











Inter-rater variability for the American Society of Anesthesiologists classification in patients undergoing hepato-pancreato-biliary surgery (MILESTONE-2): international survey among surgeons and anaesthesiologists

Simone Augustinus^{1,2,*} , Jasper P. Sijberden^{1,2,3}, Matthanja Bieze^{4,5}, Vandana Agarwal⁶, Luca A. Aldrighetti⁷ , Adnan Alseidi⁸, Francisco C. Bonofiglio⁹, Kevin C. P. Conlon¹⁰, Katia Donadello¹¹, Joris Erdmann^{1,2} , Cristina Ferrone¹², Michael Guertin¹³, Ronald Harter¹³, Maria E. Franceschetti¹⁴, Giuseppe K. Fusai¹⁵, Bas Groot Koerkamp¹⁶ , Thilo Hackert¹⁷ , Jin-Young Jang¹⁸, Thomas Kander¹⁹, Tobias Keck²⁰, Dominik Krzanicki²¹, Ho-Jin Lee²², Keith Lewis²³, Giuseppe Natalini¹⁴, Carla Nau²⁴, Timothy M. Pawlik²⁵ , Henry A. Pitt²⁶, Rafaella Reineke²⁷, Roberto Salvia²⁸, Eduardo de Santibanes²⁹, Shailesh V. Shrikhande³⁰, Martin Smith³¹, Attila Szijarto³², Bobby Tingstedt³³ , Alice C. Wei³⁴ , John Windsor³⁵ , Mohammed Abu Hilal³, Manuel Pardo³⁶, Markus W. Hollmann⁴, and Marc G. Besselink^{1,2,*} ; for the MILESTONE Study Group

¹Department of Surgery, Amsterdam UMC, Location University of Amsterdam, Amsterdam, The Netherlands

²Cancer Center Amsterdam, Amsterdam, The Netherlands

³Department of Surgery, Fondazione Poliambulanza Istituto Ospedaliero, Brescia, Italy

⁴Department of Anesthesiology, Amsterdam UMC, Location University of Amsterdam, Amsterdam, The Netherlands

⁵Department of Anesthesiology and Pain Management, Toronto General Hospital, University of Toronto, Toronto, Ontario, Canada

⁶Department of Anesthesia, Critical Care and Pain, Tata Memorial Hospital, Homi Bhabha National Institute, Mumbai, Maharashtra, India

⁷Hepatobiliary Surgery Division, Ospedale San Raffaele, Milano, Italy

⁸Division of Hepatopancreatobiliary and Endocrine Surgery, University of California, San Francisco, California, USA

⁹Department of Anesthesiology, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina

¹⁰Department of Surgery, Trinity College Dublin, Tallaght University Hospital, Dublin, Ireland

¹¹Department of Anesthesia and Intensive Care B, DSCOMI, University of Verona, University Hospital Integrated Trust of Verona, Verona, Italy

¹²Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA

¹³Department of Anesthesiology, The Ohio State University Wexner Medical Center & College of Medicine, Columbus, Ohio, USA

¹⁴Department of Anesthesia and Intensive Care, Istituto Fondazione Poliambulanza, Brescia, Italy

¹⁵Hepatobiliary Surgery and Liver Transplantation Unit, Royal Free Hospital, London, UK

¹⁶Department of Surgery, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

¹⁷Department of General, Visceral and Thoracic Surgery, University Hospital Hamburg-Eppendorf, Germany

¹⁸Department of Surgery and Cancer Research Institute, Seoul National University College of Medicine, Seoul, South Korea

¹⁹Department of Intensive and Perioperative Care, Skåne University Hospital, Lund and Lund University, Lund, Sweden

²⁰DGAV StuDoQ|Pancreas and Clinic of Surgery, UKSH Campus Lübeck, Lübeck, Germany

²¹Department of Anaesthesia, Royal Free Hospital, London, UK

²²Department of Anesthesiology and Pain Medicine, Seoul National University Hospital, Seoul, South Korea

²³Department of Anesthesiology and Perioperative Medicine, Rutgers Robert Wood Johnson Medical School, New Brunswick, New Jersey, USA

²⁴Department of Anaesthesiology and Intensive Care, University Medical Centre Schleswig-Holstein, Lübeck, Germany

²⁵Department of Surgery, The Ohio State University Wexner Medical Center, Columbus, Ohio, USA

²⁶Department of Surgery, Rutgers Cancer Institute of New Jersey, New Brunswick, New Jersey, USA

²⁷Department of Anesthesiology and Intensive Care, IRCCS San Raffaele Scientific Institute, Milan, Italy

²⁸General and Pancreatic Surgery Department, Pancreas Institute, University and Hospital Trust of Verona, Verona, Italy

²⁹Department of Surgery, Hospital Italiano, University of Buenos Aires, Buenos Aires, Argentina

³⁰Department of GI and HPB Surgery, Tata Memorial Centre, Homi Bhabha National Institute, Mumbai, India

³¹Department of Surgery, Faculty of Health Sciences, School of Clinical Medicine, University of the Witwatersrand, Johannesburg, South Africa

³²Department of Anesthesiology, I. sz. Sebészeti Klinika, Semmelweis Egyetem, Budapest, Hungary

³³Department of Surgery, Clinical Sciences Lund, Lund University, Skåne University Hospital, Lund, Sweden

³⁴Hepatopancreatobiliary Service, Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, New York, USA

³⁵HBP/Upper GI Unit, Auckland City Hospital/Department of Surgery, University of Auckland, New Zealand

³⁶Department of Anesthesia and Perioperative Care, University of California, San Francisco, California, USA

*Correspondence to: Marc G. Besselink, MD, Department of Surgery, Amsterdam UMC, University of Amsterdam, Cancer Center Amsterdam, De Boelelaan 1117 (ZH-7F), 1081 HV Amsterdam, The Netherlands (e-mail: m.g.besselink@amsterdamUMC.nl; s.augustinus@amsterdamUMC.nl).

Abstract

Background: Patients undergoing hepato-pancreato-biliary surgery are typically preoperatively assessed using the American Society of Anesthesiologists (ASA) classification, which is also used for case-mix adjustment when comparing centre outcomes. Studies determining the inter-rater variability of the ASA classification within hepato-pancreato-biliary surgery are currently lacking.

