Interobserver Reliability among Experienced and Inexperienced Anesthesia Personnel during Routine Preoperative Airway Assessment Tests

Aungsumat Wangdee, R.N., Arunotai Siriussawakul, M.D., Bencharatana Yokubol, M.D., Chanya Chansawad, Sirirat Rattana-arpa, M.D.

Department of Anesthesiology, Faculty of Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand.

ABSTRACT

Objective: The authors aimed to determine the interobserver reliability during routine preoperative airway assessments by experienced and inexperienced assessors. For the current, Thai registered-nurse-anesthetist-student training curriculum for preoperative airway assessment classes, the accuracy of airway evaluations performed by those trainees has not yet been evaluated.

Methods: An experienced anesthesiologist, a resident, a nurse anesthetist and thirty-one registered nurse anesthetist students were enrolled in a prospective observational study to assess eight obese and ten lean volunteers using seven airway assessment methods at the beginning of an academic year. The sternomental distance (SMD), the thyromental distance (TMD), the hyomental distance (HMD), the inter-incisor gap (IIG), malformation of the teeth, the Mallampati classification (MPT) and the range of neck movement (RNM) were assessed. Interobserver reliability was determined using intraclass correlation coefficient (ICC) and kappa coefficient (K).

Results: Higher strengths of interobserver agreements; and good to very good agreements were found among the two, more experienced assessors, namely, the resident and the nurse anesthetist. When seven airway assessment methods were evaluated by registered nurse anesthetist students, the interobserver agreements ranged from poor to moderate. For the subgroup analysis of lean and obese volunteers, higher reliabilities of airway assessments were found for the lean volunteers.

Conclusion: The accuracies of the airway assessments performed by the registered nurse anesthetist students were lower than those of the more-experienced groups. Thus, there is some room for improvement in the current nurse-anesthetist curriculum.

Keywords: Interobserver reliability; preoperative airway assessment; registered nurse anesthetist students (Siriraj Med J 2017;69:11-17)

INTRODUCTION

According to the Thai Anesthesia Incidents (THAI) study in Thailand, the incidence of difficult intubation and failed intubation in 2005 were 0.23% and 0.03% respectively.[1] The Department of Anesthesiology, Faculty of Medicine Siriraj Hospital, provides a two-hour lecture on preoperative airway assessment techniques in its annual training program for 30 to 40 registered nurse anesthetist students to minimize those incidents. The lecture's objective is to enhance patients' safety by using an accurate difficult-airway predictor to prompt the anesthesia provider team to anticipate adverse airway events. The reliability of measurements by assessors plays an important role in preoperative difficult-airway predictions.[2] Previous studies claimed that interobserver reliabilities depend on multiple factors related to the raters and the patients.[3] Karkouti and colleagues found excellent interobserver reliabilities for two senior residents in only two tests: mouth opening (ICC = 0.93) and chin protrusion (ICC = 0.89).[3] Hilditch et al. presented three tests by an anesthesiologist and a nurse anesthetist which also had excellent interobserver reliabilities.[4] Both papers found low interobserver reliabilities in many airway assessment tests.[3-4]

Correspondence to: Sirirat Rattana-arpa
E-mail: sirirat.raa@mahidol.ac.th
doi:10.14456/smj.2017.3

The authors aimed to compare the interobserver reliability of a resident, a nurse anesthetist and thirty-one registered nurse anesthetist students with an experienced anesthesiologist when measuring seven airway parameters.[5-9] Moreover, as the accuracy of airway evaluation is more important among obese patients, who usually have increased risk of difficult airway,[10] the authors also studied the strength of agreements when the various assessors evaluated obese versus lean volunteers.

MATERIALS AND METHODS

After receiving approval from the Siriraj Institutional Review Board (Si 516/2013), a prospective observational study was conducted in order to find out the interobserver reliabilities of the anesthesiologist and thirty-one registered nurse anesthetist students. Informed consent forms were signed by the 34 assessors enrolled in the study and the 18 adult volunteers. The volunteers, who had never undergone surgery under general anesthesia with endotracheal tube intubation, were then divided into two groups. The first comprised of ten, healthy, lean patients, while the second contained eight obese patients, each of whose BMI exceeded 30 kg/m^2.

