BJA

Reliability of the American Society of Anesthesiologists physical status scale in clinical practice

A. Sankar[1], S. R. Johnson[2,3], W. S. Beattie[4], G. Tait[4] and D. N. Wijeysundera[2,4,5]*

[1] Faculty of Medicine and [2] Institute of Health Policy Management and Evaluation, University of Toronto, Toronto, Ontario, Canada
[3] Division of Rheumatology, Department of Medicine, Toronto Western Hospital, Mount Sinai Hospital, and University of Toronto, Toronto, Ontario, Canada
[4] Department of Anesthesia, Toronto General Hospital and University of Toronto, Toronto, Ontario, Canada
[5] Li Ka Shing Knowledge Institute of St Michael's Hospital, Toronto, Ontario, Canada

* Corresponding author: Department of Anesthesia and Pain Management, Toronto General Hospital, EN 3-450, 200 Elizabeth Street, Toronto, ON, Canada M5G 2C4. E-mail: d.wijeysundera@utoronto.ca

Editor's key points

- The ASA physical status classification was designed as a measure of preoperative health status, not operative risk.
- This study found good agreement with how different anaesthetists rate a patient's ASA classification.
- This study used psychometric methods to show that the ASA classification is an indicator of perioperative risk.

Background. Previous studies, which relied on hypothetical cases and chart reviews, have questioned the inter-rater reliability of the ASA physical status (ASA-PS) scale. We therefore conducted a retrospective cohort study to evaluate its inter-rater reliability and validity in clinical practice.

Methods. The cohort included all adult patients ($\geq$18 yr) who underwent elective non-cardiac surgery at a quaternary-care teaching institution in Toronto, Ontario, Canada, from March 2010 to December 2011. We assessed inter-rater reliability by comparing ASA-PS scores assigned at the preoperative assessment clinic *vs* the operating theatre. We also assessed the validity of the ASA-PS scale by measuring its association with patients' preoperative characteristics and postoperative outcomes.

Results. The cohort included 10 864 patients, of whom 5.5% were classified as ASA I, 42.0% as ASA II, 46.7% as ASA III, and 5.8% as ASA IV. The ASA-PS score had moderate inter-rater reliability (κ 0.61), with 67.0% of patients ($n=7279$) being assigned to the same ASA-PS class in the clinic and operating theatre, and 98.6% ($n=10\,712$) of paired assessments being within one class of each other. The ASA-PS scale was correlated with patients' age (Spearman's ρ, 0.23), Charlson comorbidity index ($\rho=0.24$), revised cardiac risk index ($\rho=0.40$), and hospital length of stay ($\rho=0.16$). It had moderate ability to predict in-hospital mortality (receiver-operating characteristic curve area 0.69) and cardiac complications (receiver-operating characteristic curve area 0.70).

Conclusions. Consistent with its inherent subjectivity, the ASA-PS scale has moderate inter-rater reliability in clinical practice. It also demonstrates validity as a marker of patients' preoperative health status.

Keywords: anaesthesiology; health status; reliability and validity

Accepted for publication: 9 February 2014

The ASA physical status (ASA-PS) scale is commonly used to subjectively estimate preoperative health status. While originally created for statistical data collection and reporting in anaesthesia,[1] it is now used for allocating resources,[2] reimbursing anaesthesia services,[3] and predicting perioperative risk.[4–15]

Inter-rater reliability is important when assessing the ASA-PS.[16] Most reliability studies of the ASA-PS involved different anaesthesiologists rating *hypothetical* case scenarios. These studies found only fair inter-rater agreement (κ 0.21–0.4), thus raising concerns about the scale's reliability.[17–21] There has been little evaluation of its reliability in clinical practice. In

a multicentre study involving 1357 anaesthesia records,[22] the ASA-PS score assigned by the responsible anaesthesiologist had moderate agreement (κ 0.53) with the score assigned by another blinded anaesthesiologist who had reviewed a duplicate version of the same medical record.[22] A similar single-centre study of 430 paediatric anaesthesia records found low-to-moderate reliability (κ 0.43).[12]

Given the paucity of relevant data, we undertook a cohort study to characterize the reliability and validity of the ASA-PS scale in clinical practice. The primary objective was to evaluate the inter-rater agreement of ASA-PS scores assigned at

outpatient preoperative assessment clinics *vs* operating theatres. The secondary objectives were to assess the scale's validity as a measure of health status by measuring its association with patient characteristics, validated predictive indices [Charlson comorbidity index[23] and revised cardiac risk index (RCRI)],[11] hospital stay, complications, and mortality.

