

ORIGINAL ARTICLE OPEN ACCESS

Combined Lyon and Milan Scores Predict Gerd Management Outcome Better Than Either Score Alone or Their Individual Components

Stefano Siboni¹ | Marco Sozzi¹ | Pierfrancesco Visaggi²  | Sara Boveri³ | Benjamin D. Rogers^{4,5} | Nicola De Bortoli²  | Anthony Hobson⁶ | Brian E. Louie⁷  | Yeong Yeh Lee⁸  | Salvatore Tolone⁹ | Elisa Marabotto¹⁰ | Jordan Haworth⁶ | Megan L. Ivy⁷ | Lorenzo Cusmai¹ | Edoardo Campioli¹ | Dimitrios Theodorou¹¹ | Tania Triantafyllou¹¹ | Andrea Pasta¹⁰  | Francesco Calabrese¹⁰  | Vincent Tee⁸ | Daniele Bernardi¹ | Carlo Galdino Riva¹ | Roberta De Maron¹ | Enrique Coss-Adame¹²  | Amit Patel¹³ | Emanuele Asti¹ | Edoardo Savarino¹⁴ | C. Prakash Gyawali³ 

¹Division of General and Emergency Surgery, IRCCS Policlinico San Donato, University of Milan, Milano, Italy | ²Division of Gastroenterology, University of Pisa, Pisa, Italy | ³Laboratory of Biostatistics and Data Management, Scientific Directorate, IRCCS Policlinico San Donato, Milano, Italy | ⁴Division of Gastroenterology, Washington University School of Medicine, St. Louis, Missouri, USA | ⁵Division of Gastroenterology, Hepatology, and Nutrition, University of Louisville, Louisville, Kentucky, USA | ⁶The Functional Gut Clinic, London, UK | ⁷Division of Thoracic Surgery, Swedish Medical Center, Digestive Health Institute, Seattle, WA, USA | ⁸School of Medical Sciences and GI Function and Motility Unit, Kota Bharu, Malaysia | ⁹Division of General, Mini-Invasive and Bariatric Surgery, University of Naples, Naples, Italy | ¹⁰Gastroenterology Unit, IRCCS Policlinico San Martino, Genoa, Italy | ¹¹Foregut Surgery Unit, University of Athens School of Medicine, Athens, Greece | ¹²Gastroenterology Department, Instituto Nacional de Ciencias Médicas y Nutrición, Mexico City, México | ¹³Division of Gastroenterology, Duke University School of Medicine and the Durham Veterans Affairs Medical Center, Durham, North Carolina, USA | ¹⁴Division of Gastroenterology, Department of Surgical, Oncological and Gastroenterological Sciences, University of Padua, Padua, Italy

Correspondence: Stefano Siboni (stefano.siboni@grupposandonato.it)

Received: 16 April 2025 | **Revised:** 27 April 2025 | **Accepted:** 10 June 2025

Handling Editor: Colin Howden

Funding: The authors received no specific funding for this work.

Keywords: gastro-oesophageal reflux disease | high-resolution manometry | reflux monitoring

ABSTRACT

Background: Confidence in gastro-oesophageal reflux disease (GERD) diagnosis is crucial to improve outcomes from escalation of treatment. The Lyon score phenotypes patients through endoscopy and pH-impedance (MII-pH). The Milan score quantifies antireflux barrier through high-resolution manometry (HRM) parameters.

Aim: To explore the relationship between the Lyon and Milan scores and their combined performance in predicting clinical outcomes.

Methods: We collected clinical and follow-up data of consecutive patients with HRM and MII-pH from nine centres. Clinical improvement was defined as a 50% reduction in global symptoms. The relationship between Lyon and Milan scores and the rate of patients improved in Lyon phenotypes and Milan categories were explored. The ability of the Lyon, Milan, DeMeester scores and acid exposure time (AET) in predicting outcomes was assessed through receiver operating characteristics (ROC) analysis.

Results: Among 532 patients (50.6% female, age 50 years), 47.7% had pathologic GERD. A stepwise increase in the Milan score in Lyon phenotypes was observed. Sixty-three patients had surgical treatment, and 131 medical. Clinical improvement in Lyon phenotypes Conclusive and Severe was 81% and 83%; in Milan categories very likely and extremely likely was 88.5% and 100%. If

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2025 The Author(s). *Alimentary Pharmacology & Therapeutics* published by John Wiley & Sons Ltd.

Lyon and Milan scores were positive, improvement was 89%; if discordant, 63%; if both negative, 19% ($p < 0.001$). ROC analysis showed an AUC of 0.790 for Lyon score, 0.835 for Milan score, 0.736 for DeMeester score and 0.741 for AET.

Conclusions: The Lyon and Milan scores outperformed AET and DeMeester scores in predicting outcomes in GERD patients. When concordant, they provide optimal predictive accuracy, guiding escalation of therapy.

1 | Introduction

Gastro-oesophageal reflux disease (GERD) is a complex disorder that presents with a broad spectrum of symptoms that can overlap with other oesophageal and extra-oesophageal disorders, including disorders of gut-brain interaction (DGBI). In particular, DGBI may present with typical oesophageal symptoms, such as heartburn and chest pain, while aero-digestive and respiratory conditions may manifest extra-oesophageal symptoms, such as cough, hoarseness and shortness of breath [1]. Most of these clinical presentations prompt trials of proton pump inhibitors (PPIs) by gastroenterologists and nongastroenterologists alike. Thus, in clinical practice, PPI use is commonly adopted to establish the diagnosis of GERD based on PPI response, especially with typical symptoms [2]. However, among patients with oesophageal symptoms, this approach leads to PPI over-prescription, carrying an important economic burden without a significant improvement in overall quality of life [3]. For these reasons, confidence in a reflux-based diagnosis is crucial to improve patient outcomes, particularly from the escalation of antireflux therapy when symptoms persist.

The Lyon 2.0 consensus standardised the diagnostic approach for patients with suspected GERD based on presentation, defining precise criteria for a conclusive diagnosis as well as adjunctive evidence in borderline cases [4]. In the pursuit of predictors of clinical outcomes following GERD therapy, several possible endoscopic and reflux monitoring variables have been explored, but acid exposure time (AET) and mean nocturnal baseline impedance (MNBI) seem to have higher value than other parameters [5–8]. Some of these parameters have been utilised by the surgical community in identifying the most optimal candidates for antireflux surgery (ARS). Thus far, an abnormal reflux-monitoring study, typical reflux symptoms and symptomatic response to PPIs predict good clinical outcome after surgery [9], while body mass index (BMI) $> 35 \text{ kg/m}^2$ and elevated residual lower oesophageal sphincter (LES) pressure are associated with a suboptimal outcome [10, 11]. Historically, the DeMeester score, a composite score developed from six metrics from an ambulatory pH monitoring study, has been used to identify GERD and to select patients for surgery but it is limited by the lack of validation in randomised controlled studies [12].

The recently introduced Lyon score incorporates endoscopic and pH-impedance (MII-pH) parameters to phenotype patients with suspected GERD. This has the potential to be a useful tool to help increase confidence in GERD diagnosis and to predict treatment response [13]. In the original paper, validated on heartburn- and regurgitation-predominant populations, the Lyon score outperformed the DeMeester score in predicting typical symptom response. Moreover, the performance of the score was higher than the single components in predicting outcomes after GERD treatment.