Looking more closely at the total of 34 examiners, one was an experienced anesthesiologist, and two others were the anesthesiologist's research assistants: a second-year resident and a nurse anesthetist. The assistants were already well trained to do airway examinations of the lean and obese patients; and in a previous study, the interobserver reliabilities of the anesthesiologist and the two raters were determined to be > 0.7.[11] By contrast, the remaining examiners, 31 registered nurse anesthetist students, had only attended the lecture course without undertaking any extra-training.

All 34 examiners assessed the 18 volunteers to determine seven parameters, namely, the sternomental distance (SMD), the thyromental distance (TMD), the hyomental distance (HMD), the inter-incisor gap (IIG), malformation of the teeth, the Mallampati classification (MPT) and the range of neck movement (RNM).[6,10] The seven airway assessments took a total of five minutes per volunteer, and all volunteers and examiners had a ten-minute rest break every 20 minutes to avoid fatigue.

According to current theories and the measurement instructions given in the classroom, the parameters for the preoperative airway assessment techniques used in the present study and their measurement methods are defined as follows.[8,12] The sternomental distance (SMD) is the distance from the jugulum to the lower edge of the middle of the chin. The thyromental distance (TMD) is measured from the superior thyroid notch to the lower edge of the middle of the chin. The hyomental distance (HMD) is the distance from the body of the hyoid bone to the mentum. The inter-incisor gap (IIG) is the distance between the upper and the lower incisors when the mouth is fully open. All quantitative parameters were measured in centimeters using measuring tapes which were produced by the same manufacturer to prevent measurement bias.

As for the qualitative parameters, malformation of the teeth is determined by the presence of any dental abnormalities, such as loose teeth or missing teeth. The Mallampati classification is classified into the following four groups by using the intra-oral visibility, or the pharyngeal view, when the head is in a neutral position with the mouth fully open and the tongue protruding: the Mallampati class I is defined as the soft palate, fauces, uvula and pillars are visible; the Mallampati class II is defined as the soft palate, fauces and base of the uvula are visible; the Mallampati class III is defined as the soft palate is visible; and the Mallampati class IV is defined as only the hard palate is visible. The range of neck motion is rated by the extent of cervical spine movement when the head and the neck are fully flexed or extended.[8]

Statistical analysis

A sample of 18 volunteers was required to get an agreement among the assessors (kappa coefficient) of 0.6 with a 95% confidence interval (CI) of 0.6±0.4. Data was analyzed by using the software program, SPSS (version18, SPSS Inc., Chicago, IL, USA). The intraclass correlation coefficient (ICC) was used to analyze the results of continuous data, whereas kappa coefficient (K) was used to analyze the results of discrete data. The ICC and K ranged from -1 to 1; in which "-1" means total disagreement between the two groups, "0" means the agreement is not better than by chance, and "1" means perfect agreement. The interpretation of the values of the intraclass correlation coefficients and the kappa coefficients is in Table 1.

TABLE 1. Presents an interpretation of values of intraclass correlation (ICC) and kappa coefficient.

Kappa or ICC	Strength of agreement
≤0.20	Poor
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Good
≥0.81	Very good

RESULTS

Eighteen volunteers were enrolled in the study with no drop-outs. Table 2 presents the characteristics of the 18 volunteers. There were 11 females and 7 males, aged between 23 and 61 years of age, with a mean age 38.33±12.69 years, and a mean body mass index (BMI) 28.74±6.75 kg/m^2. All quantitative and qualitative data presented in Tables 2 and 3 were reference values, recorded by an experienced anesthesiologist.

TABLE 2. Demographic data of all volunteers (n=18) and subgroups (lean vs. obese).

