Intra- and Interrater Reliability of Ultrasound Assessment of Gastric Volume

Richelle Kruisselbrink, M.D., F.R.C.P.C., Cristian Arzola, M.D., F.R.C.P.C., Ryan Endersby, M.D., F.R.C.P.C., Cyrus Tse, B.Sc., Vincent Chan, M.D., F.R.C.P.C., F.R.C.A., Anahi Perlas, M.D., F.R.C.P.C.

ABSTRACT

Background: Gastric sonography can provide information about gastric content and volume that can help determine aspiration risk at the bedside. The primary objective of this study is to assess the intrarater and interrater reliability of a previously validated method of gastric volume assessment based on gastric antral area. The secondary objective is to evaluate the agreement between two different methods to measure gastric antral area.

Methods: Three independent raters performed a standardized gastric ultrasound assessment in healthy subjects who had been randomly allocated to ingest a predetermined volume of clear fluid (apple juice) from 0 to 400 ml. Each rater measured the gastric antral area, using twice the two-diameter method and twice the free-tracing method. The rater order was allocated at random and raters were unaware of the volume ingested and of one-another's measurements. The Guidelines for Reporting Reliability and Agreement Studies were followed for conducting and reporting this study.

Results: Twenty-two volunteers were studied. Ultrasound assessment of antral cross-sectional area and volume was found to have "nearly perfect" intrarater and interrater reliability (correlation coefficient >0.8) with maximum differences within 13%. A Bland–Altman analysis suggests that the free-tracing method and the two-diameter method are essentially equivalent, within a clinically acceptable level of agreement.

Conclusions: Ultrasound assessment of gastric volume by clinical anesthesiologists is highly reproducible with high intrarater and interrater reliability. The free-tracing method to measure antral cross-sectional area is equivalent to the two-diameter method. (ANESTHESIOLOGY 2014; 121:46-51)

ASPIRATION of gastric contents is a serious perioperative event associated with morbidity and mortality.[1–3] Patients with a "full stomach" are at increased aspiration risk when sedation or general anesthesia impair their lower esophageal sphincter tone and protective airway reflexes.[4] Gastric ultrasound assessment is the first validated noninvasive imaging tool that can provide information about the nature and volume of gastric content at the bedside.[5–8] Furthermore, several studies suggest that the cross-sectional area of the gastric antrum (antral CSA) can accurately predict gastric fluid volume.[6–8] However, the reliability of ultrasound gastric volume assessment has not yet been established. The primary aim of this study is to evaluate the interrater and intrarater reliability of ultrasound measurements of antral CSA and gastric volume. The second objective is to evaluate the agreement between the traditional two-diameter method (TDM) and a less-reported but simpler and more convenient free-tracing method (FTM) of antral CSA measurement.[7,9–11]

Materials and Methods

After obtaining Research Ethics Board approval from the University Health Network, (Toronto, Ontario, Canada) and informed consent, we conducted this prospective cross-sectional study on healthy volunteers. The study took place at the Toronto Western Hospital in Toronto, Ontario, Canada. Healthy volunteers were sought by posting study adds on community message boards within the hospital. Volunteers were enrolled between November 2011 and June 2012. Inclusion criteria were: age of 18 to 85 yr, American Society of Anesthesiologists' physical status class I to II, body mass index less than 35 kg/m², height greater than 145 cm, and the ability to understand the protocol and provide informed consent. Exclusion criteria were: pregnancy, diabetes mellitus, a history of upper gastrointestinal disease (including hiatus hernia and gastric tumors), and previous surgical procedures on the esophagus, stomach, or upper abdomen.

What We Already Know about This Topic

- We attempt to bring patients to the operating room with small quantities of gastric content to prevent aspiration.

What This Article Tells Us That Is New

- Gastric sonography can provide information about the volume of the stomach. Ultrasound assessment of gastric volume by clinical anesthesiologists is highly reproducible with high intrarater and interrater reliability.

