

Original Article

Inter-rater reliability of the American Society of Anesthesiologists physical status rating for emergency gastrointestinal surgery

Tsutomu Shichino,¹ Motohiro Hirao,² and Yoshio Haga^{3,4}

¹Department of Anesthesia, National Hospital Organization Kyoto Medical Center, Kyoto, Japan, ²Department of Surgery, National Hospital Organization Osaka National Hospital, Osaka, Japan, ³Institute for Clinical Research, National Hospital Organization Kumamoto Medical Center, Kumamoto, Japan, and ⁴Department of International Medical Cooperation, Graduate School of Medical Sciences Kumamoto University, Kumamoto, Japan

Aim: The American Society of Anesthesiologists Physical Status (ASA-PS) classification system is used worldwide and has also been incorporated into various prediction rules. However, concerns have been raised regarding inter-rater agreement in various surgical fields. Although emergency gastrointestinal surgery is relatively common and associated with high postoperative mortality, a reliability study has not yet been undertaken in this field. The aim of the present study was to investigate the inter-rater reliability of ASA-PS for emergency gastrointestinal surgery.

Methods: Three sets of scenarios were generated for each ASA-PS class (2E, 3E, and 4E) in emergency gastrointestinal surgery, resulting in nine scenarios. These scenarios described the preoperative profiles of patients in one hospital. Two or three anesthesiologists from 18 other hospitals provided scores for ASA-PS for each scenario.

Results: Fifty anesthesiologists scored the ASA-PS class. Between 66% and 90% of these anesthesiologists assigned the same ratings as the reference ratings for the individual scenarios. Inter-rater reliability was assessed using Fleiss' kappa (95% confidence interval) of 0.55 (0.54–0.56, $P < 0.001$) and an intraclass correlation coefficient (95% confidence interval) of 0.79 (0.63–0.93, $P < 0.001$).

Conclusion: The results of the present study revealed the consistency of ASA-PS ratings between anesthesiologists for emergency gastrointestinal surgery. The ASA-PS may serve as a reliable variable in the prediction rules for this field.

Key words: Digestive system surgical procedures, emergencies, health status indicators, quality assurance, reliability and validity

INTRODUCTION

THE ORIGINAL OBJECTIVE of the American Society of Anesthesiologists Physical Status (ASA-PS) classification was to create a uniform system for statistical analyses.¹ However, many modern prediction models include ASA-PS as an independent variable.^{2–7} Although the reliability of ASA-PS is important for prediction models, previous studies reported fair to moderate inter-rater agreement between different anesthesiologists.^{8–13}

Emergency gastrointestinal surgery is relatively common. According to national statistics in Japan, the number of

procedures for intestinal obstruction accounts for 44% of cases of colectomy for malignant tumors. In addition to the volume, emergency gastrointestinal surgery is associated with high risks of postoperative morbidity and mortality, leading to extensive healthcare costs.¹⁴ The preoperative prediction of a patient's prognosis is of clinical importance in treatment planning and informed consent. We previously constructed a prediction rule for postoperative mortality, designated as Calculation of Post-operative Risk in Emergency Surgery (CORES).¹⁵ The National Surgical Quality Improvement Program (NSQIP) is also used as a prediction rule in a wide variety of emergency and elective surgeries.⁶ Both CORES and NSQIP include ASA-PS as an independent variable.

As patients undergoing emergency gastrointestinal surgery present with various degrees of physiological decompensation, ASA-PS ratings in these patients may vary among anesthesiologists. Therefore, the inter-rater reliability

Corresponding: Yoshio Haga, M.D., Institute for Clinical Research, National Hospital Organization Kumamoto Medical Center, 1-5 Ninomaru, Chuo-ku, Kumamoto 8600008, Japan. E-mail: yoshio@kumamed.jp.

Received 17 May, 2016; accepted 9 Aug, 2016; online publication 13 Sep, 2016

of ASA-PS in this field urgently needs to be assessed, but has not yet been investigated. The present study was undertaken to evaluate the inter-rater reliability of ASA-PS for emergency gastrointestinal surgery.