Parameter	Volunteer group	Mean ± SD	Maximum	Minimum
Age (years)	All (n=18)	38.33 ± 12.69	61	23
Weight (kg)	Lean (n=10)	61 ± 8.9	80	52
	Obese (n=8)	93.1 ± 13.4	110	72
Height (cm)	Lean (n=10)	161.5 ± 7.8	175	145
	Obese (n=8)	161.9 ± 10.9	174	143
BMI (kg/m2)	Lean (n=10)	23.41 ± 3.09	29.74	17.96
	Obese (n=8)	35.40 ± 2.68	40.66	32.18
SMD (cm)*	Lean (n=10)	17.55 ± 2.06	20.50	13.00
	Obese (n=8)	17.60 ± 1.38	20.00	16.00
TMD (cm)*	Lean (n=10)	9.16 ± 1.05	10.20	7.00
	Obese (n=8)	9.60 ± 0.49	10.00	8.80
HMD (cm)*	Lean (n=10)	4.28 ± 0.93	5.80	2.30
	Obese (n=8)	5.43 ± 0.62	6.30	4.40
IIG (cm)*	Lean (n=10)	5.10 ± 0.72	6.00	4.00
	Obese (n=8)	5.00 ± 0.96	6.00	3.20

*Quantitative parameters measured by an experienced anesthesiologist (goal standard), body mass index (BMI), sternomental distance (SMD), thyromental distance (TMD), hyomental distance (HMD), inter-incisor gap (IIG), standard deviation (SD)

TABLE 3. Qualitative parameters of all volunteers (n=18) and subgroups (lean vs. obese) rated by an experienced anesthesiologist.

Volunteers	Teeth malformation		MPT		RNM (anteflexion)		RNM (retroflexion)	
	Present	Absent	Grade I- II	Grade III- IV	No limit	Limit	No limit	Limit
Lean (n=10)	4	6	7	3	9	1	9	1
	(40%)	(60%)	(70%)	(30%)	(90%)	(10%)	(90%)	(10%)
Obese (n=8)	0	8	5	3	8	0	8	0
	(0%)	(100%)	(62.5)	(37.5%)	(100%)	(0%)	(100%)	(0%)

Mallampati classification (MPT), range of neck movement (RNM)

The results of the interobserver agreements revealed higher strengths of agreement among the two, more-experienced people (the resident and the nurse anesthetist). Table 4 shows the overall ICC and K of airway assessment reliabilities of the second-year resident, the nurse anesthetist and the 31 registered nurse anesthetist students compared with the anesthesiologist. When the strengths of agreement between the anesthesiologist with the resident were compared, excellent interobserver agreements were found when the TMD, the HMD, teeth malformation and range of neck movement were assessed. Good agreements were also found when the SMD and the IIG assessments were done. In the comparison between the anesthesiologist and the nurse anesthetist, very good agreements were found in the SMD, the TMD, the HMD and range of neck movement assessments. Good agreements were also found when the IIG and the MPT were assessed.

Moving on to the results of the registered nurse anesthetist student group, the strengths of the interobserver agreements were good when the SMD was measured. Fair to moderate strengths of interobserver agreements were found for the assessments of the anteflexion of neck motion, the TMD, the MPT, the HMD, teeth malformation and the IIG. In the case of the evaluations of retroflexion neck movements, however, the results revealed poor strengths of interobserver agreements.

TABLE 4. The intraclass correlation coefficient (ICC) and the kappa coefficient between an experienced anesthesiologist and 3 different groups of assessors for all volunteers (n=18).