Methods: An international survey was collected and a case-vignette study was performed (November 2022–April 2023) regarding the ASA classification in patients undergoing hepato-pancreato-biliary surgery among anaesthesiologists and surgeons from (inter)national societies. The survey consisted of 23 questions and eight case-vignettes. Primary analysis included descriptive statistics and the inter-rater variability was calculated using Light's Kappa.

Results: Overall, 1283 participants from 55 countries responded: 1073 (84%) anaesthesiologists and 210 (16%) surgeons. The ASA classification was commonly used, both clinically 1003/1283 (78%) and for research 728/762 (96%). The majority of respondents ($n = 1019$, 79%) declared that ASA score impacted their perioperative strategy. There inter-rater variability was fair–moderate (Kappa 0.26–0.42) in all case-vignettes. Inter-rater variability differed within and among geographic regions for each case. Over 80% ($n = 1138$) of respondents stated that they would take the underlying disease (for example cancer) into account, but this changed the preferred ASA score within the case-vignettes by only 1%. Type of surgery changed the preferred score in the case-vignettes (13% difference). The most common suggestions to improve the ASA classification were to clarify whether type of operation should be considered, create a more extensive definition, and provide more examples.

Conclusions: Inter-rater variability was present within the ASA classification of patients undergoing hepato-pancreato-biliary surgery, which may impact perioperative strategy and hamper research results. Additional guidance to classify patients according to ASA is urgently needed. Until then, more objective measurements should be considered for case-mix adjustment within research.

Introduction

For over 60 years, the American Society of Anesthesiologists (ASA) physical status classification system has been used to facilitate bedside assessment and categorize patients' co-morbidities prior to anaesthesia¹. Combined with other factors (for example type of surgery, frailty, level of deconditioning) the classification can be used to predict perioperative risks. Nevertheless, considerable variability in ASA scoring among anaesthesiologists has been recognized. Several case-vignette studies addressing common surgical procedures in individual countries have demonstrated significant inter-rater variability^{2–5}.

Hepato-pancreato-biliary (HPB) surgery is one of the most complex types of abdominal surgery and is associated with considerable perioperative risks^{6,7}. A recent study identified a large variability in ASA classification among four large registries for pancreatic surgery; classification as ASA III/IV varied from 78% (North America), 48% (Germany), 23% (the Netherlands), to 3% (Sweden) of patients⁸. As large differences in patients' co-morbidities among these four countries are highly unlikely, this large discrepancy could partly arise from a different interpretation of the ASA classification between countries. For example, whether the type of surgery (for example more or less complex) or underlying disease (for example malignancy) is accounted for when determining the ASA score. Alternatively, other variables might explain the different ratings in ASA classification (for example reimbursement policies, experience within a centre). It is of importance to understand the reasons behind these differences as ASA score is increasingly used for case-mix adjustment when comparing surgical outcomes among centres in clinical auditing^{9–12}.

Data on the reliability and inter-rater variability of the ASA classification among patients undergoing HPB surgery are lacking. The first MILESTONE survey, conducted among surgeons and anaesthesiologists, assessed worldwide variety in fluid and pain management in liver surgery¹³. This second MILESTONE survey among anaesthesiologists and surgeons aims to define international inter-rater variability, as well as characterize potential explanations for any observed variability among patients undergoing HPB surgery. Such data will hopefully help to optimize the ASA classification and provide context for ASA-related benchmarking of postoperative outcomes.

Methods

Study design

An online survey including clinical case-vignettes was created by the multidisciplinary MILESTONE study group and followed the AAPOR best practices for survey research¹⁴. The Medical Ethical Committee of the Amsterdam UMC confirmed that the Medical Research Involving Human Subjects Act (WMO) did not apply to the study. Participants were advised that they provided informed consent to the use of their anonymized responses by completing and submitting the survey.

Population

Multiple national and international societies were invited via email to participate and disseminate the survey among their members. For anaesthesiologists, 133 national anaesthesiology societies associated with the World Federation of Societies of Anaesthesiologists (WFSA) and the Liver Intensive Care Group of Europe (LICAGE) were invited. For surgeons, the International Hepato-Pancreato-Biliary Association (IHPBA) and its three regional chapters, four minimally invasive HPB surgery societies/groups, and three regional HPB societies were asked to participate. A full list of invited societies can be found in [Supplementary Material S1](#). The co-authors were asked to distribute the survey online in their professional networks.

Survey distribution

The survey was created by the MILESTONE study group (that is all authors), a multidisciplinary group of experts on perioperative care for HPB patients. The survey was tested by the first and senior authors by completing the survey and checking for completeness and potential flaws before sending it to the societies. When societies wished to collaborate, the survey link was disseminated by email to the members between November 2022 and April 2023. Data were collected anonymously using Qualtrics XM survey software® (Qualtrics, Provo, Utah, USA).

Data collection

The survey gathered demographic information on country, sex, age, medical specialty, years of experience, and level of expertise present in the hospital. Subsequent questions included six general questions on the ASA classification, nine questions on considerations made in the ASA classification process, and

eight clinical cases (full survey: [Supplementary Material S2](#)). Within the clinical cases, four patients undergoing pancreatoduodenectomy (PD) and four patients undergoing extended left hemihepatectomy without biliary reconstruction were described. In each case a single aspect changed, which was highlighted in bold ([Supplementary Material S2](#)). The respondents were asked to score the ASA classification for each case (that is ASA I–V). Within the MILESTONE study group consisting of international experts, the aspects that were tested in the case-vignettes were based on well-known points of discussion from clinical practice. Some of these were addressed both within the case-vignettes and within the general questions, to confirm whether the respondents' opinion on how the classification should be interpreted are in line with how they classify the patients within the case-vignettes. An open question was added to determine if some discussion points were missed. Points of discussion that could not be incorporated within a clinical case were incorporated in the general questionnaire (that is financial compensation, experience in a peripheral/academic centre). No 'correct ASA scores' per clinical case were described, as there is no consensus yet within the ASA classification on the points up for discussion.

Statistical analysis

Free text was recoded into common terminology while cross-checking for entry errors by the two first authors (S.A., J.S.). For binary or categorical variables, the results were reported as counts and proportions. For continuous variables, the results were reported as the mean with standard deviation when normally distributed, or as median with interquartile range when not-normally distributed. Normality was assessed by visual inspection of histograms. Differences in baseline characteristics were assessed using the chi-square test or Fisher's exact test for categorical variables and the Mann–Whitney U test, or Student's t test for continuous variables when appropriate.

