

Assessment of anti-reflux surgery with endoscopy: a narrative review

Joseph J. Fantasia, Sarah K. Thompson[^]

Flinders University Discipline of Surgery, College of Medicine & Public Health, Flinders Medical Centre, Bedford Park, South Australia, Australia
Contributions: (I) Conception and design: SK Thompson; (II) Administrative support: None; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: JJ Fantasia; (V) Data analysis and interpretation: Both authors; (VI) Manuscript writing: Both authors; (VII) Final approval of manuscript: Both authors.

Correspondence to: Sarah K. Thompson, MD. Rm 5E221.3, Flinders Medical Centre, Bedford Park, South Australia 5042, Australia.
Email: sarah.thompson@flinders.edu.au.

Background and Objective: Gastro-esophageal reflux disease (GERD) affects a significant portion of the Australian population. Minimally invasive laparoscopic fundoplication is a highly effective treatment in appropriately selected patients. Endoscopy is an important investigation in the evaluation of pre-operative symptoms, as well as to investigate problems postoperatively. However, endoscopic assessment of post-fundoplication anatomy is not standardized and large variations in reporting are seen. The literature was examined for inter-rater reliability of endoscopic assessment of fundoplication and other upper gastrointestinal pathologies.

Methods: An electronic search was completed in Ovid MEDLINE/PubMed, CINAHL, SCOPUS, Cochrane Database of Systematic Reviews and Web of Science. Study characteristics were collated and analyzed.

Key Content and Findings: Fifty-two articles of varying quality were included in the review. The studies were reviewed and grouped based on their anatomical/pathological focus as well as the number of observers and statistical correlation coefficients used.

Conclusions: This comprehensive review has identified a shortage of literature on the inter-rater reliability of assessing fundoplication on endoscopy. Of the studies included which looked at fundoplication, six out of eight had their design focused on comparison pre- and post-operatively rather than specific reliability of assessors. This area of deficiency provides scope and opportunity for further research to improve reporting of fundoplication with endoscopy.

Keywords: Gastro-esophageal reflux disease (GERD); endoscopy; fundoplication; inter-rater reliability; agreement

Received: 28 March 2022; Accepted: 08 July 2022.

doi: 10.21037/aoe-22-11

View this article at: <https://dx.doi.org/10.21037/aoe-22-11>

Introduction

It is estimated that the prevalence of gastro-esophageal reflux disease (GERD) is 10–15% of the Australian population (1,2). Up to 30% of adults experience reflux or reflux-type symptoms, while acknowledging that prevalence

is subject to variation in definitions used in studying GERD symptoms (3).

Medical therapy is generally effective in up to 80% of patients, with approximately 20–30% of patients not responding and experiencing persistent symptoms (1).

[^] ORCID: 0000-0002-9908-6208.

Conservative measures used to address GERD are lifestyle modifications such as weight loss, stress management, and elevation of the head of the bed, with the evidence supporting weight loss as most efficacious treatment (1). Management of confirmed GERD commences with an initial 'step down' approach of four to eight weeks of acid suppression treatment [generally proton pump inhibitors (PPIs)], with doses progressively reduced to maintenance level as symptoms respond (3).

Surgical intervention is indicated in patients who have refractory reflux symptoms on maximal medical therapy (such as patients with volume reflux), those who have unwanted side-effects from anti-reflux therapy, or those who want to avoid taking lifelong medication (1,3,4). The use of minimally invasive laparoscopic fundoplication as a surgical intervention for GERD has risen dramatically since its introduction in the early 1990's (5). In appropriately selected patients, it is highly effective with a 90% satisfaction rate, and a morbidity and mortality rate of less than 0.3% (3,6).

Endoscopy is an essential investigation used in the period following fundoplication if symptoms persist or if new symptoms arise (5,7). Ideally, it is performed by the operating surgeon or an experienced gastroenterologist to assess wrap position and the presence of any hiatal hernia (8). Revisional surgery will be required in 1–5% of the population (9) and an accurate description of the distal esophagus, gastroesophageal junction (GEJ), and fundoplication is critical in determining the cause of failure (5). Endoscopy can reveal twice as many key features as radiography alone, with the objective assessment of post-fundoplication anatomy essential prior to planned revisional surgery (10-13).

However, the description of a fundoplication with endoscopy is not standardized and large variations in reporting are seen (13). A study by Juhasz *et al.* (5) reported that only 32% of post-surgical endoscopy reports mentioned a previous fundoplication! There is significant deficiency in the reporting of fundoplication anatomy on endoscopy performed by community physicians (5) and a lack of uniformity in the description of endoscopic findings (13). Whilst this may be due to a lack of universally accepted endoscopic terminology, common inadequacies in reporting are also seen, such as 'fundoplication changes noted' or 'hiatus hernia seen' without any further description or measurements.

This narrative review will examine the current literature for evidence of inter-rater reliability of endoscopic

assessment of a prior fundoplication. We will also look for evidence of inter-rater reliability of endoscopy in the assessment of other upper gastrointestinal pathology. This article has been prepared in accordance with the Narrative Review reporting checklist (available at <https://aoe.amegroups.com/article/view/10.21037/aoe-22-11/rc>).

Methods

Search strategy

Databases searched were Ovid MEDLINE/PubMed, CINAHL, SCOPUS, Cochrane Database of Systematic Reviews and Web of Science. Development of search terms was aimed to identify papers which focused on the analysis of testing reliability of fundoplication assessment with endoscopy. Inter-rater reliability was defined as the degree of agreement amongst independent observers (in this case, proceduralists performing endoscopy) in the assessment of fundoplication (or other upper gastrointestinal pathologies).

