Journal of Clinical Anesthesia 79 (2022) 110794

Contents lists available at ScienceDirect

Journal of Clinical Anesthesia

journal homepage: www.elsevier.com/locate/jclinane

Original Contribution

An evaluation of the accuracy and self-reported confidence of clinicians in using the ASA-PS Classification System

Saullo Queiroz Silveira, (MD) [a], Leopoldo Muniz da Silva, (PhD) [a,*], Ronald Figueiredo Gomes, (MD) [b], Arthur de Campos Vieira Abib, (MD) [a], Joaquim Edson Vieira, (MD) [c], Anthony M.-H. Ho, (MD) [d], Helidea de Oliveira Lima, (MSc) [e], Fernando Nardy Bellicieri, (MD) [a], Daenis Camire, (MD) [d], Rafael Sousa Fava Nersessian, (MD) [a], Glenio B. Mizubuti, (MD) [d]

[a] *Department of Anesthesiology, São Luiz Hospital - ITAIM / Rede D'Or – CMA Anestesia team, D'Or Institute for Research and Education (IDOR), São Paulo, SP, Brazil*
[b] *Department of Anesthesiology, São Luiz Hospital – Jabaquara, D'Or Institute for Research and Education (IDOR), Rede D'Or, Brazil*
[c] *Department of Anesthesiology, Faculdade de Medicina da Universidade de São Paulo (FMUSP), São Paulo, SP, Brazil*
[d] *Department of Anesthesiology and Perioperative Medicine, Queen's University, Kingston, ON, Canada*
[e] *Department of Quality and Safety - Rede D'Or, São Paulo, SP, Brazil*

A R T I C L E I N F O

Keywords:
Decision making
Surveys and questionnaires
Health status
Perioperative care
Continuing education

A B S T R A C T

Objectives: The American Society of Anesthesiologists Physical Status (ASA-PS) is a grading system routinely adopted worldwide by physicians to classify patients' overall health status. Concerns have been raised surrounding the subjectiveness of this system, potentially leading to poor inter-rater agreement/reliability. We hypothesized that physicians are overconfident when assigning ASA-PS scores and that presenting them with the ASA-PS definitions/examples would improve accuracy. We therefore evaluated participants' accuracy and self-reported confidence on the ASA-PS Classification System (1) while assigning ASA-PS according to their baseline knowledge/judgment; and (2) after a single exposure to the ASA-PS definitions/examples.
Design: Prospective before-and-after web-based study.
Participants: 272 anesthesiologists and 114 non-anesthesiologists.
Interventions: Participants voluntarily answered a web-based questionnaire consisting of 10 hypothetical cases. They were asked to assign an ASA-PS score and rate their perceived self-confidence level (20–100%) on the accuracy of their assigned score for each case both (1) before and (2) after reviewing the ASA-PS definitions/examples. The correct ASA-PS for each hypothetical case was determined by consensus among investigators.
Measurements: Participants' accuracy, self-reported confidence, and calibration of confidence on the application of ASA-PS Classification System. Agreement between measures was tested using kappa coefficient.
Results: Anesthesiologists had better accuracy than non-anesthesiologists both on initial [6(5–7) vs. 4(3–5) out of 10; $p < 0.001$] as well as subsequent [7(6–8) vs. 6(4–7); $p < 0.001$] ASA-PS score assignments. Participants' self-reported confidence was greater than their accuracy for assigned ASA-PS scores ($p < 0.001$). ASA-PS agreement between anesthesiologists and non-anesthesiologists was poor ($\kappa < 0.20$). Participants' accuracy for hypothetical cases of ASA-PS I, II, and III involving adult patients was overall greater than for ASA-PS IV, V, and III (the latter involving a neonate) for both anesthesiologists and non-anesthesiologists ($p < 0.001$).
Conclusions: Physicians tend to disagree and be overconfident when assigning ASA-PS scores. A brief consultation of the ASA-PS definitions/examples improves the accuracy for both anesthesiologists and non-anesthesiologists.

1. Introduction

The American Society of Anesthesiologists Physical Status (ASA-PS) Classification System was created in 1941 as a tool for compilation of statistical data [1]. Since then, anesthesiologists and non-anesthesiologists have used it in areas such as resource allocation, billing, and perioperative risk assessment [2–5].

Although widely adopted, the ASA-PS classification presents

* Corresponding author at: Rua Dr. Alceu de Campos Rodrigues, 229, Conj. 403, CEP 04544-000 São Paulo, SP, Brazil.
E-mail address: leopoldo.muniz@saoluiz.com.br (L.M. da Silva).

https://doi.org/10.1016/j.jclinane.2022.110794
Received 4 November 2021; Received in revised form 19 March 2022; Accepted 25 March 2022
Available online 31 March 2022
0952-8180/

important inconsistencies including high scoring variability among healthcare professionals [6–9]. Possible causes for such variability include (i) the inherent subjectivity of the system; (ii) educational factors; (iii) professional experience, and (iv) institutional characteristics [10]. Indeed, it is virtually impossible to objectively define the ASA-PS across the infinite universe of patient presentations; thus, inter-rater variability is inevitable. In response, the ASA published in 2014 examples for each category of the ASA-PS to serve as guidelines [11], which has proven beneficial as an educational tool to improve physicians' accuracy [3].

The degree of certainty/confidence about a decision is part of the decision-making process [6,12] and brings about the concept of "calibration" [12] – a heavily studied topic in psychology and decision science. In our case, *accuracy* refers to one's ability to correctly assign ASA-PS scores to hypothetical patients [5], while *confidence* relates to the individual's (i.e., assignor's) self-reported degree of certainty about the assigned ASA-PS scores. If one's confidence is appropriate for the level of accuracy, the person is said to be well calibrated. A lack of linear relationship between accuracy and confidence implies poor calibration [12,13]. Hence, "under-" and "over-confidence" are terms used to indicate poor calibration, diverging in opposite directions. Importantly, calibration does not necessarily simply reflect accuracy. For instance, one can be very well calibrated by answering a small number/proportion of questions correctly (i.e., low accuracy) so long they assign appropriately low confidence levels to their answers.