The Milan score, a novel metric based on high-resolution manometry (HRM) findings, has been recently developed and then validated in a multicentre cohort of 528 patients to quantify the disruption of the antireflux barrier (ARB) and to predict the likelihood of pathologic reflux in patients with GERD symptoms [14]. The Milan score integrates four variables, namely ineffective oesophageal motility (IEM), oesophago-gastric junction (EGJ) contractile integral (EGJ-CI), EGJ morphology and straight leg raise (SLR) manoeuvre response, to quantify the risk for definitive reflux. A Milan score of 137, corresponding to a 50% risk rate, has a sensitivity of 80.4% and specificity of 85.2%, while higher scores, representing a more impaired ARB, have lower sensitivity but higher specificity. The ability of the Milan score to predict clinical outcomes after GERD treatment has not yet been demonstrated [15].

Based on the ability of the Lyon and Milan scores to both predict conclusive GERD and the fact that the components of the scores do not overlap, we hypothesised that abnormal Lyon and Milan scores will associate with a good symptom response from escalation of medical and surgical reflux therapy in patients with GERD symptoms and that one score may complement the other.

The aim of this study was to evaluate the performance of the Lyon and Milan scores alone and in combination in predicting clinical outcomes in patients with GERD symptoms undergoing medical or surgical treatment for GERD and to compare their predictive value against that of the AET and the DeMeester score.

2 | Methods

2.1 | Study Population

Patients between 18 and 75 years who underwent HRM and MII-pH enrolled in previous studies [14, 16] at nine tertiary medical centres were included and contacted by phone to obtain follow-up information following oesophageal physiologic testing. Patients underwent pathophysiologic tests as up-front GERD evaluation, as pre-ARS work-up and if refractory to medical treatment, defined as persistent symptoms despite double-dose PPI for at least 8 weeks [2]. Information regarding change in management (escalation of medical management, antireflux surgery) was also collected upon follow-up. Exclusion criteria consisted of previous foregut surgery, morbid obesity (BMI $> 35 \text{ kg/m}^2$), eosinophilic oesophagitis, oesophageal achalasia, scleroderma, other major motility disorders, incomplete studies and incomplete symptom information.

To ensure uniformity across the nine centres, all sites adhered to standardised protocols for HRM and MII-pH monitoring, including the use of the standard Chicago Classification 4.0 protocol [17], utilisation of standard SLR methodology [14, 16] and performance of MII-pH monitoring off therapy in

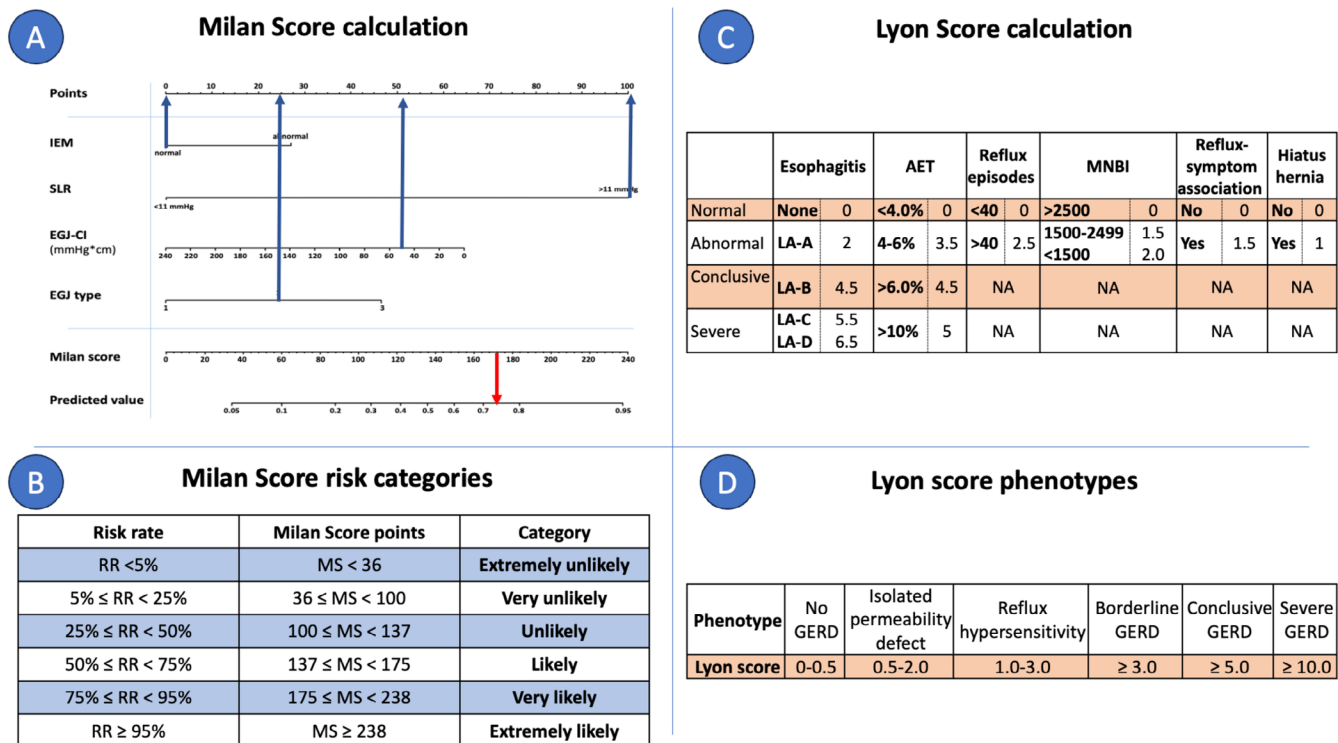


FIGURE 1 | Calculation of the Milan score (A), Milan score categories (B), calculation of the Lyon score (C) and Lyon score phenotypes (D).

unproven GERD [4]. The coordinating centre provided protocol guidelines to ensure consistency in data collection and interpretation. The two cohorts of patients were similar in terms of demographic features, as well as HRM and MII-pH characteristics, as demonstrated in the Milan score validation study [14].

The study protocol was approved by the ethical committee at the proposing centre (identifier: CET 335–2024), by institutional review boards at each centre, and was conducted following the Declaration of Helsinki.

2.2 | Oesophageal High-Resolution Manometry

All tests were performed following the Chicago Classification 4.0 (CCv4.0) standard protocol [17], with 10 swallows in the primary position (upright or recumbent) and five swallows in the secondary position. For the purpose of this study, data from HRM provocative manoeuvres such as multiple rapid swallows were not included in the analysis.

Per CCv4.0 criteria, ineffective oesophageal motility was defined as > 70% of weak swallows or ≥ 50% of failed swallows. EGJ and LES metrics assessed consisted of LES basal pressure, EGJ-CI [18], total and intra-abdominal LES length and EGJ morphology as defined by CCv4.0 [17]. The SLR manoeuvre was performed in the supine position as previously described [16]. An intra-oesophageal peak pressure increase of 11 mmHg during the manoeuvre was considered abnormal.

The Milan score was computed for each patient using the official online calculator (www.milanscore.com). The nomogram

with the corresponding points for each variable is shown in Figure 1A. Risk rate for pathologic GERD have been classified into six categories, from extremely unlikely to extremely likely (Figure 1B). A Milan score > 137 (risk rate 50%) was considered abnormal [14].

2.3 | Oesophageal pH and pH-Impedance Study

Reflux monitoring was performed off antisecretory therapy using a 24-h pH-impedance catheter. The pH-impedance studies were performed by positioning the pH electrode 5 cm above the manometrically identified LES after calibration.

After exclusion of meal periods, all pH-impedance tracings were reviewed using Wingate consensus criteria [19] and the following data were recorded: total, upright and recumbent AET, number of acid, weakly acid and weakly alkaline reflux episodes, DeMeester score, symptom index (SI) and symptom association probability (SAP). MNBI and postreflux swallow-induced peristaltic wave (PSPW) index were also extracted from pH-impedance studies [20, 21]. The MNBI was calculated manually as the average impedance value from three 10-min periods during the night, where no swallows or reflux episodes occurred in the most distal channel.