The search terms were developed around four main themes: gastroesophageal reflux disease (GERD), fundoplication, endoscopy, and reliability (*Table 1*). After removal of duplicates, 13,174 papers were screened by title and abstract. The papers that remained were then reviewed by full text to determine suitability (*Figure 1*). In addition, comprehensive inclusion and exclusion criteria were determined.

Inclusion criteria

The papers included in this review had to include the assessment of interobserver reliability/agreement during endoscopic assessment of ear/nose/throat and upper gastrointestinal pathologies. Fundoplication was the main subject of interest, although other pathologies were also included. Further inclusion criteria included primary fundoplication (not revisional fundoplication) and adult subjects (age >18 years).

Exclusion criteria

Papers that did not involve the use of endoscopy nor assessment/reliability testing were excluded. One study was excluded as it involved the re-classification of endoscopy reports with a new reporting system but did not study reliability (13). Meeting abstracts and non-English publications were also excluded. Articles published before

Table 1 Search strategy summary

Items	Specification
Date of search (specified to date, month and year)	(I) 18/02/2021; (II) 23/05/2022
Databases and other sources searched	Ovid MEDLINE/PubMed, CINAHL, Scopus, Cochrane and Web of Science
Search terms used (including MeSH and free text search terms and filters)	Gastroesophageal Reflux, Esophageal Reflux or GERD or GORD or Gastric Acid Reflux or Gastric Acid Reflux Disease or Gastroesophageal reflux or Gastro esophageal reflux or Gastro esophageal Reflux Disease or Gastro oesophageal reflux or Gastroesophageal Reflux Disease or Reflux, Gastroesophageal Fundoplication or Laparoscopic fundoplication or Open fundoplication or Nissen fundoplication or Total fundoplication or Partial fundoplication Endoscopy or Upper endoscopy or Endoscopic assessment or Endoscopic evaluation or Endoscopic Surgical Procedure or Endoscopic Surgical Procedures or Endoscopy, Surgical or Surgical Endoscopy or Surgical Procedure, Endoscopic or Surgical Procedures, Endoscopic Reliab* or Interater reliability or Intrarater reliability or Inter rater reliability or Intra rater reliability or Interobserver reliability or Intraobserver reliability or Inter observer reliability or Intra observer reliability or Inter examiner reliability or Intra examiner reliability or interobserver agreement or validity or assessment or evaluation or agreement
Timeframe	1990–2022
Inclusion and exclusion criteria (study type, language restrictions etc.)	Inclusion criteria: the papers included in this review were required to discuss interobserver reliability/agreement during endoscopic assessment of ear/nose/throat and upper gastrointestinal pathologies. Fundoplication was the main subject of interest, though other anatomy or pathology assessed on upper gastrointestinal endoscopy were also included. Further inclusion criteria were papers based on primary fundoplication (not revisional fundoplication) and studies based on adult subjects (age >18 years) Exclusion criteria: papers that did not involve use of gastroesophageal endoscopy or did not involve assessment or reliability testing were excluded. Papers that looked at reliability assessment of a questionnaire or tool were reviewed however excluded if there was no component assessing reliability of assessment on endoscopy. One study was excluded as it involved the reclassification of endoscopy reports with a new reporting system, but did not study reliability Grey literature and unpublished papers were excluded. Papers without access to full text and those not published in English were also excluded. Articles published prior to 1990 were excluded on the premise of seeking up to date evidence in addition to laparoscopic fundoplication not existing prior to this
Selection process (who conducted the selection, whether it was conducted independently, how consensus was obtained, etc.)	Selection process was conducted by JJ. Fantasia and discussed and reviewed by SK. Thompson

GERD, gastro-esophageal reflux disease; GORD, gastro-oesophageal reflux disease.

1990 were excluded because laparoscopic fundoplication did not exist prior to then.

Results

Fifty-two studies were reviewed, as seen in the PRISMA flow diagram (*Figure 1*).

Distribution of the studies

Studies varied with regard to the anatomy and/or pathology described, as outlined in *Table 2*. Eight studies (*Table 3*) were solely focused on endoscopy related to fundoplication (5,14–20). The most recent study prospectively evaluated inter-rater reliability of fundoplication assessment with