In the medical field, many examples exist in which calibration has been proven poor, with considerable levels of overconfidence, leading to misguided decision-making and perpetuation of the misconception [6,12,13]. Accordingly, there have been several studies assessing the accuracy of ASA-PS assignment [3,6–10] and the impact of examples to improve accuracy [3], but none of them have analyzed the calibration of physicians' confidence, or the impact of examples on this dimension. We hypothesized that the ASA-PS is often applied with overconfidence by physicians, potentially impacting decision-making, risk stratification, resource allocation, and scientific production; and that presenting them with the ASA-PS definitions and ASA-approved examples would improve accuracy and calibration.

We aimed to evaluate the accuracy and self-reported confidence of anesthesiologists and non-anesthesiologists on the ASA-PS Classification System at two different time points: (1) before and (2) after re-familiarization of the ASA-PS definitions and examples published by the ASA [2].

2. Methods

Following Hospital São Luiz & Rede D'or and Affiliated Teaching Hospitals Research Ethics Board approval (Protocol# 4,518,843; CAAE: 42047321.8.0000.0087), physicians were contacted by e-mail and invited to answer a web-based questionnaire. Physicians were also encouraged to disseminate the invitation email to other eligible participants. The invitation emails were initially distributed by provincial anesthesiology societies across all five Brazilian macro geographic regions, i.e., North, Northeast, Central-West, Southeast, and South. Data collection took place between January and April 2021.

Participants were divided into 2 groups: anesthesiologists and non-anesthesiologists. Non-anesthesiologists included surgical specialists (herein referred to as surgeons) and medical specialists (i.e., cardiologists, nephrologists, and pulmonologists) that commonly utilize the ASA-PS when evaluating patients perioperatively. We created a public web portal using Google forms (https://forms.gle/DmdApWBN9Jden2WA7) to ensure anonymity. All participants provided informed consent (electronically) prior to beginning the study. The first step of the research protocol consisted of 10 hypothetical clinical scenarios (see below) where participants assigned an ASA-PS according to their underlying knowledge/judgment. In the next step, participants reviewed the ASA-PS category's definitions and examples (Table 1) [11]. There was no time limit for participants to review this table, but once their review was complete, they were instructed to carry forward to the next step at which point they were not allowed to go back to the table. Then, participants received the same 10 hypothetical cases in a different/random order and were asked to (re-)assign (i.e., confirm or modify) the ASA-PS. Participants were not allowed to go back and change previous answers while responding to the second questionnaire. The correct ASA-PS for each hypothetical case was previously determined by consensus among the investigators using objective interpretation of the ASA-PS Classification System definitions and the ASA-approved examples (Table 1) [11]. All hypothetical cases were discussed by the authors including one internist (H.O.L) and 10 anesthesiologists with varying degrees of professional experience – two senior resident physicians and eight board-certified anesthesiologists with 3–32 years of clinical experience. Consensus was reached when at least 80% of these adjudicators agreed with the proposed classification. Cases 1, 3, 4, 5, 8 and 9 were adapted from previous work by Hurwitz et al. [3], whereas cases 2,

Table 1
American Society of Anesthesiologists-Physical Status (ASA-PS) Classification System.

ASA-PS	Definition	Examples, Including, but Not Limited to
I	A normal healthy patient.	Healthy, nonsmoking, no or minimal alcohol use.
II	A patient with mild systemic disease.	Mild diseases only without substantive functional limitations. Examples include (but not limited to) current smoker, social alcohol drinker, pregnancy, obesity (30 < BMI < 40), well-controlled DM or HTN, mild lung disease.
III	A patient with severe systemic disease.	Substantive functional limitations; one or more moderate to severe diseases. Examples include (but not limited to) poorly controlled DM or HTN, COPD, morbid obesity (BMI ≥40), active hepatitis, alcohol dependence or abuse, implanted pacemaker, moderate reduction of ejection fraction, ESRD undergoing regularly scheduled dialysis, premature infant PCA <60 weeks, history (>3 months) of MI, CVA, TIA, or CAD/stents.
IV	A patient with severe systemic disease that is a constant threat to life.	Examples include (but not limited to) recent (<3 months) MI, CVA, TIA, or CAD/stents, ongoing cardiac ischemia or severe valve dysfunction, severe reduction of ejection fraction, sepsis, DIC, ARDS, or ESRD not undergoing regularly scheduled dialysis.
V	A moribund patient who is not expected to survive without the operation	Examples include (but not limited to) ruptured abdominal/thoracic aneurysm, massive trauma, intracranial bleed with mass effect, ischemic bowel in the face of significant cardiac pathology or multiple organ/system dysfunction.
VI	A declared brain-dead patient whose organs are being removed for donor purposes	

The addition of "E" denotes emergency surgery. An emergency is defined as existing when delay in treatment of the patient would lead to a significant increase in the threat to life or body part. ARDS = acute respiratory distress syndrome; BMI = body mass index (kg/m^2); CAD = coronary artery disease; COPD = chronic obstructive pulmonary disease; CVA = cerebrovascular accident; DIC = disseminated intravascular coagulation; DM = diabetes mellitus; ESRD = end-stage renal disease; HTN = hypertension; MI = myocardial infarction; PCA = post-conceptual age; TIA = transient ischemic attack. Adapted from https://www.asahq.org/standards-and-guidelines/asa-physical-status-classification-system [11].

6, 7 and 10 were created by two investigators (S.Q.S. and L.M.S) and reviewed/validated by the other members of the research group to include scenarios (i.e., pregnant patients, neonates, ASA-PS IV-E and ASA-PS V-E, respectively) that were absent in the study by Hurwitz et al. [3]. These adaptations were introduced to expose responders to clinical scenarios involving all (I-V) ASA classes. After assigning an ASA-PS score for each case, participants were asked to indicate their perceived level of confidence associated with that response prior to moving to the next case. This procedure was repeated for all 10 hypothetical cases. The percentages of correct answers and confidence levels indicated by participants throughout the questionnaire were recorded. Confidence levels were selected from a dropdown menu using a 5-point scale (i.e., 20%, 40%, 60%, 80%, and 100%) [14,15] where 20% indicated significant/complete uncertainty and 100% indicated complete certainty. The electronic questionnaire was piloted to ensure user friendliness and that any technical issues were resolved prior to starting data collection. The hypothetical cases in the survey were as follows:

Case 1. A 32-year-old man presents for laparoscopic gastric bypass. He currently weighs 118 kg (body mass index [BMI] 42 kg/m^2), after an intentional weight loss of 15 kg over the past 6 months. He has gastroesophageal reflux disease (GERD) that is controlled on omeprazole. He currently walks 5 km/day and denies any chest pain or shortness of breath on exertion. Preoperative blood pressure is 118/70 mmHg, heart rate 84 beats/min. (Correct: ASA-PS III).

