

OBSTETRICS

Interrater reliability of qualitative ultrasound assessment of gastric content in the third trimester of pregnancy

C. Arzola^{1*}, J. Cubillos^{1,2}, A. Perlas³, K. Downey¹ and J. C. A. Carvalho¹

¹ Department of Anesthesia and Pain Management, Mount Sinai Hospital, University of Toronto, 600 University Avenue, Room 19-104, Toronto, ON, Canada M5G 1X5

² Department of Anesthesia, Hospital Universitario San Ignacio, Pontificia Universidad Javeriana, Carrera 7 No. 40-62, Bogota, Colombia

³ Department of Anesthesia, Toronto Western Hospital, University Health Network, University of Toronto, McLaughlin Pavilion 2-405, 399 Bathurst St, Toronto, ON, Canada M5-T 2S8

* Corresponding author. E-mail: carzola@mtsinai.on.ca

Editors' key points

- This study investigated the reliability of bedside gastric ultrasound performed by anaesthetists on third trimester pregnant women.
- They found a substantial agreement and consistency of the qualitative ultrasound assessment of gastric contents.
- These findings are encouraging for bedside aspiration risk assessment during pregnancy. Further research is warranted.

Background. Pulmonary aspiration of gastric contents in pregnant women undergoing general anaesthesia is one of the most feared complications in obstetric anaesthesia. Bedside gastric ultrasonography is a feasible imaging tool to assess the gastric content. The purpose of this study was to investigate the reliability of qualitative bedside assessment of the gastric content performed by anaesthesiologists on third trimester pregnant women.

Methods. Pregnant women (≥ 32 weeks gestational age) were randomized to undergo ultrasound (US) assessments of their stomach in a fasting state (> 8 h), or after ingestion of clear fluids only, or solid food. Three anaesthesiologists trained in gastric ultrasonography performed the assessments using a low-frequency curved-array US transducer (5–2 MHz). Primary outcome of the study was the consistency of raters in diagnosing the correct status of the gastric content, which was used to determine the interrater reliability among the three anaesthesiologists. Secondary outcomes were overall proportion of correct and incorrect diagnoses and the specific proportions of correct diagnosis across the three gastric content groups.

Results. We analysed 32 pregnant women. The interrater reliability displayed a kappa statistic of 0.74 (bias corrected 95% CI: 0.68–0.84). The overall proportion of correct diagnosis was 87.5% (84 of 96). The odds of correct diagnosis for 'solid contents' were 16.7 times the odds for 'empty', and 14.3 times for 'clear fluid'.

Conclusions. Our results show the consistency of the qualitative US assessment of gastric contents of pregnant women in the third trimester by anaesthesiologists. A kappa of 0.74 suggests substantial agreement in terms of interrater reliability for this diagnostic measurement.

Clinical trial registration. ClinicalTrials.gov identifier: NCT01564030.

Keywords: complications, aspiration; complications, regurgitation; equipment, ultrasound machines; gastrointestinal tract, preoperative aspiration; pregnancy

Accepted for publication: 11 May 2014

Pulmonary aspiration of the gastric content in pregnant women undergoing general anaesthesia is one of the most feared complications in obstetric anaesthesia.^{1–3} The resulting respiratory compromise is associated with significant morbidity and mortality,^{4–7} which are thought to depend on the nature, volume, and acidity of the aspirate.⁸

Current knowledge suggests that the gastric emptying process in healthy non-labouring pregnant women is similar to that of non-pregnant women. However, once labour begins, gastric emptying is significantly delayed, only returning to

normal function after 18 h postpartum.^{9–14} There are many urgent situations in obstetrics when the status of the gastric content becomes key information for clinical management and decision-making. In the absence of a practical bedside tool to identify the status of the gastric content, all pregnant women are labelled as full stomach and managed as such.^{15–17}

Among the many applications of point-of-care ultrasound (US), there has been an increasing interest in bedside US gastric assessment.^{18–19} Perlas and colleagues^{20–22} and Bouvet and colleagues^{23,24} have demonstrated, in non-pregnant

subjects, that bedside trans-abdominal gastric ultrasonography is a feasible imaging tool to qualitatively and quantitatively assess the gastric content in the perioperative period. Furthermore, Arzola and colleagues²⁵ have recently studied the learning curves in a group of anaesthesiologists performing the technique in male volunteers, and suggested the amount of training required to achieve competence in this new skill.