METHODS

THIS STUDY WAS carried out in the surgery/anesthesiology network in the National Hospital Organization (NHO), Japan. Nineteen NHO hospitals participated in this study. The study protocol was approved by the Central Institutional Review Board of the NHO (Tokyo, Japan) on October 3, 2014 (H26-1003001). Anesthesiologists in these hospitals had graduated from different universities and did not share any common background. Each hospital had between 2 and 17 anesthesiologists and managed 800 to 5500 surgical cases per year.

Three scenarios were constructed for ASA-PS: 2E, 3E, and 4E, based on actual cases in the first author's hospital, resulting in nine scenarios for emergency gastrointestinal surgical cases (see Appendix). The first author selected scenarios from the latest three consecutive clinical cases in each category of the ASA-PS class in the hospital in order to avoid selection bias, the surgeries for which were carried out between September 2013 and October 2014. These scenarios consisted of history, vital signs, and laboratory data and were reviewed for their compatibility by two members of the protocol committee. These scenarios were written in a Word file (Microsoft, Redmond, WA, USA) in the order of Patient 1 to Patient 9 and were delivered by e-mail to the anesthesiologists of the 18 hospitals of the NHO, except for the hospital at which the scenarios were made. Prior to their delivery, no discussion, arrangement, or training was given on ASA-PS ratings between the hospitals. Two or three anesthesiologists in each hospital provided ASA-PS scores for each scenario. The ASA-PS scores were anonymously entered into an Excel file (Microsoft, Redmond, WA, USA) and submitted to the principal investigator of the participating hospitals.

Statistical analyses were carried out using Stata 12.0 (StataCorp LP, College Station, TX, USA). A two-tailed *P*-value of 0.05 was used to define significance. We measured agreement between ASA-PS ratings assigned by the anesthesiologists using a two-way random intraclass correlation coefficient (ICC), as reported by Sankar and colleagues.¹³ Sample size was determined as 50 based on the assumption that ICC was 0.7 with a half width of 95% confidence interval of 0.2. We also analyzed the agreement between ASA-PS ratings using Fleiss' kappa.¹⁶ These indices ranged between 0 and 1.0, with a higher value indicating greater consistency. Landis and Koch characterized reliability

statistical values of 0.00–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and values exceeding 0.80 as almost perfect.¹⁷

RESULTS

THE TWO MEMBERS of the protocol committee independently qualified the case-based scenarios based on the availability of the clinical data as well as information about history. Fifty anesthesiologists from 18 hospitals scored ASA-PS ratings for the nine scenarios. The response rate was 100%. Figure 1 shows histograms of the ASA ratings obtained. Between 66% and 90% of the anesthesiologists scored the same rating for each scenario as the reference scores. The ICC (95% confidence interval) among these anesthesiologists was 0.79 (0.63–0.93, *P* < 0.001); Fleiss' kappa was 0.55 (0.54–0.56, *P* < 0.001). There was no special tendency of ASA ratings amongst the hospitals.

DISCUSSION

POSTOPERATIVE MORBIDITY AND mortality rates for emergency gastrointestinal surgery are high. The ASA-PS score is used as an independent variable in prediction rules in this setting.^{6,15} Therefore, we determined whether the ASA-PS rating was reliable between anesthesiologists. The results obtained showed that agreement between anesthesiologists was relatively high and, thus, reliability may be regarded as satisfactory. To the best of our knowledge, this is the first study to investigate the inter-rater reliability of ASA-PS in the field of emergency gastrointestinal surgery.

When ASA-PS was originally devised in 1941,¹⁸ Saklad emphasized that its aim was not to estimate "operative risk", but to facilitate the "tabulation of statistics". However, many studies have been undertaken in an attempt to determine the relationship between ASA-PS and the risk of perioperative morbidity and/or mortality.^{2–5,7,14,15,19–28} Apart from a few studies,^{25,26,28} most found a strong correlation and suggested the usefulness of ASA-PS as an independent risk factor in certain prediction models.^{4,5,14,15,19,23,24} If inter-rater reliability is guaranteed, ASA-PS has the potential to be a very useful factor in prediction models.