Case 2. A 22-year-old woman with full-term pregnancy presents to the labor and delivery unit for elective cesarean section. She currently weighs 68 kg (BMI 25 kg/m^2). Her weight prior to pregnancy was 60 kg. She is otherwise healthy and denies any comorbidities during or prior to her pregnancy and has never undergone any surgery. She takes only oral multivitamin prescribed by the obstetrician. She has never smoked or used alcohol. Blood pressure is 110/60 mmHg, heart rate 70 beats/min. (Correct ASA-PS: II).

Case 3. An 87-year-old woman presents for cataract surgery. She weighs 55 kg (BMI 20 kg/m^2). She works as a volunteer in a bookstore four times a week. She denies any comorbidity. However, she has not seen a physician in 20 years and does not take any medication. She lives alone and maintains regular domestic activities such as walking to the supermarket once a week. Physical examination is unremarkable. (Correct ASA-PS: I).

Case 4. A 69-year-old man is admitted to the operating room for an elective endovascular repair of an infra-renal abdominal aortic aneurysm (AAA). He weighs 135 kg (BMI 35 kg/m^2) and has a history of hypertension controlled with metoprolol and nifedipine. He had a myocardial infarction 6 years ago treated with two drug-eluting stents for which he takes aspirin daily. He underwent a cardiac evaluation a month ago and was deemed optimized from a cardiovascular standpoint. There are no other comorbidities and his physical examination is unremarkable. (Correct ASA-PS: III).

Case 5. A 42-year-old woman presents for umbilical hernia repair. She has a history of uncontrolled hypertension in the past leading to end-stage renal disease (ESRD). She is compliant with her hemodialysis three times per week. Her last dialysis session was 12 h ago. She denies any other end-organ damage related to her hypertension. Lately, her blood pressure has been well controlled on atenolol. She denies chest pain or shortness of breath while doing yard work. Her blood pressure is 122/84 mmHg and her potassium is 4.1 mEq/l on the day of surgery (Correct ASA-PS: III).

Case 6. A 3-week-old male born at 35 weeks of gestational age presents for inguinal hernia repair. Birth weight was adequate for gestational age and he has no other comorbidities. (Correct ASA-PS: III).

Case 7. A 20-year-old man collided his car against a fixed bulkhead at a speed of 100 km/h. He weighs 78 kg (BMI 24 kg/m^2). His wife reports that he has no known comorbidities, does not use any medications, and does not smoke or consume alcohol. At that moment, he presents Glasgow Coma Scale 5 T, blood pressure 120/60 mmHg, and heart rate 95 beats/ min. Computed tomography scan of his head demonstrates a right epidural hematoma with mass effect and midline deviation. He is taken to the operating room for decompressive craniotomy (Correct ASA-PS: V-E).

Case 8. A 53-year-old woman presents for bilateral breast reduction. She has a past medical history significant for hypertension, GERD, and tobacco use. She weighs 80 kg (BMI 29 kg/m^2). Her blood pressure is routinely well controlled (~120/80 mmHg); however, this morning it was 154/99 mmHg. She walks 5 km three times a week and denies any chest pain or shortness of breath. Her GERD is well controlled on omeprazole. She admits to smoking two to three cigarettes a day, which is down from her usual half-pack a day; she has been smoking for the past 30 years. Her physical examination is within normal limits. (Correct ASA-PS: II).

Case 9. A 56-year-old woman presents for vaginal hysterectomy for uterine fibroids. She weighs 73 kg (BMI 28 kg/m^2). She has hypertension which is well controlled on metoprolol. She has a 20 pack-years smoking history but quit smoking 5 years ago and denies any respiratory symptoms/complaints. She was recently diagnosed with noninsulin-dependent type II diabetes mellitus. Her hemoglobin A1c is 10.5%, and her fasting blood glucose is 250 mg/dl (13.9 mmol/l). (Correct ASA-PS: III).

Case 10. An 18-year-old man is admitted for exploratory laparotomy for suspected appendicitis with a 6-day history of symptoms. He weighs 77 kg (BMI 22 kg/m^2). On admission to the operating room, he presents blood pressure 85/40 mmHg, heart rate 130 beats/min, temperature 38.5 °C and respiratory rate 22 breaths/min, after having received 2 l of Ringers lactate for volume expansion. (Correct ASA-PS: IV-E).

Only completed questionnaires were included in the final analysis. As we did not have access to the number of potential participants who were approached and initially agreed to participate and the number of participants who ultimately completed the questionnaire, the completion rate (i.e., ratio of users who actually finished the questionnaire to users who agreed to participate) was unavailable. This report is in compliance with the Checklist for Reporting Results of Internet *E*-Surveys (CHERRIES) guidelines [16].

STATISTICAL CONSIDERATIONS.

The minimum required sample size was calculated using the online software G*POWER (https://g-power.apponic.com) [17], based on the estimated accuracy of assigned scores for ASA-PS Classification before (5.7 ± 1.7) and after (7.7 ± 1.7) reviewing the ASA-PS definitions/examples, with a minimum difference of 20% [3]. For a power of 80%, effect size of 0.2 and a significance level of 5%, using a Wilcoxon matched-pairs signed-rank test, we estimated that 82 participants (anesthesiologists and non-anesthesiologists) per group were necessary.