Ultrasound examinations were conducted with a low-frequency (2 to 5 MHz) curvilinear-array transducer using a Philips (CX 50) (Bothell, WA) or Sonosite (M-Turbo) system (Bothell, WA), with image-compounding technology. After an 8-h fasting period for both solids and liquids, subjects underwent a baseline gastric ultrasound examination by a certified sonographer in the supine and right lateral decubitus positions to rule out the presence of significant gastric volume at baseline. After the baseline assessment, each subject was randomized to ingest one of five predetermined volumes of apple juice (0, 100, 200, 300, or 400 ml). Randomization was performed with a computer-generated list of random numbers and concealed in opaque envelopes. A standardized scanning protocol was carried out beginning 3 min after ingestion. Subjects underwent three gastric ultrasound examinations by three independent raters in random order.

The three raters had variable proficiency levels, but all had previous experience in gastric sonography. The first rater was a certified sonographer with more than 10-yr clinical experience and 4-yr experience (>500 gastric scans) in gastric ultrasound assessment. A second rater was a clinical anesthesiologist with more than 10-yr experience in other ultrasound clinical applications and 4-yr experience (>500 previous gastric scans) in gastric sonography. The third rater was an anesthesia fellow with 3-yr experience in other ultrasound applications and 6-month experience (>50 scans) in gastric sonography.

The gastric antrum was imaged in a sagittal plane, between the left lobe of the liver and the pancreas, at the level of the aorta, with the subjects in the right lateral decubitus, as previously reported (fig. 1).[5–7] Frequent peristaltic contractions are a normal occurrence after ingestion of fluid and are readily recognized during gastric ultrasound assessment as temporary decreases in antral diameter. Raters were instructed to obtain the images between (and not during) peristaltic contractions to avoid underestimating antral CSA and gastric volume. Raters were instructed to obtain the images within 5 min. Each rater obtained six independent still images of the gastric antrum. Each image was labeled and stored for review by the rater at the end of the study day. These steps were followed to minimize the time interval between the three raters, which could negatively impact reliability due to gastric emptying. At the end of the study day, each rater independently reviewed his/her generated images and measured antral CSA using both the TDM and the FTM. The TDM is frequently reported in the literature. It consists on measuring two perpendicular diameters of the antrum (fig. 2) and calculating the antral area, assuming that the antrum has a perfect elliptical shape, using a standard formula for the surface area of an ellipse as follows:

$$CSA = \frac{AP \times CC \times \pi}{4}$$
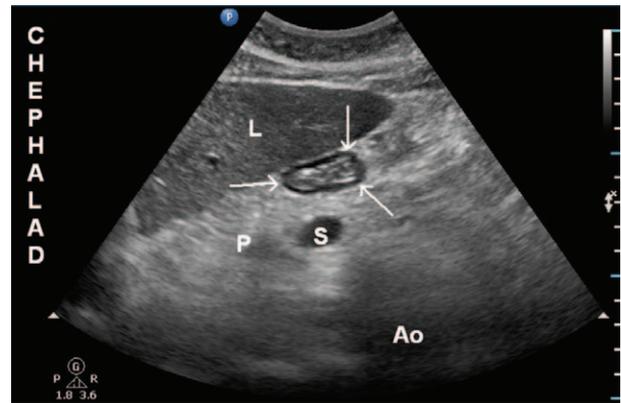
Fig. 1. Sonographic image of the gastric antrum in the epigastric area obtained in a sagittal plane using a curved-array low-frequency probe. The *arrows* signal the gastric antrum which appears empty, small, with no discernible content. Note a hypoechoic layer within the gastric wall that corresponds to the muscularis propriae. Ao = aorta; L = liver; P = pancreas; S = splenic vein.
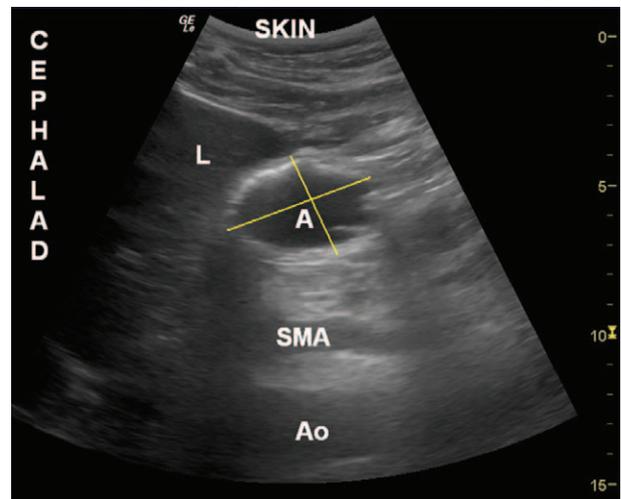
Fig. 2. The two-diameter method of gastric antral area measurement. The image was obtained in the epigastric area with the patient in the right lateral decubitus using a curved-array low-frequency probe. Note the presence of clear fluid in the stomach seen as hypoechoic-anechoic content. The two *yellow perpendicular lines* represent the antero-posterior and cranio-caudal diameters measured from serosa to serosa, used for the area calculation that assumes the antral cross-section is an *ellipse*. A = antrum; Ao = aorta; L = liver; SMA = superior mesenteric artery.