2.4 | Lyon Score

The Lyon score was calculated for each patient according to the described methodology [13]. Esophagitis and hiatal hernia on endoscopic evaluation, AET, total number of reflux episodes, MNBI and reflux-symptom association on MII-pH were

collected and classified as normal, abnormal, conclusive or severe, and a single score was assigned to each variable accordingly. Off-PPI endoscopic variables have been collected for the previous studies and were already part of the dataset [14, 16]. Oesophagitis was graded according to the Los Angeles classification as follows [22]. When not reported, the presence of hiatal hernia was corroborated with HRM findings.

The sum of the single components generated the final Lyon score (Figure 1C) and the study population was stratified into Lyon phenotypes (Figure 1D). A Lyon score > 6.5 was considered pathologic, as per original paper [13].

2.5 | Medical and Surgical Treatment Escalation

During the follow-up period, all patients underwent medical or surgical GERD management escalation as directed by their individual treating physicians and surgeons. Medical treatment escalation consisted of double-dose PPI for at least 2 months, plus antacids and alginate-based compounds if needed. Surgical treatment included partial/total fundoplication or magnetic sphincter augmentation (MSA).

2.6 | Symptom Assessment

At the time of HRM and MII-pH studies, each patient completed the global symptom severity (GSS) score, a previously reported visual analog scale (VAS) assessing the overall severity of oesophageal symptoms on a scale ranging from 0 (no symptoms) to 100 (maximum symptoms) [5, 23]. Moreover, the GERD-Q [24], the GERD health-related quality of life (GERD-HRQL) [25] and the reflux symptom index (RSI) [26], were administered at time of presentation. Demographic, clinical and pathophysiological (HRM and MII-pH) data were extracted from the electronic patient record, including the Milan and Lyon scores, AET and DeMeester score.

All patients who agreed to be followed up at each single centre were prospectively contacted by phone to assess whether they underwent medical vs. surgical treatment and to re-administer the questionnaires. Data on medical treatment (dose and response to PPI), surgical procedure (date and type of operation), as well as follow-up GSS were collected. Successful outcome was defined as a 50% reduction in GSS. Objective GERD was defined as AET > 6% on MII-pH test, Los Angeles grades B, C or D oesophagitis, biopsy-proven long segment Barrett's oesophagus or peptic stricture on endoscopy. In case of inconclusive evidence of GERD (AET 4%–6%), Lyon 2.0 adjunctive criteria (total number of reflux episodes > 80/day, a MNBI < 1500 Ω , or reflux-symptom association) were used to support or refute pathological GERD. GERD was excluded when AET < 4% [4]. Abnormal DeMeester score was defined as > 14.72 [27].

2.7 | Statistical Analysis

Continuous variables are reported as median and interquartile range (IQR), while categorical variables as numbers and

percentages. The normality of continuous variables was assessed using the Shapiro–Wilk test. The nonparametric Kruskal–Wallis test was used to compare continuous variables, while categorical variables were compared using the Chi-Square test or Fisher test as appropriate. GERD and no-GERD groups were compared in terms of demographic and pathophysiologic variables. The relationship between Lyon and Milan scores was investigated through comparative analysis including Spearman correlation and Bland–Altman plots. Clinical improvement was explored in terms of Lyon score phenotypes, Milan score categories and the concordance between the scores. Receiver operating characteristics (ROC) analysis with the area under the curve (AUC) assessment was performed for all the different diagnostic criteria to assess the ability to determine a clinical improvement, as well as for the combined Lyon and Milan scores, standardised by means of logistic regression. AUC significance was tested against a null hypothesis of no predictive ability (AUC = 0.5) using the DeLong test. Pairwise comparisons of AUCs were adjusted for multiple comparisons using Bonferroni correction.

A two-tailed $p < 0.05$ was considered significant for all statistical tests. Statistical analyses were performed using SAS version 9.4 (SAS Institute, Cary, North Carolina).

3 | Results

Among the 532 included patients, 254 had pathologic GERD. The median age was 50 years, 50.6% were females and median BMI was 24.6 kg/m². Patients with pathologic GERD were significantly older, more frequently males and had a higher BMI. As expected, endoscopic findings and symptom burden were significantly worse in the GERD group (Table 1).

HRM and ambulatory reflux monitoring characteristics are summarised in Table 2. Patients with GERD had a higher rate of hiatal hernia, a shorter LES total and intra-abdominal length, a lower EGJ-CI and a higher rate of IEM and positive SLR. All MII-pH parameters were significantly different between the GERD and no-GERD groups. No patients were diagnosed with distal oesophageal spasm, hypercontractile oesophagus or EGJ-outflow obstruction.

There was a statistically significant, strong positive correlation between the Lyon and Milan scores ($r = 0.697$, $p < 0.001$). We observed a stepwise increase of Milan score values in Lyon phenotypes (from 54.4 in the phenotype No GERD, to 82 in Borderline GERD, 153 in Conclusive GERD and 184 in Severe GERD) in the overall population ($n = 532$) (Figure S1). The Bland–Altman analysis showed a mean difference of 0.001 between the standardised Milan and Lyon scores, with a standard deviation of 0.778. The 95% limits of agreement ranged from –1.525 to 1.525. The mean difference was not significantly different from zero ($p = 1.000$). The Bland–Altman plot (Figure S2) displayed no evident proportional bias, indicating moderate agreement between the standardised scores.

Out of 194 patients with follow-up data, 131 patients opted for medical treatment and 63 underwent ARS, of whom 30 patients had partial fundoplication, 19 had MSA and 14 had 360° degree

TABLE 1 | Demographic characteristics of the overall patient population stratified by GERD diagnosis. Continuous values are expressed as median [IQR].

	Total (<i>n</i> = 532)	No GERD (<i>n</i> = 278)	GERD (<i>n</i> = 254)	<i>p</i>
Male, <i>n</i> (%)	258 (49.4)	118 (43.1)	140 (56.5)	0.003
Age, (years)	50 [21]	48 [25]	52 [19]	0.007
BMI, (Kg/m ²)	24.6 [6.0]	23.4 [5.5]	25.3 [5.4]	<0.001
Symptom duration, (months)	28 [48]	24 [48]	36 [60]	0.003
HRQL, (points)	15 [14]	13 [15]	16.5 [14.0]	0.004
RSI, (points)	9 [12]	9 [13]	8 [12]	0.673
GERD-Q(A), (points)	7 [5]	7 [5]	8 [5]	<0.001
GERD-Q(B), (points)	2 [3]	2 [3]	2 [3]	0.677
GSS, (points)	70 [50]	60 [50]	77.5 [38]	0.025
PPI use, <i>n</i> (%)	419 (81.5)	210 (77.5)	209 (86.0)	0.017
PPI response				0.005
No benefit, <i>n</i> (%)	130 (31.7)	77 (37.6)	53 (25.9)	
Partial benefit, <i>n</i> (%)	177 (43.2)	89 (43.4)	88 (42.9)	
Full benefit, <i>n</i> (%)	103 (25.1)	39 (19.0)	64 (31.2)	
Endoscopic findings				
Hiatal hernia, <i>n</i> (%)	299 (59.4)	123 (47.1)	176 (72.7)	<0.001
Barrett's oesophagus, <i>n</i> (%)	26 (5.3)	0 (0.0)	26 (10.8)	<0.001
Oesophagitis, <i>n</i> (%)	145 (28.9)	39 (15.0)	99 (41.1)	<0.001
Oesophagitis Los Angeles grade				<0.001
Grade A, <i>n</i> (%)	80 (54.4)	39 (100.0)	41 (41.8)	
Grade B, <i>n</i> (%)	41 (27.9)	0 (0.0)	41 (41.8)	
Grade C, <i>n</i> (%)	10 (6.8)	0 (0.0)	10 (10.2)	
Grade D, <i>n</i> (%)	6 (4.1)	0 (0.0)	6 (6.1)	

Abbreviations: BMI, body mass index; GERD, gastro-oesophageal reflux disease; GERD-Q, gastro-oesophageal reflux disease questionnaire; HRQL, health-related quality of life; PPI, proton pump inhibitors; RSI, reflux symptom index.

fundoplication. Therefore, further statistical analysis was performed in 194 patients. The flow chart of the study population is shown in Figure S3. Follow-up time was comparable between the groups (22.8 vs. 24.3 months, $p=0.312$). Clinical, HRM and MII-pH data are presented in Table 3.