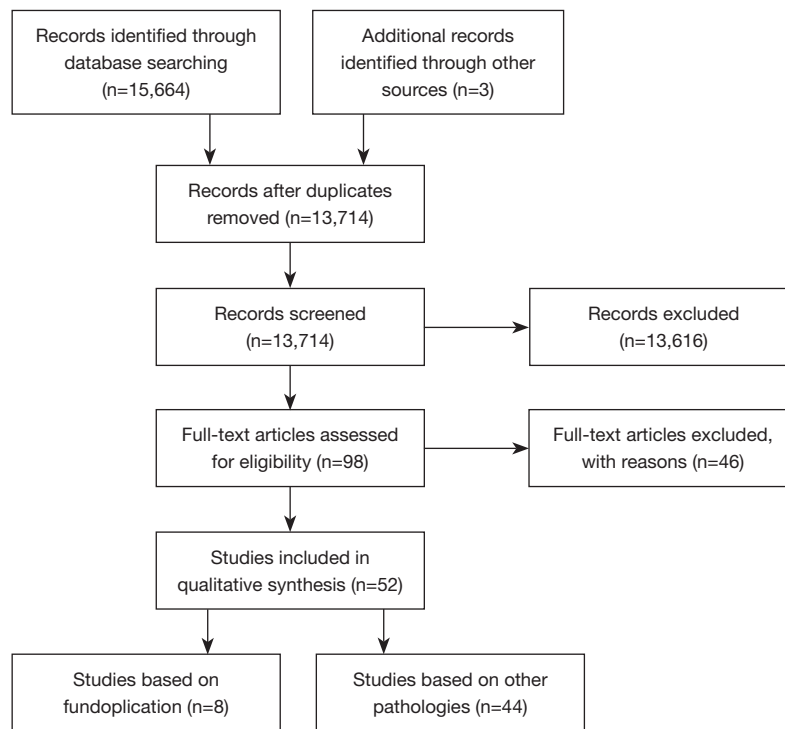


Figure 1 PRISMA flow diagram.

Table 2 Overview of the anatomy/pathologies identified in the studies

Anatomy/pathology	Studies
Fundoplication	Juhasz 2011 (5); Cadière 2009 (14); Csendes 2019 (15); Muls 2013 (16); Petersen 2012 (17); Song 2022 (18); Testoni 2015 (19); Witteman 2013 (20)
Barrett's esophagus	Alvarez Herrero 2009 (21); de Groof 2019 (22); Kato 2017 (23); Lee 2010 (24); Silva 2011 (25); Subramaniam 2020 (26); Trindade 2017 (27); Vahabzadeh 2012 (28)
Laryngopharyngeal reflux	Belafsky 2001 (29); Branski 2002 (30); Chang 2015 (31); Kelchner 2007 (32); Lechien 2020 (33); Musser 2011 (34)
Esophagitis	Armstrong 1996 (35); Bytzer 1993 (36); Kusano 1999 (37); Lundell 1999 (38); Ma 2022 (39); Pandolfino 2002 (40); Rath 2004 (41); Wasielica-Berger 2018 (42)
Secretions/swallowing/oropharyngeal function	Borders 2020 (43); Kaneoka 2013 (44); Leder 2005 (45); Logemann 1999 (46); Miles 2019 (47); Mortensen 2016 (48); Naubauer 2015 (49); Pilz 2016 (50); Pisegna 2018 (51); Starmer 2021 (52); Tohara 2010 (53); Warnecke 2020 (54)
Velopharyngeal insufficiency	Miler 2019 (55); Sie 2008 (56); Yoon 2006 (57)
Inter arytenoid assessment	Coppess 2019 (58)
Laryngopharyngeal sensory discrimination threshold	Cunningham 2007 (59)
Esophageal varices	D'Antiga 2015 (60)
Radiation induced GI toxicity	Lin 2021 (61)
Gastric mucosa atrophy	Miwata 2015 (62)
Rhinosinusitis	Parhar 2014 (63)
laryngeal, hypopharyngeal, and oropharyngeal lesions	Zwakenberg 2016 (64)

GI, gastrointestinal.

Table 3 Overview of fundoplication related studies

Author	Raters/ endoscopists	Sample size/ patients or videos	Classification/grading	Statistical analysis
Juhasz 2011 (5)	Not specified	208 reports	Report review looked for Esophageal caliber and contents Pathologic changes GEJ & relation to fundoplication Ease in traversing the fundoplication Crural impression if any, symmetry Competence of fundoplication, slippage Gastric contents, evidence of recurrent hiatal hernia	9 categories of findings Chi-square test (P value)
Cadière 2009 (14)	Not specified	14 patients	Hill grade (GEV) Valve length (apex of the fundus to the valve lip) Wrap circumference	Comparing length and circumference Mann-Whitney test (P value)
Csendes 2019 (15)	Not specified	150 patients	Esophagitis, cardia type (based on Hill grade)	Fisher's exact test
Muls 2013 (16)	Not specified	66 patients	Hill grade (GEV) Hiatal hernia grade Esophagitis grade	Spearman test (P value)
Petersen 2012 (17)	2	23 videos	Hill grade (GEV) Wrap circumference (0–360 deg) Quality grade of wrap	Mean scores compared Mann-Whitney U test (P value)
Song 2022 (18)	31	20 image sets	Questions based on endoscopic wrap assessment—appearance, position, integrity	Krippendorff alpha Cohen's Kappa
Testoni 2015 (19)	Not specified	50 patients	Hill grade + Jobe length (GEV) Presence of hiatal hernia, esophagitis (LA grading system)	Wilcoxon's test Mann-Whitney tests Fisher's exact test
Witteman 2013 (20)	Not specified	15 patients	Hiatal hernia Esophagitis (LA grading scale) Appearance of the fundoplication	Simply presented as categorical counts

GEJ, gastroesophageal junction; GEV, gastroesophageal valve; LA, Los Angeles classification.

endoscopy (18). One study examined the reporting of fundoplication retrospectively (5), while the remaining six studies evaluated the reporting of fundoplication after surgery in a prospective manner, expressed with p-values indicating statistically significant change.

Of the remaining 44 studies (Table 4) included in the broader search of the literature, 8 were on Barrett's esophagus (21-28), 6 were on laryngopharyngeal reflux (29-34), 8 were on esophagitis (35-42), 12 were on secretions or swallowing/oropharyngeal function (43-54), 3 on velopharyngeal insufficiency (55-57) and the remainder on various ear/nose/throat and upper gastrointestinal pathologies (58-64).