Statistical analysis compared the difference between percentage of correct answers (accuracy) and self-reported confidence level (confidence). Over- and under-confidence were defined as a positive and negative difference between confidence and accuracy, respectively. The proportions of correct answers for each case with definitions and examples were determined and compared using McNemar's test for paired proportions. The median number of correct answers before and after reviewing the ASA-PS definitions and examples were compared using Wilcoxon matched-pairs signed-rank test.

Self-reported confidence was compared using Wilcoxon matched-pairs signed-rank test. The Mann-Whitney *U* test was used to compare differences between the median of accuracy versus the median of self-reported confidence in each group (anesthesiologists and non-anesthesiologists). The agreement between anesthesiologists and non-anesthesiologists was performed using weighted-kappa coefficient ranging between 0 and 1 [$\kappa > 0.80$ (almost perfect); $\kappa = 0.61$ to 0.80

(substantial); κ = 0.41 to 0.60 (moderate); κ = 0.21 to 0.40 (fair); κ < 0.20 (poor)]. *P*-value <0.05 represented statistical significance. Statistical analysis was performed using STATA 16.0 software (Stata Corp, College Station, TX). The research data related to this submission has been published in Mendeley Data (https://data.mendeley.com/datasets/6m689bchn4/1; DOI: 10.17632/6m689bchn4.1). The files associated with this dataset are licensed under an attribution non-commercial 3.0 Unported license (CC BY NC 3.0).

3. Results

A total of 386 physicians (272 anesthesiologists and 114 non-anesthesiologists) completed the entire twin sets of questionnaire. Most participants were < 40 years-old (70.1%), male (58.5%), and practiced at teaching hospitals (69.9%). Nearly half (47.7%) had ≤4 years, 20.5% had 5–9 years, and 31.8% had ≥10 years of experience. Among non-anesthesiologists, 49% and 51% were medical specialists and surgeons, respectively (Table 2).

The percentages of the different ASA-PS classes (I—V) assigned by participants (i.e., anesthesiologists and non-anesthesiologists) according to each hypothetical case before and after reviewing the ASA-PS definitions/examples are shown in Fig. 1. There was no improvement in accuracy among our entire cohort after (compared to before) reviewing the ASA-PS definitions/examples in 3 cases: Case 4 – elective endovascular infra-renal AAA repair (ASA-PS III); Case 5 – controlled hypertension, ESRD undergoing regularly scheduled dialysis (ASA-PS III); and Case 8 – controlled hypertension, tobacco use (ASA-PS II). A similar finding was observed for Case 9 – uncontrolled diabetes, tobacco use (ASA-PS III) among anesthesiologists, and Case 10 – acute abdomen for urgent exploratory laparotomy (ASA-PS IV-E) among non-anesthesiologists.

There was a significant decrease in accuracy among participants >60 years-old ($p = 0.001$) and those with >10 years of clinical experience ($p = 0.001$). For non-anesthesiologists, there was no difference in accuracy between medical specialists and surgeons before ($p = 0.79$) or after ($p = 0.67$) reviewing the ASA-PS definitions and examples. Improved accuracy was observed for participants practicing at teaching hospitals ($p = 0.05$) after reviewing definitions and examples. There was no improvement in accuracy ($p = 0.41$) among participants from different geographic regions in Brazil after reviewing definitions and examples (Supplementary Table 1).

For the purpose of constructing the graphs, we noted the accuracy of each individual assigned ASA-PS score matched to each self-reported confidence level. On further analysis of anesthesiologists vs. non-anesthesiologists, low kappa values were observed further confirming the inter-observer variability in the application of the ASA-PS by providers from different groups. Even after participants had reviewed the ASA-PS definitions/examples, poor agreement was observed in all cases (Supplementary Fig. 1).

Participants' accuracy for hypothetical ASA-PS I, II, and III cases involving adult patients was 70.1% before and 71% after exposure to the ASA-PS definitions/examples for anesthesiologists, and 51.7% and 70.8%, respectively, for non-anesthesiologists, whereas for cases ASA-PS IV, V and III (Case 6 involving a neonate) it was, respectively, 24.1% and 44% for anesthesiologists, and 22.2% and 51.9% for non-anesthesiologists. Ultimately, the overall participants' accuracy for hypothetical ASA-PS I, II, and III involving adult patients was greater than for ASA-PS IV, V and III (neonate) for both anesthesiologists and non-anesthesiologists both before and after reviewing the ASA-PS definitions/examples ($p < 0.001$) (Supplementary Fig. 2).

On initial assessment (i.e., prior to reviewing the ASA-PS definitions/examples), of the 272 anesthesiologists, 230 (84.6%) were overconfident, 38 (14.0%) were under-confident, and four (1.4%) were perfectly calibrated; whereas of the 114 non-anesthesiologists, 97 (85.1%) were overconfident, 16 (14.5%) were under-confident, and one (0.9%) was perfectly calibrated. When participants reported 100% confidence in their assigned ASA-PS, the actual observed accuracy before and after reviewing the ASA-PS definitions/examples was 58.9% and 73.8%, respectively, for anesthesiologists; and 38.2% and 60.7%, respectively, for non-anesthesiologists. Finally, disagreement was observed in the assigned ASA-PS between providers when evaluating the overall results (κ < 0.20) (Fig. 2; Supplementary Tables 3 and 4).

Fig. 3 shows participants' accuracy and self-reported confidence levels while assigning ASA-PS to the 10 hypothetical cases before and after reviewing the ASA-PS definitions/examples. The observed improvement in accuracy was unrelated to the studied groups (anesthesiologists vs. non-anesthesiologists). Overall, anesthesiologists scored higher than non-anesthesiologists both before [6(5–7) vs. 4(3–5), $P < 0.001$] and after [7(6–8) vs. 6(4–7), $p < 0.001$] reviewing the ASA-PS definitions/examples. Additionally, participants were poorly calibrated, i.e., their overall confidence levels were greater than their accuracy (i.e., overconfidence) for the observed results. The magnitude of overconfidence before and after reviewing the ASA-PS definitions/examples was 21.8% (95% CI, 15.5–27.3%) and 24.3% (95% CI, 19.5–29.7%), $p = 0.54$, respectively, for anesthesiologists; and 21% (95% CI, 14.6–29.4%) and 27.3% (95% CI, 19.9–36%), $p = 0.27$, for non-anesthesiologists.