Several studies have reported the reliability or validity of ASA-PS ratings since Owens *et al.* investigated the consistency of ratings in 1978.⁸ Most of these studies used hypothetical case scenarios for the different anesthesiologists' ratings, and revealed only fair inter-rater agreement. Although the case scenarios may have been made deliberately in these studies, the purpose of verifying consistency was not achieved. For example, a wide range of replies were

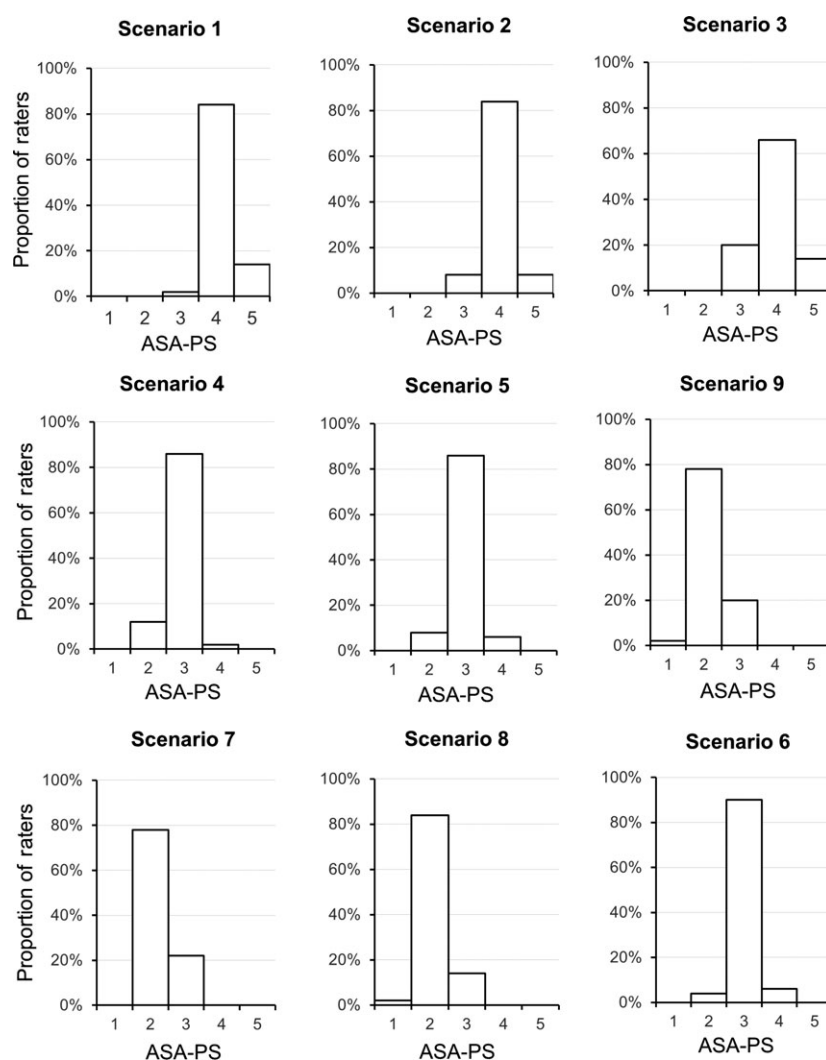


Fig. 1. Distribution of American Society of Anesthesiologists Physical Status (ASA-PS) ratings assigned to nine case-based scenarios by 50 anesthesiologists in the surgery/anesthesiology network of the National Hospital Organization, Japan. The study investigated the inter-rater reliability of ASA-PS for emergency gastrointestinal surgery.

supplied for the fourth case scenario by Hynes and Lawler (subarachnoid hemorrhage),⁹ with all five possible ASA grades. Their scenarios were only descriptive and included few of the objective data needed to accurately determine ASA-PS. A prospective cross-over study designed to determine the degree of agreement for ASA grades between two different anesthesiologists showed a moderate agreement ($\kappa = 0.53$),²⁹ with a retrospective cohort study comparing ASA-PS scores assigned at a preoperative assessment clinic versus the operating theatre showing higher reliability ($\kappa = 0.61$).¹³ These two studies were based on “real world” clinical practices.