Fundoplication study findings

Song *et al.* 2022 (18) evaluated the accuracy of the

endoscopic assessment of Nissen fundoplication integrity. Thirty-one participants (gastroenterology fellows, subspecialists and foregut surgeons) scored fundoplication anatomy from 20 image sets. They found that diagnostic confidence was considerably varied with poor inter-rater agreement (low to no agreement), and a Krippendorff's alpha less than 0.3. Intra-rater reliability between paired images varied from none to moderate agreement (kappa range =0 to 0.67).

The study by Juhasz *et al.* was a retrospective review study. It involved the assessment of 208 endoscopy reports, performed by general endoscopists outside the specialist center and by upper gastrointestinal surgeons from within the specialist center (5). The authors found inadequate reporting of fundoplication by general endoscopists, with an alarming 68% failing to report the presence of a fundoplication (P<0.05).

Table 4 Overview of all other anatomy/pathology related studies

Anatomy/pathology	Author	Raters/ endoscopists	Sample size/ patients or videos	Classification/grading	Inter-observer	Intra-observer
Barrett's esophagus	Alvarez Herrero 2009 (21)	8	200 images	Simplified classification (regularity of mucosal and vascular patterns)	Kappa	Kappa
	de Groof 2019 (22)	6	40 images	VAS for macroscopic appearance, surface relief, lesions, delineations	Paired <i>t</i> -test, Wilcoxon signed rank test, McNemar test, Spearman's rank	
	Kato 2017 (23)	4	248 images	NBI classification (narrow band imaging)	Kappa	Kappa
	Lee 2010 (24)	34	21 videos	Prague C & M Criteria; location of GEJ and diaphragmatic hiatus	ICC	–
	Silva 2011 (25)	9	84 videos	Grading system (Kanzas, Amsterdam or Nottingham); Prediction—Histological, certainty, time	Cohen's kappa, ICC	–
	Subramaniam 2020 (26)	10	50 images	BLINC	Kappa	Kappa
	Trindade 2017 (27)	8	120 images	Mucosal assessment	Kappa	
	Vahabzadeh 2012 (28)	18	18 videos	Prague C & M Criteria; location of GEJ and diaphragmatic hiatus	ICC	–
Laryngo-pharyngeal reflux	Belafsky 2001 (29)	2	40 patients	8 items—RFS	Pearson product moment	Pearson product moment
	Branski 2002 (30)	5	120 videos	degree of erythema and degree of edema for inter-arytenoid pachydermia; likelihood/severity of LPRD	ICC	Kendall bivariate
	Chang 2015 (31)	10	30 videos	8 items—RFS	Leiss's ICC, Multi rater kappa	–
	Kelchner 2007 (32)	4	30 videos	8 items—RFS	ICC, McNemar's statistic, Log linear regression	–
	Lechien 2020 (33)	5	106 videos	3 parts—RSA	Kendall's Concordance	Spearman's rank
	Musser 2011 (34)	3	36 videos	8 items—RFS	Cohen's kappa, ICC	Cohen's kappa, ICC
Esophagitis	Armstrong 1996 (35)	59	123 images/ videos	LA classification	Cohen kappa	
	Bytzer 1993 (36)	3	150 patients	Savary-Miller classification	Kappa	
	Kusano 1999 (37)	21	50 images	LA classification	Kappa	
	Lundell 1999 (38)	46	22 videos	Endoscopic classification	Kappa	–
	Ma 2022 (39)	2	42 videos	EREFS	ICC	ICC
	Pandolfino 2002 (40)	9	235 images	LA and Hetzel-dent classification	Kappa	Kappa
	Rath 2004 (41)	9	60 patients	LA and Savary-Miller classification; MUSE scoring systems	Kappa	
	Wasielica-Berger 2018 (42)	4	56 images	LA classification	Kappa	

Table 4 (continued)

Table 4 (continued)

Anatomy/pathology	Author	Raters/ endoscopists	Sample size/ patients or videos	Classification/grading	Inter-observer	Intra-observer
Secretions/ swallowing/ oropharyngeal function	Borders 2020 (43)	4	125 videos	Presence, absence or inability to rate LAR	Fleiss' kappa, Cohen's kappa	ICC Cohen's kappa
	Kaneoka 2013 (44)	4	63 patients	Boston Residue and Clearance Scale during FEES	ICC	ICC
	Leder 2005 (45)	3	20 patients	FEES	Kappa	
	Logemann 1999 (46)	2	3 patients	FEES	Paired t-test	
	Miles 2019 (47)	28	10 videos	Secretion Scale	ICC	ICC
	Mortensen 2016 (48)	2	33 patients	Swallowing Assessment of Saliva Scale	Kappa	
	Naubauer 2015 (49)	20	261	Yale Pharyngeal Residue Severity Rating Scale during FEES	Kappa	Kappa
	Pilz 2016 (50)	2	60 videos	Four ordinal FEES variables	Linear weighted kappa coefficient	Linear weighted kappa coefficient
	Pisegna 2018 (51)	44	81 videos	FEES	ICC	Kappa
	Starmer 2021 (52)	3	100 patients	Dynamic imaging grade	Quadratic weighted kappa	Quadratic weighted kappa
Velopharyngeal insufficiency	Tohara 2010 (53)	9	10 patients	16 points—FEES	Cohen's kappa	Cohen's kappa
	Warnecke 2020 (54)	33	10 videos	Categorical variables	Krippendorff alpha	Light's K
	Miler 2019 (55)	16	50 videos	Golding Kushner Scale	ICC, Fleiss' kappa	
Rhinosinusitis	Sie 2008 (56)	16	50 videos	Golding Kushner Scale	ICC, kappa coefficient	ICC
	Yoon 2006 (57)	6	50 videos	Golding Kushner Scale	ICC, kappa coefficient	ICC
	Parhar 2014 (63)	5	50 images	P-J staging	ICC	Fleiss' kappa
Inter arytenoid assessment	Coppess 2019 (58)	4	30 videos	Interarytenoid assessment protocol	Cohen kappa	Cohen kappa
Laryngo- pharyngeal sensory discrimination threshold	Cunningham 2007 (59)	3	27 patients	LPSDT	Spearman Rank	Spearman Rank
Esophageal varices	D'Antiga 2015 (60)	10	100 images	Classification A and B Scales	Fleiss' Kappa, Cohen Kappa	
Radiation induced GI toxicity	Lin 2021 (61)	2	19 patients	Toxicity scoring system	Kappa, Gwet's AC1	
Gastric mucosa atrophy	Miwata 2015 (62)	12	91 patients	Kimura-Takemoto Classification	Kappa	Kappa
Laryngeal, oro/hypo pharyngeal lesions	Zwakenberg 2016 (64)	12	100 images	Lesion assessment	Fleiss' Kappa	Cohen Kappa