The primary analysis included descriptive statistics. The preferred ASA score within a case was defined as the ASA score that was chosen by the highest percentage of participants. To determine differences among regions, and among anaesthesiologists and surgeons, chi-square and Fisher's exact tests were used as appropriate. To determine the clinical relevance, the absolute largest difference (ALD) within the preferred scores among regions and anaesthesiologists and surgeons was evaluated. A $\geq 10\%$ difference among raters was considered clinically relevant. The inter-rater variability was calculated using Light's Kappa (Kappa for more than two raters) and defined as the K-value. Inter-rater variability describes the degree of agreement among independent observers who rate or assess the same concept. Interpretation of the K-value was based on the Kappa value interpretation as determined by Landis and Koch: poor (≤ 0), slight (0.01–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), or perfect agreement (0.81–1.00)¹⁵. A two-tailed P of < 0.05 was considered statistically significant. Results were analysed using RStudio (version 4.2.1).

Results

Overall, 17 of 133 invited anaesthesiology societies (12.8%) and 9 of 10 invited surgical societies (90%) disseminated the survey to their members, including the ASA. In total, 1729 participants responded. Following exclusion of incomplete responses, 1283 participants (74.2%) were included in the analytic cohort. Respondents worked in 55 different countries: 601 in North America (47%), 423

in Europe (33%), 150 in the Asia-Pacific (12%), 96 in South America (7%), and 11 in Africa (0.9%). Respondents included 1073 anaesthesiologists (84%) among whom the majority worked in a general anaesthesiology unit with an interest in HPB ($n = 588$; 46%); among the 210 surgeons (16%), the majority worked in a dedicated HPB unit without liver transplant surgery ($n = 89$; 42%). Baseline characteristics are reported in [Table 1](#).

Practical use

Most respondents used the ASA classification ($n = 1003$, 78%), and some of these used additional prediction tools to determine patients' perioperative risks ($n = 255$; 20%). When queried about their perception of the ASA classification as an objective system, most respondents ($n = 362$, 28%) provided a rating of 7/10 ([Fig. 1](#)). The median score of all respondents was 6.0 (i.q.r. 5.0–7.0, [Fig. 1](#)), which was higher among anaesthesiologists versus surgeons (7.0 versus 6.0, $P < 0.001$). Seventy-five percent ($n = 1019$) of respondents reported that patient ASA classification could lead to a change in the perioperative strategy (anaesthesiologists, 73% ($n = 845$) versus surgeons, 86% ($n = 174$; [Fig. 2](#)).

A variance in the additional prediction tools employed was observed (full list: [Supplementary Material S3](#)). The most frequently used scores were the STOP-BANG score (Sleep Apnea)¹⁶, Revised Cardiac Risk Index (Lee Criteria)¹⁷, MET score (functional capacity)¹⁸, Frailty Index¹⁹, and the ACS NSQIP surgical risk score²⁰. In addition, clinical assessment tools such as medical history and functional tests were used. Among the 2% ($n = 57$) of respondents who did not use the ASA score, most reported that they solely used a patient's medical history ([Supplementary Material S4](#)). Anaesthesiology departments scored the ASA classification ($n = 1272$, 99%), with the assessment performed by a medical specialist ($n = 943$, 73%), a resident ($n = 213$, 17%), a specialized nurse ($n = 23$, 2%), or all the above equally ($n = 104$, 8%). Among respondents involved in research ($n = 762$), the majority 'always' used the ASA classification within research ($n = 369$, 48%), followed by 'often' ($n = 212$, 28%), 'sometimes' ($n = 147$, 29%), and 'never' ($n = 34$, 4%).

Variables considered for ASA classification

When classifying patients according to ASA, most respondents ($n = 1054$, 82%) did not take the type of operation (complex versus less complex) and the complexity of anaesthetic management (for example fibre optic intubation, $n = 981$, 76%) into account. In contrast, nearly all respondents ($n = 1136$, 89%) considered cancer as a co-morbidity within scoring, particularly when the clinical condition of the patient was impaired ([Fig. 3](#)). A minority of the participants believed that academic centres score ASA lower ($n = 329$, 25%) and non-academic centres score ASA higher ($n = 280$, 22%) ([Fig. 4](#)). This was slightly more prevalent in the 499 participants working in an anaesthesiology-dedicated HPB unit ($n = 145$, 32% believe academic centres score lower, and $n = 126$, 28% believe peripheral centres score higher, $P < 0.001$) than in 588 participants working in a general anaesthesiology unit ($n = 134$, 23% believe academic centres score lower, and $n = 115$, 20% believe peripheral centres score higher, $P < 0.001$). Half the participants stated that the ASA classification should be used for financial reimbursement from health insurance companies ([Fig. 4](#)). No consensus was observed on the highest percentage of disagreement among ASA score that would be acceptable in clinical practice (for example 'less than 10%' $n = 507$ (40%) versus 'less than 20%' $n = 404$ (30%) versus 'less than 5%' $n = 269$ (21%)). The five most-reported suggestions to reduce undesired variability in the ASA classification and reflect the true perioperative risk were as follows: take type of operation into

Table 1 Baseline characteristics

	Total (n = 1283)	Anaesthesiologist (n = 1073)	Surgeon (n = 210)	P
Sex				<0.001
Female	409 (32)	384 (36)	25 (12)	
Male	867 (68)	683 (64)	184 (88)	
Non-binary/third gender	7 (0.5)	6 (0.6)	1 (0.5)	
Age, years, median (i.q.r.)	48.0 (40.0–57.0)	48.0 (40.0–58.0)	49.5 (41.0–56.0)	0.724
Continent				<0.001
Africa	11 (0.8)	10 (0.9)	1 (0.5)	
Asia-Pacific	150 (11.7)	105 (9.8)	45 (21)	
Europe	423 (32.9)	280 (26)	143 (68)	
North America	601 (46.8)	586 (55)	15 (7)	
South America	96 (7.5)	91 (8.5)	5 (2)	
Unknown	2 (0.1)	1 (<0.1)	1 (0.5)	
Hospital type				<0.001
Academic (central)	685 (53.4)	509 (47)	176 (84)	
Non-academic, teaching	311 (24.2)	279 (26)	8 (4)	
Non-academic, non-teaching	287 (22.4)	285 (27)	26 (12)	
Years registered as specialist, median (i.q.r.)	17.0 (9.0–26.0)	17.0 (9.0–26.0)	19.5 (10.0–25.0)	0.535
Level of HPB surgery expertise				<0.001
Dedicated HPB unit, no transplant	273 (21.3)	184 (17)	89 (42)	
Dedicated HPB unit, transplant	343 (26.7)	264 (25)	79 (38)	
General surgery, interest HPB	420 (32.7)	380 (35)	40 (19)	
No HPB surgery	233 (18.1)	231 (22)	2 (1)	
Unknown	14 (1.1)	14 (1.3)	0 (0)	
Level of HPB anaesthesiology expertise				<0.001
Dedicated HPB unit, no transplant	116 (9.0)	87 (8)	29 (14)	
Dedicated HPB unit, transplant	333 (25.9)	259 (24)	74 (35)	
General anaesthesiology, interest HPB	588 (46)	493 (46)	95 (45)	
No HPB surgeries performed	225 (17.5)	220 (21)	5 (2)	
Unknown	21 (1.6)	14 (1.3)	7 (3)	