In the present study, most anesthesiologists graded a case into the same class in all scenarios. We tested inter-rater agreement using two measurements, Fleiss’ kappa and ICC. Fleiss’ kappa is a conservative measure of inter-rater reliability for categorical ratings that removes the portion of agreement due to chance. An interpretation of Fleiss’ kappa has not yet been established because the number of categories and subjects affects the magnitude of the value. Ihejirika and colleagues¹⁶ analyzed the inter-rater reliability of ASA-PS using nine scenarios, similar to the present study, and obtained moderate agreement ($\kappa = 0.51$). They concluded that substantial agreement strength for reliability was

achieved.¹⁶ The present study achieved a higher value for $F-\kappa$ (0.55) than that reported by Ihejirika and colleagues.¹⁶ The ICC is a measure of inter-rater reliability for quantitative ratings. However, it has sometimes been used to assess ASA-PS.¹³ The value in the current study of 0.79 means “substantial”. Based on these results, we concluded that the inter-rater agreement of ASA-PS for emergency gastrointestinal surgery may be substantial.

There were some limitations in the present study. First, we did not use actual clinical patients, but case-based scenarios. The scenarios lacked radiographic images and electrocardiogram data. These limited data may influence the decisions by anesthesiologists. Second, this was not a large volume cohort study, using only nine case scenarios. Furthermore, each scenario may be relatively easy to judge ASA-PS ratings. We selected scenarios from the latest three consecutive clinical cases in each category of the ASA-PS class in one hospital in order to prevent selection bias. However, the consistency of this study suggests that anesthesiologists are more likely to grade in a similar manner, at least for typical cases. Finally, we did not use the ASA IE scenario, which narrows the range of potential options and, thus, biases reliability upwards.

In conclusion, the ASA-PS rating was substantial between anesthesiologists for emergency gastrointestinal surgery. Thus, ASA-PS may serve as a reliable index in prediction rules for this field.

CONFLICT OF INTEREST

NONE.

ACKNOWLEDGMENTS

THIS STUDY WAS supported by a grant for NHO network clinical research (H26-Surgery-001). The authors wish to thank all the anesthesiologists who assigned ASA-PS ratings for the scenarios. The authors are also grateful to colleagues at the Clinical Research Center, NHO Nagoya Medical Center for their support with the statistical analyses. Finally, the authors acknowledge Dr. Koji Oba, Department of Biostatistics, School of Public Health, Graduate School of Medicine, The University of Tokyo, for his kind advice on biostatistics.

REFERENCES

- 1 Driggs RD, Lamont A, Eckenhoff JE. The role of anesthesia in surgical mortality. *JAMA* 1961; 178: 261–6.

- 2 Glance LG, Lustik SJ, Hannan EL *et al.* The surgical mortality probability model: derivation and validation of a simple risk prediction rule for noncardiac surgery. *Ann. Surg.* 2012; 255: 696–702.
- 3 Davenport D, Bowe EA, Henderson WG, Khuri SF, Mentzer RM. National Surgical Quality Improvement Program (NSQIP) risk factors can be used to validate American Society of Anesthesiologists Physical Status Classification (ASA PS) levels. *Ann. Surg.* 2006; 243: 636–41.
- 4 Dalton JE, Kurz A, Turan A, Mascha EJ, Sessler DI, Saager L. Development and validation of a risk quantification index for 30-day postoperative mortality and morbidity in noncardiac surgical patients. *Anesthesiology* 2011; 114: 1336–44.
- 5 Gupta PK, Gupta H, Sundaram A, *et al.* Development and validation of a risk calculator for prediction of cardiac risk after surgery. *Circulation* 2011; 124: 381–7.
- 6 Gupta PK, Franck C, Miller WJ, Gupta H, Forse RA. Development and validation of a bariatric surgery morbidity risk calculator using the prospective, multicenter NSQIP dataset. *J. Am. Coll. Surg.* 2011; 212: 301–9.
- 7 Haga Y, Ikei S, Ogawa M. Estimation of Physiologic Ability and Surgical Stress (E-PASS) as a new prediction scoring system for postoperative morbidity and mortality following elective gastrointestinal surgery. *Surg. Today* 1999; 29: 219–25.
- 8 Owens WD, Felts JA, Spitznagel EL Jr. ASA physical status classifications: a study of consistency of ratings. *Anesthesiology* 1978; 49: 239–43.
- 9 Hynes SR, Lawler PGP. An assessment of the consistency of ASA physical status classification allocation. *Anaesthesia* 1995; 50: 195–9.
- 10 Mak PHK, Campbell RCH, Irwin MG. The ASA physical status classification: inter-observer consistency. *Anesth Intensive Care*. 2002; 30: 633–40.
- 11 Aronson CW, McAuliffe MS, Miller K. Variability in the American Society of Anesthesiologists physical status classification scale. *AANA J.* 2003; 71: 265–74.
- 12 Ranta S, Hynymen M, Tammisto T. A survey of the ASA physical status classification: significant variation in allocation among Finnish anaesthesiologists. *Acta Anaesthesiol. Scand.* 1997; 41: 629–32.
- 13 Sankar A, Johnson SR, Beattie WS, Tait G, Wijesundera DN. Reliability of the American Society of Anesthesiologists physical status scale in clinical practice. *Br. J. Anaesth.* 2014; 113: 424–32.
- 14 Kawashima Y, Seo N, Morita K, *et al.* Anesthesia-related mortality and morbidity in Japan (1999). *J. Anesth.* 2002; 16: 319–31.
- 15 Miyazaki N, Haga Y, Matsukawa H, *et al.* The development and validation of the Calculation of post-Operative Risk in Emergency Surgery (CORES) model. *Surg. Today* 2014; 44: 1443–56.