VAS, Visual Analog Scale; GEJ, gastroesophageal junction; GI, gastrointestinal; NBI, narrow band imaging; GEJ, gastroesophageal junction; BLINC, Blue Light Imaging for Barrett's Neoplasia Classification; RFS, Reflux Finding Score; LPRD, laryngo-pharyngeal reflux disease; RSA, Reflux Sign Assessment; LA, Los Angeles classification; EREFS, Endoscopic Reference Score: edema, rings, exudate, furrows, stricture; MUSE, metaplasia, ulcer, stricture, erosion; LAR, laryngeal-adductor reflex; FEES, fiberoptic endoscopic evaluation of swallowing; LPSDT, laryngopharyngeal sensory discrimination threshold test; ICC, intraclass correlation coefficient.

The study by Petersen *et al.* was designed to evaluate the fundoplication constructed following an experimental transoral endoscopic procedure using the EsophyX™ device (17). The study involved 2 independent investigators reporting on intraoperative videos of 23 patients, on two occasions (pre- and post-transoral fundoplication). Assessment of the fundoplication involved 3 features; Hill grade [defined as Type I—prominent fold of tissue along the lesser curvature next to the endoscope; Type II—fold is less prominent and there are periods of opening and rapid closing around the endoscope; Type III—fold is not prominent and the endoscope is not tightly gripped by the tissue; Type IV—no fold, and the lumen of the esophagus is open, hiatal hernia is always present (65)], estimated measure of circumference of the wrap, and overall quality grade of the wrap. The authors did not look at inter-rater reliability.

Csendes *et al.* evaluated the objective appearance of a Nissen fundoplication fifteen years postoperatively with endoscopy (15), while remaining four studies evaluated the fundoplication constructed endoscopically (EsophyX™) (14,16,19,20). None of studies disclosed how many raters assessed the fundoplication. Based on the design methods, the assessments were made by the endoscopist performing the procedure, rather than by video review, as was the case in the study by Petersen *et al.* (17). All of the remaining studies (14,16,19,20) reported similar results supporting feasibility of the procedure.

It is important to note that the recent study by Song *et al.* is the only study with a robust design to evaluate inter-rater reliability of fundoplication assessment with endoscopy (18). The authors of the remaining seven studies presented comparative statistical analysis in the form of P values, which are representative of a statistical comparison and association rather than specific inter-rater reliability.

Non-fundoplication study findings

The 44 remaining studies, whilst not assessing fundoplication per se, were designed to determine the inter-rater and/or intra-rater reliability in the assessment and reporting of other upper gastrointestinal pathologies on endoscopy (Table 4). The studies all had a similar design methodology, although they each describe different assessment tools, leading to the use of different statistical analyses of reliability.

The correlation coefficients used to determine reliability were; Kappa, Krippendorff alpha, Light's kappa, Cohen's

kappa, Intra-class correlation coefficient, Fleiss' kappa, Kendall's W and Spearman Rank correlation coefficient.

Discussion

The aim of this narrative review was to examine the current literature for evidence of inter-rater reliability in the assessment of fundoplication assessment with endoscopy. Our secondary aims were to look for any literature on reliability measures when assessing upper gastrointestinal pathology (i.e., not necessarily fundoplication).

We confirmed our suspicion that there is a paucity of data on the reliability of fundoplication assessment with endoscopy. Juhasz *et al.* (5) made the same observation in their study, finding a low percentage of general endoscopists routinely identified the presence of a fundoplication in their endoscopy report. They recommended a standardized proforma to assess not only the presence of a fundoplication, but its integrity and anatomical features in their more recent paper (13). To our knowledge, however, this assessment tool has not been validated externally.

The methodological design varied considerably across the 52 studies included in our narrative review. The initial eight studies which examined fundoplication included a small number of raters, which is associated with low statistical power and reduced veracity of findings (66). These studies were not powered to evaluate reliability between raters, rather they were designed to compare the appearance of the fundoplication pre- and post-intervention. In contrast, the subsequent forty-four studies were designed to examine the reliability between raters when assessing upper gastrointestinal pathology. These studies therefore had a higher number of raters to allow for a more powerful calculation of reliability, and all of these studies used statistical coefficients to determine inter-rater reliability.