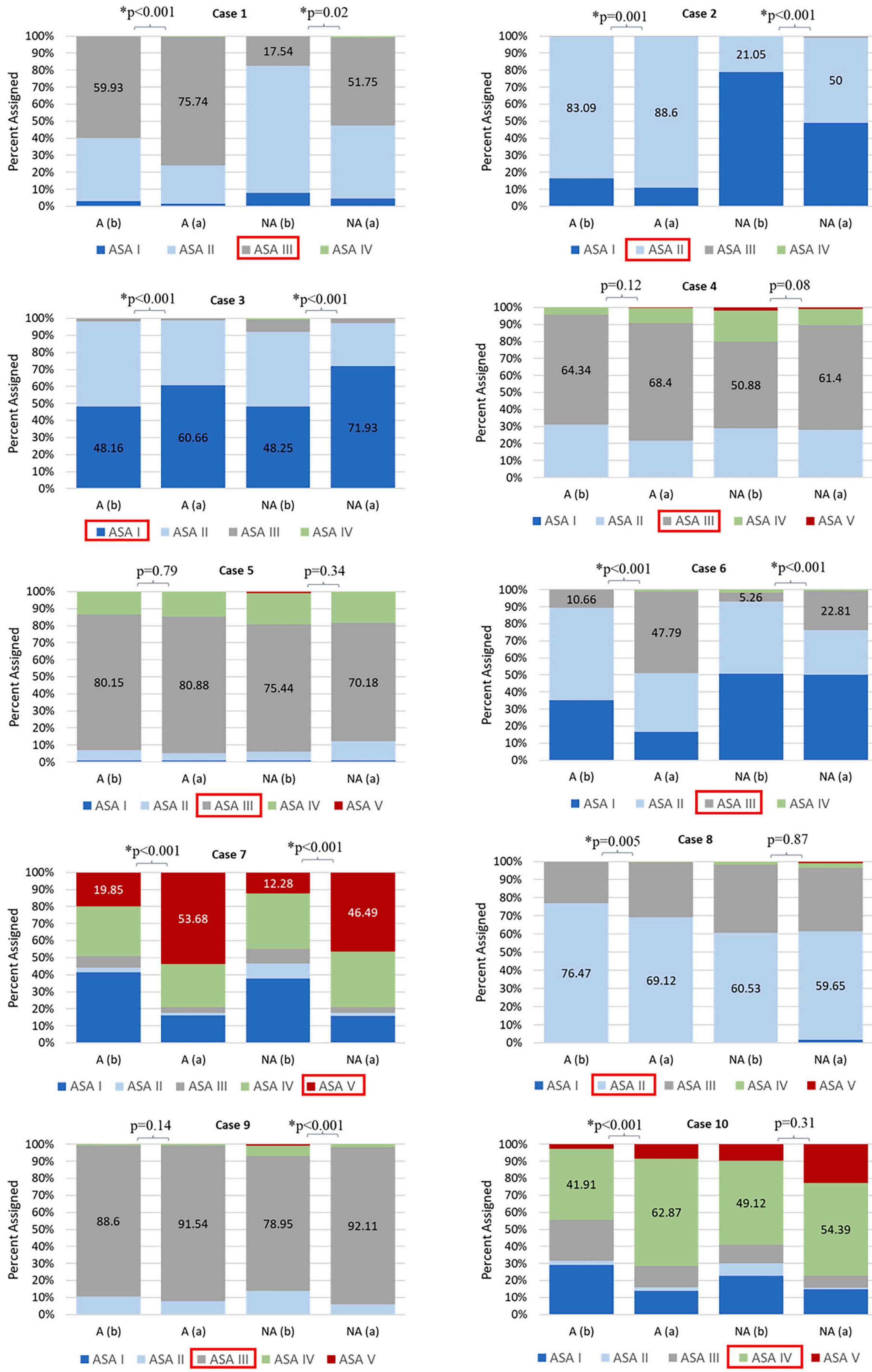
Table 2
Participants' characteristics.

Characteristics	n	%
Specialties		
Anesthesiologists	272	70.47
Surgeons	56	14.51
Clinicians	58	15.03
Gender		
Male	218	56.48
Female	168	43.52
Age (Years)		
<30	50	12.95
30–39	224	58.03
40–49	61	15.80
50–59	31	8.03
≥60	20	5.18
Professional experience		
≤4 years	184	47.67
5–9 years	79	20.47
≥10 years	123	31.87
Geographic Regions		
North	10	2.59
Northeast	48	12.44
Midwest	20	5.18
Southeast	276	71.50
South	32	8.29
Working in teaching hospitals		
Yes	270	69.95
No	116	30.05

4. Discussion

Our results confirmed our primary hypothesis, i.e., doctors (both anesthesiologists and non-anesthesiologists) are often overconfident while utilizing the ASA-PS Classification System. Indeed, overconfidence is common among physicians, medical students, and other healthcare professionals while performing a variety of tasks/activities [12–14,18–21]. Juslin et al. [14] summarized the data from numerous studies involving 29 samples of 2-alternative general knowledge questions and found a mean magnitude of overconfidence of 10%. Accordingly, our results corroborate this "baseline" overconfidence in medical decision-making, with an even higher magnitude of overconfidence for both anesthesiologists (21.8%) and non-anesthesiologists (24.3%) when assigning ASA-PS scores. Indeed, similar rates of overconfidence (26.4%) have been demonstrated across many anesthesiology-related areas such as perioperative neuromuscular monitoring [12].