Overall, clinical improvement was achieved in greater proportions of patients who underwent ARS over medical treatment (78% vs. 47%, $p<0.001$). Clinical improvement was significantly higher in Conclusive GERD (79.3%) and Severe GERD (83.0%) compared with other Lyon score phenotypes (Figure S4a). Milan score categories showed a stepwise increase in the rate of patients with good clinical outcomes (0% in extremely unlikely, 30.8% very unlikely, 64.7% unlikely, 80.6% likely, 88.5% very likely, 100% extremely likely) (Figure S4b).

In the overall cohort ($n=532$), 303 patients (56.9%) had a positive Lyon score, and 238 (45%) had a positive Milan score. In the follow-up cohort, 105 patients (54.1%) had a positive Lyon score and 85 (43.8%) had a positive Milan score. When the Lyon

and Milan scores were both positive (74 patients), the rate of improvement was 89%; when only one score was positive (43 patients) 63%; and when both were negative (77 patients) 19% ($p<0.001$) (Figure 2). In the subgroup of patients undergoing surgery ($n=63$), when the Lyon and Milan scores were concordant and pathologic, 91.4% had good outcomes after ARS; when only one score was positive, it was 84.6%; and when both were negative, it was 40.0%.

The performance in predicting response to treatment was assessed through ROC curves, and we found an AUC of 0.792 for Lyon score, 0.835 for Milan score, 0.745 for AET and 0.739 for DeMeester score (Figure 3). The Lyon and Milan scores outperformed the AET and DeMeester scores (Lyon vs. AET, $p=0.048$; Lyon vs. DeMeester, $p=0.035$; Milan vs. AET, $p=0.005$; Milan vs. DeMeester, $p=0.003$), but their predictive ability was similar ($p=0.149$). These results have been confirmed after applying the Bonferroni correction for multiple comparisons. The logistic regression combined Lyon + Milan score achieved an AUC of 0.843, significantly higher than the Lyon score ($p=0.014$), but

TABLE 2 | HRM and pH characteristics of the overall patient population stratified by GERD diagnosis. Continuous values are expressed as median [IQR].

	Total (<i>n</i> = 532)	No GERD (<i>n</i> = 278)	GERD (<i>n</i> = 254)	<i>p</i>
HRM				
EGJ type				<0.001
1, <i>n</i> (%)	279 (52.4)	189 (68)	90 (35.4)	
2, <i>n</i> (%)	166 (31.2)	76 (27.3)	90 (35.4)	
3, <i>n</i> (%)	87 (16.4)	13 (4.7)	74 (29.1)	
Hiatal hernia, <i>n</i> (%)	253 (48.1)	87 (31.9)	166 (65.9)	<0.001
Hiatal hernia size, (cm)	0.0 [1.7]	0.0 [0.7]	1.3 [2.6]	<0.001
LES total length, (cm)	2.0 [0.8]	2.2 [0.9]	1.9 [0.7]	<0.001
LES intrabdominal length, (cm)	0.1 [1.2]	0.7 [1.5]	0.0 [0.6]	<0.001
EGJ- CI, (mmHg×cm)	31.0 [33.6]	42.8 [37.1]	23.7 [27.0]	<0.001
DCI, (mmHg×cm×sec)	1029 [1186]	1248 [1268.0]	806.4 [957]	<0.001
Patients with IEM, <i>n</i> (%)	123 (23.1)	46 (16.5)	77 (30.3)	<0.001
SLR manoeuvre, (mmHg)	8.7 [19.0]	3.2 [7.6]	18.9 [17.3]	<0.001
Positive SLR, <i>n</i> (%)	236 (44.6)	40 (14.4)	196 (77.8)	<0.001
Milan score, points	109 [125]	61 [36.5]	158 [92]	<0.001
Milan score > 137, <i>n</i> (%)	238 (45)	39 (14.1)	199 (79.0)	<0.001
pH				
Acid exposure time, (%)	5.1 [9.3]	1.7 [2.5]	11.2 [10.1]	<0.001
AET > 6%, <i>n</i> (%)	239 (44.9)	0 (0.0)	238 (93.7)	<0.001
DeMeester score, (points)	19.7 [34.6]	7.1 [9.1]	41.8 [38.1]	<0.001
DeMeester score > 14.72, <i>n</i> (%)	296 (55.7)	48 (17.3)	248 (97.6)	<0.001
Total reflux episodes, (<i>n</i>)	37 [37]	30 [30]	46.5 [36]	<0.001
Acid reflux episodes, (<i>n</i>)	25 [27]	15 [19]	36 [31]	<0.001
Weakly acid reflux episodes, (<i>n</i>)	8 [12]	8.5 [11]	8 [13]	0.660
MNBI, (Ω)	2070 [2355]	3230 [1686]	1130 [1121]	<0.001
MNBI < 1500 Ω, <i>n</i> (%)	190 (35.8)	21 (7.6)	169 (66.5)	<0.001
PSPW index, (%)	46 [40]	66.8 [29.0]	34 [19]	<0.001
Reflux symptom association, <i>n</i> (%)	87 (22.2)	32 (15.8)	55 (28.9)	0.002
Lyon score, points	6.5 [8.4]	2 [4.5]	10 [4.0]	<0.001

Abbreviations: AET, acid exposure time; EGJ, oesophago-gastric junction; EGJ-CI, oesophago-gastric junction contractile integral; GERD, gastro-oesophageal reflux disease; IEM, ineffective oesophageal motility; LES, lower oesophageal sphincter; MNBI, mean nocturnal baseline impedance; PSPW, postreflux swallow-induced peristaltic wave; SLR, straight leg raise.

not than the Milan score ($p=0.468$). Optimal thresholds and associated sensitivity, specificity, positive and negative predictive values are shown in Table 4. The Lyon score and the Milan score performed better than any of the individual components (Table 4).

All the patients with intermediate Lyon score phenotypes (Reflux hypersensitivity and Borderline GERD, 40 patients) and a positive Milan score (5, 12.5%) exhibited improvement after treatment (100% specificity).

4 | Discussion

In this study, we investigated if the combination of abnormal Lyon and Milan scores could predict outcomes from escalation of medical and surgical reflux therapy in patients with GERD symptoms. We utilised this methodology since the components of the Lyon and Milan scores do not overlap, and since each score has demonstrated promise in predicting symptom outcomes following GERD management. Therefore, by analysing reflux testing and clinical longitudinal outcome data

TABLE 3 | Demographic, HRM and MII-pH characteristics of the patient population with available follow up, stratified by GERD diagnosis. Continuous values are expressed as median [IQR].