If applicable to pregnant women, this tool may contribute to guide clinical practice and decision making in both elective and emergency situations in obstetric anaesthesia. However, it is unknown whether the competence achieved by anaesthesiologists in non-pregnant subjects can be transferred to the obstetric population. The purpose of this study was to investigate the reliability of qualitative bedside US assessment of the gastric content performed by anaesthesiologists on third trimester pregnant women.

Methods

After approval by the Research Ethics Board of Mount Sinai Hospital (Toronto, Canada), we conducted this observational cohort study of a group of anaesthesiologists performing qualitative US assessment of the gastric content in term pregnant women. Written informed consent was obtained from all the participants (anaesthesiologists and third trimester pregnant women). The study was registered at ClinicalTrials.gov (NCT01564030). We followed the Guidelines for Reporting Reliability and Agreement Studies (GRRAS statement) in conducting and reporting our investigation.²⁶

The participating *raters* consisted of three anaesthesiologists previously trained in qualitative US assessment of gastric content in human male volunteers²⁵ followed by 1-year of experience practicing the technique regularly. The inclusion criteria for the *rated subjects* were non-labouring pregnant women ≥ 32 weeks gestational age; ≥ 18 years; ASA physical status I–III; weight 50–120 kg; height ≥ 150 cm; ability to understand the rationale of the study assessments and written informed consent. Exclusion criteria were diabetes, a history of upper gastrointestinal disease (including hiatus hernia), and previous surgical procedures on the oesophagus, stomach, or upper abdomen.

Subjects were recruited on the day preceding the study assessments. Following an overnight fast of at least 8 h, subjects were randomized into one of the three groups: (a) *Empty Group*: to remain fasted until completion of the assessments; (b) *Clear Fluids Group*: to drink 250 ml of apple juice 5 min before the assessments; (c) *Solid Contents Group*: to eat a meal comprising a small muffin and coffee with cream or milk between 5 and 15 min before the assessments. Randomization was performed in blocks of six subjects (two subjects per group) using a computer-generated list of random numbers. Group allocation was concealed from the raters using sealed opaque envelopes, which were prepared and kept by an independent research assistant.

Assessments

The anaesthesiologists were asked to make a qualitative US diagnosis among three distinct gastric content groups

(*empty, clear fluids, or solid contents*) completing the diagnostic task in no more than 10 min. The time to complete the task was recorded from the first contact of the US probe with the patient's skin until the diagnosis was made and declared to the research assistant, which included the US examinations in both supine and lateral decubitus. At each session, the US examinations were started 5 min after the subjects completed the ingestion. The assessment sessions were paused every 15 min for a top-up of clear fluids (100 ml) to the subjects allocated to clear fluids, which was done in a blinded fashion. The purpose of this 'top-up' of clear fluids was to maintain similar examination conditions to all raters because of possible rapid gastric emptying of clear fluids in a fasted subject. The three raters carried out independent assessments in a random sequence. They were blinded to group allocation and to the other raters' findings. The rated subjects represented independent data points.

A standardized technique was performed with a portable US system equipped with a 5–2 MHz curved-array transducer (M-Turbo™, SonoSite Canada, Inc., Markham, ON, Canada). Subjects were first placed supine and then in the right lateral decubitus, always in a 45° semi-recumbent position. In both of these positions, fluid or semi-fluid content gravitates preferentially to the antrum, and air or gas is displaced proximally towards the body or fundus, thus facilitating antral sonography.²⁰ The examination focused on the antrum, which is the portion of the stomach most amenable to US imaging because of the consistent shape, location, and least amount of air content. Both positions were part of the same continuous diagnostic process, and we did not consider them two different diagnostic modalities. The gastric antrum was imaged in a sagittal to right parasagittal plane between the left lobe of the liver and the pancreas, at the level of the aorta or inferior vena cava (Fig. 1). A detailed description of the US technique and ultrasonography characteristics of gastric antrum content has been reported recently.²⁷