Overall accuracy improved (along with increased self-reported



(caption on next page)

Fig. 1. Percentages of American Society of Anesthesiologists-Physical Status (ASA-PS) assigned by studied groups, i.e., anesthesiologists (A) and non-anesthesiologists (NA), according to each hypothetical case before (b) and after (a) reviewing the ASA-PS definitions and examples. The correct ASA-PS is highlighted (red box) on each case's legends. Proportion of correct answers before (b) and after (a) exposure to the ASA-PS definitions and examples; *p-value <0.05 (Non-parametric McNemar's test). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

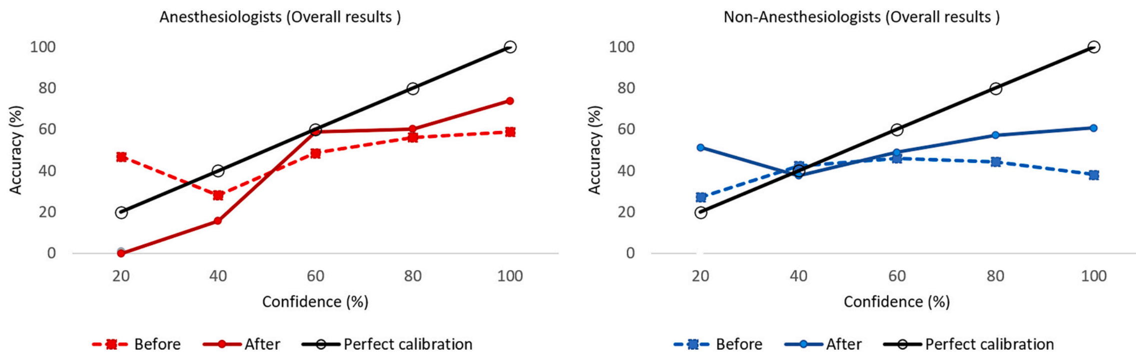


Fig. 2. Calibration curves for anesthesiologists and non-anesthesiologists before and after reviewing the ASA-PS definitions and examples. The black lines represent perfect calibration where accuracy values perfectly match confidence values (open circles). The red and blue lines represent participants' accuracy in comparison to their self-reported confidence levels in a 5-point scale, i.e., from 20 to 100% in 20% increments/intervals (filled circles/squares). The degree of agreement between the two measures was evaluated by the Kappa coefficient: Before (Kappa = 0.06, p = 0.07); After (Kappa = 0.11, p = 0.04). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

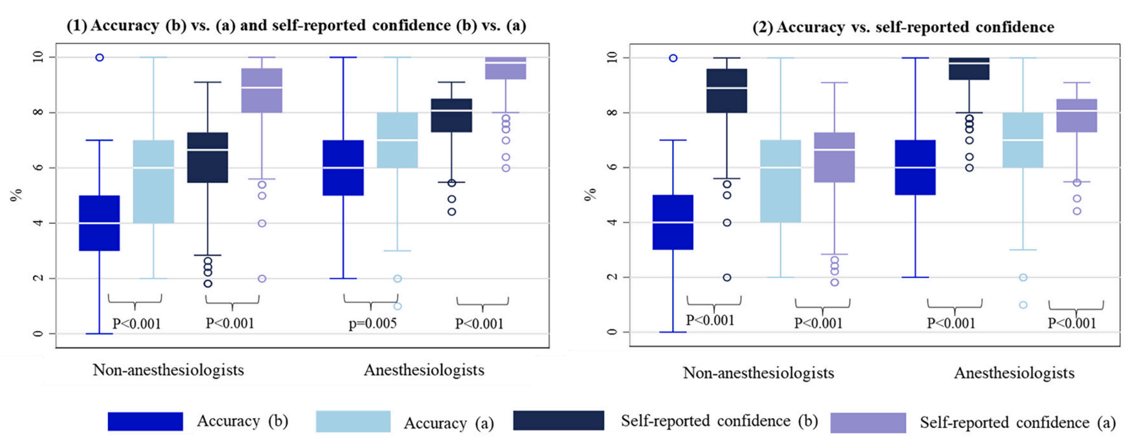


Fig. 3. Boxplot: Median (25–75% percentile) of anesthesiologists' and non-anesthesiologists' accuracy and self-reported confidence levels while assigning American Society of Anesthesiologists Physical Status (ASA-PS) to 10 hypothetical cases before and after reviewing the ASA-PS definitions and examples. The median marks the mid-point of the data and is shown by the line that divides the box into two parts. The upper and lower whiskers represent scores outside the middle 50% (i.e., the lower 25% of scores and the upper 25% of scores). Outliers are shown as circles. (1) Analysis between values of accuracy and self-reported confidence levels observed for anesthesiologists and non-anesthesiologists before (b) and after (a) reviewing ASA-PS definitions and examples (Wilcoxon matched-pairs signed-rank test). (2) Analysis between values of accuracy and self-reported confidence observed for anesthesiologists and non-anesthesiologists (Mann-Whitney U test).

confidence levels) after participants were exposed to the ASA-approved examples, which is in line with previous findings [3]. We therefore speculate that exposure to ASA-approved examples may have given participants the impression of better understanding the classification system, thereby increasing their susceptibility to overconfidence.

Calibration is not affected by level of expertise [15]. Accordingly, there was no significant difference in (over-)confidence levels by anesthesiologists (experts) compared to non-anesthesiologists (non-experts) across our entire (i.e., both before and after exposure do the ASA-PS definitions/examples) investigation.

Hurwitz et al. [3] asked participants to assign ASA-PS to 10 hypothetical cases (all being adult patients within ASA-PS I, II and, III) using (i) only the ASA-PS Classification System definitions and, subsequently, (ii) the (then newly) ASA-approved examples, and found that participants had >70% accuracy in 3 out of 10 (30%) cases prior to, and 7 out of 10 (70%) cases after being exposed to the ASA-approved examples. In comparison, the present investigation included cases from all ASA-PS classes (I-V), one of which being a neonate (Case 6).

Overall, participants' accuracy for ASA-PS I, II, and III involving adult patients was greater than for ASA-PS IV, V and III (the latter involving a neonate – Case 6) for both anesthesiologists and non-anesthesiologists (Supplementary Fig. 2). Lack of familiarity with neonatal anesthesia/physiology [22] and/or failure to understand the (less commonly used) ASA-PS definitions, i.e., classes IV and V, may have contributed to this finding. As for Case 7 (ASA-PS V-E), most participants surprisingly classified it as ASA-PS I-E. Indeed, prior to the trauma, the patient would correctly fit within the ASA-PS I classification, however, respondents clearly disregarded the morbidity associated with the patient's ongoing (moribund) condition (i.e., traumatic brain injury with ongoing intracranial hypertension) exposing him to imminent risk of death (ASA-PS V). Simply put, the indication for emergency surgery in Case 7 originates from the same life-threatening condition that puts this patient in the ASA-PS class VE. Caution therefore must be exercised to avoid such a systematic error (highlighted by our findings) of not taking into consideration the ongoing (physiologic/pathologic) processes leading to the indication of an emergency surgical intervention while assigning the ASA-PS.

