

Inter-Reader Reproducibility of Quantitative Esophageal Transit Time on Preoperative Barium Esophagram

Peter J Fernández, PhD

University of Wisconsin School of Medicine and Public Health

Sean P Duminie, MD

University of Wisconsin School of Medicine and Public Health

Adam De Jesus, MD

University of Wisconsin School of Medicine and Public Health

Brooklynn Rylee Rickett

University of Wisconsin School of Medicine and Public Health

David Harris, MD

University of Wisconsin School of Medicine and Public Health

Amber Shada, MD

University of Wisconsin School of Medicine and Public Health

Anne Lidor, MD, MPH

University of Wisconsin School of Medicine and Public Health

Matthew R Smith, MD

msmith20@uwhealth.org

University of Wisconsin School of Medicine and Public Health

Research Article

Keywords: Reproducibility of Results, Gastrointestinal Transit, Fluoroscopy, Esophageal Motility Disorders, Preoperative Care

Posted Date: October 18th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11225317/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Purpose

A companion surgical study linked esophageal transit time (ETT) with manometric findings and postoperative symptom burden. This study assessed inter-reader reproducibility of quantitative, frame-based ETT on preoperative barium esophagram for antireflux surgery and evaluated a pragmatic visual binary classification (< 6 versus $\geq$ 6 seconds).

Methods

A randomly selected subcohort of 66 preoperative esophagrams were drawn from a parent antireflux surgery cohort and these motility series were independently reviewed by three blinded readers. ETT was defined as transit of the trailing bolus edge from the first-rib landmark to the tubulovesicular junction. Readers first classified transit visually using a mental stopwatch, then performed frame-based measurement. Quantitative reliability was assessed using two-way random-effects, absolute-agreement intraclass correlation coefficients (ICCs) and measurement-error estimates. Fleiss kappa assessed inter-reader classification agreement and Cohen kappa assessed within-reader agreement between methods.

Results

Mean quantitative ETT was 5.15 ± 1.00 , 5.34 ± 1.21 , and 5.27 ± 1.04 seconds for Readers 1, 2, and 3, respectively. Single-reader reliability was good (ICC(A,1) = 0.777; 95% CI, 0.688–0.848), and reliability of the three-reader mean was excellent (ICC(A,3) = 0.913; 95% CI, 0.869–0.944). The absolute standard error of measurement was 0.52 seconds, the 95% repeatability coefficient was 1.43 seconds, and the mean absolute pairwise reader difference was 0.49 seconds. At 6 second binary threshold, inter-reader agreement was moderate for dichotomized quantitative measurements (Fleiss κ = 0.572; unanimous agreement, 75.8%) and qualitative classifications (Fleiss κ = 0.432; unanimous agreement, 62.1%). Within-reader agreement between methods varied (Cohen κ = 0.208–0.603; agreement, 69.7%–81.8%).

Conclusion

Quantitative, frame-based ETT demonstrated good single-reader reproducibility and excellent reliability when readers' measurements were averaged. At the 6-second threshold, quantitative and qualitative classifications both demonstrated moderate inter-reader agreement, although agreement was higher for the quantitative classification. Qualitative classification may offer a practical preliminary estimate during cine esophagram review.

Introduction

Gastroesophageal reflux disease is common, and antireflux surgery can provide durable symptom control in appropriately selected patients [1–4]. Nevertheless, recurrent or persistent symptoms occur in a clinically important minority of patients, making careful preoperative evaluation central to patient selection [4]. Symptoms, endoscopy, esophagram, pH monitoring, and high-resolution manometry (HRM) provide complementary information; HRM remains important for excluding major motor disorders, while barium esophagram adds anatomic and dynamic information about esophageal motility and bolus clearance [3–9].

Timed barium studies, radionuclide transit studies, and radiographic-manometric comparisons have previously supported transit assessment as an objective marker of esophageal emptying and motor function [10–16]. Using fluoroscopy, esophageal transit time (ETT) is defined as the time required for the trailing edge of a swallowed barium contrast bolus to travel from the first-rib landmark at the thoracic inlet to the tubulovesicular junction (TVJ) in a prone oblique position without gravity assist. In a companion antireflux-surgery cohort, longer preoperative ETT was associated with more failed contractions, weaker distal contractions, and greater long-term patient-reported symptom burden [17].

