GH, et al. Clinical and in vitro analysis of determinants of gastroesophageal competence. A study of the principles of antireflux surgery. *Am J Surg*. 1979;137:39–46, [http://dx.doi.org/10.1016/0002-9610\(79\)90008-4](http://dx.doi.org/10.1016/0002-9610(79)90008-4).
29. Joelsson BE, DeMeester TR, Skinner DB, et al. The role of the esophageal body in the antireflux mechanism. *Surgery*. 1982;92:417–24. PMID: 7101132.
30. DeMeester TR, Lafontaine E, Joelsson BE, et al. Relationship of a hiatal hernia to the function of the body of the esophagus and the gastroesophageal junction. *J Thorac Cardiovasc Surg*. 1981;82:547–58. PMID: 7278346.
31. Wernly JA, DeMeester TR, Bryant GH, et al. Intra-abdominal pressure and manometric data of the distal esophageal sphincter. Their relationship to gastroesophageal reflux. *Arch Surg*. 1980;115:534–9, <http://dx.doi.org/10.1001/archsurg.1980.01380040156028>.
32. Hoshino M, Sundaram A, Mittal SK. Role of the lower esophageal sphincter on acid exposure revisited with high-resolution manometry. *J Am Coll Surg*. 2011;213:743–50, <http://dx.doi.org/10.1016/j.jamcollsurg.2011.09.002>.
33. Gyawali CP, Zerbib F, Bhatia S, et al. Chicago Classification update (V4.0): technical review on diagnostic criteria for ineffective esophageal motility and absent contractility. *Neurogastroenterol Motil*. 2021;33:e14134, <http://dx.doi.org/10.1111/nmo.14134>.
34. Yadlapati R, Kahrilas PJ, Fox MR, et al. Esophageal motility disorders on high-resolution manometry: Chicago Classification version 4.0©. *Neurogastroenterol Motil*. 2021;33:e14058, <http://dx.doi.org/10.1111/nmo.14058>.
35. Tolone S, de Cassan C, de Bortoli N, et al. Esophagogastric junction morphology is associated with a positive impedance-pH monitoring in patients with GERD. *Neurogastroenterol Motil*. 2015;27:1175–82, <http://dx.doi.org/10.1111/nmo.12606>.
36. Luther J, Zarro S, Sagaram M, et al. Intermittent hiatus hernia on high-resolution manometry associates with abnormal reflux burden similar to persistent hiatus hernia. *Am J Gastroenterol*. 2023;118:2071–4, <http://dx.doi.org/10.14309/ajg.0000000000002353>.
37. Akimoto S, Singhal S, Masuda T, et al. Classification for esophagogastric junction (EGJ) complex based on physiology. *Dis Esophagus*. 2017;30:1–6, <http://dx.doi.org/10.1093/dote/dox048>.
38. Masuda T, Mittal SK, Kovacs B, et al. Thoracoabdominal pressure gradient and gastroesophageal reflux: insights from lung transplant candidates. *Dis Esophagus*. 2018;31, <http://dx.doi.org/10.1093/dote/doy025>.
39. Siboni S, Sozzi M, Kristo I, et al. The Milan score: a novel manometric tool for a more efficient diagnosis of gastroesophageal reflux disease. *United European Gastroenterol J*. 2024;12:552–61, <http://dx.doi.org/10.1002/ueg2.12565>.
40. Ferrari D, Siboni S, Sozzi M, et al. The Milan Score predicts objective gastroesophageal reflux disease in patients with type 2 esophagogastric junction. *Neurogastroenterol Motil*. 2025:e14987, <http://dx.doi.org/10.1111/nmo.14987>.
41. Siboni S, Sozzi M, Visaggi P, et al. The Milan Score is an effective manometric tool to predict gastroesophageal reflux in patients with laryngopharyngeal symptoms. *Neurogastroenterol Motil*. 2025;2:e70015, <http://dx.doi.org/10.1111/nmo.70015>.
42. Sozzi M, Siboni S, Cusmai L, et al. A new manometric tool to evaluate the efficacy of laparoscopic anti-reflux surgery: first postoperative application of the Milan score. *Ann Esophagus*. 2024;7:25, <http://dx.doi.org/10.21037/aoe-24-32>.

