



Inter-rater and intra-rater reliability of the airway diameter measured by sonography

Julio César Gómez-Tamayo¹ · Jorge Iván Puerta-Guarín¹ · Claudia Marcela Rojas-Camejo¹ · Juan Pablo Caicedo¹ · José Andrés Calvache^{1,2}

Received: 5 July 2017 / Accepted: 6 December 2017
© Società Italiana di Ultrasonologia in Medicina e Biologia (SIUMB) 2018

Abstract

Introduction Sonography has gained a broad field of applications in current anesthetic practice. During airway management, it could have an important role, because it allows real-time measurement of anatomical aspects and a functional evaluation with high safety profile. However, assessment of the reliability and precision of these measurements is critical if sonography is used as a clinical diagnostic tool. The aim of the study was to estimate inter-rater and intra-rater reliability of the airway diameter measured by sonography in healthy volunteers.

Methods We conducted a cross-sectional study during 2015 on healthy adult volunteers. Using sonography, transverse internal diameter of subglottic space was measured twice by two blinded sonographers.

Results Agreement was classified as very good for intra-rater measurements (ICC = 0.84) and good for the inter-rater measurements (ICC = 0.71). The 95% CI of agreement limits for the intra-rater measurements were inferior to those for the inter-rater measurements.

Conclusions Ultrasound measurement is a reliable and precise method to measure the internal subglottic diameter of the airway. Ultimately, this method may provide clinicians valuable information regarding airway diameter in adults and may help to guide treatment options.

Keywords Airway ultrasound · Airway · Precision · Reliability · Sonography · Subglottic space

Sommario

Scopo L'ecografia ha acquisito nel tempo un ruolo importante in diverse applicazioni nella pratica anestesiológica corrente. Durante la gestione delle vie aeree, potrebbe avere un ruolo importante in quanto consente di misurare in tempo reale gli aspetti anatomici e una valutazione funzionale con un elevato profilo di sicurezza. Tuttavia, la valutazione dell'affidabilità e della precisione di queste misurazioni è fondamentale se l'ecografia viene utilizzata come strumento diagnostico clinico. Lo scopo di questo studio è di stimare l'affidabilità inter e intra-osservatore del diametro delle vie aeree misurate ecograficamente su volontari sani.

Metodi Nel corso del 2015 abbiamo condotto uno studio trasversale su volontari adulti sani. Utilizzando l'ecografia, il diametro interno trasversale dello spazio subglottico è stato misurato due volte da due ecografisti in doppio cieco.

Risultati L'accordo inter e intra-osservatore è stato classificato rispettivamente come buono (CCI = 0.71) e molto buono (CCI = 0.84). I limiti della misura di concordanza intra-osservatore con IC del 95% erano inferiori rispetto a quelli inter-osservatore.

Conclusioni La misura dello spazio subglottico con l'ecografia è un metodo affidabile e preciso per misurare il diametro subglottico interno della via aerea. In definitiva, questo metodo può fornire ai medici informazioni preziose riguardo il diametro delle vie aeree negli adulti e può aiutare a guidare le opzioni di trattamento.

✉ Julio César Gómez-Tamayo
juliogota@gmail.com

Extended author information available on the last page of the article

Introduction

Ultrasonography has gained a broad field of applications in current anesthetic practice. It has an important role, because it allows real-time anatomical and functional evaluation with a high safety profile [1].

Like other imaging techniques, the reliability of ultrasound measurement depends on two properties: validity and reproducibility. While validity expresses the degree to which the phenomenon of interest is actually measured (proximity to truth or lack of bias), reproducibility indicates the extent to which the same values are obtained when a measure is achieved repeatedly under similar conditions [2]. Measurement variability and agreement among them may determine their usefulness in clinical practice [3].

Some authors have explored the importance of sonography as a tool in airway evaluation and management, primarily in pediatric population [4, 5]. Most have focused on adjusting the diameter of the tracheal tube to use or on comparing the measurement with the traditional methods, such as the formulas [4–7]. Considering that the airway is a dynamic structure and sonographic measurements depend on the evaluator, these may change over time or vary according to sonographer skills.

The objective of this study was to determine the variability and reproducibility of the sonographic measurement of the subglottic diameter in healthy volunteers, as an important step in defining its clinical applicability.