Clinical usefulness nevertheless depends on whether independent readers can reproduce the quantitative measurement and whether a faster qualitative classification during cine playback is sufficiently reproducible for practice. Conventional qualitative esophagram has shown limited sensitivity and specificity for dysmotility compared with HRM, whereas selected videoesophagram and marshmallow-swallow protocols have shown potential as screening approaches [18–20]. However, the reproducibility of a continuous ETT measurement obtained from a standard single-swallow cine esophagram has not been established, nor has agreement between quantitative measurement and a pragmatic visual threshold classification been evaluated.

The purpose of this observer-agreement study was to quantify inter-reader reproducibility of continuous, quantitative, frame-based ETT across three blinded readers and to evaluate agreement of a pragmatic visual binary classification at a 6-second ETT threshold during standard cine playback.

Materials and methods

Study Design and Population

This retrospective reader study was derived from a previously reported surgical cohort, which screened adults who underwent Nissen or Toupet fundoplication, with or without concurrent hiatal hernia repair, at the study institution between 2014 and 2025 [17]. Patients were eligible for the parent analysis if they had a preoperative esophagram and completed at least one postoperative health-related quality-of-life assessment at 6 weeks, 1 year, 3 years, or 5 years; patients with achalasia or prior operative foregut surgery were excluded [17]. The screening cohort comprised 392 preoperative esophagrams, including 310 performed at the study institution and 82 performed at outside institutions. Exclusions were categorized as unavailable images, prior surgery, absence of a motility series, or insufficient imaging.

Imaging was considered insufficient when the motility cine included multiple swallows, began after the bolus had passed the thoracic inlet, or ended before the bolus reached the distal landmark. After exclusions, 183 examinations remained in the eligible cohort (Fig. 1). The multireader analysis comprised a randomly selected subcohort of 66 examinations with a complete prone-oblique single-swallow motility cine adequate for ETT measurement. No clinical outcomes or manometry variables were analyzed in the present study; those analyses are reported in the companion surgical study [17].

Fluoroscopic Acquisition

Readers evaluated the prone-oblique single-swallow motility series, which minimizes gravitational assistance and permits observation of bolus passage through the thoracic esophagus. The cine was reviewed in the picture archiving and communication system (PACS), where the displayed frame number was used for landmark recording. Acquisition frame rate was easily obtained from the DICOM header and recorded. Variable frame rates were included in the subcohort (4 at 3 fps, 43 at 8 fps, 8 at 10 fps, 1 at 13 fps, and 10 at 15 fps).

Esophageal Transit Time (ETT) Measurement

Qualitative ETT Classification

For the qualitative assessment, each reader reviewed the cine at standard realtime playback speed and classified transit as < 6 seconds or ≥ 6 seconds based on visual assessment and mental timing, without counting frames. The 6-second threshold was selected a priori as a pragmatic approximation of the 5.375-second cutoff derived in the companion cohort [17] because a whole-second threshold was considered more practical for visual estimation during routine cine playback.

Quantitative ETT Measurement

For each examination, readers recorded the first cine frame in which the trailing edge of the contrast bolus passed the first-rib landmark at the thoracic inlet and the first frame in which the trailing edge reached the TVJ, which is the proximal aspect of the esophageal vestibule (Fig. 2). The TVJ is used as the distal anatomic landmark because the lower esophageal vestibule serves as a transient reservoir for larger volumes and its emptying does not necessarily reflect primary esophageal motility. Frame numbers and acquisition frame rate were recorded. Quantitative ETT was calculated using the following formula, yielding a continuous measurement in seconds:

$$\text{ETT (s)} = \frac{|\text{frame}_{TVJ} - \text{frame}_{inlet}|}{\text{frame rate}}$$

where TVJ refers to the tubulovesibular junction (see Supplemental example videos).

Reader Assessment

The three readers included 3 radiologists with 7, 9, and 10 years of fluoroscopic gastrointestinal imaging experience. All readers independently reviewed all 66 examinations and were blinded to clinical information and postoperative outcomes. During the same review session for each examination, the reader first performed the qualitative classification during standard playback and then recorded the quantitative start and end frames.

Statistical Analysis

Study design and reporting were guided by the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) [21]. Inter-reader agreement of the qualitative binary classification and the dichotomized quantitative classification was assessed with Fleiss kappa and the proportion of examinations for which all three readers agreed [22]. Within each reader, agreement between qualitative and quantitative classifications was assessed with Cohen kappa and percent agreement [23].