where AP represents the antero-posterior antral diameter, CC represents the cranio-caudal antral diameter, and π is 3.1416.

The FTM consists on measuring the antral area by using the free-tracing caliper of the ultrasound unit (fig. 3). This method is simpler to apply and does not need an intermediate calculation. It would therefore be a preferable method to use at the bedside if it is proven to be accurate.

The first set of three images was used by each rater to generate one data point for the TDM (an average of three

measurements) and one data point for the FTM (an average of three measurements). Similarly, the second set of three images was used by each rater to obtain a second data point for both TDM and FTM as an average of three measurements in both cases. Using three images per data point is a standard practice in gastric sonography and has been previously reported by several authors.[7,8,10] Each rater generated a total of four measurements per subject (twice with the TDM and twice with the FTM). The examiners were blinded to the volume ingested and unaware of the other raters' findings.

The antral CSA was used to calculate total gastric volume based on a previously validated model as follows:

$$\text{Volume (ml)} = 27.0 + 14.6 \times \text{Right-lat CSA (cm}^2) - 1.28 \times \text{age (yr)}$$

where Right-lat CSA is the antral CSA measured in the right lateral decubitus position.

We followed the Guidelines for Reporting Reliability and Agreement Studies in conducting and reporting our investigation.[12]

Statistical Analysis

We estimated the sample size required to test the null hypothesis H_0: $\rho = \rho_0$ versus the alternative hypothesis H_1: $\rho > \rho_0$, where ρ is the "true" reliability coefficient and ρ_0 a specified value for ρ. We aimed to demonstrate an intrarater reliability coefficient of 0.8 and an interrater reliability coefficient of 0.6 (denoted by ρ_{intra} and ρ_{inter}) which have been previously considered to express "almost perfect" and "substantial" reliability, respectively.[13] Referencing the graphs by Eliasziw et al.,[13] we estimated that 22 subjects, 3 raters, and 2 measurements per rater would result in a power larger than 80%, with a significance level of $\alpha = 0.05$.

The intraclass correlation for random-effects models based on repeated-measures ANOVA[14] was used to evaluate intrarater and interrater reliability as initially described by Shrout and Fleiss.[15] In addition, we estimated the absolute and relative differences between the two measurements using the same method (TDM or FTM) among raters as an indicator of interrater agreement.[16]

The level of agreement between the TDM and the FTM was estimated with a Bland–Altman[16,17] analysis to place the magnitudes of the differences in a clinical context. This analysis plots the difference between the two methods against the mean of both methods for each subject. The 95% limits of agreement for the differences were also calculated. The assumption of normal distribution of the differences was checked by the Shapiro–Wilkinson test for normal data. We also estimated proportions of agreement between the two methods among raters within specific limits.[12,18]