Materials and methods

After approval from the Institutional Review Board at Hospital Universitario San José (Popayán, Colombia) (Reference: 103-1.5.32, 09/10/2014), we conducted a cross-sectional study between March and December 2015.

Three members of the research team were trained to measure the subglottic diameter through ultrasound by an experienced radiology specialist. They were guided during three training rounds with four consecutive volunteers each one (12 volunteers in total). Each training-round was done during different times and these volunteers were not included as participants of the study. To measure the subglottic airway diameter, the radiologist and members of the research team followed the instructions and methodology published by Schramm et al. [4, 5]. During spontaneous ventilation, the minimal transverse diameter of the subglottic airway was measured by the trainer with full explanation of the procedure of image acquisition. Measurements were performed with a 6–13 MHz linear transducer connected to the portable ultrasound (Sonosite® M-Turbo, Bothell, WA, USA) at 10 MHz as definite resolution. Initially, the transducer was located in the midline of the neck in sagittal position to identify the cricoid cartilage and the first tracheal ring. After that, the transducer was rotated transversely (short axis) to locate the cricoid cartilage, which is a hypoechoic arch, because it is an air-mucosal interface. Then, the transducer advanced in cephalic direction to just above the cricoid cartilage, an anatomical site corresponding to the subglottic space [7] (Fig. 1). The transverse internal diameter of this structure was measured (in millimeters) on the reference of the internal borders (Fig. 2).

For the study phase, healthy volunteers over 18 years of age with ASA physical status 1 or 2 were included. All volunteers agreed to participate and signed a written informed consent form prior to data collection. Exclusion criteria for this study were ASA physical status greater than 2 and a history of having predictors of difficult airway, given by: an obvious visual alteration of the external anatomy of the neck, circumference of the neck larger than 40 cm, previous exposure to radiation of the head and neck, or limitation of

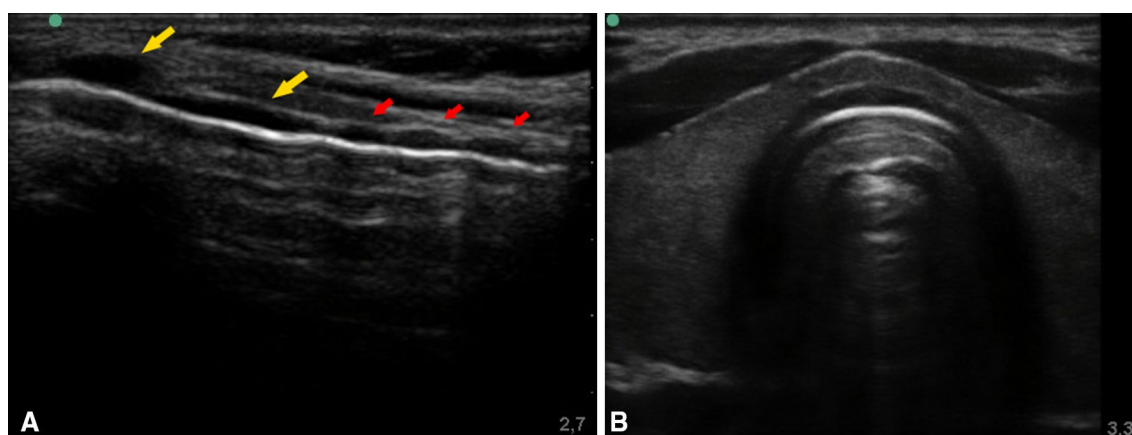


Fig. 1 **a** Long axis of the trachea showing thyroid cartilage, cricoid cartilage (yellow arrows), and tracheal rings (red arrows). **b** Short-axis sonogram of the anatomical site corresponding to the subglottic space



Fig. 2 Assessment and measurement of the transverse internal diameter (dashed line)

cervical mobilization (extension) [8]. In addition, volunteers with obvious airway malformations and/or medical conditions impeding or altering ultrasound measurement, such as cervical collar, tracheostomy, tracheomalacia, masses or tumors, were also excluded.