Continuous quantitative ETT reliability was assessed using a two-way random-effects, absolute-agreement intraclass correlation coefficient. The single-measure ICC (ICC(A,1), equivalent to ICC(2,1)) was the primary estimate and represented the reliability of one reader's measurement; ICC(A,3), equivalent to ICC(2,3), represented the reliability of the mean of three readers. Confidence intervals were calculated using the McGraw–Wong F-distribution method [24–26], and ICCs were interpreted using published thresholds [26]. Absolute standard error of measurement and the 95% repeatability coefficient summarized measurement error.

Mean absolute reader differences and pair-specific Bland–Altman bias and 95% limits of agreement were used to summarize continuous agreement [27]. Each pair-specific analysis contained one independent difference per examination. Acquisition-frame-rate robustness was assessed by correlating frame rate with examination-level mean absolute pairwise reader difference and by repeating the ICC analysis after excluding 3-frame-per-second examinations. Proportional-bias, magnitude-dependent-dispersion, and detailed frame-rate analyses were considered exploratory and are reported in the supplement. Analyses were reproduced directly from the supplied workbook using Python 3.12 (pandas 2.2 and SciPy 1.17).

Results

Cohort analysis

Of the 392 examinations screened, 209 (53.3%) were excluded and 183 (46.7%) were included in the eligible parent cohort (Fig. 1). Among 310 examinations performed at the study institution, 129 (41.6%) were excluded: 12 had no available images, 14 were excluded for prior surgery, 86 lacked a motility series, and 17 had insufficient imaging. Among 82 outside-institution examinations, 80 (97.6%) were excluded: one for prior surgery, 74 for absence of a dynamic motility series, and 5 for insufficient

imaging. The eligible cohort therefore comprised 181 internal and 2 outside examinations. Overall, absence of a motility series accounted for 160 of 209 exclusions (76.6%); the other reasons were insufficient imaging (22/209, 10.5%), prior surgery (15/209, 7.2%), and unavailable images (12/209, 5.7%). A randomly selected subcohort of 66 examinations underwent multireader assessment.

Reader Measurements

Mean ETT was 5.15 ± 1.00 seconds for Reader 1, 5.34 ± 1.21 seconds for Reader 2, and 5.27 ± 1.04 seconds for Reader 3. Across all reader measurements, the mean was 5.25 ± 1.09 seconds, and values ranged from 2.5 to 8.6 seconds (Table 1).

Table 1
Summary of esophageal transit time measurements and intraclass correlation coefficients

Metric	Value
Reader 1 ETT, mean $\pm$ SD	5.15 ± 1.00 s
Reader 2 ETT, mean $\pm$ SD	5.34 ± 1.21 s
Reader 3 ETT, mean $\pm$ SD	5.27 ± 1.04 s
All measurements, mean $\pm$ SD	5.25 ± 1.09 s (range, 2.5–8.6)
Single-reader absolute agreement, ICC(A,1)	0.777 (95% CI, 0.688–0.848)
Mean-of-three-reader agreement, ICC(A,3)	0.913 (95% CI, 0.869–0.944)
Absolute standard error of measurement	0.52 s
95% repeatability coefficient	1.43 s
Mean absolute pairwise difference	0.49 s

Inter-Reader Reproducibility

Single-reader absolute-agreement reliability was 0.777 (95% CI, 0.688–0.848), whereas reliability of the three-reader mean was 0.913 (95% CI, 0.869–0.944). These values corresponded to good and excellent reliability, respectively, according to commonly used thresholds. The absolute standard error of measurement was 0.52 seconds, corresponding to a 95% repeatability coefficient of 1.43 seconds. The mean absolute difference across reader pairs was 0.49 seconds.

Mean reader-pair bias ranged from -0.19 to $+0.07$ seconds, indicating small average systematic offsets. Pair-specific 95% limits of agreement are shown in Fig. 3; differences in individual examinations occasionally approached 3 seconds.

Acquisition frame rate was not associated with examination-level mean absolute pairwise disagreement ($\rho = 0.002$; $P = .988$). Excluding the four examinations acquired at 3 frames per second left reliability

essentially unchanged ($ICC(A,1) = 0.766$; $ICC(A,3) = 0.908$). Detailed frame-rate analyses are provided in the supplement.

Inter-Reader Agreement of Quantitative and Qualitative Classifications

At the 6-second threshold, inter-reader agreement was moderate for the dichotomized quantitative classification (Fleiss $\kappa = 0.572$; all three readers agreed in 50 of 66 examinations [75.8%]) and the qualitative classification (Fleiss $\kappa = 0.432$; all three readers agreed in 41 of 66 examinations [62.1%]). Agreement was higher for the quantitative classification.