Medical specialists and surgeons tend to disagree with (and downgrade) the ASA-PS classification in comparison with anesthesiologists [23]. In our studied cohort, anesthesiologists had better accuracy than non-anesthesiologists at the two analyzed moments (i.e., before and after reviewing the ASA-PS definitions/examples), whereas there was no demonstrable difference in accuracy between medical specialists and surgeons. As previously shown by Hurwitz et al. [3], exposing participants to the ASA-PS definitions/examples proved effective at improving accuracy (Table 2), especially for non-anesthesiologists whose median accuracy improved by 2 points, while anesthesiologists improved by 1 point.

Several studies have demonstrated significant inconsistency among anesthesiologists when it comes to ASA-PS classification agreement, with kappa values ranging from fair ($\kappa = 0{,}2$–04) [6,8,9,24–26], moderate ($\kappa = 0{,}4$–06) [7,27], good ($\kappa = 0{,}6$–08) [28], and excellent ($\kappa = 0{,}8$–1) [10]. While excellent agreement has been (rarely) demonstrated among anesthesiologists, the same investigators found poor agreement between internists and anesthesiologists [10]. In the present investigation, the general kappa between anesthesiologists and non-anesthesiologists was poor both before and after participants were exposed to the ASA-PS definitions/examples. This lack of agreement both among anesthesiologists as well as between anesthesiologists and non-anesthesiologists is likely in part due to (i) the inherent subjectivity of the classification system, (ii) factors related to education and professional experience, and (iii) institutional characteristics [10]. Interestingly, the only study that found an excellent agreement among anesthesiologists was conducted at a military institution [10].

Among general practitioners and general internists, knowledge level is inversely correlated with time since graduation, and higher among academic staff [29]. Research has also shown that patients of doctors further from medical school graduation have poorer outcomes, with a 4.5% increase in relative risk for mortality for each decade since graduation [30]. Against this background, our finding that there was a significant decrease in accuracy among respondents >60 years-old and with >10 years of work experience is interesting. In our study, since most of the participants were < 40 years-old (70.9%), with <10 years of practice since graduation (68.2%), and were affiliated to academic hospitals (69.9%), our results should be applied with caution to older professionals in general. A second limitation pertains to the fact that we only exposed participants to the ASA-PS definitions/examples, rather than being more encompassing with discussions as seen in previous assessments [10]. Nevertheless, our study design and results highlight the importance of using examples to improve practitioners' understanding of the ASA-PS Classification System. Thirdly, we did not collect data on participants' prior knowledge and clinical experience with the use of the ASA-PS Classification System; hence, it is plausible that several participants had already come across some of the hypothetical cases (adapted from Hurwitz et al. [3]) used in the current investigation, which could have biased our results. Finally, we were unable to estimate the total number of invited participants (due to the recruitment method used) and thus we cannot report on the response rate. Instead, we only had access to (and therefore analyzed) the responses for those who completed the questionnaire. It is possible, therefore, that exclusion of incomplete responses may have biased our findings. For example, it is conceivable that those with low confidence or incomplete knowledge might have started the questionnaire but declined or failed to complete it.

In conclusion, practitioners from different specialties (anesthesiologists, medical specialists, and surgeons) demonstrated overconfidence and disagreement while utilizing the ASA-PS Classification System. A brief exposure to the ASA-PS definitions and ASA-approved examples markedly improved the accuracy of practitioners from all fields. Overconfidence seems to be a common phenomenon in medical practice.

Contributors

SQS and LMS designed and supervised the study. LMS, RSFN, RFG coordinated the development of the project. LMS, SQS, FNB, GBM wrote the first draft of this manuscript. ACVA, LMS, SQS contributed to data acquisition and analysis. GBM, AMHH, DC, LMS contributed to drafting and revising the manuscript. All authors revised it critically and approved the final version for publication.

Patient consent for publication

Not required.

Funding statement

Support was provided solely from institutional and/or departmental sources. No external funding was obtained for carrying out this investigation.

Authorship statement

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript. Furthermore, each author certifies that this material or similar material has not been and will not be submitted to or published in any other publication before its appearance in Journal of Clinical Anesthesia.

Authorship contributions

SQS and LMS designed and supervised the study. LMS, RSFN, RFG coordinated the development of the project. LMS, SQS, FNB, GBM wrote the first draft of this manuscript. ACVA, LMS, SQS contributed to data acquisition and analysis. GBM, AMHH, DC, LMS contributed to drafting and revising the manuscript. All authors revised it critically and approved the final version for publication.

Ethics statements

This investigation was approved by the Institutional Research Ethics Committee (Protocol # 4,518,843; CAAE: 42047321.8.0000.0087). Patient consent for publication not required.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

CMA anesthesia team for providing us with the opportunity to pilot this initiative.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinane.2022.110794.

References

[1] Saklad M. Grading of patients for surgical procedures. Anesthesiology 1941;2:281–4.
[2] Abouleish AE, Leib ML, Cohen NH. ASA provides examples to each ASA physical status class. ASA Newsl 2015;79:38–9.