Volunteers were placed in supine position with the head in extension during spontaneous breathing. Measurements were performed with a 6–13 MHz linear transducer from a portable ultrasound (Sonosite® M-Turbo, Bothell, WA, USA) using a deep of 2.7 and 3.3 cm for the long and short axes, respectively. Other settings were: focus on at level around 1.5 and 2 cm, gain level less than 50%, frequency 10 MHz, and tissue harmonics on.

Three researchers (second-year anesthesiology residents) performed the measurements. Two performed the procedure twice (time 1 and time 2), and a third one recorded and collected the results and duration for each assessment. Each measurement was blinded to the sonographer to avoid information bias. The roles of researchers were reassigned successively at random on each subject included. In addition, age, gender, and weight were collected for each participant.

Sample size was calculated to contrast null hypothesis ($H_0: \rho = \rho_0$) with the alternative hypothesis ($H_1: \rho > \rho_0$), where ρ is the value for the intra-class correlation coefficient (ICC) and ρ_0 is a value predetermined for the ICC. According to an a priori hypothesis, it was expected to estimate an intra-rater ICC = 0.8 and inter-rater ICC = 0.6, which were previously considered as a substantial reliability (consistency) [9]. Using the figures reported by Eliasziw et al. [10], a minimum of 35 subjects, two evaluators and two measurements for each, was enough to reach a statistical power superior to 80% with a significance level $\alpha = 0.05$.

Data analysis was made using descriptive statistical summary measures [mean \pm SD, median (IQR),

frequencies, and proportions] according to their distribution and nature. Ultrasound measurements were expressed in millimeters and their normality was evaluated through the Shapiro–Wilkinson test. For each one, coefficient of variation and 95% CI were reported to describe measurement precision and are reported in units of the measurement itself.

To evaluate intra-rater and inter-rater reliability, ICC was calculated from a random effects model using analysis of variance (ANOVA) for repeated measurements [11]. In addition, we reported 95% CI for each ICC. Reliability was qualified with the ICC as: bad or null (ICC < 0.20), poor (ICC = 0.21–0.40), moderate (ICC = 0.41–0.60), good (ICC = 0.61–0.80), or very good (ICC = 0.81–1.00) [12, 13].

Bland–Altman plots were constructed to graphically analyze agreement between measurements. This method of displaying data plots the difference between two measurements for each subject against their mean [14, 15]. The 95% CI of agreement limits were calculated. A priori, clinical usefulness of the ultrasound airway measurement was predetermined if these limits of agreement range from ± 5 mm.

All statistical calculations were performed using R (R Foundation, Vienna, Austria) [16].

Results

The study recruited 100 participants of which 64 were females and 36 were males, with a mean age of 34 years (1 SD = 10.6 years) and median of 30 years (IQR = 25–42). The mean weight for the participants was 67 kg (1 SD = 12 kg). Table 1 shows statistical summaries and the distribution of each measurement performed by the evaluators at the time one, time two, and their variability.

Inter-rater and intra-rater reliability is shown in Table 2. Agreement was classified as very good for intra-rater measurements (ICC = 0.84) and good for intra-rater measurements (ICC = 0.71).

Bland–Altman analysis is presented in Fig. 3 and 95% CI agreement limits are shown in Table 3. The 95% CI agreement limits for the intra-rater measurements were inferior to those for the inter-rater measurements.

Discussion

The main results of the present study indicate that variability of independent measures of subglottic diameter is low. Second, very good intra-rater measurement agreement exists and agreement limits are narrow, within a 5-mm range.

Table 1 Statistical summaries of each measurement

	Mean \pm 1 SD	Variation coefficient	95% CI	Median (IQR)	P value*
Evaluator one Measurement one	8.12 \pm 1.59	21.7	7.8–8.4	7.6 (2, 6, 8, 9)	0.000
Evaluator one Measurement two	8.04 \pm 1.53	19	7.7–8.3	7.8 (6.9–9.0)	0.038
Evaluator two Measurement one	7.97 \pm 1.23	15.5	7.7–8.2	7.7 (7.0–8.5)	0.024
Evaluator two Measurement two	7.86 \pm 1.22	12	7.6–8.1	7.7 (7.0–8.45)	0.000

Time one and time two ($n = 100$)

*Kolmogorov–Smirnov normality test

Table 2 Intra-class correlation coefficient for each evaluator and each measurement (intra-rater and inter-rater reliability) ($n = 100$)