Values are n (%) unless indicated otherwise. Bold numbers indicate statistical significance. i.q.r., interquartile range; HPB, hepato-pancreato-biliary.

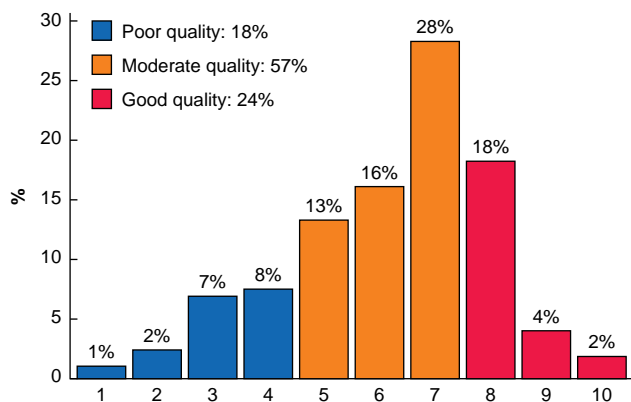


Fig. 1 Quality of American Society of Anesthesiologists (ASA) classification according to respondents

Score 1–10 for quality of the ASA classification. 1: I consider ASA a poor classification system, non-objective with very poor interobserver agreement, 10: I consider ASA an excellent classification system, highly objective and excellent interobserver agreement.

account ($n = 145$, 12.5%), more extensive ASA definition ($n = 137$, 11.9%), more examples ($n = 94$, 8.1%), take performance status into account ($n = 77$, 6.7%), and a further ASA subdivision ($n = 73$, 6.3%) (Table 2). Among all respondents, 1156 (90%) mentioned at least one factor that should be added to the definition of ASA classification (Supplementary Material S5).

Clinical case-vignettes

The most commonly chosen option across all case-vignettes was ASA III or IV (Table 2). Within the pancreatoduodenectomy case-vignettes, the most commonly chosen option was ASA II. However, it became ASA III when either the patient had coronary artery bypass grafting (CABG) 2 years prior (case 2) or

the patient's condition did not recover after biliary stenting (case 3). Changing the age (83 instead of 55 years old, case 4) did not make a difference, as the preferred ASA score remained II. For all case-vignettes that included a patient undergoing an extended left hemihepatectomy, the preferred ASA score was III, irrespective of diabetes mellitus, robotic instead of open procedure, or a benign diagnosis.

The inter-rater variability was fair/moderate within all clinical cases (Kappa range 0.26–0.42, Table 3). Six Kappa values all corresponded to 'fair agreement', and in case 2 'moderate agreement' was reached (Kappa 0.42). The case in which moderate agreement was reached described a patient who underwent pancreatoduodenectomy and had severe co-morbidities (CABG 2 years prior).

A clinically relevant difference ($\geq 10\%$) in ASA score among regions was noted in all case-vignettes (ALD range: 12–36%). Among North American respondents, in 8/8 (100%) cases the most commonly chosen ASA score was III. ASA score III was most commonly chosen in 3/8 (38%) cases in the Asia-Pacific, 4/8 (50%) in Europe, 4/8 (50%) in South America (50%), and 7/8 (88%) in Africa; the rest of the cases were scored ASA II (Supplementary Material S6). Significant inter-rater variability within regions was detected as well (Supplementary Material S8). A clinically relevant difference among anaesthesiologists and surgeons was also noted for 6/8 cases (ALD range: 23–30%, Supplementary Material S7). This finding was noted in all four of the hemihepatectomy case-vignettes. Surgeons scored lower (ASA II instead of ASA III).

Discussion

This international survey and case-vignette study identified a considerable inter-rater variability regarding the ASA

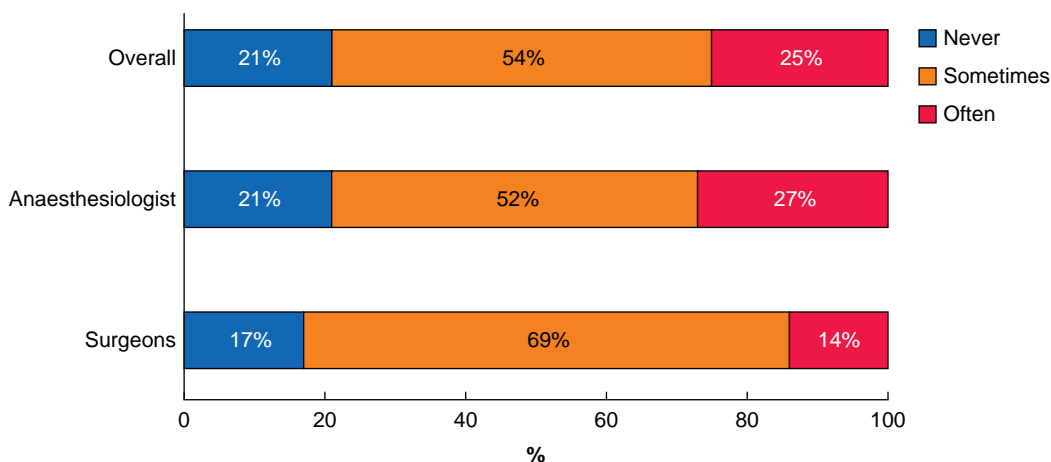


Fig. 2 Does the American Society of Anesthesiologists (ASA) score assigned to a patient change your perioperative strategy?