If the fluid content was observed only in the lateral decubitus, but not in the supine position, it was considered to be residual gastric secretion (Grade 1 antrum) and the diagnosis was of an empty stomach. In contrast, if fluid was observed in both supine and lateral decubitus, the diagnosis was of a stomach containing clear fluids (Grade 2 antrum). The antrum is completely empty if no fluid is observed in both positions (Grade 0 antrum).²¹

Study outcomes

The primary outcome was the consistency of raters in diagnosing the correct status of the gastric content, which was used to determine the interrater reliability among the three anaesthesiologists. Secondary outcomes included the overall proportion of correct and incorrect US diagnoses, and the specific proportions of correct diagnosis per gastric content groups.

Statistical analysis

We followed the statistical methods recommended by the Guidelines for Reporting Reliability and Agreement Studies

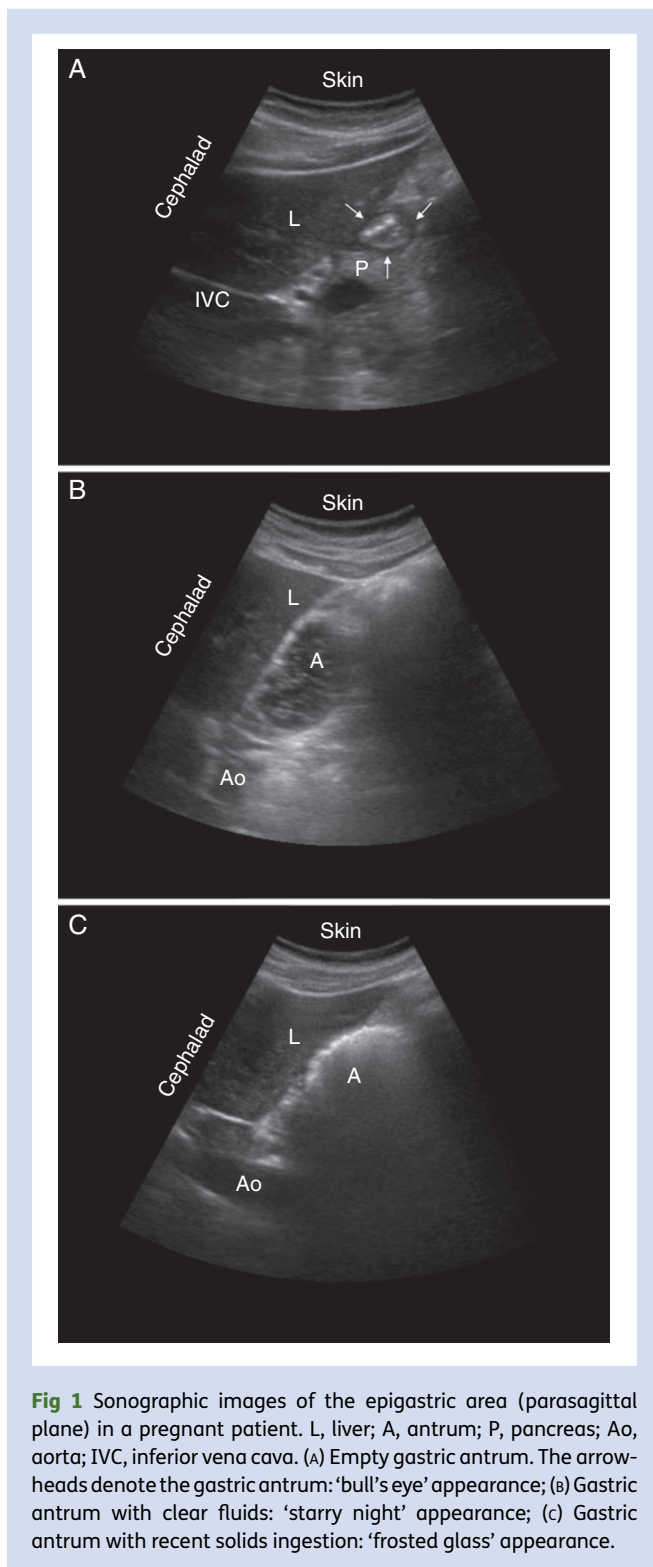


Fig 1 Sonographic images of the epigastric area (parasagittal plane) in a pregnant patient. L, liver; A, antrum; P, pancreas; Ao, aorta; IVC, inferior vena cava. (A) Empty gastric antrum. The arrowheads denote the gastric antrum: 'bull's eye' appearance; (B) Gastric antrum with clear fluids: 'starry night' appearance; (C) Gastric antrum with recent solids ingestion: 'frosted glass' appearance.

(GRRAS) statement.²⁶ To determine interrater reliability, we considered *reliability* with the implicit idea of 'trustworthiness' beyond mere agreement alone. In this context, the primary outcome was analysed through Cohen's kappa, which represents the consistency of raters in diagnosing the correct status of the gastric content. Furthermore, to express the relative importance of different types of disagreement (e.g. empty

vs solid), we performed a weighted kappa analysis using linear and quadratic weighting.²⁸ We adopted the classification by Landis and Koch, in which a test statistic >0.8 indicates near-perfect agreement, $0.61-0.8$ substantial agreement, $0.41-0.61$ moderate agreement, $0.21-0.40$ fair agreement, $0.00-0.2$ slight agreement, and <0.00 poor agreement.²⁹