	Intra-class correlation coefficient (ICC)	95% CI
Evaluator 1 (Measurement 1 and 2) Intra-rater	0.84	0.78–0.89
Evaluator 2 (Measurement 1 and 2) Intra-rater	0.84	0.78–0.89
Measurement 1 (Evaluator 1 and 2) Inter-rater	0.71	0.58–0.81
Measurement 2 (Evaluator 1 and 2) Inter-rater	0.71	0.57–0.80

Scientific interest exists in measuring the airway anatomic variability [17], which may also have clinical usefulness. Recent studies show low correlation between tomographic-based measurements of the airway with age (or other anthropometric variables of easy evaluation), demonstrating low performance of an isolated tracheal size clinical measurement [18].

Ultrasonography has demonstrated high correlation and agreement with airway measurements made by magnetic resonance imaging (precision = 0.33 mm), also with clinical applicability at the patient's bedside [7]. However, operator dependence of ultrasound as an imaging modality has been highlighted as potentially problematic for both qualitative

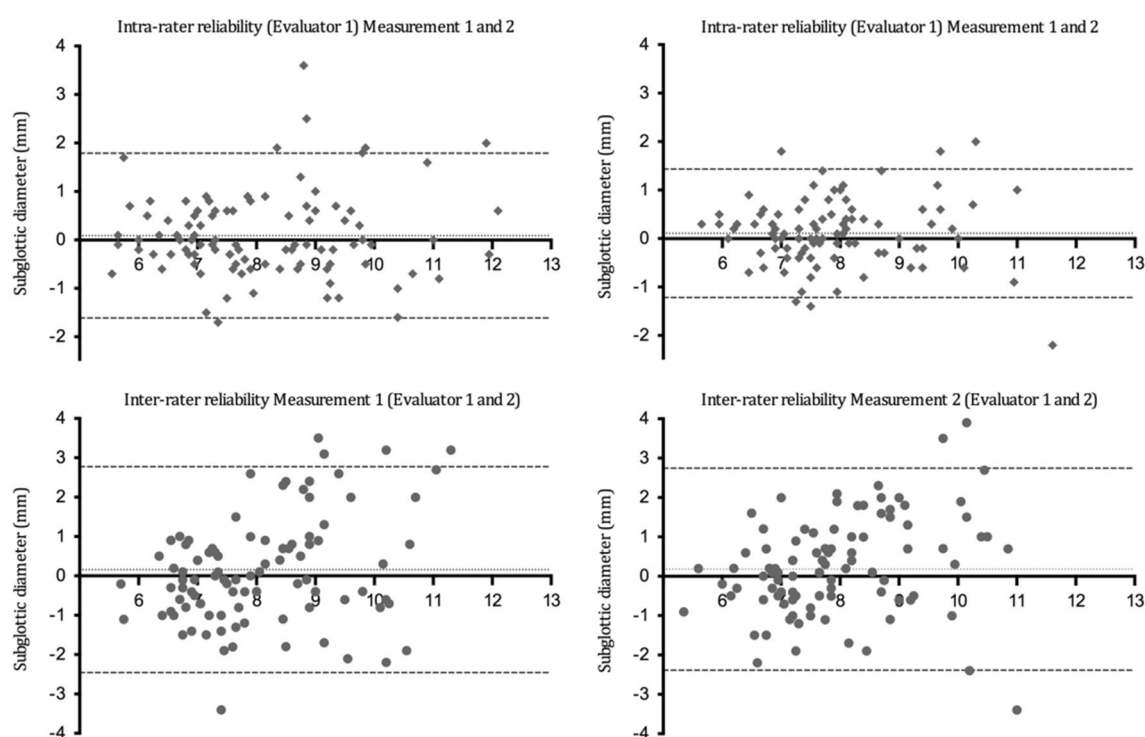
**Fig. 3** Bland–Altman plots for intra-rater agreement and inter-rater agreement ($n = 100$)