	Total (n = 194)	No GERD (n = 104)	GERD (n = 87)	p
Male, n (%)	83 (44)	39 (37)	44 (52)	0.036
Age, (years)	49 [20]	48 [19]	51 [23]	0.066
BMI, (Kg/m ²)	24.6 [5.4]	23.5 [5.4]	25.4 [5.0]	0.001
GERD-Q, (points)	8.0 [4.0]	7.0 [4.0]	9.0 [6.0]	0.019
GSS, (points)	70 [40]	60 [30]	78 [30]	0.032
PPI use, n (%)	151 (83)	80 (78)	71 (88)	0.103
PPI full benefit, n (%)	36 (25)	12 (16)	24 (36)	0.025
Endoscopic findings				
Hiatal hernia, n (%)	93 (53)	36 (38)	57 (70)	<0.001
Oesophagitis, n (%)	44 (25)	16 (17)	28 (35)	0.005
Oesophagitis Los Angeles grade				0.024
Grade A, n (%)	26 (58)	15 (83)	11 (41)	
Grade B, n (%)	15 (33)	3 (17)	12 (44)	
Grade C, n (%)	3 (6.7)	0 (0)	3 (11)	
Grade D, n (%)	1 (2.2)	0 (0)	1 (3.7)	
HRM				
EGJ type 1, n (%)	96 (49)	65 (61)	31 (36)	<0.001
EGJ-CI, (mmHg×cm)	27 [37]	37 [42]	22 [26]	<0.001
Patients with IEM, n (%)	47 (24)	17 (16)	30 (34)	0.003
Positive SLR, n (%)	84 (44)	17 (16)	67 (79)	<0.001
Milan score > 137, n (%)	85 (44)	16 (15)	69 (81)	<0.001
pH				
Acid exposure time, (%)	4 [9]	2 [2]	12 [11]	<0.001
DeMeester score, (points)	17 [33]	8 [9]	42 [37]	<0.001
Number of reflux episodes > 80, n (%)	19 (9.8)	2 (1.9)	17 (20)	<0.001
MNBI < 1500 Ω, n (%)	64 (33)	11 (10)	53 (61)	<0.001
Reflux symptom association, n (%)	33 (22)	8 (10)	25 (35)	<0.001
Lyon score, points	5.5 [7.9]	2.0 [4.5]	10.0 [4.0]	<0.001

Abbreviations: BMI, body mass index; EGJ, oesophago-gastric junction; EGJ-CI, oesophago-gastric junction contractile integral; GERD-Q, gastro-oesophageal reflux disease questionnaire; GSS, global symptom severity; IEM, ineffective oesophageal motility; LES, lower oesophageal sphincter; MNBI, mean nocturnal baseline impedance; PPI, proton-pump inhibitors; SLR, straight leg raise.

from patients from multiple institutions undergoing patho-physiologic evaluation, we demonstrated that the Lyon and Milan scores are superior to the traditional diagnostic metrics as AET and DeMeester score in predicting outcome after the escalation of antireflux therapy. The outcome measure, the GSS questionnaire, has been already validated as an effective tool to predict outcome after medical and surgical treatment [5, 27], while other questionnaires (i.e., GERD-Q) have been extensively used to stratify GERD symptoms but not to assess outcomes.

The ROC analysis demonstrated a higher diagnostic accuracy of the Milan score (0.835) and Lyon score (0.792) compared to AET (0.745) and DeMeester score (0.739). This superiority in performance can be attributed to the comprehensive nature of these novel scores. Unlike traditional metrics that primarily focus on a single physiological parameter, the Lyon and Milan scores integrate multiple diagnostic variables. Their ability to predict pathologic GERD is further confirmed by the ability to predict improvement after therapy, as demonstrated by the high rate of patients improved in the conclusive and severe

GERD Lyon phenotypes (79.3 and 83% respectively) and in likely, very likely and extremely likely Milan score categories (80.6%, 88.5% and 100%).

A particularly intriguing aspect of our study was the strong positive correlation between the Lyon and Milan scores ($r=0.697$, $p<0.001$). Also, the stepwise increase of Milan score values across Lyon score phenotypes further supports the complementary nature of these diagnostic tools. From the 'No GERD' phenotype (Milan score 54.4) to the 'Severe GERD' phenotype (Milan score 184), we observed a consistent progression that aligns with disease severity. The Bland–Altman analysis revealed moderate agreement between the standardised Lyon and Milan scores, suggesting that the scores are not fully interchangeable and supporting the complementary use of the scores.

Our data suggest that these scores capture complementary point of views of GERD pathophysiology. The Milan score integrates

and weights four HRM parameters that represent different well-demonstrated mechanisms of the ARB: the EGJ type provides a description of the synergy between the LES and the crural diaphragm (CD) [28, 29], the EGJ-CI incorporates LES length and pressure [18, 30], IEM offers assessment of clearance provided by oesophageal body motility [31, 32] and SLR is able to effectively document the dynamic ability of the EGJ to prevent backflow pressure from the stomach when intra-abdominal pressure increases [16, 33–35]. This score is therefore a valuable tool to quantify ARB disruption. On the other side, the Lyon score focuses on the consequences of a disrupted ARB, namely the duration of oesophageal acid exposure (AET), its impact on mucosal integrity, represented by MNBI, the number of reflux episodes, reflux-symptom association and presence of erosive changes on endoscopy [13]. Even if hiatal hernia is present in both the score, measured as EGJ morphology in the Milan score, and as endoscopic hiatal hernia in the Lyon score, these scores naturally complement each other. The first offers the underlying mechanism of a disrupted ARB, while the second its consequences. Combined, they provide a comprehensive assessment of the multifactorial pathophysiology at play in the development of GERD symptoms and oesophageal end-organ damage. We also believe that the fact that hiatal hernia is present in both scores is not an issue since the diagnostic accuracy of the scores outperformed the single components (Table 4). The low sensitivity (14.4%) and AUC (0.553) of the oesophagitis component in the Lyon score likely result from excluding LA Grade A oesophagitis, per Lyon 2.0 criteria, which consider Grade A less specific for GERD due to its prevalence in asymptomatic individuals.

The complementary nature of the Lyon and Milan scores is demonstrated by the high rate (89%) of clinical improvement when both scores were positive, indicating that a multidimensional diagnostic approach can significantly enhance clinical decision-making. When only one score was positive, the rate of clinical improvement was 63%, probably indicating that these

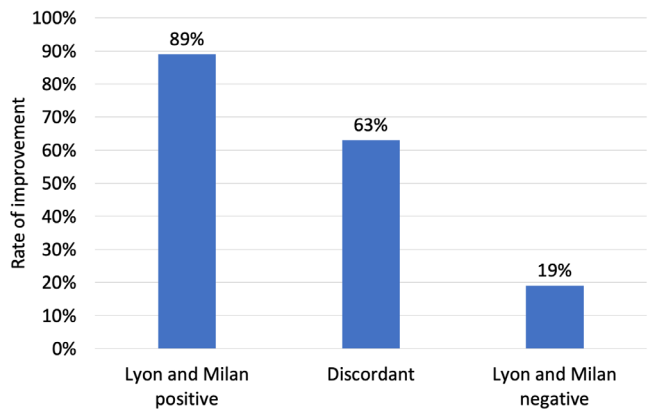


FIGURE 2 | Proportion of patients with clinical improvement according to the concordance between Milan and Lyon scores.