The concordance correlation coefficient between the two methods was calculated.[14,19] This coefficient measures the variation of their linear relation from an ideal 45° line through the origin (perfect agreement). It measures how far each observation deviates from the line that best fits the data (precision), and also how far this line deviates from the 45° line through the origin (accuracy).[20] Although precision is expressed by the Pearson correlation coefficient, which as a sole measure could potentially over-estimate agreement, the accuracy is expressed by a bias correction factor.[16] To visually represent what concordance correlation coefficient evaluates, the TDM measurements were plotted against the FTM measurements, and the line that best fit the data, compared with the line of perfect agreement. Statistical analyses were performed using Stata/IC 12.0 for Mac (Stata Corp, College Station, TX).

Results

Twenty-two volunteers (7 women and 15 men) were enrolled and completed the study. There were no missing data. Demographics are presented in table 1. Ultrasound assessment of gastric volume showed "nearly perfect" overall interrater reliability with intraclass correlation coefficient of 0.96 for both the TDM and FTM (table 2). The interrater reliability remained substantial for each gastric volume level between 0 and 400 ml, even though this secondary calculation may not be sufficiently powered as only four to five subjects were randomized to any given volume level (table 2). Similarly, the intrarater reliability was nearly perfect for all three sonographers with intraclass correlation coefficients of 0.96 to 0.99 (table 3). To place these results in a clinical context, the median absolute difference from mean values was
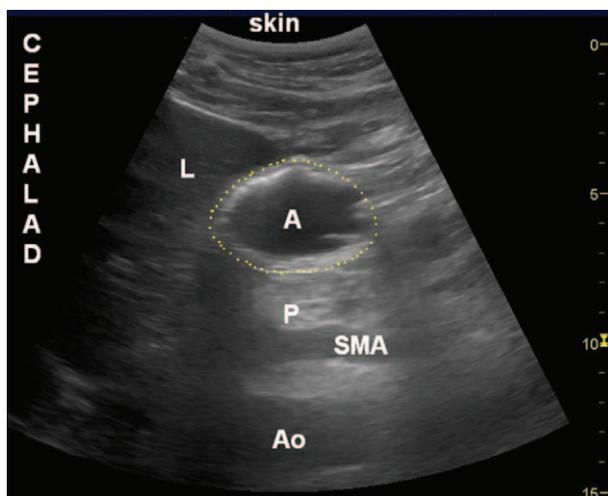
Fig. 3. The free-tracing method of gastric antral area measurement. The image was obtained in the epigastric area with the patient in the right lateral decubitus using a curved-array low-frequency probe. Note the presence of clear fluid in the stomach seen as hypoechoic-anechoic content. The *yellow dotted line* surrounding the gastric antrum represents a manual tracing of the outer layer of the gastric wall (serosa) which allows direct area measurement using the equipment internal caliper. A = antrum; Ao = aorta; L = liver; P = pancreas; SMA = superior mesenteric artery.

Table 1. Demographics

	Mean ± SD (n = 22)	Min–max (n = 22)
Age (yr)	28 ± 7.2	22–45
Weight (kg)	71.7 ± 11.6	50–84
Height (cm)	172.2 ± 6.4	162–182
BMI (kg/cm^2)	24.1 ± 2.9	18.8–28.1

BMI = body mass index.

Table 2. Interrater Reliability

	Two-diameter ICC (95% CI)	Free-tracing ICC (95% CI)
All volumes, ml	0.96 (0.91–0.98)	0.96 (0.92–0.98)
0	0.98	0.98
100	0.93	0.93
200	0.68	0.80
300	0.97	0.96
400	0.99	0.98

ICC = intraclass correlation coefficient.

Table 3. Intrarater Reliability

	Two-diameter Method ICC (95% CI)	Free-tracing Method ICC (95% CI)
Rater 1	0.97 (0.94–0.99)	0.97 (0.94–0.99)
Rater 2	0.98 (0.95–0.99)	0.99 (0.97–0.99)
Rater 3	0.96 (0.91–0.98)	0.96 (0.90–0.98)

ICC = intraclass correlation coefficient.