The secondary outcomes evaluated the overall proportion of correct and incorrect diagnosis and also the specific proportions of correct and incorrect diagnosis across the various gastric content groups. We assessed the effect of the gastric content groups on the duration of US examinations, which was determined using a GEE approach (generalized estimating equations method) accounting for the correlated data.

Furthermore, we compared the odds of a correct diagnosis among gastric content groups. This analysis was conducted using multiple logistic regression models with the GEE approach for correlated data. The covariates included for the full model were the gastric content group, BMI, weight, and gestational age. The results (odds of correct diagnosis) were based on the final model derived by backward procedures.

Our sample size of 32 subjects was calculated considering four possible categories (empty, clear fluids, solid contents, inconclusive). According to Cicchetti and Fleiss,^{30,31} the sample size (n) when the number of categories (k) ranges from 3 to 10 is determined by $n=2k^2$ (total=32 subjects). Contrary to Carp and colleagues,⁹ who were unable to identify the stomach in 40% of the subjects, our raters were able to visualize the antrum and to make a diagnosis in all subjects, without the need of using the inconclusive category. Therefore, the analysis was still carried out in the entire sample size of 32 subjects but considering only three categories (empty, clear fluid, solid).

Descriptive statistical methods were used to describe the study population. The statistical analyses were performed using SAS 9.2 (SAS Institute, Inc., Cary, NC), R 10.2 (<http://www.r-project.org/>) and STATA/IC for Mac, Release 13.0 (StataCorp, College Station, TX, USA). A significance level of <0.05 was used without multiple comparison adjustment.

Results

The study recruitment was conducted from April to June 2012. The cohort of three anaesthesiologists completed all the US examinations. The allocation of subjects to the gastric content groups was 'empty' ($n=10$), 'clear fluids' ($n=10$), and 'solid contents' ($n=12$). The subjects' characteristics were: mean age of 32 (range 20–44) years, mean (SD) weight of 91 (18) kg, height of 164 (7) cm, BMI of 34 (6) kg m^{-2} , and gestational age of 35 (1) weeks. There were no significant differences among subjects' characteristics across the three gastric content groups (P -values ≥ 0.05 for all comparisons).