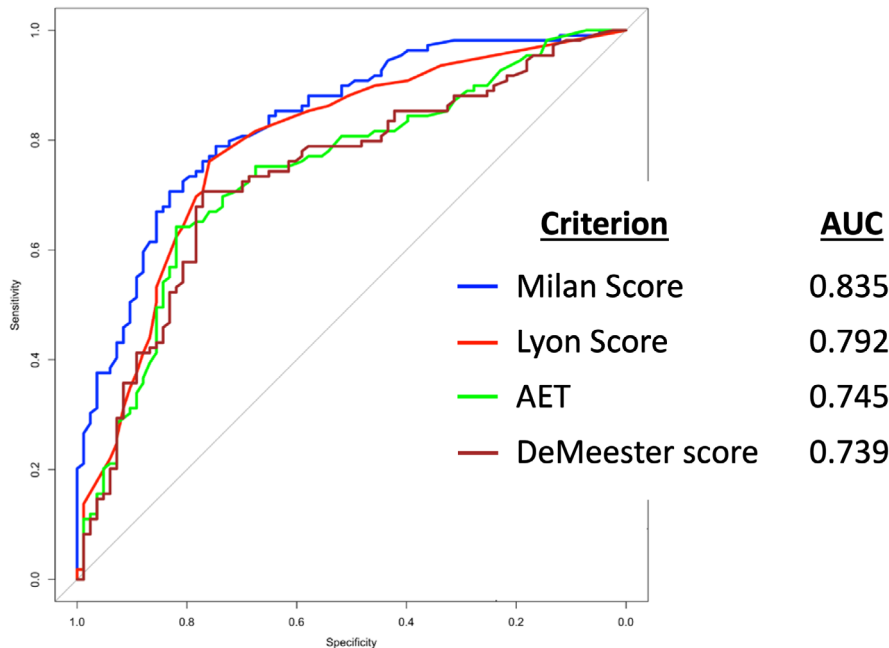


FIGURE 3 | Receiver operating characteristic (ROC) curve of the Lyon score, Milan score, acid exposure time (AET) and DeMeester score in predicting clinical improvement after antireflux therapy.

TABLE 4 | Performance characteristics of Lyon score, Milan score, acid exposure time (AET), DeMeester score and the different components of the Lyon and Milan scores.

	Optimal threshold	AUC, (95% CI)	Sensitivity	Specificity	Positive predictive value	Negative predictive value	<i>p</i>
Lyon score	4.75	0.792 (0.726–0.858)	76.6	75.9	81.0	70.8	<0.001
Milan score	136	0.835 (0.778–0.891)	67.0	85.5	85.9	66.4	<0.001
AET	5.8%	0.745 (0.674–0.815)	64.9	81.9	82.1	61.8	<0.001
DeMeester score	16.9	0.739 (0.668–0.811)	70.3	77.1	76.0	64.4	<0.001
Lyon score components							
Oesophagitis Los Angeles grade		0.553 (0.494–0.611)	14.4	96.4			
AET		0.732 (0.664–0.799)	62.2	81.9			
Reflux episodes		0.630 (0.566–0.695)	44.7	78.3			
MNBI		0.734 (0.667–0.801)	75.7	66.3			
RSA		0.575 (0.525–0.625)	23.4	91.6			
Hiatus hernia		0.642 (0.575–0.710)	62.2	66.3			
Milan score components							
EGJ Type		0.667 (0.600–0.735)	62.2	65.1			
IEM		0.575 (0.517–0.633)	30.6	84.3			
EGJ-CI		0.712 (0.639–0.785)	73.0	67.5			
Positive SLR		0.747 (0.688–0.807)	65.1	84.3			

Abbreviations: EGJ, oesophago-gastric junction; EGJ-CI, oesophago-gastric junction contractile integral; IEM, ineffective oesophageal motility; MNBI, mean nocturnal baseline impedance; RSA, reflux symptom association; SLR, straight leg raise.

patients have less severe GERD. Patients with both negative scores and good outcomes (19% in the overall population and 40% in the surgical group) probably represent the cohort with a placebo effect of PPI and/or ARS, or those whose reflux evaluation suffered from day-to-day variability and either over- or under-diagnosed true GERD [36, 37]. The combined score, integrating the Lyon and Milan scores via logistic regression, achieved an AUC of 0.843, superior to the Lyon score alone (AUC=0.790, $p=0.014$). Given that the Lyon score parameters are routinely collected for GERD diagnosis, incorporating the Milan score enhances the predictive accuracy. This improvement underscores the complementary diagnostic value of the Milan score, potentially refining GERD classification in clinical practice.

Interestingly, in the surgical group, the positivity of one or two scores predicted surgical outcomes with similar proportions (84.6% vs. 91.4%, $p=0.731$) indicating the better performance of surgery (84.6%) compared to PPI (63%) when at least one score is abnormal; this suggests that carefully selected patients might benefit more from surgical intervention.

From this perspective, the Milan score is demonstrated to be a valuable HRM-based tool that corroborates evidence from endoscopy and ambulatory reflux monitoring, represented by the Lyon score, to predict clinical outcomes following medical or surgical treatment in patients with suspected GERD. In this regard, the Milan score could be particularly useful in the evaluation of candidacy for ARS, since patients undergoing presurgical

evaluation are more likely to have HRM performed, making the Milan score a clinically useful tool to select patients for surgery. Moreover, the Milan score may function as an integral part of evidence supporting or refuting true GERD in patients with inconclusive GERD as per the Lyon Consensus 2.0 [4]. Although the subgroup of patients within the Borderline GERD phenotype was small (40 patients), all those with a pathologic Milan score showed good outcomes after treatment (100% specificity). Of note, this finding is based on a small subgroup of patients and requires further analysis to confirm this exploratory result.

We think that the Milan score should not be used as standalone diagnostic tool to select patients for GERD medical treatment or ARS. In clinical practice, it is important that a definitive diagnosis is reached before invasive procedures or acid suppression therapy escalation are proposed to the patients. Based on the results of this study, the Lyon score should be used to confirm GERD based on endoscopy and MII-pH, while the Milan score to assess antireflux barrier dysfunction. Patients with concordant positive Lyon and Milan scores are ideal candidates for ARS, as they exhibit definitive GERD with significant barrier dysfunction. Those with only one positive score (63% improvement) may benefit from surgery but require careful evaluation of symptom severity and PPI response. In patients with inconclusive Lyon phenotypes, a positive Milan score supports GERD diagnosis and predicts treatment response (100% specificity in our cohort). Both scores can be integrated into practice, as MII-pH, endoscopy and HRM are standard in tertiary centres.

The need for escalation of GERD management, particularly ARS, is directly related to both GERD pathophysiology (quantified as disruption of the antireflux barrier by the Milan score) and consequences of disruption of the antireflux barrier (manifest as a high Lyon score). Either score alone is enough when findings are conclusive, such as a sizeable hiatus hernia or advanced grade oesophagitis. When findings are borderline or inconclusive, scoring systems are useful as they allow all abnormal or borderline metrics to carry weight in the determination of conclusive GERD. The combined use of the Lyon and Milan score addresses three key GERD scenarios. In refractory GERD, concordant positive scores (89% improvement) identify patients likely to benefit from escalation to double-dose PPI or ARS, while concordant negative scores (19% improvement) suggest non-GERD causes, such as gut–brain interaction disorders. To increase confidence in conclusive diagnosis, they provide a holistic view of the antireflux barrier disruption and its consequences, providing physicians with adjunctive tools to guide treatment escalation. In tailoring antireflux procedures, the Milan score HRM parameters (e.g., IEM, EGJ-CI) have the potential to guide procedure selection and to predict surgical success (91.4% improvement with concordant positive scores). Further studies are needed to explore the ability of the Milan score to tailor ARS.

The strengths of our study lie in the assessment of outcomes in a large multicentre cohort of patients that improves the generalizability and robustness of the findings across varied clinical settings, with objective physiologic data and outcome measurements using validated questionnaires. Moreover, patients were evaluated at a medium-term follow-up (24.3 months) and both medical and surgical outcomes were taken into account.