Within-reader agreement between qualitative and quantitative binary classifications varied ($\kappa = 0.208$ – 0.603 ; overall agreement, 69.7%–81.8%) (Table 2).

Table 2
Agreement between each reader's qualitative and quantitative classifications at the 6-second threshold.

Reader	κ	Agreement
Reader 1	0.480	78.8%
Reader 2	0.603	81.8%
Reader 3	0.208	69.7%

Discussion

This reader analysis found that quantitative, frame-based ETT, calculated using standardized anatomic landmarks, was reproducible across three independent readers. Absolute agreement was good for a single reader and excellent for the mean of three readers. When ETT was categorized at the 6-second threshold as a practical threshold, both the quantitative and qualitative classifications demonstrated moderate inter-reader agreement, although agreement was higher with the quantitative classification. Agreement between the two approaches varied by reader. Although the performance of the qualitative assessment was more reader-dependent than the quantitative measurement, its moderate reproducibility supports its potential use as a practical real-time estimate. Some of the larger absolute reader differences occurred at longer ETT values; however, these disagreements may be less likely to alter the < 6-versus- $\geq$ 6-second categorization when all measurements remain well above the threshold.

Prior studies have used different approaches to evaluate barium esophagram as a preoperative assessment of esophageal function. O'Rourke et al. evaluated routine esophagram against HRM as the reference standard for detecting dysmotility [18]. Although this approach measures concordance with HRM, it does not establish whether esophagram findings predict clinically meaningful postoperative outcomes, because HRM is a physiologic comparator rather than a postoperative clinical endpoint.

Alicuben and colleagues examined whether video esophagram findings would alter surgical planning [19], whereas Caudell et al. evaluated a structured marshmallow-swallow protocol as a potential alternative to HRM [20]. In contrast, the companion surgical study evaluated quantitative ETT obtained from a standard esophagram sequence in relation to both manometric findings and postoperative symptom burden [17], while the present study evaluates the reproducibility of that measurement across independent readers. A principal practical advantage of ETT is that it can be measured from a routine single-swallow motility cine acquired with standard fluoroscopic equipment, rather than requiring a specialized examination protocol. This may make ETT more generalizable across referral settings, provided that the complete cine sequence and frame-rate metadata are preserved and transferred.

This distinction is important because barium esophagram and HRM provide related but different information. HRM remains the reference standard for diagnosing specific motor disorders and continues to have an important role before antireflux surgery [6]. ETT is not intended to reproduce the detailed physiologic information provided by HRM. Rather, it offers an objective measure of bolus transit that can be obtained from a routinely acquired, noninvasive fluoroscopic sequence. The companion surgical study found associations between longer ETT, manometric abnormalities, and postoperative symptom burden [17]. It also raises the possibility that ETT could eventually contribute to preoperative risk stratification and help identify patients who may derive limited additional information from invasive physiologic testing; this hypothesis requires prospective validation. The present analysis complements the surgical study by showing that the imaging measurement underlying this proposed approach is not dependent on a single reader.

The quantitative and qualitative approaches may serve different roles in clinical practice. Quantitative measurement provides a reproducible value when precision is required. Qualitative classification can be performed during routine cine review and demonstrated moderate inter-reader agreement, suggesting that it may provide a useful initial assessment. Clear cases may be recognized qualitatively, while examinations near the clinical threshold or cases in which the result could affect management should undergo formal frame-based quantification. This combined approach may reduce the additional effort required to incorporate ETT into routine interpretation. Implementation also depends on acquisition and image availability. Quantitative ETT requires preservation of the complete motility cine, frame numbers, and DICOM frame-rate information. The technical prerequisite for ETT measurement is often missing from transferred studies in our experience, an important consideration for referral centers and may necessitate repeating the study. In this screening cohort, most outside examinations and a large proportion of internal examinations were excluded from the analysis, although these totals also reflect exclusion due to prior surgery. Absence of a motility series was the leading reason for exclusion. These findings may limit the immediate feasibility of ETT across routine referral studies; broader adoption will require consistent acquisition, preservation, and transfer of the complete dynamic prone-oblique single-swallow motility sequence.