9.5 ml (p25–p75: 3 to 22 ml), whereas the median relative difference was 2.7% (p25–p75: 1.1 to 5.2%) when using the TDM. The FTM displayed differences of similar magnitudes. The maximum relative difference within each method was not higher than 13%.

Regarding the secondary outcome, a Bland–Altman analysis showed a mean observed difference of –0.33 cm^2 with an SD of 0.88 cm^2 (table 4). The upper and lower values of the 95% limits of agreement band were 1.4 and –2.07 cm^2, which correspond to volumes of 13.1 and –20.5 ml, respectively (fig. 4). Ninety-two percent, 96%, and 100% of the values fell within an absolute difference to the mean observation of 15, 20, and 25 ml, respectively. In other words, the TDM and FTM yield similar volume assessments within a difference of 15 ml in 92% of cases, within 20 ml in 96% of cases, and within 25 ml in 100% of cases. For greater clarity, in figure 1, we can appreciate that, for example, for a volume of 300 ml in the stomach, the

difference between the two methods is not greater than 25 ml. Thus, the two methods are essentially equivalent. Furthermore, the concordance correlation coefficient was nearly perfect (0.995) with high precision and accuracy (Pearson correlation coefficient of 0.996 and bias of 0.999) (table 4 and fig. 5).

Discussion

We previously validated a mathematical model for calculating gastric volume (up to 500 ml) based on an ultrasound assessment of gastric antral CSA.[7] Gastric sonography is a novel point-of-care application of diagnostic ultrasound that allows the clinical anesthesiologist to evaluate a patient's gastric content and volume and thus aspiration risk at the bedside, and help guide anesthetic and airway management. However, for this new diagnostic tool to be clinically applicable, it needs to be not only valid (accurate under ideal study conditions) but also reliable (*i.e.*, reproducible) with low intrarater and interrater variability.

The results confirm our hypothesis that measurements of gastric antral CSA using bedside ultrasound are highly reliable both within the same rater and among raters. The median relative difference between measurements was only 2.7% with a maximum relative difference not greater than 13%. Because fasted individuals have baseline gastric volumes of up to approximately 1.5 ml/kg (approximately 100 ml for the average adult) without a significant aspiration risk, the absolute volume differences observed of 9.5 ml (interquartile range of 3 to 22 ml) are well within clinically acceptable margins of error.[21,22]

Intrarater reliability is often higher than interrater reliability in many studies evaluating diagnostic tools, because one potential source of variance, the rater, is eliminated.[10,23–26] However, in the current study, both intrarater and interrater reliability was similarly high. This may be a result of a rigorous definition and standardization of the scanning protocol used by all three raters. Lack of such standardization has been implicated in less optimal reproducibility measurements in other studies.[10] Four fundamental components of our protocol included: locating the antrum in cross-section in the sagittal epigastric plane that coincides with the long-axis view of the aorta, placing the patient in the right lateral decubitus position, taking measurements between peristaltic contractions when the antrum is at rest, and measuring the antrum from serosa to serosa including the full thickness of the gastric wall. All of these factors affect antral size as well as the sensitivity of the prediction model particularly at low gastric volumes.[6]

Table 4. Agreement between TD and FT Methods: Bland–Altman Analysis

TD *vs.* FT	Mean Difference	SD	95% Limits of Agreement	CCC	Precision (Pearson)	Bias
CSA (cm^2)	−0.33	0.88	−2.1 to 1.4	0.995	0.996	0.999
Volume (ml)	−3.7	8.6	−20.5 to 13.1			