- 16 Ihejirika RC, Thakore RV, Sathiyakumar V, Ehrenfeld JM, Obremskey WT, Sethi MK. An assessment of the inter-rater reliability of the ASA physical status score in the orthopaedic trauma population. *Injury*. 2015; 46: 542–6.
- 17 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–74.
- 18 Saklad M. Grading of patients for surgical procedures. *Anesthesiology* 1941; 2: 281–4.
- 19 Wolters U, Wolf T, Stutzer H, Schroder T. ASA classification and perioperative variables as predictors of postoperative outcome. *Br. J. Anaesth.* 1996; 77: 217–22.
- 20 Vacanti CJ, VanHouten RJ, Hill RC. A statistical analysis of the relationship of physical status to postoperative mortality in 68,388 cases. *Anesth. Analg.* 1970; 49: 564–6.
- 21 Prause G, Ratzenhofer-Komenda B, Pierer G, Smolle-Juttner F, Glanzer H, Smolle J. Can ASA grade or Goldman's cardiac risk index predict peri-operative mortality? A study of 16,227 patients. *Anaesthesia* 1997; 52: 203–6.
- 22 Marx GF, Mateo CV, Orkin LR. Computer analysis of postanesthetic deaths. *Anesthesiology* 1973; 39: 54–8.
- 23 Prause G, Offiner A, Ratzenhofer-Komenda B, Vicenzi M, Smolle J, Smolle-Juttner F. Comparison of two preoperative indices to predict perioperative mortality in non-cardiac thoracic surgery. *Eur. J. Cardiothorac. Surg.* 1997; 11: 670–5.
- 24 Dimick JB, Osborne NH, Hall BL, Ko C, Birkmeyer JD. Risk-adjustment for comparing hospital quality with surgery: how many variables are needed? *J. Am. Coll. Surg.* 2010; 210: 503–8.
- 25 Wolters U, Wolf T, Stutzer H, Schroder T, Pichlmaier. Risk factors, complications, and outcome in surgery: a multivariate analysis. *Eur. J. Surg.* 1997; 163: 563–8.
- 26 Chijiwa K, Yamaguchi K, Yamashita H, Ogawa Y, Yoshida J, Tanaka M. ASA physical status and age are not factors predicting morbidity, mortality, and survival after pancreatoduodenectomy. *Am. Surg.* 1996; 62: 701–5.
- 27 Hightower CE, Riedel BJ, Feig BW, *et al.* A pilot study evaluating predictors of postoperative outcomes after major abdominal surgery: physiological capacity compared with the ASA physical status classification system. *Br. J. Anaesth.* 2010; 104: 465–71.
- 28 Gilbert K, Larocque BJ, Patrick LT. Prospective evaluation of cardiac risk indices for patients undergoing noncardiac surgery. *Ann. Intern. Med.* 2000; 133: 356–9.
- 29 Cuvillon P, Nouvellon E, Marret E, Albaladejo P, Fortier LP, Fabbro-Perray P, Malinovsky JM, Ripart J. American Society of Anesthesiologists' physical status system: a multicentre Francophone study to analyse reasons for classification disagreement. *Eur J Anaesthesiol* 2011; 28: 742–7.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix. Case-based scenarios to assess the inter-rater reliability of the American Society of Anesthesiologists Physical Status rating for emergency gastrointestinal surgery.