This study has several limitations. It was a retrospective, single-center analysis of a subset of examinations with complete and measurable cine sequences. Each examination was interpreted only

once by each reader; therefore, intra-reader repeatability was not evaluated. The qualitative analysis used a single 6-second threshold, selected as a pragmatic approximation of the 5.375-second cutoff derived in the companion surgical study [17]. Quantitative classification served as the operational reference for the qualitative comparison rather than an independent clinical gold standard. The study therefore establishes measurement reproducibility but does not independently establish diagnostic accuracy, prognostic validity, or equivalence to HRM. Because ETT was derived from discrete acquisition frames, temporal resolution imposes a lower bound on measurement precision that varies with acquisition frame rate; remaining reader disagreement likely reflects both frame discretization and differences in identification of the bolus edge and anatomic landmarks. Future work should evaluate inter- and intra-reader reproducibility across institutions, reader experience levels, and acquisition systems. Standardized landmark instructions, preservation of cine metadata, and automated or PACS-integrated timing tools may further improve reproducibility and adoption. Prospective multicenter studies should externally validate the pragmatic 6-second threshold and determine whether ETT adds value beyond existing preoperative assessment.

Conclusion

Quantitative, frame-based ETT was reproducible across three independent readers, with good reliability for a single reader and excellent reliability when measurements were averaged. In a companion study, this measurement was associated with strength and failure of contractions on manometry and long-term patient-reported symptom burden following antireflux surgery. Quantitative, frame-based ETT demonstrated good single-reader and excellent three-reader reliability, supporting its use as a reproducible measurement on routine cine esophagram. Visual classification at a 6-second threshold showed moderate agreement and may provide a rapid preliminary estimate, whereas frame-based measurement remains preferable when greater precision is required.

Declarations

Author Contribution

P.J.F. and M.R.S. analyzed the data, constructed the figures, and prepared the manuscript. M.R.S. and S.P.D. collected and prepared the data. D.H., A.S., and A.L. assembled the study cohort. All authors (P.J.F., S.P.D., A.D.J., B.R.R., D.H., A.S., A.L., and M.R.S.) reviewed the final manuscript.

Acknowledgement

The authors would like to acknowledge Madalyn Owen and Haley Alvarez-Garcia for managing the REDCap database.

Data Availability

All data supporting the findings of this study are available within the paper and its Supplementary Information.

References

1. Katz PO, Dunbar KB, Schnoll-Sussman FH, Greer KB, Yadlapati R, Spechler SJ (2022) ACG clinical guideline for the diagnosis and management of gastroesophageal reflux disease. *Am J Gastroenterol* 117:27–56. <https://doi.org/10.14309/ajg.0000000000001538>
2. McKinley SK, Dirks RC, Walsh D et al (2021) Surgical treatment of GERD: systematic review and meta-analysis. *Surg Endosc* 35:4095–4123. <https://doi.org/10.1007/s00464-021-08358-5>
3. Slater BJ, Collings A, Dirks R et al (2023) Multi-society consensus conference and guideline on the treatment of gastroesophageal reflux disease (GERD). *Surg Endosc* 37:781–806. <https://doi.org/10.1007/s00464-022-09817-3>
4. Patti MG, Allaix ME, Fisichella PM (2015) Analysis of the causes of failed antireflux surgery and the principles of treatment: a review. *JAMA Surg* 150:585–590. <https://doi.org/10.1001/jamasurg.2014.3859>
5. Bello B, Zoccali M, Gullo R, Allaix ME, Herbella FA, Gasparaitis A, Patti MG (2013) Gastroesophageal reflux disease and antireflux surgery—what is the proper preoperative work-up? *J Gastrointest Surg* 17:14–20. <https://doi.org/10.1007/s11605-012-2057-5>
6. Gyawali CP, Carlson DA, Chen JW, Patel A, Wong RJ, Yadlapati RH (2020) ACG clinical guidelines: clinical use of esophageal physiologic testing. *Am J Gastroenterol* 115:1412–1428. <https://doi.org/10.14309/ajg.0000000000000734>
7. Baker ME, Einstein DM (2014) Barium esophagram: does it have a role in gastroesophageal reflux disease? *Gastroenterol Clin North Am* 43:47–68. <https://doi.org/10.1016/j.gtc.2013.11.008>
8. Yadlapati R, Kahrilas PJ, Fox MR et al (2021) Esophageal motility disorders on high-resolution manometry: Chicago classification version 4.0. *Neurogastroenterol Motil* 33:e14058. <https://doi.org/10.1111/nmo.14058>
9. Vaezi MF, Pandolfino JE, Yadlapati RH, Greer KB, Kavitt RT (2020) ACG clinical guidelines: diagnosis and management of achalasia. *Am J Gastroenterol* 115:1393–1411. <https://doi.org/10.14309/ajg.0000000000000731>
10. Blonski W, Jacobs J, Feldman J, Richter JE (2025) The history and use of the timed barium esophagram in achalasia, esophagogastric junction outflow obstruction, and esophageal strictures. *Neurogastroenterol Motil* 37:e14928. <https://doi.org/10.1111/nmo.14928>
11. Duy L, Clayton S, Morimoto N, Wang S, DiSantis D (2025) Beyond visualizing the bird beak: esophagram, timed barium esophagram and manometry in achalasia and its 3 subtypes. *Abdom Radiol (NY)* 50:1488–1497. <https://doi.org/10.1007/s00261-024-04554-8>
12. Neyaz Z, Gupta M, Ghoshal UC (2013) How to perform and interpret timed barium esophagogram. *J Neurogastroenterol Motil* 19:251–256. <https://doi.org/10.5056/jnm.2013.19.2.251>