CCC = concordance correlation coefficient; CSA = antral cross-sectional area; FT = free-tracing; TD = two-diameter.
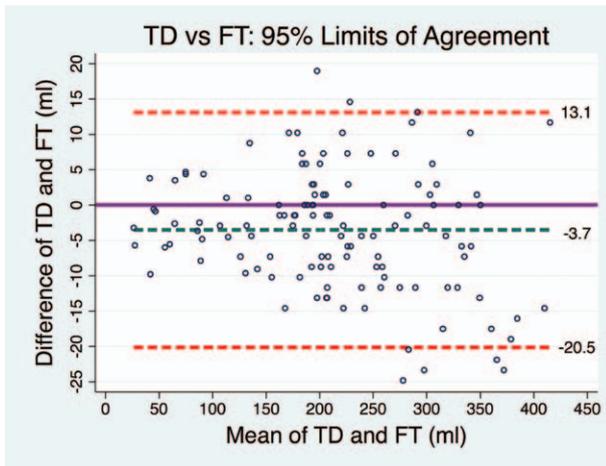
Fig. 4. Bland–Altman analysis. The *x* axis represents the mean of the two-diameter (TD) and free-tracing (FT) methods. The *y* axis represents the difference between the two volumes (the volume calculated by the TD method minus the volume calculated by the FT method). The *purple line* represents perfect agreement. The *green line* represents the observed mean agreement (–3.7 ml), and the *red dashed lines* represent the 95% limits of agreement (lower limit: –20.0 ml, upper limit: 10.9 ml).
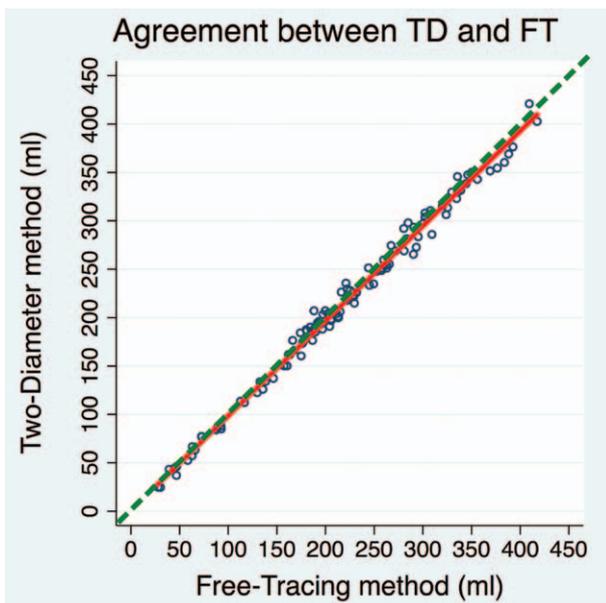
Fig. 5. Agreement between the two-dimeter (TD) method and the free-tracing (FT) method. The *green dashed line* represents ideal perfect agreement between the two methods. The *solid red line* is the "best-fit" for the observed data (correlation line). This best-fit line deviates little from the line of perfect concordance (high accuracy, 0.999), and there is no major dispersion around the best-fit line (precision 0.996).

A strength of our reliability assessment lies in the varied training and expertise of the three raters. Other studies assessing inter- and intrarater reliability have created idealized conditions for the reliability assessment which limited the generalizability of their results.[23] The results suggest that gastric ultrasound assessment is reproducible

not only by very experienced sonographers but also when performed by clinical anesthesiologists marginally exceeding the number of previous scans required to achieve 95% accuracy in qualitative gastric sonography (in this case, just over 50 previous gastric scans and 6 months' experience).[27] However, as with any new diagnostic tool, there are many aspects of training that remain to be defined. A previous study suggested a minimum of 33 gastric examinations followed by feedback are required to achieve a 95% accuracy rate in qualitative gastric sonography (differentiating an empty stomach from clear fluid or solid content).[27] Due to the additional steps required for a quantitative volume assessment as evaluated in this study, we expect a greater number of scans would be required to achieve a similar success rate. This remains to be studied.