Our study has also some limitations. First, symptoms were evaluated with a self-administered questionnaire that, although internationally validated, has only a fair correlation with objective GERD. Even though symptom response was collected prospectively with patient questionnaires, the current data set was collected from retrospective analysis of existing data. Nevertheless, the approach to assessing symptoms and symptom outcome is moving towards patient-reported outcomes, and we believe that validated questionnaires are of higher value than physician assessment of symptoms benefits. However, objective data from repeat ambulatory reflux monitoring would have provided useful adjunctive information, but were not available for this study. Moreover, telephone administration may have introduced minor variability, even though this modality is increasingly recognised as a reliable method for patient-reported outcomes [23, 36] and the use of trained personnel and standardised scripts minimised the bias. Second, the surgical treatment consisted of three different operations (partial and total fundoplication, MSA) and patients were evaluated after surgery at different follow-up intervals. We did not have enough patients within each ARS option for comparative assessment of benefit between these operations. The study population of 532 patients, essential for the comparison between the Lyon and the Milan scores in a large set of patients, dropped to 194 patients that actually completed the follow-up. The substantial loss to follow-up is a limitation that may introduce selection bias, as patients who completed follow-up may differ in motivation or other unmeasured factors. However, comparison of baseline characteristics between

the follow-up cohort and the overall population showed no significant differences (Table S1). The large initial sample size and multicentre design help mitigate this limitation, but future studies with higher follow-up rates are needed to confirm our findings. Also, we were forced to set a cut-off for the Lyon and Milan scores, even though the idea behind them is to provide a risk rate, not to define a positive or negative result of the score. This inevitable dicotomization led to a high rate of patients with good outcomes in the ‘unlikely’ Milan score category (risk rate 25%–50%). The study included patients undergoing pathophysiologic tests for different reasons. This diversity reflects real-world clinical practice but complicates interpretation of treatment outcomes, particularly regarding refractory GERD. Lastly, in our cohort there were patients who underwent medical and even surgical treatment in the absence of objective GERD evidence according to modern standards; while not typically acceptable from our point of view, this unfortunately reflects a real-world clinical scenario in some settings. We feel data from these patients helped solidify the value of documenting objective GERD parameters before considering definitive GERD management options, especially ARS.

In conclusion, the Lyon and Milan scores outperformed AET and DeMeester score in predicting outcomes in GERD patients. When concordant, they provide optimal predictive accuracy, guiding escalation of therapy. Moreover, the Milan score complements existing Lyon 2.0 adjunctive metrics (MNBI, reflux symptoms association, total number of reflux episodes) in predicting outcome in patients with inconclusive GERD.

Author Contributions

Stefano Siboni: conceptualization, investigation, writing – original draft, writing – review and editing, methodology, validation, data curation. **Marco Sozzi:** conceptualization, writing – original draft, writing – review and editing, investigation. **Pierfrancesco Visaggi:** writing – original draft, writing – review and editing, methodology. **Sara Boveri:** formal analysis, data curation, investigation, conceptualization, writing – review and editing. **Benjamin D. Rogers:** writing – review and editing, methodology, validation, supervision. **Nicola De Bortoli:** supervision, writing – review and editing, writing – original draft, methodology. **Anthony Hobson:** writing – original draft, writing – review and editing, investigation. **Brian E. Louie:** writing – original draft, writing – review and editing, supervision. **Yeong Yeh Lee:** supervision, writing – review and editing, writing – original draft. **Salvatore Tolone:** writing – original draft, writing – review and editing, investigation. **Elisa Marabotto:** investigation, writing – original draft, writing – review and editing, supervision. **Jordan Haworth:** investigation, writing – original draft, writing – review and editing. **Megan L. Ivy:** investigation, writing – original draft, writing – review and editing. **Lorenzo Cusmai:** investigation, data curation, writing – original draft, writing – review and editing. **Edoardo Campioli:** data curation, investigation, writing – original draft, writing – review and editing. **Dimitrios Theodorou:** supervision, writing – review and editing, writing – original draft. **Tania Triantafyllou:** investigation, writing – review and editing, writing – original draft. **Andrea Pasta:** writing – original draft, writing – review and editing, investigation. **Francesco Calabrese:** investigation, writing – original draft, writing – review and editing. **Vincent Tee:** investigation, writing – original draft, writing – review and editing. **Daniele Bernardi:** investigation, writing – original draft, writing – review and editing. **Carlo Galdino Riva:** investigation, writing – original draft, writing – review and editing. **Roberta De Maron:** investigation, writing – review and editing, writing – original draft. **Enrique Coss-Adame:** supervision, writing – original draft, writing – review and

editing. **Amit Patel:** writing – review and editing, writing – original draft, supervision. **Emanuele Asti:** writing – original draft, writing – review and editing, supervision. **Edoardo Savarino:** conceptualization, writing – original draft, writing – review and editing, supervision. **C. Prakash Gyawali:** conceptualization, writing – original draft, writing – review and editing, supervision.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