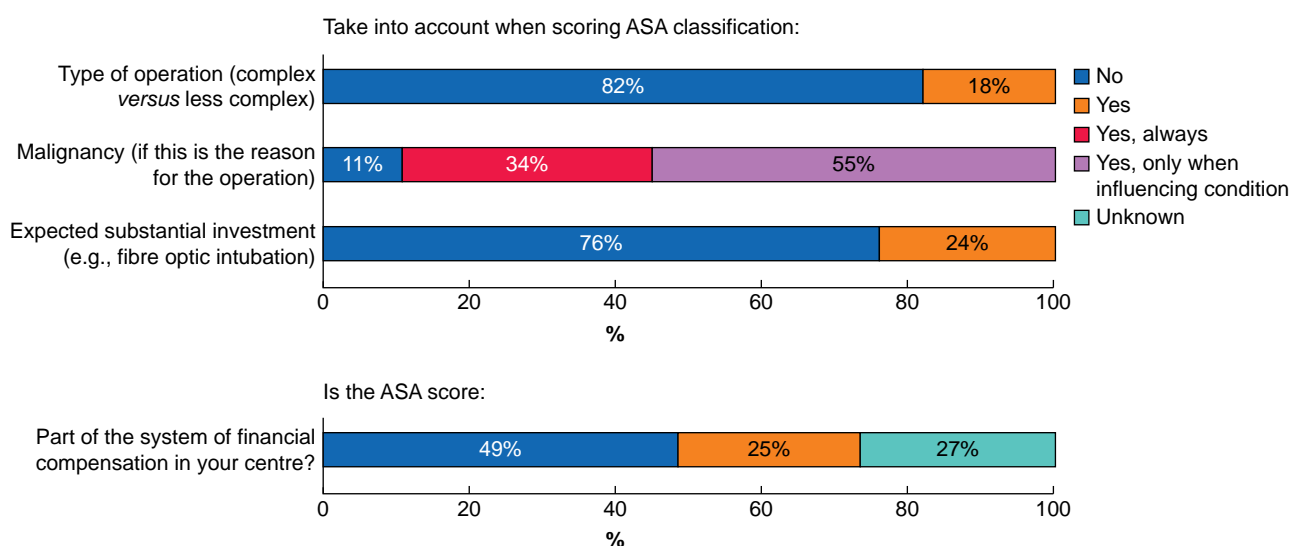


Fig. 3 Considerations made when classifying patients according to American Society of Anesthesiologists (ASA) score

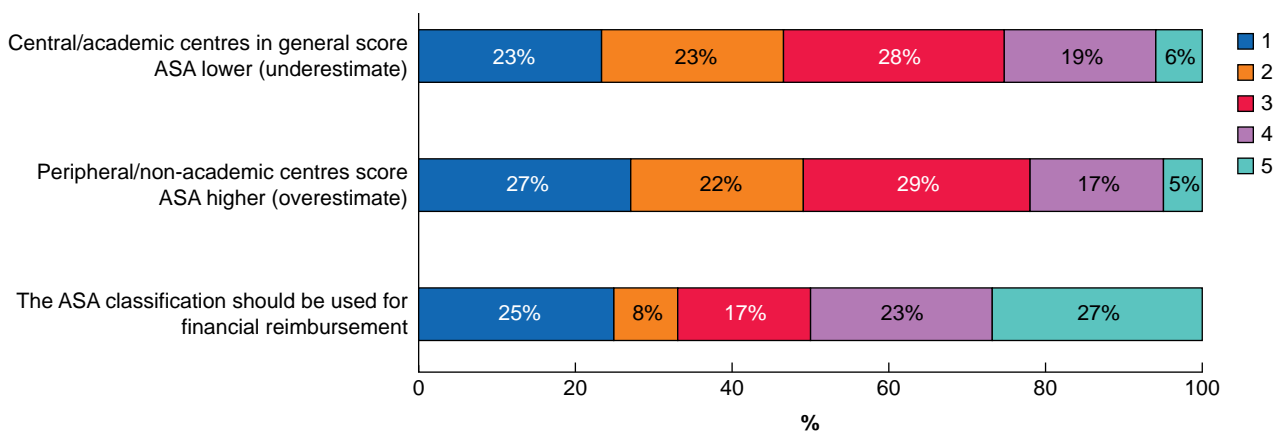


Fig. 4 Considerations made in the American Society of Anesthesiologists (ASA) scoring process
Rated on a scale of 1-5, with 1 strongly disagree, and 5 strongly agree.

classification of patients undergoing HPB surgery. Most respondents used the ASA classification (78%), among whom 79% adapted their perioperative strategy according to its results. Over 80% of respondents stated that they would take the

underlying disease (for example cancer) into account for ASA scoring, but this barely changed the ASA score within the case-vignettes (1% difference in most reported score). Only 11% stated that they would take the type of operation into account

Table 2 Top 10 suggestions to improve the inter-rater variability of the ASA classification

Suggestions to add to the ASA classification	No. of times reported*
1. Include type of procedure	145
2. More extensive definition	137
3. Provide more examples	94
4. Include performance status (functional status)	77
5. Further subdivision	73
6. Include frailty	70
7. Include age	49
8. Include more co-morbidities	40
9. Other scores	38
10. Include BMI/weight	34

*Multiple suggestions can be made by one respondent. ASA, American Society of Anesthesiologists; BMI, body mass index.

(for example complex surgery); nevertheless, this did change the preferred score in the case-vignettes (13% difference).

The inter-rater variability in ASA classification among patients undergoing common (less-complex) surgical procedures has been confirmed in surveys from Finland, Italy, the United Kingdom, and the United States²⁻⁵. The Kappa values reported in these studies, ranging from 0.21 to 0.40, were comparable to those in the present international study^{2,4}. In addition to the heterogeneity in ASA scoring in hypothetical cases, substantial variation in the classification of patients exists in daily clinical practice²¹. Suggestions have been made to improve the ASA classification, such as adding institutional-specific or ASA-approved examples to the classification, which have been proven to decrease variability in scoring^{22,23}. Examples were added to paediatric and obstetric patients specifically, as the ASA score is more variable for these subgroups^{24,25}. These additions are not yet implemented for patients undergoing HPB surgery, even though considerable differences exist in this patient group. HPB surgery is highly complex surgery, and often includes frail patients who typically undergo surgery for cancer. The benefits from surgery must be carefully weighted relative to the perioperative risk, which is often determined by various risk factors including the ASA score.