Primary outcomes

The interrater reliability demonstrated substantial agreement (kappa statistic of 0.74, bias corrected 95% CI: 0.68–0.84). Similarly, the reliability between each rater and the other two raters combined was substantial or near-perfect (rater 1: 0.66; rater 2: 0.83; rater 3: 0.75; P -values <0.0001). To apply

greater emphasis to disagreements in diagnosing the ‘solid contents’ group, we further analysed the results through weighted kappa and the interrater reliability was not significantly altered (kappa: 0.65; 0.81; 0.78; *P*-values < 0.0001).

Secondary outcomes

There was a pool of 96 possible outcome diagnoses: 30 ‘empty’, 30 ‘clear fluids’, and 36 ‘solid contents’. The overall proportion of correct diagnosis was 87.5% (84/96). An incorrect diagnosis occurred more frequently in subjects of greater weight, greater BMI, and more advanced gestational age (Table 1). The duration of US examinations in the ‘clear fluids’ group was shorter than the duration in the ‘empty’ group (least square mean [95% CI]: 3.1 [2.6–3.6] min and 4.6 [4.3–5.0] min, respectively; *P*<0.05); and the duration in the ‘clear fluids’ group was also shorter than the ‘solid contents’ group (3.1 [2.6–3.6] min and 3.9 [3.3–4.5] min, respectively; *P*<0.05). The duration between the ‘empty’ and the ‘solid contents’ groups was not statistically different. Nevertheless, the length of the US examination was not significantly associated with the incidence of an incorrect diagnosis (*P*=0.38; Table 1).

Although the specific proportions of correct and incorrect diagnosis across the three gastric content groups displayed no statistical differences (Table 2; *P*=0.052), when comparing the odds of correct diagnosis through multivariable analyses adjusted by BMI, weight, and patient age, the odds of correct diagnosis for ‘solid contents’ were 16.7 times the odds of correct diagnosis for ‘empty’, and 14.3 times the odds of correct diagnosis

for ‘clear fluid’ (Table 3). Finally, the odds of correct diagnosis were not affected with increasing gestational age when adjusted by BMI.

Discussion

Our results confirm the consistency of the qualitative US assessment of gastric contents of pregnant women in the third trimester by anaesthesiologists. A kappa of 0.74 means that the raters account for 74% of the agreement over and above what would be expected by chance alone. This value suggests ‘substantial agreement’ in terms of interrater reliability for this diagnostic measurement.

We were able to visualize the antrum in all studied subjects, and therefore the raters were in a position to make a diagnosis of the gastric content. This is in contrast with previous studies that looked at the gastric content of labouring and non-labouring pregnant women, and also in the postpartum period.^{9 32} With the US technology available at the time, those studies were not able to visualize the empty stomach in many of their subjects. While they consistently identified solid contents in the stomach immediately after and for several hours after a full lunch, an empty stomach was not identifiable unless ingestion of water could be seen in real-time streaming into the stomach to indicate its location and make it visible on US. They concluded that the reduced ability of US to identify the stomach in parturients would severely limit the clinical utility of this technique in this population. Subsequently, gastric US assessment continued to be used in pregnant women mainly as a tool to study gastric emptying under various circumstances.^{10 11 33} The current US technology has allowed further advancement in refining and broadening the scope of this diagnostic tool.^{20–24 27} Furthermore, the feasibility of training anaesthesiologists to incorporate this tool as part of their skill set has been recently demonstrated.²⁵ While our study shows the reproducibility and consistency of this relevant finding, it also confirms and validates its use in the pregnant population.

Although we had anticipated more technical difficulties to identify the stomach in pregnant women because of the presence of the gravid uterus, these difficulties were effectively overcome throughout the assessment period. Of note, women with increased BMI and gestational age were more prone to

Table 1 Patient characteristics in subjects with correct and incorrect US diagnosis. *The *P*-values were based on Student’s *t*-test. No adjustment for the correlated data. US, ultrasound

	Correct diagnosis (n=84), mean (sd)	Incorrect diagnosis (n=12), mean (sd)	<i>P</i> -value*
Age (years)	32 (7)	28(8)	0.06
Weight (kg)	89 (17)	109 (11)	<0.0001
Height (cm)	164 (7)	167 (5)	0.20
BMI (kg m ⁻²)	33 (6)	39 (4)	0.0004
Gestational age (weeks)	34 (1)	35 (2)	0.03
Duration of US (min)	3.8 (1.7)	4.3 (1.7)	0.38