Statistical coefficients are a critical component of reliability studies because they determine the level of agreement between different evaluations from a response variable (67). Whilst this may seem similar to comparing results on two occasions (with P values), the important distinction between definitions is the 'level of agreement'. This agreement is important in providing evidence of the closeness of results, rather than simply an expression of results. Established correlation coefficients therefore demonstrate reliability, where higher values demonstrate greater reliability and a smaller error of between subject variability (68).

Conclusions

This narrative review has identified a paucity of literature on the reliability of endoscopic assessment of laparoscopic fundoplication. Of the eight studies which looked at fundoplication, one was designed to assess reliability (albeit with a low number of assessors), and seven had their design focused on comparison pre-and post-operatively rather than the specific reliability of assessors. The remaining 44 studies in our review confirmed that reliability studies are feasible when using endoscopy to assess other gastrointestinal pathologies. These studies will help provide a framework for the development of inter-rater reliability studies in the assessment of fundoplication with endoscopy.

It is important that general endoscopists achieve a high standard when reporting the presence and appearance of a fundoplication. These reports form part of the patient's medical record and are essential in the patient's individualized care. In addition to documentation of the location of the squamocolumnar junction, and findings in the lower esophagus (e.g., ulcerative esophagitis/Barrett's esophagus), the endoscopist should describe the appearance of the fundoplication (i.e., intact, partially intact, or disrupted), the position of the wrap (i.e., above, at, or below the level of the diaphragm), and the presence or not of a hiatus hernia (in the retroflexed position). Images should also be taken to support the description and to serve as a baseline for future comparison. It is equally important that a robust classification system is developed and/or validated for universal reporting.

Acknowledgments

The authors thank Dr. Aliese Millington for her assistance in the development of the database search strategy.

Funding: None.

Footnote

Reporting Checklist: The authors have completed the Narrative Review reporting checklist. Available at <https://aoe.amegroups.com/article/view/10.21037/aoe-22-11/rc>

Peer Review File: Available at <https://aoe.amegroups.com/article/view/10.21037/aoe-22-11/prf>

Conflicts of Interest: Both authors have completed the ICMJE uniform disclosure form (available at <https://aoe.amegroups.com/article/view/10.21037/aoe-22-11/coif>).

SKT serves as an unpaid editorial board member of Annals of Esophagus from September 2019 to August 2023. The other author has no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Keung C, Hebbard G. The management of gastro-oesophageal reflux disease. *Aust Prescr* 2016;39:6-10.
2. Miller G, Wong C, Pollack A. Gastro-oesophageal reflux disease (GORD) in Australian general practice patients. *Aust Fam Physician* 2015;44:701-4.
3. Piterman L, Nelson M, Dent J. Gastro-oesophageal reflux disease--current concepts in management. *Aust Fam Physician* 2004;33:987-91.
4. Fisichella PM, Patti MG. GERD procedures: when and what? *J Gastrointest Surg* 2014;18:2047-53.
5. Juhasz A, Sundaram A, Hoshino M, et al. Endoscopic assessment of failed fundoplication: a case for standardization. *Surg Endosc* 2011;25:3761-6.
6. Dent J, Brun J, Fendrick A, et al. An evidence-based appraisal of reflux disease management--the Genval Workshop Report. *Gut* 1999;44 Suppl 2:S1-16.
7. Galvani C, Fisichella PM, Gorodner MV, et al. Symptoms are a poor indicator of reflux status after fundoplication for gastroesophageal reflux disease: role of esophageal functions tests. *Arch Surg* 2003;138:514-8; discussion 518-9.
8. Kaiser LR, Jamieson G, Thompson SK. *Operative Thoracic Surgery*. 6 ed. Boca Raton: Boca Raton: CRC Press; 2018.
9. Furnée EJ, Draaisma WA, Broeders IA, et al. Surgical reintervention after failed antireflux surgery: a systematic review of the literature. *J Gastrointest Surg* 2009;13:1539-49.