Table 3 Bland-Altman analysis

	Difference	1 SD	95% agreement limits
Evaluator 1 (Measurement 1 and 2) Intra-rater	0.08	0.86	– 1.61 to 1.78
Evaluator 2 (Measurement 1 and 2) Intra-rater	0.10	0.67	– 1.21 to 1.43
Measurement 1 (Evaluator 1 and 2) Inter-rater	0.15	1.33	– 2.45 to 2.77
Measurement 2 (Evaluator 1 and 2) Inter-rater	0.17	1.30	– 2.38 to 2.74

Differences among measurements, variability, and 95% CI agreement limits for intra-rater (evaluator 1 and 2) and for inter-rater (measurement 1 and 2) ($n = 100$)

and quantitative anatomic measurements. Our study demonstrates that variability of ultrasound airway measurement (95% CI) has ranges < 0.6 cm for each time and for independent evaluators, which can be graded as low.

To our knowledge, this is the first study to evaluate intra-rater and inter-rater agreement of ultrasound repeated airway measurements. Our results support that, for trained evaluators, ultrasound agreement is very good. Several studies have been published regarding ultrasonography agreement in other anatomic sites: some static (liver, prostate) [19, 20] and other dynamic (cava vein) [21, 22]. Although airway diameter may be modified by respiratory dynamics [23] and considering that our protocol did not include a specific static measurement time, the agreement is good. It is important to highlight that machine settings should be adjusted properly to avoid an increased artifacts formation, leading to incorrect measures. In addition, every pathologic condition (i.e., tracheal stenosis and granulomas) may increase disturbing artifacts. In particular, secretions could interfere with ultrasound beam reflection and cause artifacts that would alter measure limits.

The influence of the orotracheal tube diameter on ventilatory mechanics is historically recognized (Hagen–Pousiella law). Dead-space addition leads to increased respiratory work [24]. Based on this premise, we recommend [25] using the highest possible internal diameter for the endotracheal tube, with low resistance and which allows overcoming the mechanical effect on the tracheal mucosa. Shapiro et al. demonstrated that in adults with spontaneous breathing, the respiratory work significantly increases when the orotracheal tube diameter decreases from 7 to 6 mm [26].

The agreement limits for ultrasound measurements are within a 5-mm range. This finding explains why the intra-rater or inter-rater reliability influence upon orotracheal tube diameter selection may be low, when considering that the minimal difference for changing a tube is 5 mm (superior or inferior).

Clinical applicability of ultrasound during orotracheal tube selection has been assessed by several authors: Kim EJ et al., Shibasaki M. et al, and Schramm C et al. evaluated the

usefulness of airway ultrasonography measurement in selecting an appropriate orotracheal tube in pediatric populations. These authors documented narrow agreement limits and high predictive capability of ultrasound in selecting the orotracheal tube, superior to age- and height-based formulas [4–6].

Regarding clinical applicability, our study shows that subglottic diameter ultrasound measurements have high intra-rater and inter-rater reproducibility in adults. This study did not consider pediatric population, which might be a limitation but also an opportunity for further studies.

This study adds to the body of literature on using sonography to assess airway diameter. Our study demonstrates that with standardized positioning and a short prior training, sonography becomes a precise and reliable tool in assessing subglottic diameter.

Compliance with ethical standards

Conflict of interest All the authors declare that they have no conflict of interest.

Ethical approval All procedures performed in this study were in accordance with the ethical standards of The Institutional Research Committee and with the 1964 Helsinki declaration and its later amendments.

Informed consent All volunteers agreed to participate and signed a written informed consent form prior to data collection.


Sources funding None.

References

1. Wanderer JP, Rathmell JP (2015) Anesthesiologists and ultrasonography: a historical window. *Anesthesiology* 123(3):A23
2. Pita Fernández S, Pérttega Díaz S, Rodríguez Maseda E (2003) La fiabilidad de las mediciones clínicas. El análisis de concordancia para variables numéricas. *Cuadernos de atención primaria* 10(4):290–296
3. Cortés-Reyes E, Rubio-Romero JA, Gaitán-Duarte H (2010) Métodos estadísticos de evaluación de la concordancia y la reproducibilidad de pruebas diagnósticas. *Rev Colomb Obstet Ginecol* 61(3):247–255