1. J. R. Lechien, M. F. Vaezi, W. W. Chan, et al., “The Dubai Definition and Diagnostic Criteria of Laryngopharyngeal Reflux: The IFOS Consensus,” *Laryngoscope* 134, no. 4 (2024): 1614–1624, <https://doi.org/10.1002/lary.31134>.
2. R. Yadlapati, C. P. Gyawali, J. E. Pandolfino, and CGIT GERD Consensus Conference Participants, “AGA Clinical Practice Update on the Personalized Approach to the Evaluation and Management of GERD: Expert Review,” *Clinical Gastroenterology and Hepatology* 20, no. 5 (2022): 984–994, <https://doi.org/10.1016/j.cgh.2022.01.025>.
3. D. O. Francis, J. A. Rymer, J. C. Slaughter, et al., “High Economic Burden of Caring for Patients With Suspected Extraesophageal Reflux,” *American Journal of Gastroenterology* 108, no. 6 (2013): 905–911, <https://doi.org/10.1038/ajg.2013.69>.
4. C. P. Gyawali, R. Yadlapati, R. Fass, et al., “Updates to the Modern Diagnosis of GERD: Lyon Consensus 2.0,” *Gut* 73, no. 2 (2024): 361–371, <https://doi.org/10.1136/gutjnl-2023-330616>.
5. A. Patel, G. S. Sayuk, and C. P. Gyawali, “Parameters on Esophageal pH-Impedance Monitoring That Predict Outcomes of Patients With Gastroesophageal Reflux Disease,” *Clinical Gastroenterology and Hepatology* 13, no. 5 (2015): 884–891.
6. A. Rengarajan, E. Savarino, M. Della Coletta, M. Ghisa, A. Patel, and C. P. Gyawali, “Mean Nocturnal Baseline Impedance Correlates With Symptom Outcome When Acid Exposure Time Is Inconclusive on Esophageal Reflux Monitoring,” *Clinical Gastroenterology and Hepatology* 18, no. 3 (2020): 589–595.
7. A. Patel, D. Wang, N. Sainani, G. S. Sayuk, and C. P. Gyawali, “Distal Mean Nocturnal Baseline Impedance on pH-Impedance Monitoring Predicts Reflux Burden and Symptomatic Outcome in Gastro-Oesophageal Reflux Disease,” *Alimentary Pharmacology & Therapeutics* 44, no. 8 (2016): 890–898.
8. L. Frazzoni, M. Frazzoni, N. de Bortoli, et al., “Postreflux Swallow-Induced Peristaltic Wave Index and Nocturnal Baseline Impedance Can Link PPI-Responsive Heartburn to Reflux Better Than Acid Exposure Time,” *Neurogastroenterology and Motility* 29, no. 11 (2017): e13116.
9. G. M. Campos, J. H. Peters, T. R. DeMeester, et al., “Multivariate Analysis of Factors Predicting Outcome After Laparoscopic Nissen Fundoplication,” *Journal of Gastrointestinal Surgery* 3, no. 3 (1999): 292–300.
10. H. F. Warren, L. M. Brown, M. Mihura, A. S. Farivar, R. W. Aye, and B. E. Louie, “Factors Influencing the Outcome of Magnetic Sphincter Augmentation for Chronic Gastroesophageal Reflux Disease,” *Surgical Endoscopy* 32, no. 1 (2018): 405–412.
11. R. Salvador, G. Capovilla, M. Santangelo, et al., “Manometric Identikit of a Functioning and Effective Fundoplication for Gastroesophageal Reflux Disease in the High-Resolution Manometry ERA,” *United European Gastroenterology Journal* 12, no. 6 (2024): 749–760.
12. L. F. Johnson and T. R. DeMeester, “Twenty-Four-Hour pH Monitoring of the Distal Esophagus. A Quantitative Measure of Gastroesophageal Reflux,” *American Journal of Gastroenterology* 62 (1974): 325–332.
13. C. P. Gyawali, L. Marchetti, B. D. Rogers, 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,” *American Journal of Gastroenterology* 120, no. 5 (2025): 1009–1018, <https://doi.org/10.14309/ajg.0000000000003083>.
14. S. Siboni, M. Sozzi, I. Kristo, et al., “The Milan Score: A Novel Manometric Tool for a More Efficient Diagnosis of Gastro-Esophageal Reflux Disease,” *United European Gastroenterology Journal* 12, no. 5 (2024): 552–561, <https://doi.org/10.1002/ueg2.12565>.
15. F. Zerbib, “High Resolution Manometry as a Diagnostic Tool for Gastro-Esophageal Reflux: A Step Forward but Not Yet Ready for Clinical Use,” *United European Gastroenterology Journal* 12, no. 5 (2024): 537–538, <https://doi.org/10.1002/ueg2.12573>.
16. S. Siboni, I. Kristo, B. D. Rogers, et al., “Improving the Diagnostic Yield of High-Resolution Esophageal Manometry for GERD: The “Straight Leg-Raise” International Study,” *Clinical Gastroenterology and Hepatology* 21, no. 7 (2023): 1761–1770.
17. R. Yadlapati, P. J. Kahrilas, M. R. Fox, et al., “Esophageal Motility Disorders on High-Resolution Manometry: Chicago Classification Version 4.0,” *Neurogastroenterology and Motility* 33 (2021): e14058.
18. F. Nicodème, M. Pipa-Muniz, K. Khanna, P. J. Kahrilas, and J. E. Pandolfino, “Quantifying Esophagogastric Junction Contractility With a Novel HRM Topographic Metric, the EGJ-Contractile Integral: Normative Values and Preliminary Evaluation in PPI Non-Responders,” *Neurogastroenterology and Motility* 26, no. 3 (2014): 353–360.
19. C. P. Gyawali, B. Rogers, M. Frazzoni, et al., “Inter-Reviewer Variability in Interpretation of pH-Impedance Studies: The Wingate Consensus,” *Clinical Gastroenterology and Hepatology* 19 (2021): 1976–1978.
20. M. Frazzoni, E. Savarino, N. de Bortoli, 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,” *Clinical Gastroenterology and Hepatology* 14 (2016): 40–46.
21. M. Frazzoni, N. de Bortoli, L. Frazzoni, S. Tolone, V. Savarino, and E. Savarino, “Impedance-pH Monitoring for Diagnosis of Reflux Disease: New Perspectives,” *Digestive Diseases and Sciences* 62, no. 8 (2017): 1881–1889.
22. L. R. Lundell, J. Dent, J. R. Bennett, et al., “Endoscopic Assessment of Oesophagitis: Clinical and Functional Correlates and Further Validation of the Los Angeles Classification,” *Gut* 45, no. 2 (1999): 172–180, <https://doi.org/10.1136/gut.45.2.172>.
23. W. W. Chan, M. Schroeder, A. Richardson, et al., “Validation of Esophageal Global Symptom Severity as a Patient Reported Outcome for Evaluation of Reflux Symptoms,” *American Journal of Gastroenterology* (2022): 10–14309. Online ahead of print.
24. R. Jones, O. Junghard, J. Dent, et al., “Development of the GerdQ, a Tool for the Diagnosis and Management of Gastro-Oesophageal Reflux Disease in Primary Care,” *Alimentary Pharmacology & Therapeutics* 30 (2009): 1030–1038.
25. V. Velanovich, “The Development of the GERD-HRQL Symptom Severity Instrument,” *Diseases of the Esophagus* 20 (2007): 130–134.
26. P. C. Belafsky, G. N. Postma, and J. A. Koufman, “Validity and Reliability of the Reflux Symptom Index (RSI),” *Journal of Voice* 16 (2002): 274–277.
27. L. F. Johnson and T. R. DeMeester, “Development of the 24-Hour Intraesophageal pH Monitoring Composite Scoring System,” *Journal of Clinical Gastroenterology* 8, no. Suppl 1 (1986): 52–58.
28. T. R. DeMeester, E. Lafontaine, B. E. Joelsson, et al., “Relationship of a Hiatal Hernia to the Function of the Body of the Esophagus and the

Gastroesophageal Junction,” *Journal of Thoracic and Cardiovascular Surgery* 82 (1981): 547–558.

29. P. J. Kahrilas, R. K. Mittal, S. Bor, et al., “Chicago Classification Update (v4.0): Technical Review of High-Resolution Manometry Metrics for EGJ Barrier Function,” *Neurogastroenterology and Motility* 33, no. 10 (2021): e14113, <https://doi.org/10.1111/nmo.14113>.

30. T. R. DeMeester, J. A. Wernly, G. H. Bryant, A. G. Little, and D. B. Skinner, “Clinical and In Vitro Analysis of Determinants of Gastroesophageal Competence. A Study of the Principles of Antireflux Surgery,” *American Journal of Surgery* 137 (1979): 39–46.

31. B. E. Joelsson, T. R. DeMeester, D. B. Skinner, E. LaFontaine, P. F. Waters, and G. C. O’Sullivan, “The Role of the Esophageal Body in the Antireflux Mechanism,” *Surgery* 92, no. 2 (1982): 417–424.

32. C. P. Gyawali, S. Roman, A. J. Bredenoord, et al., “International GERD Consensus Working Group. Classification of Esophageal Motor Findings in Gastro-Esophageal Reflux Disease: Conclusions From an International Consensus Group,” *Neurogastroenterology and Motility* 29, no. 12 (2017): e13904.

33. S. Siboni, L. Bonavina, B. D. Rogers, et al., “Effect of Increased Intra-Abdominal Pressure on the Esophagogastric Junction: A Systematic Review,” *Journal of Clinical Gastroenterology* 56, no. 10 (2022): 821–830, <https://doi.org/10.1097/MCG.0000000000001756>.

34. J. A. Wernly, T. R. DeMeester, G. H. Bryant, C. I. Wang, R. B. Smith, and D. B. Skinner, “Intra-Abdominal Pressure and Manometric Data of the Distal Esophageal Sphincter. Their Relationship to Gastroesophageal Reflux,” *Archives of Surgery* 115 (1980): 534–539.

35. B. D. Rogers, A. Rengarajan, I. A. Ali, S. L. Hasak, V. Hansalia, and C. P. Gyawali, “Straight Leg Raise Metrics on High-Resolution Manometry Associate With Esophageal Reflux Burden,” *Neurogastroenterology and Motility* 32, no. 12 (2020): e13929.

36. F. Cremonini, D. C. Ziogas, H. Y. Chang, et al., “Meta-Analysis: The Effects of Placebo Treatment on Gastro-Oesophageal Reflux Disease,” *Alimentary Pharmacology & Therapeutics* 32, no. 1 (2010): 29–42, <https://doi.org/10.1111/j.1365-2036.2010.04315.x>.

37. K. L. Greason, D. L. Miller, C. Deschamps, et al., “Effects of Anti-reflux Procedures on Respiratory Symptoms,” *Annals of Thoracic Surgery* 73, no. 2 (2002): 381–385, [https://doi.org/10.1016/s0003-4975\(01\)03407-5](https://doi.org/10.1016/s0003-4975(01)03407-5).

Supporting Information

Additional supporting information can be found online in the Supporting Information section.