10. Iqbal A, Awad Z, Simkins J, et al. Repair of 104 failed anti-reflux operations. *Ann Surg* 2006;244:42-51.
11. Jaiwala J, Massey B, Staff D, et al. Post-fundoplication symptoms: the role for endoscopic assessment of fundoplication integrity. *Gastrointest Endosc* 2001;54:351-6.
12. Jobe BA, Kahrilas PJ, Vernon AH, et al. Endoscopic appraisal of the gastroesophageal valve after antireflux surgery. *Am J Gastroenterol* 2004;99:233-43.
13. Mittal SK, Juhasz A, Ramanan B, et al. A proposed classification for uniform endoscopic description of surgical fundoplication. *Surg Endosc* 2014;28:1103-9.
14. Cadière GB, Van Sante N, Graves JE, et al. Two-year results of a feasibility study on antireflux transoral incisionless fundoplication using EsophyX. *Surg Endosc* 2009;23:957-64.
15. Csendes A, Orellana O, Cuneo N, et al. Long-term (15-year) objective evaluation of 150 patients after laparoscopic Nissen fundoplication. *Surgery* 2019;166:886-94.
16. Muls V, Eckardt AJ, Marchese M, et al. Three-year results of a multicenter prospective study of transoral incisionless fundoplication. *Surg Innov* 2013;20:321-30.
17. Petersen RP, Filippa L, Wassenaar EB, et al. Comprehensive evaluation of endoscopic fundoplication using the EsophyX™ device. *Surg Endosc* 2012;26:1021-7.
18. Song EJ, Yadlapati R, Chen JW, et al. Variability in endoscopic assessment of Nissen fundoplication wrap integrity and hiatus herniation. *Dis Esophagus* 2022;35:doab078.
19. Testoni PA, Testoni S, Mazzoleni G, et al. Long-term efficacy of transoral incisionless fundoplication with Esophyx (Tif 2.0) and factors affecting outcomes in GERD patients followed for up to 6 years: a prospective single-center study. *Surg Endosc* 2015;29:2770-80.
20. Witteman BP, Kessing BF, Snijders G, et al. Revisional laparoscopic antireflux surgery after unsuccessful endoscopic fundoplication. *Surg Endosc* 2013;27:2231-6.
21. Alvarez Herrero L, Curvers WL, Bansal A, et al. Zooming in on Barrett oesophagus using narrow-band imaging: an international observer agreement study. *Eur J Gastroenterol Hepatol* 2009;21:1068-75.
22. de Groof AJ, Swager AF, Pouw RE, et al. Blue-light imaging has an additional value to white-light endoscopy in visualization of early Barrett's neoplasia: an international multicenter cohort study. *Gastrointest Endosc* 2019;89:749-58.
23. Kato M, Goda K, Shimizu Y, et al. Image assessment of Barrett's esophagus using the simplified narrow band imaging classification. *J Gastroenterol* 2017;52:466-75.
24. Lee CK, Chung IK, Lee SH, et al. Endoscopic partial resection with the unroofing technique for reliable tissue diagnosis of upper GI subepithelial tumors originating from the muscularis propria on EUS (with video). *Gastrointest Endosc* 2010;71:188-94.
25. Silva FB, Dinis-Ribeiro M, Vieth M, et al. Endoscopic assessment and grading of Barrett's esophagus using magnification endoscopy and narrow-band imaging: accuracy and interobserver agreement of different classification systems (with videos). *Gastrointest Endosc* 2011;73:7-14.
26. Subramaniam S, Kandiah K, Schoon E, et al. Development and validation of the international Blue Light Imaging for Barrett's Neoplasia Classification. *Gastrointest Endosc* 2020;91:310-20.
27. Trindade AJ, Inamdar S, Smith MS, et al. Volumetric laser endomicroscopy in Barrett's esophagus: interobserver agreement for interpretation of Barrett's esophagus and associated neoplasia among high-frequency users. *Gastrointest Endosc* 2017;86:133-9.
28. Vahabzadeh B, Seetharam AB, Cook MB, et al. Validation of the Prague C & M criteria for the endoscopic grading of Barrett's esophagus by gastroenterology trainees: a multicenter study. *Gastrointest Endosc* 2012;75:236-41.
29. Belafsky PC, Postma GN, Koufman JA. The validity and reliability of the reflux finding score (RFS). *Laryngoscope* 2001;111:1313-7.
30. Branski RC, Bhattacharyya N, Shapiro J. The reliability of the assessment of endoscopic laryngeal findings associated with laryngopharyngeal reflux disease. *Laryngoscope* 2002;112:1019-24.
31. Chang BA, MacNeil SD, Morrison MD, et al. The Reliability of the Reflux Finding Score Among General Otolaryngologists. *J Voice* 2015;29:572-7.
32. Kelchner LN, Horne J, Lee L, et al. Reliability of speech-language pathologist and otolaryngologist ratings of laryngeal signs of reflux in an asymptomatic population using the reflux finding score. *J Voice* 2007;21:92-100.
33. Lechien JR, Rodriguez Ruiz A, Dequanter D, et al. Validity and Reliability of the Reflux Sign Assessment. *Ann Otol Rhinol Laryngol* 2020;129:313-25.
34. Musser J, Kelchner L, Neils-Strunjas J, et al. A comparison of rating scales used in the diagnosis of extraesophageal reflux. *J Voice* 2011;25:293-300.
35. Armstrong D, Bennett JR, Blum AL, et al. The endoscopic assessment of esophagitis: a progress report on observer agreement. *Gastroenterology* 1996;111:85-92.
36. Bytzer P, Havelund T, Hansen JM. Interobserver variation in the endoscopic diagnosis of reflux esophagitis. *Scand J*