4. Schramm C, Knop J, Jensen K, Plaschke K (2012) Role of ultrasound compared to age-related formulas for uncuffed endotracheal intubation in a pediatric population. *Paediatr Anaesth* 22(8):781–786
5. Kim EJ, Kim SY, Kim WO, Kim H, Kil HK (2013) Ultrasound measurement of subglottic diameter and an empirical formula for proper endotracheal tube fitting in children. *Acta Anaesthesiol Scand* 57(9):1124–1130
6. Shibasaki M, Nakajima Y, Shimizu F, Shime N (2010) Prediction of pediatric endotracheal tube size by ultrasonography. *Anesthesiology* 113:819–824
7. Lakhal K, Delplace X, Cottier J-P, Tranquart F, Sauvagnac X, Mercier C et al (2007) The feasibility of ultrasound to assess subglottic diameter. *Anesth Analg* 104(3):611–614
8. Lee S, Lee J, Kim T et al (2011) The relationship between the predictors of obstructive sleep apnea and difficult intubation. *Korean J Anesthesiol* 60(3):173–178
9. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ et al (2011) Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 64(1):96–106
10. Eliasziw M, Young SL, Woodbury MG, Fryday-Field K (1994) Statistical methodology for the concurrent assessment of interrater and intra-rater reliability: using goniometric measurements as an example. *Phys Ther* 74:777–788
11. Shrout PE, Fleiss J (1979) Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 86:420–428
12. Barnhart HX, Haber MJ, Lin LI (2007) An overview on assessing agreement with continuous measurements. *J Biopharm Stat* 17(4):529–569
13. Bland JM, Altman DG (2003) Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet Gynecol* 22(1):85–93
14. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307–310
15. Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. *Stat Methods Med Res* 8:135–160
16. Team RC. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013. ISBN:3-900051-07-0; 2014
17. Griscom NT, Wohl ME (1985) Dimensions of the growing trachea related to body height. Length anteroposterior and transverse diameters, cross-sectional area and volume in subjects younger than 20 years of age. *Am Rev Respir Dis* 131(6):840–844
18. Karmakar A, Pate MB, Solowski NL, Postma GN, Weinberger PM (2015) Tracheal size variability is associated with sex: implications for endotracheal tube selection. *Ann Otol Rhinol Laryngol* 124(2):132–136
19. Childs JT, Esterman AJ, Thoires KA (2014) Ultrasound measurements of the liver: an intra and inter-rater reliability study. *Aust J Ultrasound Med* 17:113–119
20. Eri LM, Thomassen H, Brennhovd B, Håheim LL (2002) Accuracy and repeatability of prostate volume measurements by transrectal ultrasound. *Prostate Cancer Prostatic Dis* 5(4):273–278
21. Saul T, Lewiss RE, Langsfeld A, Radeos MS, Del Rios M (2012) Inter-rater reliability of sonographic measurements of the inferior vena cava. *J Emerg Med* 42(5):600–605
22. Akkaya A, Yesilaras M, Aksay E, Sever M, Atilla OD (2013) The interrater reliability of ultrasound imaging of the inferior vena cava performed by emergency residents. *Am J Emerg Med* 31(10):1509–1511
23. Schwab RJ, Gefter WB, Pack AI, Hoffman EA (1993) Dynamic imaging of the upper airway during respiration in normal subjects. *J Appl Physiol* 74(4):1504–1514
24. Bolder PM, Healy TEJ, Bolder AR, Beatty PCW, Kay B (1986) The extra work of breathing through adult endotracheal tubes. *Anesth Analg* 65(8):853–859
25. Plummer AL, Gracey DR (1989) Consensus conference on artificial airways in patients receiving mechanical ventilation. *CHEST* 96(1):178–180
26. Shapiro M, Wilson RK, Casar G, Bloom K, Teague RB (1986) Work of breathing through different sized endotracheal tubes. *Crit Care Med* 14(12):1028–1031

Affiliations

Julio César Gómez-Tamayo¹  · Jorge Iván Puerta-Guarín¹ · Claudia Marcela Rojas-Camejo¹ · Juan Pablo Caicedo¹ · José Andrés Calvache^{1,2} 

¹ Department of Anesthesiology, Hospital Universitario San José, Universidad del Cauca, Carrera 6 No. 10N-142, Third Floor, Popayán, Colombia

² Department of Anesthesiology, Erasmus University Medical Centre, Rotterdam, The Netherlands