	ICC (mean±SD)				Kappa (mean±SD)			
	SMD	TMD	HMD	IIG	Teeth malformation	MPT	RNM (anteflexion)	RNM (retroflexion)
Resident (n=1)	0.71†	0.87*	0.93*	0.77 †	0.83*	0.57	1*	1*
Nurse anesthetist (n=1)	0.93*	0.91*	0.96*	0.80†	0.56	0.61†	1*	1*
Nurse students (n=31)	0.63 ± 0.18	0.23 ± 0.27	0.41 ± 0.19	0.50 ± 0.16	0.43 ± 0.41	0.38 ± 0.25	0.21 ± 0.41	0.15 ± 0.44

*) very good reliability, †) good reliability, sternomental distance (SMD), thyromental distance (TMD), hyomental distance (HMD), inter-incisor gap (IIG), Mallampati classification (MPT), range of neck movement (RNM), standard deviation (SD)

Table 5 presents the interobserver agreements for the subgroup populations. As for the resident, very good agreements were found for five parameters for the lean-volunteer group: the SMD, the TMD, the HMD, anteroflexion and retroflexion of neck. Good agreements were also found when the IIG and the MPT of lean-volunteer group were assessed. The same results also found when the nurse anesthetist assessed the lean volunteers.

In the case of the obese group, however, the ICCs of three of the parameters were lower: the SMD, the TMD and the HMD, when those two experienced assessors evaluated the volunteers. On the other hand, the strengths of agreement for the IIG measurements for the obese group were higher than those for the lean group. As for the discrete data, there was perfect agreement between all experienced assessors when the range of neck motion of the obese volunteers was evaluated (K=1.0). The experienced assessors also attained perfect agreements when evaluating the teeth malformations of the obese patients, but only fair to good agreements were found for the lean-patient measurements. The strengths of agreements for the obese patients' Mallampati classifications were lower than those for the lean patients.

TABLE 5. The intraclass correlation coefficient (ICC) and the kappa coefficient between an experienced anesthesiologist and 3 different groups of assessors for lean (n=10) and obese volunteers (n=8).

Volunteer		ICC (mean±SD)			Kappa (mean±SD)				
		SMD	TMD	HMD	IIG	Teeth malformation	MPT	RNM (anteflexion)	RNM (retroflexion)
Resident (n=1)	Lean	0.95*	0.93*	0.99*	0.73†	0.78b	0.74†	1*	1*
	Obese	0.54	0.65†	0.71b	0.83*	1*	0.39	1*	1*
Nurse anesthetist (n=1)	Lean	0.94*	0.95*	0.97*	0.74†	0.40	1*	1*	1*
	Obese	0.91*	0.63†	0.85†	0.89†	1*	0.14	1*	1*
Nurse students (n=31)	Lean	0.68 ± 0.21	0.41 ± 0.24	0.48 ± 0.23	0.62 ± 0.20	0.40 ± 0.36	0.39 ± 0.40	0.22 ± 0.41	0.19 0.39
	Obese	0.57 ± 0.27	-0.10 ± 0.22	0.06 ± 0.34	0.38 ± 0.29	0.52 ± 0.51	0.31 ± 0.31	1 ± 0*	1 ± 0*

*) very good reliability, †) good reliability, sternomental distance (SMD), thyromental distance (TMD), hyomental distance (HMD), inter-incisor gap (IIG), Mallampati classification (MPT), range of neck movement (RNM), standard deviation (SD)

Turning to the registered nurse anesthetist student group, the overall strengths of agreements measured in this group were lower than the experienced group. The ICCs of these four parameters for the lean patients – the SMD, the TMD, the HMD and the IIG– were higher than those for the obese patients. The moderate strengths of agreements were found when the MPT and teeth malformation of both lean and obese volunteers were assessed. When the range of neck motion of the lean volunteers was evaluated by student nurses, the results showed poor to moderate strength of agreements. In contrast, very good agreements were found when these assessors evaluated the range of neck motion of the obese volunteers (K=1.0).

DISCUSSION

Interobserver reliability has three primary influences, which are related to the assessors, the volunteers and technical errors.4 The lecture on preoperative airway assessment techniques for registered nurse anesthetist students provides anatomical landmarks and other technical information in order to reduce errors arising from the application of inconsistent techniques and inaccurate methodologies. Nevertheless, low strengths of agreements for the preoperative airway assessments were still found among the student nurses who participated in this airway assessment study.