Table 2 Gastric content group in subjects with correct and incorrect US diagnosis. *The *P*-value was based on Fisher’s exact test

Gastric group	Correct diagnosis (N=84), n (%)	Incorrect diagnosis (N=12), n (%)	<i>P</i> -value*
Empty	25 (29.8)	5 (41.7)	0.052
Clear fluids	24 (28.6)	6 (50.0)	
Solid	35 (41.7)	1 (8.3)	

Table 3 Multivariable analysis: adjusted odds ratios for correct diagnosis according to gastric content groups. Multiple logistic regression model with GEE approach (generalized estimating equations method) for correlated data. The covariates included for the full model were gastric content group, BMI, weight, and gestational age. The results reported are based on the final model derived by backward procedures. OR, odds ratio; CI, confidence interval

	Adjusted OR (95% CI)	<i>P</i> -value
Solid vs empty	16.7 (2–100)	0.01
Solid vs fluid	14.3 (1.9–100)	0.01
Fluid vs empty	1.06 (0.15–7.70)	0.95

result in incorrect diagnosis. However, when the odds of a correct diagnosis were adjusted by BMI, the effect of gestational age was not statistically significant. Further research is warranted to determine the effect of anthropometrics and patients characteristics on gastric ultrasonography. Nevertheless, some particularities of the examination in pregnant women are worth describing. All portions of the stomach tend to be displaced upwards and more to the right when compared with the non-pregnant subjects. Furthermore, positioning the patient in a complete right lateral decubitus seems to be more crucial, as a slightly less tilted position rendered worse examination conditions. However, we believe that an adequate understanding and familiarization with the scanning process in the adult non-pregnant population is a prerequisite before attempting this more challenging setting.

The performance of any diagnostic instrument requires valid and reproducible assessments. While validity relates to results corresponding to the true state, reproducibility addresses the reliability and agreement in terms of obtaining similar results of a stable phenomenon assessed by different individuals.³⁴ In our study, the interrater reliability established the consistency of raters in differentiating among different gastric content groups under similar assessment conditions. Several factors can influence the reliability of a diagnostic test as evaluated by the kappa statistic. These factors include the prevalence of the 'condition', the presence of bias in the assessments and possible non-independence of ratings. We tried to minimize the influence of these factors as follows: (a) the prevalence of the target condition was controlled by using a comparable number of rated subjects allocated into the three distinct gastric content groups and the adherence to a strict protocol; (b) bias was controlled by proper allocation concealment and randomization of the raters and study subjects; furthermore the 95% confidence interval was calculated with statistical bias correction and we performed weighting analysis to determine possible rating trends; (c) moreover, the strength of our results is also based on the independent US examinations and rated subjects. The overall and individual rater magnitudes of kappa are within the range of substantial agreement with a narrow confidence interval, even with a plausible 'true' value in the area of near-perfect agreement ($\kappa=0.84$).

The clinical implications of misdiagnosis are critical for the decision-making process. One important limitation of the unweighted kappa is that all rating disagreements are statistically treated equally, whereas weighted kappa allocates different weights to the disagreements according to their relative importance in the clinical context. We therefore further analysed the ratings by weighting the categories based on the clinical relevance of misdiagnosis.³⁵ We assumed that the misdiagnosis of a solid content would be the most relevant in clinical practice. Even in doing so, the resulting weighted kappa remained stable, confirming the adequate discrimination and consistency of the measurements. Furthermore, the multivariable analysis adjusted by patient characteristics confirmed the favourable odds of having a correct diagnosis in the *solid content* group in comparison with the *empty* and *clear fluid* groups.