13. Ott DJ, Richter JE, Chen YM, Wu WC, Gelfand DW, Castell DO (1987) Esophageal radiography and manometry: correlation in 172 patients with dysphagia. *AJR Am J Roentgenol* 149:307–311. <https://doi.org/10.2214/ajr.149.2.307>
14. Aksglaede K, Funch-Jensen P, Vestergaard H, Thommesen P (1992) Diagnosis of esophageal motor disorders: a prospective study comparing barium swallow, food barium mixture, and continuous swallows with manometry. *Gastrointest Radiol* 17:1–4. <https://doi.org/10.1007/BF01888495>
15. Blackwell JN, Hannan WJ, Adam RD, Heading RC (1983) Radionuclide transit studies in the detection of oesophageal dysmotility. *Gut* 24:421–426. <https://doi.org/10.1136/gut.24.5.421>
16. Madigan KE, Smith JS, Evans JK, Clayton SB (2022) Elevated average maximum intrabolus pressure on high-resolution manometry is associated with esophageal dysmotility and delayed esophageal emptying on timed barium esophagram. *BMC Gastroenterol* 22:78. <https://doi.org/10.1186/s12876-022-02165-5>
17. De Jesus A, Rickett BR, Smith M, Duminie SP, Harris D, Shada A, Lidor A (2026) Predictive value of preoperative esophagram in postoperative outcomes after antireflux surgery. *Surg Endosc*. <https://doi.org/10.1007/s00464-026-13126-4>
18. O'Rourke AK, Lazar A, Murphy B, Castell DO, Martin-Harris B (2016) Utility of esophagram versus high-resolution manometry in the detection of esophageal dysmotility. *Otolaryngol Head Neck Surg* 154:888–891. <https://doi.org/10.1177/0194599816629379>
19. Alicuben ET, Bildzukewicz N, Samakar K, Katkhouda N, Dobrowolsky A, Sandhu K, Lipham JC (2019) Routine esophageal manometry is not useful in patients with normal videoesophagram. *Surg Endosc* 33:1650–1653. <https://doi.org/10.1007/s00464-018-6456-x>
20. Caudell CW, Covil EP, Gilpin JW, Hodgins B, Ewing A, Kothari SN (2022) Can the marshmallow esophagram replace high-resolution manometry as an appropriate screening for esophageal motility prior to anti-reflux surgery? *Am J Surg* 224:1366–1369. <https://doi.org/10.1016/j.amjsurg.2022.10.036>
21. Kottner J, Audigé L, Brorson S et al (2011) Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 64:96–106. <https://doi.org/10.1016/j.jclinepi.2010.03.002>
22. Fleiss JL (1971) Measuring nominal scale agreement among many raters. *Psychol Bull* 76:378–382. <https://doi.org/10.1037/h0031619>
23. Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20:37–46. <https://doi.org/10.1177/001316446002000104>
24. Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 86:420–428. <https://doi.org/10.1037/0033-2909.86.2.420>
25. McGraw KO, Wong SP (1996) Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1:30–46. <https://doi.org/10.1037/1082-989X.1.1.30>
26. Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 15:155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>