Assessor factors can affect interrater reliabilities.[4] The results of the presented study indicate that the two, more-experienced assessors tended to perform more accurate airway assessments than the inexperienced learners (the registered nurse anesthetist students). The lower reliabilities among those student nurses probably resulted from multiple factors. For example, some of the student nurses may have lacked the ability to handle a complex physical examination task which required them to integrate their cognitive and psychomotor skills.[12] In addition, because the student nurses' basic anatomical knowledge was not as developed as that of medical students or doctors, they may sometimes have been unable to correctly identify the relevant anatomical landmarks.[12] In the case of the presented study, the upper part of the sternum or the jugulum is easier to identify than the hyoid and thyroid bones. Thus, good strengths of agreement were generally found between the anesthesiologist and the student nurses for the SMD

measurements, in contrast to moderate strengths for the HMD measurements, and only fair strengths for the TMD measurements. Furthermore, the complicated test – the Mallampati classification – requires a greater assessor skill level to recognize discrete anatomical landmarks. In keeping with that, the presented study showed that the interrater reliabilities for the student nurses were only fair.[4] To increase interrater reliabilities, Adamus et al, recommended that anesthesia providers use simple parameters to achieve higher levels of accuracy for predictions of difficult intubations.[12] Additionally, a more intensive and standardized training program, the use of diagrams and one-to-one teaching with patients should be introduced for all trainees.[4]

Volunteer factors, such as misunderstandings of complicated maneuvers and volunteer fatigue, can also reduce test reliabilities.[4] Tests that require volunteer compliance (for example, the IIG, the Mallampati classification and the range of neck motion) may produce lower reliabilities, if volunteers do not fully understand an assessor's instructions or are fatigued.[3] To reduce such errors, the proposed airway maneuvers were clearly described to all volunteers before the examination period.[3] From the authors' data, the strengths of agreement between the anesthesiologist and the two skilled assessors were good to very good. At the same time, there were poor to fair strengths of agreement between the student nurses and the anesthesiologist. It can therefore be inferred that the assessors' skill levels affect the test reliabilities after the volunteer factors are excluded.

Turning to the subgroup analysis, the more-experienced assessors displayed good competency performing the preoperative airway assessments of both the lean and the obese volunteers.[13] Higher strengths of agreement were found for the lean-volunteer assessments, especially for the continuous data. This is because lean people have prominent anatomical structures which are easy to identify.[10] As for the categorical data, perfect agreements on neck motion evaluation among all groups of assessors were found for the obese-volunteer examinations due to none of the obese volunteers had any neck deformities. In contrast, lower strengths of agreement for the lean volunteers' neck movements were found among the student nurse group. Those lower strengths of agreement may have been caused by ambiguous measurement instructions having been given to the student nurses in the classroom, or volunteer performance inconsistency.[4]

The results suggest the effectiveness and the reliability of lectures for clinical psychomotor skills are presently inadequate.[13] Since clinical competence is a complex construct, practice is essential to improve a learner's performance. The overall results suggest that there are opportunities for enhancement in the current curriculum for registered nurse anesthetist students; this could include initiatives such as providing more active learning sessions in simulation theaters, undertaking one-to-one teaching and conducting trainee-performance evaluations.[13]

In terms of limitations, the presented study could have enrolled larger numbers of anesthesiologists, residents and nurse anesthetists to maximize interrater reliabilities. Interrater reliabilities could then have been evaluated between two people in the same group as well as between different groups to recheck the reliabilities of all assessors. Secondly, the study was conducted among the student nurses at the beginning of the academic year without any subsequent re-evaluation before their graduation. Since the transition from a novice to an expert clearly occurs over time, a re-evaluation before graduation should reveal how student nurses' airway assessment reliabilities are influenced by time.