The secondary objective of the current study was to compare a well-established and frequently-used method to measure antral CSA (TDM) with the less often used but simpler and more convenient FTM. We confirmed our hypothesis that the FTM is equivalent to the TDM. All differences between the two methods fell within 25 ml which is clinically inconsequential. The FTM of area measurement relies on single dimensional measurements and requires no geometrical assumptions.[11] By contrast, the TDM extrapolates area from the product of linear diameters and relies on the assumption that a cross-section of the gastric antrum is either a perfect circle or ellipse.[5,9,10] The fact that the two methods were essentially identical may be explained by the observation that an antral cross-section is usually close to a perfect circle or ellipse.[7–9] An advantage of the FTM is that it can be calculated with the technical capabilities of most portable ultrasound units and does not require an intermediate step of area calculation using the formula of the area of an ellipse. This makes the FTM more attractive and user-friendly for daily clinical application.

There are several limitations to our study. First, this method of gastric volume assessment has been validated for nonpregnant adults with normal gastric anatomy and a body mass index of up to 40 kg/m^2. It is therefore not immediately applicable to other patient populations such as children, parturients, or patients with previous gastric surgery. Second, an inherent limitation of any study involving gastric sonography is the dynamic nature of the organ. Peristaltic contractions and gastric emptying start immediately after clear fluid ingestion and this may add an element of variability between successive measurements, artificially lowering the resulting reliability. To minimize this possible confounding factor, we used apple juice (rather than water) as the caloric content in the apple juice prolongs gastric emptying time.[28] In addition, we minimized the time over which the three successive ultrasound scans were performed as described in the Materials and Methods. To further reduce the influence of gastric emptying on interrater reliability, the order of the raters was allocated at random.

Many questions regarding the clinical applicability and implementation of ultrasound gastric volume assessment

remain. These are related to cost-effectiveness, the time and training required to achieve competence, and its applicability to specific patient subgroups as previously discussed. Further studies are required to establish whether the information obtained from this new diagnostic tool can improve the accuracy of aspiration risk assessment and thus improve clinical decision making and patient outcome.

In summary, ultrasound assessment of gastric CSA and gastric volume by clinical anesthesiologists is highly reproducible with high intrarater and interrater reliability. The FTM to measure gastric antral CSA is equivalent to the TDM.

Acknowledgments

Competing Interests

Dr. Chan received equipment and research support from SonoSite (Bothell, Washington), Philips (Bothell, Washington), and BK Medical (Peabody, Massachusetts). The other authors declare no competing interests.

Correspondence

Address correspondence to Dr. Perlas: Department of Anesthesia, Toronto Western Hospital, 399 Bathurst Street, Toronto, Ontario, M5T 2S8, Canada. anahi.perlas@uhn.on.ca. Information on purchasing reprints may be found at www.anesthesiology.org or on the masthead page at the beginning of this issue. ANESTHESIOLOGY's articles are made freely accessible to all readers, for personal use only, 6 months from the cover date of the issue.

References

1. Lienhart A, Auroy Y, Péquignot F, Benhamou D, Warszawski J, Bovet M, Jougla E: Survey of anesthesia-related mortality in France. ANESTHESIOLOGY 2006; 105:1087–97

2. Lockey DJ, Coats T, Parr MJ: Aspiration in severe trauma: A prospective study. Anaesthesia 1999; 54:1097–8

3. Warner MA, Warner ME, Weber JG: Clinical significance of pulmonary aspiration during the perioperative period. ANESTHESIOLOGY 1993; 78:56–62

4. Cotton BR, Smith G: The lower oesophageal sphincter and anaesthesia. Br J Anaesth 1984; 56:37–46

5. Perlas A, Davis L, Khan M, Mitsakakis N, Chan VW: Gastric sonography in the fasted surgical patient: A prospective descriptive study. Anesth Analg 2011; 113:93–7

6. Perlas A, Chan VW, Lupu CM, Mitsakakis N, Hanbidge A: Ultrasound assessment of gastric content and volume. ANESTHESIOLOGY 2009; 111:82–9

7. Perlas AM, Nicholas M, Lui L, Cino M, Haldipur N, Davis L, Cubillos J, Chan VW: Validation of a mathematical model of ultrasound-determined gastric volume by gastroscopic examination. Anesth Analg 2013; 116:357–63