[3] Hurwitz EE, Simon M, Vinta SR, Zehm CF, Shabot SM, Minhajuddin A, et al. Adding examples to the ASA-physical status classification improves correct assignment to patients. Anesthesiology 2017;126:614–22. https://doi.org/10.1097/ALN.0000000000001541.
[4] Skaga NO, Eken T, Søvik S, Jones JM, Steen PA. Pre-injury ASA physical status classification is an independent predictor of mortality after trauma. J Trauma 2007;63:972–8. https://doi.org/10.1097/TA.0b013e31804a571c.
[5] Glance LG, Lustik SJ, Hannan EL, Osler TM, Mukamel DB, Qian F, 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. https://doi.org/10.1097/SLA.0b013e31824b45af.
[6] Owens WD, Felts JA, Spitznagel Jr EL. ASA physical status classifications: a study of consistency of ratings. Anesthesiology 1978;49(4):239–43.
[7] Cuvillon P, Nouvellon E, Marret E, Albaladejo P, Fortier LP, Fabbro-Perray P, et al. American Society of Anesthesiologists' physical status system: a multicentre francophone study to analyse reasons for classification disagreement. Eur J Anaesthesiol 2011;28:742–7. https://doi.org/10.1097/EJA.0b013e328348fc9d.
[8] Mak PH, Campbell RC, Irwin MG. American Society of Anesthesiologists. The ASA physical status classification: inter-observer consistency. American Society of Anesthesiologists. Anaesth Intensive Care 2002;30:633–40. https://doi.org/10.1177/0310057X0203000516.
[9] Riley R, Holman C, Fletcher D. Inter-rater reliability of the ASA physical status classification in a sample of anaesthetists in Western Australia. Anaesth Intensive Care 2014;42:614–8. https://doi.org/10.1177/0310057X1404200511.
[10] Knuf KM, Manohar CM, Cummings AK. Addressing inter-rater variability in the ASA-PS classification system. Mil Med 2020;185:e545–9. https://doi.org/10.1093/milmed/usz433.
[11] ASA Physical Status Classification System, original approval – October 15, 2014. Available at: https://www.asahq.org/standards-and-guidelines/asa-physical-status-classification-system. (Accessed October 3, 2021).
[12] Naguib M, Brull SJ, Hunter JM, Kopman AF, Fülesdi B, Johnson KB, et al. Anesthesiologists' overconfidence in their perceived knowledge of neuromuscular monitoring and its relevance to all aspects of medical practice: an international survey. Anesth Analg 2019;128:1118–26. https://doi.org/10.1213/ANE.0000000000003714.
[13] Arkes HR, Dawson NV, Speroff T, Harrell Jr FE, Alzola C, Phillips R, et al. The covariance decomposition of the probability score and its use in evaluating prognostic estimates. SUPPORT Investigators Med Decis Making 1995;15:120–31. https://doi.org/10.1177/0272989X9501500204.
[14] Juslin P, Winman A, Olsson H. Naive empiricism and dogmatism in confidence research: a critical examination of the hard-easy effect. Psychol Rev 2000;107:384–96. https://doi.org/10.1037/0033-295x.107.2.384.
[15] Lichtenstein S, Fischhoff B. Do those who know more also know more about how much they know? Organ Behav Hum Perform 1977;20:159–83. https://doi.org/10.1016/0030-5073(77)90001-0.
[16] Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet E-surveys (CHERRIES). J Med Internet Res 2012;14:e8. https://doi.org/10.2196/jmir.6.3.e34.
[17] Faul F, Erdfelder E, Lang A-G, Buchner A. G*power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007;39:175–91. https://doi.org/10.3758/bf03193146.
[18] Barnsley L, Lyon PM, Ralston SJ, Hibbert EJ, Cunningham I, Gordon FC, et al. Clinical skills in junior medical officers: a comparison of self- reported confidence and observed competence. Med Educ 2004;38:358–67. https://doi.org/10.1046/j.1365-2923.2004.01773.x.
[19] Hodges B, Regehr G, Martin D. Difficulties in recognizing one's own incompetence: novice physicians who are unskilled and unaware of it. Acad Med 2001;76:S87–9. https://doi.org/10.1097/00001888-200110001-00029.
[20] Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. JAMA 2006;296:1094–102. https://doi.org/10.1001/jama.296.9.1094.
[21] Gordon MJ. A review of the validity and accuracy of self assessments in health professions training. Acad Med 1991;66:762–9. https://doi.org/10.1097/00001888-199112000-00012.
[22] Drzymalski DM, Gao W, Moss DR, Liao BT, Shore-Lesserson L, Podovei M. Factors associated with neonatal resuscitation knowledge and comfort across academic anesthesia institutions. J Matern Fetal Neonatal Med 2020;9:1–7. https://doi.org/10.1080/14767058.2020.1843018.
[23] Curatolo C, Goldberg A, Maerz D, Lin HM, Shah H, Trinh M. ASA physical status assignment by non-anesthesia providers: do surgeons consistently downgrade the ASA score preoperatively? J Clin Anesth 2017;38:123–8. https://doi.org/10.1016/j.jclinane.2017.02.002.
[24] Haynes SR, Lawler PGP. An assessment of the consistency of ASA physical status classification allocation. Anaesthesia 1995;50:195–9. https://doi.org/10.1111/j.1365-2044.1995.tb04554.x.
[25] Ranta S, Hynynen 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.
[26] Aronson WL, McAuliffe MS, Miller K. Variability in the American Society of Anesthesiologists Physical Status classification scale. AANA J 2003;71:265–74.
[27] Malviya S, Voepel-Lewis T, Chiravuri SD, Gibbons K, Chimbira WT, Nafiu OO, et al. Does an objective system-based approach improve assessment of perioperative risk in children? A preliminary evaluation of the 'NARCO'. Br J Anaesth 2011;106:352–8. https://doi.org/10.1093/bja/aeq398.
[28] Sankar A, Johnson SR, Beattie WS, Tait G, Wijeysundera DN. Reliability of the American Society of Anesthesiologists physical status scale in clinical practice. Br J Anaesth 2014;113:424–32. https://doi.org/10.1093/bja/aeu100.
[29] Zupanic M, Kreuer J, Bauer D, Nouns ZM, Ehlers JP, Fischer MR. Spontaneously retrievable knowledge of German general practitioners depending on time since graduation, measured with the progress test medicine. GMS. J Med Educ 2020;37:Doc49. https://doi.org/10.3205/zma001342.
[30] Norcini JJ, Boulet JR, Opalek A, Dauphinee WD. Patients of doctors further from medical school graduation have poorer outcomes. Med Educ 2017;51:480–9. https://doi.org/10.1111/medu.13276.