43. Johnson LF, DeMeester TR. Development of the 24-hour intraesophageal pH monitoring composite scoring system. *J Clin Gastroenterol*. 1986;8:52–8, <http://dx.doi.org/10.1097/00004836-198606001-00008>.
44. Latorre-Rodríguez AR, Mittal SK, Simmonds H, et al. pHoEnix score: development and validation of a novel approach to decrease the number of inconclusive GERD diagnoses. *Surg Endosc*. 2024;38:6880–93, <http://dx.doi.org/10.1007/s00464-024-11105-1>.
45. Schwameis K, Lin B, Roman J, et al. Is pH testing necessary before antireflux surgery in patients with endoscopic erosive esophagitis? *J Gastrointest Surg*. 2018;22:8–12, <http://dx.doi.org/10.1007/s11605-017-3554-3>.
46. Xu D-T, Feng G-J, Zhao L-L, et al. Relationship among symptoms, mucosal injury, and acid exposure in gastroesophageal reflux disease. *Chin Med J (Engl)*. 2013;126:4430–4. PMID: 24286401.
47. Schlottmann F, Andolfi C, Herbella FA, et al. GERD: presence and size of hiatal hernia influence clinical presentation, esophageal function, reflux profile, and degree of mucosal injury. *Am Surg*. 2018;84:978–82. PMID: 29981634.
48. Neto RML, Herbella FAM, Schlottmann F, et al. Does DeMeester score still define GERD? *Dis Esophagus*. 2019;32:doy118, <http://dx.doi.org/10.1093/dote/doy118>.
49. Lechien J-R, Mouawad F, Bobin F, et al. Review of management of laryngopharyngeal reflux disease. *Eur Ann Otorhinolaryngol Head Neck Dis*. 2021;138:257–67, <http://dx.doi.org/10.1016/j.anorl.2020.11.002>.
50. Kahrilas P, Anastasiou F, Bredenoord AJ, et al. Proton pump inhibitors: rational use and use-reduction. *The Windsor Workshop*. *Dig Dis*. 2024;42:211–20, <http://dx.doi.org/10.1159/000538399>.
51. Ribolsi M, Luca-Guarino MP, Balestrieri P, et al. The results from up-front esophageal testing predict proton pump inhibitor response in patients with chronic cough. *Am J Gastroenterol*. 2021;116:2199–206, <http://dx.doi.org/10.14309/ajg.0000000000001389>.
52. Krause AJ, Kaizer AM, Carlson DA, et al. Validated clinical score to predict gastroesophageal reflux in patients with chronic laryngeal symptoms: COuGH RefluX. *Clin Gastroenterol Hepatol*. 2024;22:1200–9.e1, <http://dx.doi.org/10.1016/j.cgh.2024.01.021>.
53. Spantideas N, Drosou E, Bougea A, et al. Proton pump inhibitors for the treatment of laryngopharyngeal reflux. A systematic review. *J Voice*. 2020;34:918–29, <http://dx.doi.org/10.1016/j.jvoice.2019.05.005>.
54. Francis DO, Rymer JA, Slaughter JC, et al. High economic burden of caring for patients with suspected extraesophageal reflux. *Am J Gastroenterol*. 2013;108:905–11, <http://dx.doi.org/10.1038/ajg.2013.69>.
55. Fuchs HF, Müller DT, Berlth F, et al. Simultaneous laryngopharyngeal pH monitoring (Restech) and conventional esophageal pH monitoring—correlation using a large patient cohort of more than 100 patients with suspected gastroesophageal reflux disease. *Dis Esophagus*. 2018;31, <http://dx.doi.org/10.1093/dote/doy018>.
56. Kurylo CM, Eastwood D, Blumin JH, et al. Correlation of esophageal mean nocturnal baseline impedance with markers of laryngopharyngeal reflux. *Laryngoscope*. 2023;133:1927–32, <http://dx.doi.org/10.1002/lary.30421>.
57. Krause AJ, Greytak M, Kaizer AM, et al. Diagnostic yield of ambulatory reflux monitoring systems for evaluation of chronic laryngeal symptoms. *Am J Gastroenterol*. 2024;119:627–34, <http://dx.doi.org/10.14309/ajg.0000000000002557>.