Figures

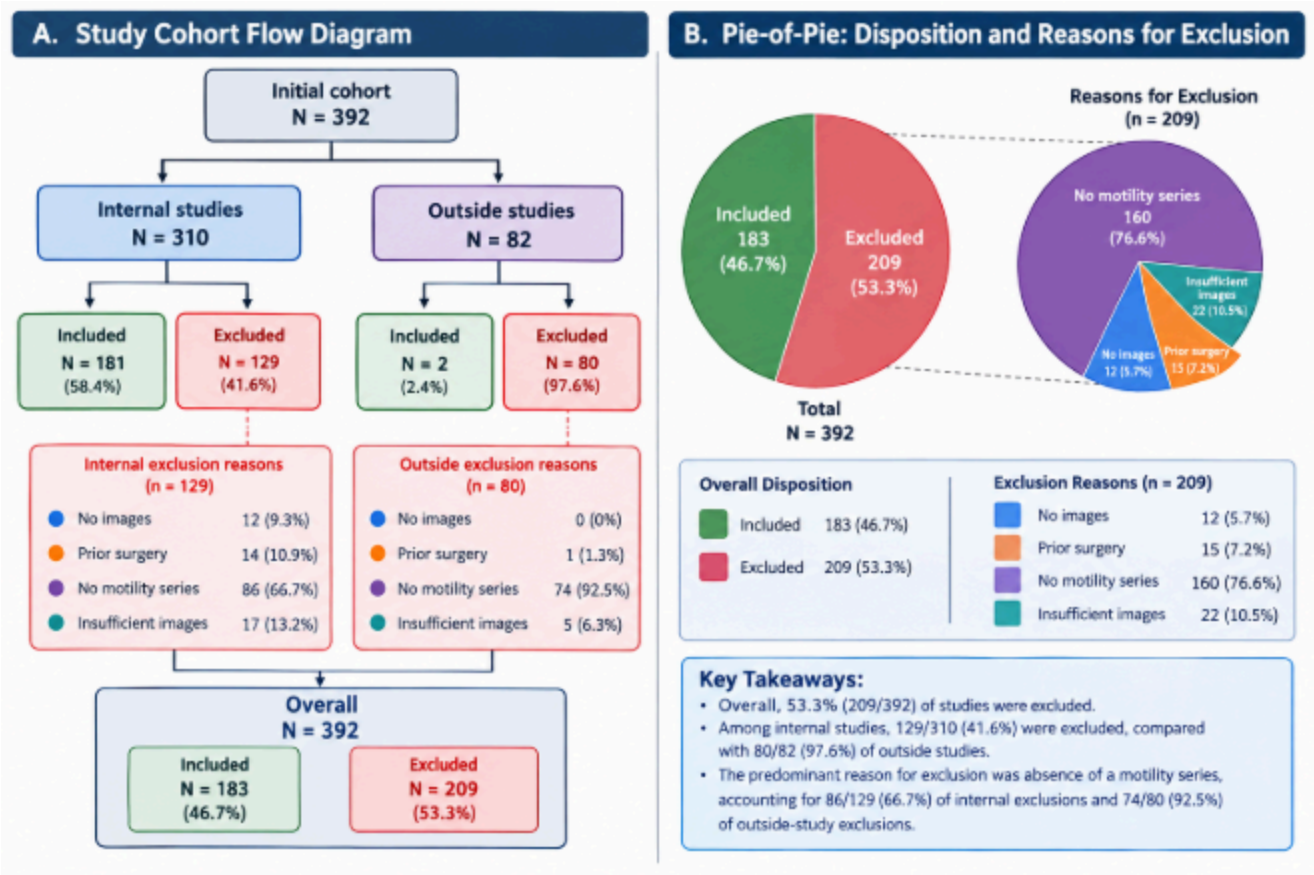


Figure 1

Breakdown of the larger study cohort. (A) Flow diagram showing internal and outside examinations and reasons for exclusion. (B) Overall disposition and distribution of reasons among the 209 excluded examinations. From the included cohort, a subcohort of 66 examinations were selected for this reproducibility study.



Figure 2

Fluoroscopic motility study demonstrating esophageal transit time (ETT) measurement. The patient is in a prone-oblique position for the dynamic motility series, which minimizes gravitational assistance. Swallowed barium contrast appears as dense signal in the esophagus, stomach, and small bowel. The first and last frames illustrate the proximal and distal measurement endpoints at the first-rib landmark and tubulovesicular junction (TVJ), respectively. The frame number was recorded when the trailing edge of the bolus reached each landmark. ETT was calculated as the difference between the TVJ and inlet frame numbers divided by the acquisition frame rate. A small amount of proximal escape is present and within normal limits.