- Gastroenterol 1993;28:119-25.
37. Kusano M, Ino K, Yamada T, et al. Interobserver and intraobserver variation in endoscopic assessment of GERD using the "Los Angeles" classification. *Gastrointest Endosc* 1999;49:700-4.
 38. 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.
 39. Ma C, Bredenoord AJ, Dellon ES, et al. Reliability and responsiveness of endoscopic disease activity assessment in eosinophilic esophagitis. *Gastrointest Endosc* 2022;95:1126-1137.e2.
 40. Pandolfino JE, Vakil NB, Kahrilas PJ. Comparison of inter- and intraobserver consistency for grading of esophagitis by expert and trainee endoscopists. *Gastrointest Endosc* 2002;56:639-43.
 41. Rath HC, Timmer A, Kunkel C, et al. Comparison of interobserver agreement for different scoring systems for reflux esophagitis: Impact of level of experience. *Gastrointest Endosc* 2004;60:44-9.
 42. Wasielica-Berger J, Kemona A, Kiśluk J, et al. The added value of magnifying endoscopy in diagnosing patients with certain gastroesophageal reflux disease. *Adv Med Sci* 2018;63:359-66.
 43. Borders JC, O'Dea MB, McNally E, et al. Inter- and Intra-Rater Reliability of Laryngeal Sensation Testing with the Touch Method During Flexible Endoscopic Evaluations of Swallowing. *Ann Otol Rhinol Laryngol* 2020;129:565-71.
 44. Kaneoka AS, Langmore SE, Krisciunas GP, et al. The Boston Residue and Clearance Scale: preliminary reliability and validity testing. *Folia Phoniatri Logop* 2013;65:312-7.
 45. Leder SB, Acton LM, Lisitano HL, et al. Fiberoptic endoscopic evaluation of swallowing (FEES) with and without blue-dyed food. *Dysphagia* 2005;20:157-62.
 46. Logemann JA, Rademaker AW, Pauloski BR, et al. Interobserver agreement on normal swallowing physiology as viewed by videoendoscopy. *Folia Phoniatri Logop* 1999;51:91-8.
 47. Miles A, Hunting A. Development, intra- and inter-rater reliability of the New Zealand Secretion Scale (NZSS). *Int J Speech Lang Pathol* 2019;21:377-84.
 48. Mortensen J, Jensen D, Kjaersgaard A. A validation study of the Facial-Oral Tract Therapy Swallowing Assessment of Saliva. *Clin Rehabil* 2016;30:410-5.
 49. Neubauer PD, Rademaker AW, Leder SB. The Yale Pharyngeal Residue Severity Rating Scale: An Anatomically Defined and Image-Based Tool. *Dysphagia* 2015;30:521-8.
 50. Pilz W, Vanbelle S, Kremer B, et al. Observers' Agreement on Measurements in Fiberoptic Endoscopic Evaluation of Swallowing. *Dysphagia* 2016;31:180-7.
 51. Pisegna JM, Borders JC, Kaneoka A, et al. Reliability of Untrained and Experienced Raters on FEES: Rating Overall Residue is a Simple Task. *Dysphagia* 2018;33:645-54.
 52. Starmer HM, Arrese L, Langmore S, et al. Adaptation and Validation of the Dynamic Imaging Grade of Swallowing Toxicity for Flexible Endoscopic Evaluation of Swallowing: DIGEST-FEES. *J Speech Lang Hear Res* 2021;64:1802-10.
 53. Tohara H, Nakane A, Murata S, et al. Inter- and intra-rater reliability in fibroptic endoscopic evaluation of swallowing. *J Oral Rehabil* 2010;37:884-91.
 54. Warnecke T, Muhle P, Claus I, et al. Inter-rater and test-retest reliability of the "standardized endoscopic swallowing evaluation for tracheostomy decannulation in critically ill neurologic patients". *Neurol Res Pract* 2020;2:9.
 55. Miller C, Bly R, Cofer S, et al. Multicenter Interrater Reliability in the Endoscopic Assessment of Velopharyngeal Function Using a Video Instruction Tool. *Otolaryngol Head Neck Surg* 2019;160:720-8.
 56. Sie KC, Starr JR, Bloom DC, et al. Multicenter interrater and intrarater reliability in the endoscopic evaluation of velopharyngeal insufficiency. *Arch Otolaryngol Head Neck Surg* 2008;134:757-63.
 57. Yoon PJ, Starr JR, Perkins JA, et al. Interrater and intrarater reliability in the evaluation of velopharyngeal insufficiency within a single institution. *Arch Otolaryngol Head Neck Surg* 2006;132:947-51.
 58. Coppess S, Padia R, Horn D, et al. Standardizing Laryngeal Cleft Evaluations: Reliability of the Interarytenoid Assessment Protocol. *Otolaryngol Head Neck Surg* 2019;160:533-9.
 59. Cunningham JJ, Halum SL, Butler SG, et al. Intraobserver and interobserver reliability in laryngopharyngeal sensory discrimination thresholds: a pilot study. *Ann Otol Rhinol Laryngol* 2007;116:582-8.
 60. D'Antiga L, Betalli P, De Angelis P, et al. Interobserver Agreement on Endoscopic Classification of Oesophageal Varices in Children. *J Pediatr Gastroenterol Nutr* 2015;61:176-81.
 61. Lin D, Moningi S, Abi Jaoude J, et al. Development of an Objective Scoring System for Endoscopic Assessment of Radiation-Induced Upper Gastrointestinal Toxicity. *Cancers (Basel)* 2021;13:2136.
 62. Miwata T, Quach DT, Hiyama T, et al. Interobserver and intraobserver agreement for gastric mucosa atrophy. *BMC*

- Gastroenterol 2015;15:95.
63. Parhar HS, Thamboo A, Habib AR, et al. The interrater and intrarater reliability of the Philpott-Javer staging system based on level of training. *Otolaryngol Head Neck Surg* 2014;150:538-41.
 64. Zwakenberg MA, Dikkers FG, Wedman J, et al. Narrow band imaging improves observer reliability in evaluation of upper aerodigestive tract lesions. *Laryngoscope* 2016;126:2276-81.
 65. Hansdotter I, Björ O, Andreasson A, et al. Hill classification is superior to the axial length of a hiatal hernia for assessment of the mechanical anti-reflux barrier at the gastroesophageal junction. *Endosc Int Open* 2016;4:E311-7.
 66. Button KS, Ioannidis JP, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 2013;14:365-76.
 67. Bujang MA, Baharum N. Guidelines of the minimum sample size requirements for Kappa agreement test. *Epidemiology, biostatistics, and public health* 2017;14. doi: 10.2427/12267.
 68. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med* 1998;17:101-10.

doi: 10.21037/aoe-22-11

Cite this article as: Fantasia JJ, Thompson SK. Assessment of anti-reflux surgery with endoscopy: a narrative review. *Ann Esophagus* 2022.