Our study presents some limitations. Firstly, there was no baseline US examination to rule out residual gastric contents or protocol violation before randomization. Although the internal validity may be compromised, gastric emptying in non-labouring pregnant women after 8–10 h fasting should most likely render an empty stomach as previously demonstrated. Secondly, stability of the attribute being rated is critical when repeated ratings are conducted.³⁵ Although we provided a top-up of fluids after 15 min in the clear fluid group in an attempt to provide similar conditions to all raters, this may have introduced bias in the assessments before and after this time point despite the randomization of the examination sequence. Thirdly, our three raters were anaesthesiologists already trained in bedside gastric US; therefore, the results may not be extrapolated to other physicians. Finally, although our study results show consistency of the technique, they have to be understood in the context of strict study conditions such as equally distributed gastric content groups (prevalence), standardized solid meal, US examinations after recent ingestion. These conditions may not necessarily resemble all types of challenging clinical scenarios that require an accurate diagnosis.

In conclusion, we confirmed the consistency of the qualitative bedside US assessment of the gastric content of pregnant women in the third trimester performed by anaesthesiologists trained in the technique. While this finding is encouraging, at the present time, the idea of using gastric US as a clinical tool for bedside aspiration risk assessment still demands further investigation.

Authors' contributions

C.A. helped design the study, conduct the study, analyse the data, and write the manuscript. J.C. helped design the study, conduct the study, analyse the data, and write the manuscript. A.P. helped design the study, conduct the study, analyse the data, and write the manuscript. K.D. helped design the study and conduct the study. J.C.A.C. helped design the study, conduct the study, analyse the data, and write the manuscript.

Declaration of interest

A.P. is an associate editor of the journal *Regional Anesthesia and Pain Medicine*. None of the other authors has financial or personal relationships or affiliations that could influence (or bias) this research.

Funding

This work was supported by MSH-UHN AHSC AMO AFP Innovation Fund (2012–2013, Mount Sinai Hospital-University Health Network Academic Health Sciences Centres Academic Medical Association Alternate Funding Plan) Ontario, Canada.

References

- 1 Mendelson C. The aspiration of stomach contents into the lungs during obstetric anesthesia. *Am J Obstet Gynecol* 1946; **52**: 191–205