58. Vardar R, Varis A, Bayrakci B, et al. Relationship between history, laryngoscopy and esophagogastroduodenoscopy for diagnosis of laryngopharyngeal reflux in patients with typical GERD. *Eur Arch Otorhinolaryngol*. 2012;269:187–91, <http://dx.doi.org/10.1007/s00405-011-1748-y>.
59. Belafsky PC, Postma GN, Koufman JA. Validity and reliability of the reflux symptom index (RSI). *J Voice*. 2002;16:274–7, [http://dx.doi.org/10.1016/s0892-1997\(02\)00097-8](http://dx.doi.org/10.1016/s0892-1997(02)00097-8).
60. Geeratragool T, Maneerattanaporn M, Prapruetkit J, et al. Association between laryngopharyngeal reflux clinical scores and esophageal multichannel intraluminal impedance pH monitoring interpretation according to Lyon Consensus 2.0. *Dis Esophagus*. 2025;38:doae098, <http://dx.doi.org/10.1093/dote/doae098>.
61. Wong M-W, Hsiao S-H, Wang J-H, et al. Practical application of the COuGH RefluX score for patients with laryngopharyngeal symptoms: real-world assessment in predicting gastroesophageal reflux disease and proton-pump inhibitor response. *Am J Gastroenterol*. 2024, <http://dx.doi.org/10.14309/ajg.0000000000003240>.
62. Visaggi P, Del Corso G, Gyawali CP, et al. Ambulatory pH-impedance findings confirm that grade B esophagitis provides objective diagnosis of gastroesophageal reflux disease. *Am J Gastroenterol*. 2023;118:794–801, <http://dx.doi.org/10.14309/ajg.0000000000002173>.
63. Gyawali CP, Marchetti L, Rogers BD, et al. The Lyon score: a novel reflux scoring system based on the Lyon Consensus 2.0 that associates with treatment outcome from antireflux therapy. *Am J Gastroenterol*. 2024, <http://dx.doi.org/10.14309/ajg.0000000000003083>.
64. Lundell LR, Dent J, Bennett JR, et al. Endoscopic assessment of oesophagitis: clinical and functional correlates and further validation of the Los Angeles classification. *Gut*. 1999;45:172–80, <http://dx.doi.org/10.1136/gut.45.2.172>.
65. Frazzoni M, Savarino E, de Bortoli N, et al. Analyses of the post-reflux swallow-induced peristaltic wave index and nocturnal baseline impedance parameters increase the diagnostic yield of impedance-pH monitoring of patients with reflux disease. *Clin Gastroenterol Hepatol*. 2016;14:40–6, <http://dx.doi.org/10.1016/j.cgh.2015.06.026>.
66. Gyawali CP, Rogers B, Frazzoni M, et al. Inter-reviewer variability in interpretation of pH-impedance studies: the Wingate Consensus. *Clin Gastroenterol Hepatol*. 2021;19:1976–8.e1, <http://dx.doi.org/10.1016/j.cgh.2020.09.002>.
67. Roman S, Gyawali CP, Savarino E, et al. Ambulatory reflux monitoring for diagnosis of gastro-esophageal reflux disease: update of the Porto consensus and recommendations from an international consensus group. *Neurogastroenterol Motil*. 2017;29:1–15, <http://dx.doi.org/10.1111/nmo.13067>.
68. Patel A, Sayuk GS, Gyawali CP. Parameters on esophageal pH-impedance monitoring that predict outcomes of patients with gastroesophageal reflux disease. *Clin Gastroenterol Hepatol*. 2015;13:884–91, <http://dx.doi.org/10.1016/j.cgh.2014.08.029>.
69. Davis TA, Gyawali CP. Refractory gastroesophageal reflux disease: diagnosis and management. *J Neurogastroenterol Motil*. 2024;30:17–28, <http://dx.doi.org/10.5056/jnm23145>.

70. Dhar A, Maw F, Dallal HJ, et al. Side effects of drug treatments for gastro-oesophageal reflux disease: current controversies. *Frontline Gastroenterol.* 2020;13:45–9, <http://dx.doi.org/10.1136/flgastro-2019-101386>.
71. Patti MG, Allaix ME, Fisichella PM. Analysis of the causes of failed antireflux surgery and the principles of treatment: a review. *JAMA Surg.* 2015;150:585–90, <http://dx.doi.org/10.1001/jamasurg.2014.3859>.