Methods

After research ethics board approval, we conducted a retrospective cohort study of consecutive adults aged ≥18 yr who underwent elective non-cardiac surgery from March 2010 to December 2011 at the University Health Network (Toronto, Ontario, Canada), a quaternary care medical centre offering all adult surgical services except trauma and obstetrics. The cohort included all individuals who underwent elective non-cardiac surgery within 30 days after outpatient assessment at the institutional preoperative assessment clinics. The research ethics board waived the requirement for written informed consent for this study.

Data sources

At the preoperative clinics, nurses document histories using a structured electronic questionnaire (Clinical Anesthesia Information System PreOp Clinic, Adjuvant Informatics, Flamborough, Ontario, Canada) that captures age, sex, comorbidities, and medications in a linkable data set.[24] Each record includes an ASA-PS score assigned by the anaesthesiologist in the clinic (Table 1). Case records from the clinic database were linked to the Enterprise Electronic Data Warehouse (EDW), which captures all information recorded by the hospital electronic charting system (MISYS EPR; Quadramed Corporation, Reston, VA, USA). The EDW includes information on surgeries, laboratory tests, in-hospital medications, hospital length-of-stay, in-hospital mortality, and *International Classification of Diseases* 10th Revision (ICD-10) diagnostic codes. Documented surgical information includes an ASA-PS score assigned by the anaesthesiologist in the operating theatre.

The primary variables of interest were ASA-PS scores assigned in the preoperative clinics and operating theatres. Patients' age, sex, surgery, preoperative creatinine concentration, hospital length of stay, in-hospital 30 day mortality, and

Table 1 Description of ASA-PS classes

ASA-PS class	Description
Class I	A normal healthy patient
Class II	A patient with mild systemic disease
Class III	A patient with severe systemic disease
Class IV	A patient with severe systemic disease that is a constant threat to life
Class V	A moribund patient who is not expected to survive without operation
Class VI	A declared brain-dead patient whose organs are being removed for donation

postoperative myocardial injury (troponin I concentration exceeding 0.30 µg litre^{-1}) were captured from the EDW. We ascertained specific comorbidities using the clinic data set (hypertension, coronary artery disease, heart failure, diabetes, cerebrovascular disease, asthma, chronic obstructive pulmonary disease) and EDW (Charlson comorbidity index).[23 25] We calculated the RCRI score using information from the EDW (surgical procedure and preoperative creatinine concentration) and clinic data set (other comorbidities).[11]

Contextual factors

Several factors should be considered when comparing ASA-PS ratings in operating theatres *vs* preoperative clinics at the University Health Network. *First*, it is possible that an individual patient received care from the same anaesthesiologist in the clinic and operating theatre. Such scenarios were uncommon given the ~65 consultant anaesthesiologists at the institution. *Secondly*, anaesthesiologists in operating theatres were not blinded to ASA-PS assessments performed in the clinics. Blinding was not feasible since clinic assessments are part of routine clinical care. Nonetheless, anaesthesiologists in operating theatres typically pay little attention to the clinic rating, which is reported as a single non-highlighted line in an extensive computer-generated report. *Thirdly*, anaesthesiologists in operating theatres (but not clinics) receive financial premiums from the government health insurance plan to provide anaesthetic care to ASA-PS class III or class IV patients. These premiums were paid by to the anaesthesiologists' group practice plan, and hence shared among all its members.

Analysis

Analyses were performed using STATA version 13 (StataCorp Inc., Lakeway, TX, USA) and the R statistical language.[26] A two-tailed *P*-value of <0.05 was used to define statistical significance.

Reliability refers to the reproducibility of an instrument, with inter-rater reliability referring to the application of the ASA-PS scale to the same group of patients by different raters. We measured agreement of ASA-PS ratings assigned in the clinic *vs* operating theatre using the intra-class correlation coefficient (ICC) and Cohen's weighted κ. Landis and Koch[27] characterize reliability statistic values of 0–0.20 as 'slight', 0.21–0.4 as 'fair', 0.41–0.60 as 'moderate', 0.61–0.80 as 'substantial', and values exceeding 0.80 as 'almost perfect'. McHorney and Tarlov[28] have also suggested that the ICC for measures applied to individual patients should exceed 0.90. We conducted two sensitivity analyses. *First*, we re-calculated the ICC after excluding successive randomly selected patients whose raters agreed on ASA-PS classification. This analysis assessed how the lack of blinding impacted our results. This process was repeated until the ICC approached values reported by previous blinded studies.[12 17–22] *Secondly*, we modified our study cohort such that the number of patients who were 'up-coded' in a financially advantageous manner (i.e. from class II to III, or from class III to IV) was equal to the number 'down-coded' in a financially disadvantageous manner (i.e. from class III to II, or from class

IV to III). Any excess 'up-coded' patients were classified instead as having identical ratings in the clinic and operating theatre. The ICC was re-calculated in this hypothetical cohort to assess the influence of financial incentives for anaesthesiologists in operating theatres to assign patients to ASA-PS class III or IV.