- 2 Hawkins JL, Koonin LM, Palmer SK, Gibbs CP. Anesthesia related deaths during obstetric delivery in the United States, 1979–1990. *Anesthesiology* 1997; **86**: 277–84
- 3 McClure J, Cooper G. Centre for Maternal and Child Enquiries (CMACE). Saving Mothers' Lives: reviewing maternal deaths to make motherhood safer: 2006–08. The Eighth Report on Confidential Enquiries into Maternal Deaths in the United Kingdom. *BJOG* 2011; **118** (Suppl. 1): 1–203. Chapter 8: Anaesthesia
- 4 Warner MA, Warner ME, Weber JG. Clinical significance of pulmonary aspiration during the perioperative period. *Anesthesiology* 1993; **78**: 56–62
- 5 Lienhart A, Auroy Y, Pequignot F, et al. Survey of anesthesia-related mortality in France. *Anesthesiology* 2006; **105**: 1087–97
- 6 Neelakanta G, Chikyarappa A. A review of patients with pulmonary aspiration of gastric contents during anesthesia reported to the Departmental Quality Assurance Committee. *J Clin Anesth* 2006; **18**: 102–7
- 7 Ng A, Smith G. Gastroesophageal reflux and aspiration of gastric contents in anesthetic practice. *Anesth Analg* 2001; **93**: 494–513
- 8 Marik PE. Aspiration pneumonitis and aspiration pneumonia. *N Engl J Med* 2001; **344**: 665–71
- 9 Carp H, Jayaram A, Stoll M. Ultrasound examination of the stomach contents of parturients. *Anesth Analg* 1992; **74**: 683–7
- 10 Wong CA, Loffredi M, Ganchiff JN, Zhao J, Wang Z, Avram MJ. Gastric emptying of water in term pregnancy. *Anesthesiology* 2002; **96**: 1395–400
- 11 Wong CA, McCarthy RJ, Fitzgerald PC, et al. Gastric emptying of water in obese pregnant women at term. *Anesth Analg* 2007; **105**: 751–5
- 12 Kubli M, Scrutton MJ, Seed PT, O'Sullivan G. An evaluation of isotonic 'sport drinks' during labor. *Anesth Analg* 2002; **94**: 404–8
- 13 Gin T, Cho AM, Lew JK, et al. Gastric emptying in the postpartum period. *Anaesth Intensive Care* 1991; **19**: 521–4
- 14 Chiloiro M, Darconza G, Piccioli E, et al. Gastric emptying and orocecal transit time in pregnancy. *J Gastroenterol* 2001; **36**: 538–43
- 15 American Society of Anesthesiologists Committee on Standards and Practice Parameters. Practice guidelines for preoperative fasting and use of pharmacologic agents to reduce the risk of pulmonary aspiration: Application to healthy patients undergoing elective procedures. *Anesthesiology* 2011; **114**: 495–511
- 16 American Society of Anesthesiologists Task Force on Obstetric Anesthesia. Practice Guidelines for Obstetric Anesthesia. *Anesthesiology* 2007; **106**: 843–63
- 17 Smith I, Kranke P, Murat I, et al. Perioperative fasting in adults and children: guidelines from the European Society of Anesthesiology. *Eur J Anaesthesiol* 2011; **28**: 556–69
- 18 Johnson DW, Oren-Grinberg A. Perioperative point-of-care ultrasonography: the past and the future are in anesthesiologists' hands. *Anesthesiology* 2011; **115**: 460–62
- 19 Moore CL, Copel JA. Point-of-care ultrasonography. *N Engl J Med* 2011; **364**: 749–57
- 20 Perlas A, Chan VW, Lupu CM, Mitsakakis N, Hanbidge A. Ultrasound assessment of gastric content and volume. *Anesthesiology* 2009; **111**: 82–9
- 21 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
- 22 Perlas A, Mitsakakis N, Liu L, et al. Validation of a mathematical model for ultrasound assessment of gastric volume by gastroscopic examination. *Anesth Analg* 2013; **116**: 357–63
- 23 Bouvet L, Miquel A, Chassard D, Boselli E, Allaouchiche B, Benhamou D. Could a single standardized ultrasonographic measurement of antral area be of interest for assessing gastric contents? A preliminary report. *Eur J Anaesthesiol* 2009; **26**: 1015–9
- 24 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
- 25 Arzola C, Carvalho JCA, Cubillos J, Ye XY, Perlas A. Anesthesiologists' learning curves for bedside qualitative ultrasound assessment of gastric content: a cohort study. *Can J Anesth* 2013; **60**: 771–9
- 26 Kottner J, Audigé L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011; **64**: 96–106
- 27 Cubillos J, Tse C, Chan V, Perlas A. Bedside ultrasound assessment of gastric content: an observational study. *Can J Anesth* 2012; **59**: 416–23
- 28 Kraemer HC, Periyakoil VS, Noda A. Kappa coefficients in medical research. *Stat Med* 2002; **21**: 2109–29
- 29 Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessments of majority agreement among multiple observers. *Biometrics* 1977; **33**: 363–74
- 30 Cicchetti DV, Fleiss JL. Comparison of the null distribution of weighted kappa and the C ordinal statistic. *Appl Psychol Meas* 1977; **1**: 195–201
- 31 Cicchetti DV. Testing the normal approximation and minimal sample size requirements of weighted kappa when the number of categories is large. *Appl Psychol Meas* 1981; **5**: 101–4
- 32 Jayaram A, Bowen MP, Deshpande S, Carp HM. Ultrasound examination of the stomach contents of women in the postpartum period. *Anesth Analg* 1997; **84**: 522–6
- 33 O'Sullivan G. Gastric emptying during pregnancy and the puerperium. *Int J Obstet Anesth* 1993; **2**: 216–24
- 34 Fletcher RH, Fletcher SW. Chapter 2: Abnormality. In: *Clinical Epidemiology, the Essential*, 4th Edn. Baltimore: Lippincott Williams and Wilkins, 2005; 17–34
- 35 Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005; **85**: 257–68

Handling editor: M. M. R. F. Struys