8. Bouvet L, Mazoit JX, Chassard D, Allaouchiche B, Boselli E, Benhamou D: Clinical assessment of the ultrasonographic measurement of antral area for estimating preoperative gastric content and volume. ANESTHESIOLOGY 2011; 114:1086–92

9. Bolondi L, Bortolotti M, Santi V, Calletti T, Gaiani S, Labò G: Measurement of gastric emptying time by real-time ultrasonography. Gastroenterology 1985; 89:752–9

10. Darwiche G, Almér LO, Björgell O, Cederholm C, Nilsson P: Measurement of gastric emptying by standardized real-time ultrasonography in healthy subjects and diabetic patients. J Ultrasound Med 1999; 18:673–82

11. Awadh AM, Prefumo F, Bland JM, Carvalho JS: Assessment of the intraobserver variability in the measurement of fetal cardiothoracic ratio using ellipse and diameter methods. Ultrasound Obstet Gynecol 2006; 28:53–6

12. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, Roberts C, Shoukri M, Streiner DL: Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol 2011; 64:96–106

13. Eliasziw M, Young SL, Woodbury MG, Fryday-Field K: Statistical methodology for the concurrent assessment of interrater and intrarater reliability: Using goniometric measurements as an example. Phys Ther 1994; 74:777–88

14. Lin L, Hedayat AS, Sinha B, Yang M: Statistical methods in assessing agreement: Models, issues, and tools. J Am Stat Assoc 2002; 97:257–70.

15. Shrout PE, Fleiss JL: Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 1979; 86:420–8.

16. Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1:307–10

17. Bland JM, Altman DG: Measuring agreement in method comparison studies. Stat Methods Med Res 1999; 8:135–60

18. de Vet HC, Terwee CB, Knol DL, Bouter LM: When to use agreement *versus* reliability measures. J Clin Epidemiol 2006; 59:1033–9

19. Lin L: A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989; 45:255–68.

20. King TS, Chinchilli VM: A generalized concordance correlation coefficient for continuous and categorical data. Stat Med 2001; 20:2131–47

21. Read MS, Vaughan RS: Allowing pre-operative patients to drink: Effects on patients' safety and comfort of unlimited oral water until 2 hours before anaesthesia. Acta Anaesthesiol Scand 1991; 35:591–5

22. Maltby JR: Gastric fluid volume and pH in elective patients following unrestricted oral fluid until three hours before surgery. Can J Anesth 1991; 38:425–9

23. Blackwell SC, Grobman WA, Antoniewicz L, Hutchinson M, Bannerman CG: Interobserver and intraobserver reliability of the NICHD 3-Tier Fetal Heart Rate Interpretation System. Am J Obstet Gynecol 2011; 205:378.e1–5

24. Lauder AJ, Inda DJ, Bott AM, Clare MP, Fitzgibbons TC, Mormino MA: Interobserver and intraobserver reliability of two classification systems for intra-articular calcaneal fractures. Foot Ankle Int 2006; 27:251–5

25. Simpson JM, Cook A: Repeatability of echocardiographic measurements in the human fetus. Ultrasound Obstet Gynecol 2002; 20:332–9

26. Bernstein J, Adler LM, Blank JE, Dalsey RM, Williams GR, Iannotti JP: Evaluation of the Neer system of classification of proximal humeral fractures with computerized tomographic scans and plain radiographs. J Bone Joint Surg Am 1996; 78:1371–5

27. Arzola C, Carvalho JC, Cubillos J, Ye XY, Perlas A: Anesthesiologists' learning curves for bedside qualitative ultrasound assessment of gastric content: A cohort study. Can J Anaesth 2013; 60:771–9

28. Spahn TW, Wessels A, Grosse-Thie W, Mueller MK: Assessment of pre-gastroscopy fasting period using ultrasonography. Dig Dis Sci 2009; 54:621–6