We categorized each patient as being assigned (i) the same ASA-PS score assigned in the clinic and operating theatre, (ii) a lower score in the operating theatre, and (iii) a higher score in the operating theatre. We compared the characteristics of the categories using the χ^2 test for categorical variables, and analysis of variance or the Kruskal–Wallis test for continuous variables. We also used multivariable logistic regression to determine the adjusted association of patient and surgery characteristics with inter-rater disagreement. The dependent variable was *any* disagreement in ASA-PS scores, while the predictor variables included age, surgery, and comorbidities. In the primary analysis, individual comorbidities were considered as separate predictor variables, while a sensitivity analysis instead considered the total number of concurrent systemic diseases as a predictor variable.

In the primary analysis, we assessed the validity of ASA-PS ratings in the clinic, while ratings in operating theatres were assessed in a secondary analysis. Both construct and criterion validity were evaluated. *Construct validity* refers to whether the ASA-PS scale behaves like a measure of preoperative physical status.[29] For example, individuals with poorer physical status are likely to be older and have more comorbidity. We used descriptive statistics to characterize strata defined by ASA-PS rating. Categorical variables were described using counts and proportions, while continuous variables were described using means, standard deviations, medians, and inter-quartile ranges. We compared characteristics of these strata using the χ^2 test for categorical variables, and analysis of variance or the Kruskal–Wallis test for continuous variables. The correlation of ASA-PS rating with age was further assessed using Spearman's ρ.

We also evaluated the criterion validity of the ASA-PS scale, which consists of concurrent and predictive validity. *Concurrent validity* refers to whether the scale correlates with other indices of health status measured at approximately the same time. Spearman's ρ was used to assess the correlation of ASA-PS ratings with the Charlson comorbidity index and RCRI scores. *Predictive validity* describes whether the ASA-PS scale predicts future-related events. For example, patients with poor health status are more likely to suffer postoperative morbidity and mortality. We used the area-under-the-curve (AUC) of the receiver-operating characteristic curve to separately measure the discrimination of ASA-PS ratings for the outcomes of in-hospital 30 day mortality and myocardial injury. Additionally, the correlation of ASA-PS ratings with hospital length-of-stay was measured using Spearman's ρ.

We used all available data from our databases within the study time frame (March 2010–December 2011). To place the available sample size in context, we estimated the sample size required to measure a plausible degree of inter-rater reliability with acceptable precision. The sample size required to measure an ICC of 0.41 (moderate agreement) with a lower

two-sided 95% confidence interval (CI) excluding an ICC of 0.21 (fair agreement) with 90% power was 175.

Results

The cohort consisted of 10 864 patients (Table 2), of whom 5.5% (*n*=602) were assigned to ASA-PS class I, 42.0% (*n*=4562) assigned to class II, 46.7% (*n*=5073) assigned to class III, and 5.8% (*n*=627) assigned to class IV in the preoperative clinic.

Inter-rater reliability

The agreement between ASA-PS scores assigned in the preoperative clinic *vs* operating theatre is presented in Table 3 and Figure 1. Approximately 67% of individuals (*n*=7279) were assigned to the same ASA-PS class in the clinic and operating theatre, while 98.6% (*n*=10 712) of paired assessments were within one ASA-PS class of each other. Approximately 21% (*n*=2245) were assigned to a higher ASA-PS class in the operating theatre, while 12% (*n*=1340) were assigned to a lower class. Inter-rater reliability measured by the one-way ICC was 0.61 (95% CI, 0.60–0.62), while the weighted κ statistic was 0.61 (95% CI, 0.60–0.62). The calculated ICC approached values seen in prior unblinded studies if one-third to one-half of all cases of inter-rater agreement were excluded (Supplementary Fig. S1). When the study cohort was modified to remove the effects of any financial incentives for ASA-PS classification (Supplementary Table S1), the re-calculated ICC increased to 0.68 (CI, 0.67–0.69).

In unadjusted analyses, inter-rater disagreement was associated with patient age, surgery, specific comorbidities (i.e. coronary artery disease, peripheral vascular disease, hypertension, asthma, cancer), and Charlson comorbidity index scores (Table 4). After multivariable adjustment, factors significantly associated with inter-rater disagreement were age, surgical procedure, hypertension, and malignancy (Table 5). Surgical procedures that were significantly less likely to be associated with inter-rater disagreement were general surgery, neurosurgery, orthopaedic, and urological procedures. In a sensitivity analysis, increased burden of comorbidity was associated with lower odds of inter-rater disagreement (Supplementary Table S2).