Within all case-vignettes included in this survey, the most commonly chosen ASA score was either ASA II or ASA III, whereas ASA I and ASA IV-V were almost never used. If certain values are barely used, the Kappa score can get less reliable, a phenomenon known as the 'Kappa paradox'²⁶. Within the case-vignettes, the decision to use ASA II or III seemed often inconsistent. For example, in the fourth case-vignette ASA II was scored 46%, and ASA III 44%. This distinction is most important, as many clinical studies that use ASA as a case-mix adjustment make a subdivision between patients scored either ASA I-II or ASA III. Suggestions mentioned within the survey to decrease inter-rater variability included the creation of a more extensive definition (137 respondents) or a further subdivision (77 respondents), which could better differentiate ASA II from ASA III. However, it should be taken into account that the ASA classification was designed as a pragmatic bedside tool and has to stay easy-to-use in daily clinical practice²⁷.

Adding examples to the ASA definition may decrease inter-rater variability, whereas the classification remains feasible in daily clinical practice^{24,25}. It seems that the extent of surgery and the surgical approach is already considered when scoring the ASA classification (60% of the participants classified case 7 robotic hemihepatectomy as ASA III, and this decreased to 47% when

an open hemihepatectomy was performed). Benign versus malignant disease does not seem to influence the ASA score (case 5 colorectal liver metastasis: 47% ASA III; case 8 hepatocellular adenoma: 46% ASA III), even though the vast majority of respondents (89%) stated that a malignancy should be considered for ASA classification. These data illustrate the lack of consensus and confirm the need for an extended definition or addition of examples to the classification (for example including remarks on the impact of the type of surgery, malignant versus benign disease).

ASA scoring differs among regions, which could be due to financial incentives associated with ASA classification in some countries, given that 50% of the participants report that the ASA score influences financial compensation in their country. Convincing evidence on this topic is scarce. One study from the USA examined variation in ASA scores relative to insurance eligibility²⁸. This study reported no up-coding of ASA score with changes in payer incentives.

The Simplified Acute Physiology Score II (SAPS II) was reported to have the highest power to predict mortality, but it is rarely used in clinical practice²⁹. The Charlson Co-morbidity Index (CCI) has been reported to have lower discriminative power to predict mortality, but might be more suitable for case-mix adjustment within research, as less heterogeneity could be expected (especially among countries) due to its more explicit criteria²⁹. The ASA classification is easy to use and associated with multiple postoperative outcomes including postoperative morbidity and mortality³⁰. Improvements made in the ASA classification, such as the addition of examples, have been proven to reduce inter-rater variability. Until this has been achieved, especially when comparing outcomes among countries, researchers should consider using more comprehensive measures such as the CCI for case-mix adjustment.

These results should be interpreted in light of some limitations. Among 133 (inter)national anaesthesiology societies, only 17 (12%) disseminated the survey. Unfortunately, the exact response rate is unknown, as the number of members within the different societies could not be shared. Most respondents were from Europe/North America (80%), limiting the generalization. Answers to the questions on 'the considerations made during the ASA classification process' were very heterogeneous, making it hard to draw any firm conclusion. The heterogeneity in the answers does represent the current lack of consensus on the considerations made in the decision-making of the ASA score (that is, which parameters should be considered). Almost all surgeons who responded performed HPB surgery (99%), but only 78% of all anaesthesiologists worked with HPB patients (of which the minority worked in a dedicated HPB unit: 32%), which might have influenced the results. Future research should focus on assessing the clinical and socioeconomic impact of the observed variability, achieving consensus on how the ASA classification should be used (for example take indication for surgery and type of surgery into account or not) to improve consistent use and interpretation of the classification, as well as seeking ways to adjust ASA to make it more applicable to the clinical setting (for example extending the main classification, or adding HPB surgery-specific examples).

This study demonstrated variability in interpretation and use of the ASA classification by a large cohort of international anaesthesiologists and surgeons caring for HPB surgical patients. A need exists for an improved, more objective, ASA classification to facilitate harmonized clinical decision-making and minimize inappropriate conclusions from studies

Table 3 Clinical cases and inter-rater variability

Type of surgery	Case	Co-morbidities and clinical condition	ASA score	Agreement*	
Pancreatoduodenectomy	1	Diagnosis: distal bile duct carcinoma	ASA 1	71 (6)	Fair agreement (K = 0.31)
		55 years old, BMI 30 kg/m ²	ASA 2	730 (57)	
		Biliary stasis and a general decline in condition (since 3 months)	ASA 3	445 (35)	
		After stenting recovered condition	ASA 4	35 (3)	
		No other medical history	ASA 5	2 (0)	
	2	Diagnosis: distal bile duct carcinoma	ASA 1	3 (0)	Moderate agreement (K = 0.42)
		55 years old, BMI 30 kg/m ²	ASA 2	283 (22)	
		Biliary stasis and a general decline in condition (since 3 months)	ASA 3	892 (70)	
		After stenting recovered condition	ASA 4	99 (8)	
		CABG 2 years ago, no cardiac complaints, moderate left and right ventricle function	ASA 5	6 (0)	
	3	Diagnosis: distal bile duct carcinoma	ASA 1	1 (0)	Fair agreement (K = 0.38)
		55 years old, BMI 30 kg/m ²	ASA 2	32 (2)	
Biliary stasis and a general decline in condition (since 3 months)		ASA 3	803 (63)		
After stenting the condition did recover		ASA 4	435 (34)		
+ kidney failure stage 3 (eGFR 40)		ASA 5	12 (1)		
4	Diagnosis: distal bile duct carcinoma	ASA 1	39 (3)	Fair agreement (K = 0.26)	
	83 years old, BMI 30 kg/m²	ASA 2	589 (46)		
	Biliary stasis and a general decline in condition (since 3 months)	ASA 3	561 (44)		
	After stenting recovered condition	ASA 4	85 (7)		
	No other medical history	ASA 5	9 (1)		
Hemihepatectomy	5	Diagnosis: 1 metachronous colorectal liver metastasis	ASA 1	47 (4)	Fair agreement (0.26)
		73 years old, open procedure	ASA 2	544 (42)	
		No neoadjuvant therapy, no other liver disease	ASA 3	601 (47)	
		Preoperative work-up aorta valve stenosis (AVA 1.2 cm ²)	ASA 4	87 (7)	
		Runs 5 miles 2x a week, no other medical history	ASA 5	4 (0)	
	6	Diagnosis: 1 metachronous colorectal liver metastasis	ASA 1	3 (0)	Fair agreement (K = 0.35)
		73 years old, open procedure	ASA 2	334 (26)	
		No neoadjuvant therapy, no other liver disease	ASA 3	810 (63)	
		Preoperative work-up aorta valve stenosis (AVA 1.2 cm ²)	ASA 4	133 (10)	
		Runs 5 miles 2x a week, insulin-dependent diabetes mellitus	ASA 5	3 (0)	
	7	Diagnosis: 1 metachronous colorectal liver metastasis	ASA 1	9 (1)	Fair agreement (K = 0.33)
		73 years old, robotic procedure	ASA 2	385 (30)	
		No neoadjuvant therapy, no other liver disease	ASA 3	773 (60)	
		Preoperative work-up aorta valve stenosis (AVA 1.2 cm ²)	ASA 4	11 (9)	
		Runs 5 miles 2x a week, no other medical history	ASA 5	5 (0)	
	8	Diagnosis: hepatocellular adenoma	ASA 1	48 (4)	Fair agreement (K = 0.27)
73 years old, open procedure		ASA 2	565 (44)		
No neoadjuvant therapy, no other liver disease		ASA 3	594 (46)		
Preoperative work-up aorta valve stenosis (AVA 1.2 cm ²)		ASA 4	72 (6)		
Runs 5 miles 2x a week, no other medical history		ASA 5	4 (0)		