CONCLUSION

The airway assessments performed by the student nurses to predict difficult intubations in the volunteers were generally less accurate than those of the experienced anesthesiologist, the resident and the nurse anesthetist. A more intensive and standardized training program should be introduced for all trainees to improve patients' safety. Future study should re-evaluate the seven airway measurement methods at the end of academic year in order to find out the interrater reliabilities among experienced anesthesiologists and student nurses, so this study results will represent the accuracy of difficult airway prediction of 1-year experienced trainees. This end of academic year findings will reflect the quality of airway assessment training program and lead to training curriculum development.

ACKNOWLEDGMENTS

The authors thank Suwannee Suraseranivongse, M.D. and Saowaphak Lapmahapaisan, M.D. for their manuscript reviews, and Nichapat Sookri, a research assistant, for her invaluable help with the associated paperwork.

REFERENCES

1. Chanchayanon T, Suraseranivonge S, Chau-in W. The Thai Anesthesia Incidents Study (THAI study) of difficult intubation: a qualitative analysis. J Med Assoc Thai 2005;88(7):S62-7.=
2. Rosenstock C, Gillesberg I, Gatke MR, Levin D, Kristensen MS, Rasmussen LS. Inter-observer agreement of tests used for prediction of difficult laryngoscopy/tracheal intubation. Acta Anaesthesiologica Scandinavica 2005;49(8):1057-62.
3. Karkouti K, Rose DK, Ferris LE, Wigglesworth DF, Meisami-Fard T, Lee H. Inter-observer reliability of ten tests used for predicting difficult tracheal intubation. Can J Anesth 1996;43(6): 554-49.
4. Hilditch WG, Kopka A, Crawford JM, Asbury AJ. Interobserver reliability between a nurse and anaesthetist of tests used for predicting difficult tracheal intubation. Anaesthesia 2004;59(9): 881-4.
5. Merah NA, Wong DT, Ffoulkes-Crabbe DJ, Kushimo OT, Bode CO. Modified Mallampati test, thyromental distance and inter-incisor gap are the best predictors of difficult laryngoscopy in West Africans. Can J Anaesth 2005;52(3):291-96.
6. Ittichaikulthol W, Chanpradub S, Amnoundetchakorn S, Arayajarernwong N, Wongkum W. Modified Mallampati test and thyromental distance as a predictor of difficult laryngoscopy in Thai patients. J Med Assoc Thai 2010;93(1):84-89.
7. Khan ZH, Mohammadi M, Rasouli MR, Farrokhnia F, Khan RH. The diagnostic value of the upper lip bite test combined with sternomental distance, thyromental distance, and interincisor distance for prediction of easy laryngoscopy and intubation: a prospective study. AnesthAnalg 2009;109(3):822-24.
8. Reynolds J, Marsh D, Koller H, Zenenr J, Bannister G. Cervical range of movement in relation to neck dimension. Eur Spine J 2009;18(6):863-38.
9. Huh J, Shin HY, Kim SH, Yoon TK, Kim DK. Diagnostic predictor of difficult laryngoscopy: the hyomental distance ratio. Anesth Analg 2009;108(2):544-48.
10. Langeron O, Birenbaum A, Le Sache F, Raux M. Airway management in obese patient. Minerva Anestesiologica 2014;80(3):382-92.
11. Siriussawakul A, Wangdee A, Chatsiriphatthana A, Komoltri C, Dejarkom S, Jirachaipitak S, et al. Modified Mallampati Test, Neck Circumference and Short Inter-Incisor Gap Are Risk Factors for Troublesome Intubation in Thai Obese Patients. Siriraj Med J 2014;66(6):202-9.
12. Adamus M, Jor O, Vavreckova T, Hrabalek L, Zapletalova J, Gabrhelik T, et al. Inter-observer reproducibility of 15 tests used for predicting difficult intubation. Biomed Pap Med Fac Univ Palacky, Olomouc Czech Repub 2011;155(3):275-81.
13. Wass V, Archer J. Assessing learners. In: Dornan T, Mann K, Scherpbier A, Spencer J, editors. Medical Education: Theory and Practice. 1st edition. China: Elsevier Ltd; 2011. p.229-55.