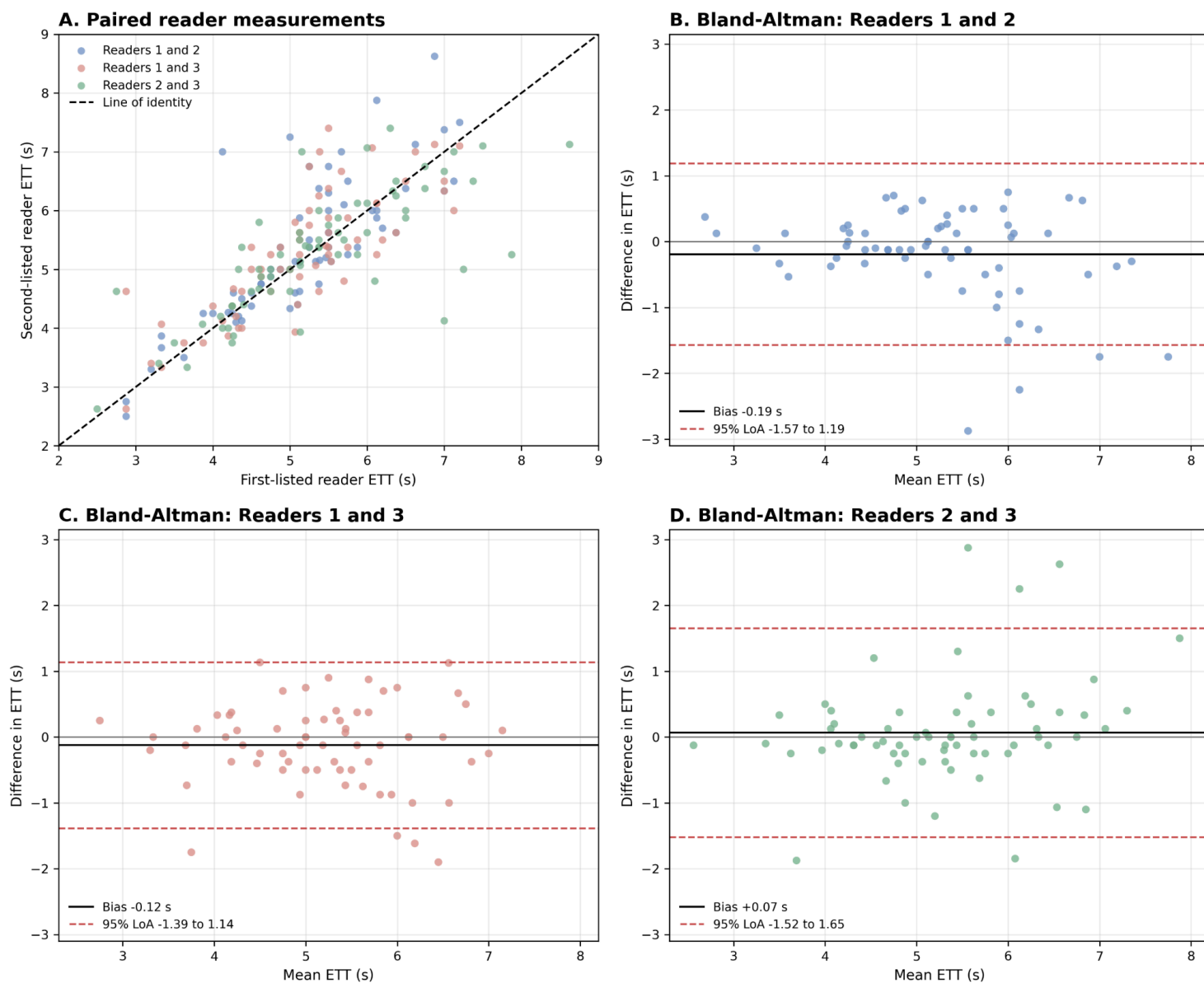


Figure 3

Inter-reader reproducibility of quantitative, frame-based esophageal transit time. (A) Paired reader measurements with the line of identity. (B-D) Pair-specific Bland–Altman plots of first-listed reader minus second-listed reader against pair mean. Black lines show mean bias, and red dashed lines show 95% limits of agreement. Each pair-specific analysis contains one difference per examination

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ETTnormal3timedcine.mp4](#)
- [ETTvestibule1timedcine.mp4](#)
- [ETTnormal2timedcine.mp4](#)

- [ETTnormal1timedcine.mp4](#)
- [ETTnormal4timedcine.mp4](#)
- [ETTvestibule2timedcine.mp4](#)
- [ETTdoubleallowcine.mp4](#)
- [ETTsupplementalstatisticalanalysesfinal.docx](#)