Values are n (%). Bold text indicates change in the clinical condition per case and the preferred American Society of Anesthesiologists (ASA) score. *Based on the Kappa value interpretation determined by Landis and Koch¹⁵. BMI, body mass index.

employing incorrect ASA classifications. This could be achieved by adding examples to the classification and raising awareness on the topic. Until then, more objective measurements should be considered for case-mix adjustment within research in patients undergoing HPB surgery.

Funding

The authors have no funding to declare.

Acknowledgements

We want to thank the collaborating societies for disseminating the survey to their members: Liver Intensive Care Groups of Europe, American Society of Anesthesiologists, SBA Sociedade Brasileira de Anestesiologia, Canadian Anesthesiologists' Society, Estonian Society of Anaesthesiologists, Société Française d'Anesthésie et de Réanimation, Deutsche Gesellschaft für Anaesthesiologie und

Intensivmedizin e.V. (DGAI), Hellenic Society of Anaesthesiology, Indian Society of Anaesthesiologists, Società Italiana di Anestesia, Analgesia, Rianimazione e Terapia Intensiva, Korean Society of Anesthesiologists, Nederlandse Vereniging voor Anesthesiologie, New Zealand Society of Anaesthetists, Macedonia Society of Anaesthesiologists, Oman Society of Anesthesia & Critical Care, South African Society of Anaesthesiologists, Swedish Society for Anaesthesiology and Intensive Care Medicine (SFAI), International Hepato-Pancreato-Biliary Association, European-African Hepato-Pancreato-Biliary Association (EAHPBA), Americas Hepato-Pancreato-Biliary Association (AHPBA), Asia-Pacific Pancreato-Biliary Association (AP-HPBA), International Laparoscopic Liver Society (ILLS), European Consortium on Minimally Invasive Pancreatic Surgery (E-MIPS), European Minimally Invasive Liver Surgical Group (E-MILS), Dutch Pancreatic Cancer Group (DPCG), and the Turkish HPB society. S.A. and J.P.S. shared first authorship. M.A.H., M.P., M.W.H., and M.G.B. shared senior authorship.

Author contributions

Simone Augustinus (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Validation, Writing—original draft, Writing—review & editing), Jasper P. Sijberden (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Validation, Writing—original draft, Writing—review & editing), Matthanja Bieze (Conceptualization, Investigation, Methodology, Project administration, Writing—review & editing), Vandana Agarwal (Conceptualization, Methodology, Writing—review & editing), Michael Guertin (Conceptualization, Methodology, Writing—review & editing), Ronald Harter (Conceptualization, Methodology, Writing—review & editing), Maria E. Franceschetti (Conceptualization, Methodology, Writing—review & editing), Thomas Kander (Conceptualization, Methodology, Writing—review & editing), Dominik Krzanicki (Conceptualization, Methodology, Writing—review & editing), Keith Lewis (Conceptualization, Methodology, Writing—review & editing), Giuseppe Natalini (Conceptualization, Methodology, Writing—review & editing), Carla Nau (Conceptualization, Methodology, Writing—review & editing), Henry Pitt (Conceptualization, Methodology, Writing—review & editing), Manuel Pardo (Conceptualization, Methodology, Supervision, Writing—review & editing), Luca Aldrighetti (Conceptualization, Methodology, Writing—review & editing), Adnan Alseidi (Conceptualization, Methodology, Writing—review & editing), Francisco Bonofiglio (Conceptualization, Methodology, Writing—review & editing), Kevin Conlon (Conceptualization, Methodology, Writing—review & editing), Katia Donadello (Conceptualization, Methodology, Writing—review & editing), Joris Erdmann (Conceptualization, Methodology, Writing—review & editing), Cristina R. Ferrone (Conceptualization, Methodology, Writing—review & editing), Giuseppe Fusai (Conceptualization, Methodology, Writing—review & editing), Bas Grootkoerkamp (Conceptualization, Methodology, Writing—review & editing), Thilo Hackert (Conceptualization, Methodology, Writing—review & editing), Jin-Young Jang (Conceptualization, Methodology, Writing—review & editing), Tobias Keck (Conceptualization, Methodology, Writing—review & editing), Hojin Lee (Conceptualization, Methodology, Writing—review & editing), Timothy Pawlik (Conceptualization, Methodology, Writing—review & editing), Raffaella Reineke (Conceptualization, Methodology, Writing—review & editing), Roberto Salvia (Conceptualization, Methodology, Writing—review & editing), Eduardo de Santibanes (Conceptualization, Methodology, Writing—review & editing), Shailesh Shrikhande (Conceptualization, Methodology, Writing—review & editing), Martin Smith (Conceptualization, Methodology, Writing—review & editing), Attila Szijarto (Conceptualization, Methodology, Writing—review & editing), Bobby Tingstedt (Conceptualization, Methodology, Writing—review & editing), Alice Wei (Conceptualization, Methodology, Writing—review & editing), John Windsor (Conceptualization, Methodology, Writing—review & editing), Mohammed Abu Hilal (Conceptualization, Methodology, Supervision, Writing—review & editing), Markus Hollmann (Conceptualization, Methodology, Supervision, Writing—review & editing), and Marc Besselink (Conceptualization, Methodology, Resources, Writing—review & editing)

Disclosure

The authors declare no conflict of interest.

Supplementary material

Supplementary Material is available at *BJS Open* online.

Data availability

Data can be made available upon reasonable request.

References

1. American Association of Anesthesiologists. ASA Physical Status Classification System 2020. Available from: <https://www.asahq.org/standards-and-practice-parameters/statement-on-asa-physical-statusclassification-system> (accessed 1 October 2024)
2. De Cassai A, Boscolo A, Tonetti T, Ban I, Ori C. Assignment of ASA-physical status relates to anesthesiologists' experience: a survey-based national-study. *Korean J Anesthesiol* 2019;**72**:53–59
3. Haynes SR, Lawler PG. An assessment of the consistency of ASA physical status classification allocation. *Anaesthesia* 1995;**50**: 195–199
4. Mak PH, Campbell RC, Irwin MG. The ASA physical status classification: inter-observer consistency. *American Society of Anesthesiologists. Anaesth Intensive Care* 2002;**30**:633–640
5. 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**: 629–632
6. Bilimoria KY, Bentrem DJ, Lillemoe KD, Talamonti MS, Ko CY. Assessment of pancreatic cancer care in the United States based on formally developed quality indicators. *J Natl Cancer Inst* 2009;**101**:848–859
7. van Wijk L, van der Snee L, Buis CI, Hentzen JEK, Haveman ME, Klaase JM. A prospective cohort study evaluating screening and assessment of six modifiable risk factors in HPB cancer patients and compliance to recommended prehabilitation interventions. *Perioper Med (Lond)* 2021;**10**:5
8. Mackay TM, Gleeson EM, Wellner UF, Williamsson C, Busch OR, Groot Koerkamp B et al. Transatlantic registries of pancreatic surgery in the United States of America, Germany, The Netherlands, and Sweden: comparing design, variables, patients, treatment strategies, and outcomes. *Surgery* 2021;**169**:396–402
9. Pitt HA, Kilbane M, Strasberg SM, Pawlik TM, Dixon E, Zyromski NJ et al. ACS-NSQIP has the potential to create an HPB-NSQIP option. *HPB (Oxford)* 2009;**11**:405–413
10. van Rijssen LB, Koerkamp BG, Zwart MJ, Bonsing BA, Bosscha K, van Dam RM et al. Nationwide prospective audit of pancreatic surgery: design, accuracy, and outcomes of the Dutch pancreatic cancer audit. *HPB (Oxford)* 2017;**19**:919–926
11. Tingstedt B, Andersson B, Jönsson C, Formichov V, Bratlie SO, Öhman M et al. First results from the Swedish National Pancreatic and Periampullary Cancer Registry. *HPB (Oxford)* 2019;**21**:34–42
12. Wellner UF, Klinger C, Lehmann K, Buhr H, Neugebauer E, Keck T. The pancreatic surgery registry (StuDoQ|pancreas) of the German society for general and visceral surgery (DGAV)—presentation and systematic quality evaluation. *Trials* 2017;**18**:163
13. Mungroop TH, Geerts BF, Veelo DP, Pawlik TM, Bonnet A, Lesurtel M et al. Fluid and pain management in liver surgery (MILESTONE): a worldwide study among surgeons and anesthesiologists. *Surgery* 2019;**165**:337–344
14. Pitt SC, Schwartz TA, Chu D. AAPOR reporting guidelines for survey studies. *JAMA Surg* 2021;**156**:785–786

15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–174
16. Chung F, Yegneswaran B, Liao P, Chung SA, Vairavanathan S, Islam S et al. STOP questionnaire: a tool to screen patients for obstructive sleep apnea. *Anesthesiology* 2008;**108**:812–821
17. Lee TH, Marcantonio ER, Mangione CM, Thomas EJ, Polanczyk CA, Cook EF et al. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. *Circulation* 1999;**100**:1043–1049
18. Jetté M, Sidney K, Blümchen G. Metabolic equivalents (METS) in exercise testing, exercise prescription, and evaluation of functional capacity. *Clin Cardiol* 1990;**13**:555–565
19. Peña FG, Theou O, Wallace L, Brothers TD, Gill TM, Gahbauer EA et al. Comparison of alternate scoring of variables on the performance of the frailty index. *BMC Geriatr* 2014;**14**:25
20. Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmieciak TE, Ko CY et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. *J Am Coll Surg* 2013;**217**:833–42.e1–3
21. Sankar A, Johnson SR, Beattie WS, Tait G, Wijeyesundera DN. Reliability of the American Society of Anesthesiologists physical status scale in clinical practice. *Br J Anaesth* 2014;**113**:424–432
22. 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**:614–622
23. Abouleish AE, Vinta SR, Shabot SM, Patel NV, Hurwitz EE, Krishnamurthy P et al. Improving agreement of ASA physical status class between pre-anesthesia screening and day of surgery by adding institutional-specific and ASA-approved examples: a quality improvement project. *Perioper Med (Lond)* 2020;**9**:34
24. Aplin S, Baines D, De Lima J. Use of the ASA physical status grading system in pediatric practice. *Paediatr Anaesth* 2007;**17**:216–222
25. Barbeito A, Muir HA, Gan TJ, Reynolds JD, Spahn T, White WD et al. Use of a modifier reduces inconsistency in the American Society of Anesthesiologists physical status classification in parturients. *Anesth Analg* 2006;**102**:1231–1233
26. Derksen BM, Bruinsma W, Goslings JC, Schep NWL. The Kappa paradox explained. *J Hand Surg Am* 2024;**49**:482–485
27. Horvath B, Kloesel B, Todd MM, Cole DJ, Prielipp RC. The evolution, current value, and future of the American Society of Anesthesiologists physical status classification system. *Anesthesiology* 2021;**135**:904–919
28. Schonberger RB, Dutton RP, Dai F. Is there evidence for systematic upcoding of ASA physical status coincident with payer incentives? A regression discontinuity analysis of the national anesthesia clinical outcomes registry. *Anesth Analg* 2016;**122**:243–250
29. Yurtlu DA, Aksun M, Ayvat P, Karahan N, Koroglu L, Aran GÖ. Comparison of risk scoring systems to predict the outcome in ASA-PS V patients undergoing surgery: a retrospective cohort study. *Medicine (Baltimore)* 2016;**95**:e3238
30. Mayhew D, Mendonca V, Murthy BVS. A review of ASA physical status—historical perspectives and modern developments. *Anaesthesia* 2019;**74**:373–379