Validity

The ASA-PS classes assigned in the clinic differed significantly with respect to age, sex, surgery, comorbidities, and composite comorbidity index scores (Table 2). In general, individuals in higher ASA-PS classes were likely to be older males with more comorbid disease and higher composite comorbidity index scores. These same individuals had longer stays in hospital, and also higher risks of postoperative mortality and myocardial injury (Table 2). The ASA-PS rating in the clinic was correlated with age (Spearman's ρ, 0.23; CI, 0.21–0.25), Charlson comorbidity index score (Spearman's ρ, 0.24; CI, 0.22–0.26), and RCRI score (Spearman's ρ, 0.40; CI, 0.38–0.42). The rating had moderate discrimination for predicting 30 day in-hospital mortality (AUC 0.69; CI, 0.62–0.76) and myocardial injury (AUC 0.70; CI, 0.65–0.75). It was weakly correlated with hospital length of stay (Spearman's ρ, 0.16; CI, 0.15–0.18).

Table 2 Characteristics of study cohort, stratified by ASA-PS rating. ENT, ear–nose–throat; SD, standard deviation. *Defined as preoperative dialysis requirement or preoperative creatinine concentration exceeding 176 μmol litre^{-1} (2.0 mg dl^{-1})

	ASA-PS rating assigned in the preoperative assessment clinic				P-value
	I (n=602)	II (n=4562)	III (n=5073)	IV (n=627)	
Patient characteristics					
Age (yr), mean (range)	44.0 (18–99)	58.0 (18–95)	61.0 (18–95)	66.0 (18–94)	<0.001
Male sex	290 (48.2%)	2107 (46.2%)	2591 (51.1%)	414 (66.0%)	<0.001
Surgical service					
ENT surgery	95 (15.8%)	656 (14.4%)	514 (10.1%)	98 (15.6%)	
General surgery	105 (17.4%)	590 (12.9%)	998 (19.7%)	118 (18.8%)	
Gynaecology	49 (8.1%)	416 (9.1%)	367 (7.2%)	46 (7.3%)	
Neurosurgery	11 (1.8%)	271 (5.9%)	592 (11.7%)	20 (3.2%)	
Ophthalmology	0 (0.0%)	2 (0.0%)	2 (0.0%)	0 (0.0%)	<0.001
Orthopaedic surgery	126 (20.9%)	1330 (29.2%)	1179 (23.2%)	27 (4.3%)	
Plastic surgery	35 (5.8%)	169 (3.7%)	80 (1.6%)	5 (0.8%)	
Thoracic surgery	13 (2.2%)	257 (5.6%)	585 (11.5%)	121 (19.3%)	
Urology	168 (27.9%)	840 (18.4%)	565 (11.1%)	57 (9.1%)	
Vascular surgery	0 (0.0%)	31 (0.7%)	191 (3.8%)	135 (21.5%)	
Comorbid disease					
Coronary artery disease	5 (0.8%)	316 (6.9%)	992 (19.6%)	229 (36.5%)	<0.001
Congestive heart failure	1 (0.2%)	9 (0.2%)	121 (2.4%)	78 (12.4%)	<0.001
Peripheral vascular disease	0 (0.0%)	32 (0.7%)	163 (3.2%)	99 (15.8%)	<0.001
Cerebrovascular disease	2 (0.3%)	86 (1.9%)	376 (7.4%)	99 (15.8%)	<0.001
Hypertension	16 (2.7%)	1574 (34.5%)	2736 (53.9%)	409 (65.2%)	<0.001
Diabetes	2 (0.3%)	386 (8.5%)	1143 (22.5%)	185 (29.5%)	<0.001
Renal insufficiency*	0 (0%)	5 (0.1%)	118 (2.3%)	47 (7.5%)	<0.001
Chronic obstructive pulmonary disease	0 (0%)	132 (2.9%)	430 (8.5%)	110 (7.5%)	<0.001
Asthma	25 (4.2%)	389 (8.5%)	629 (12.4%)	66 (10.5%)	<0.001
Rheumatic disease	0 (0.0%)	27 (0.6%)	87 (1.7%)	14 (2.3%)	<0.001
Peptic ulcer disease	0 (0.0%)	17 (0.4%)	36 (0.7%)	8 (1.3%)	0.003
Liver disease	4 (0.7%)	70 (1.5%)	168 (3.3%)	35 (5.6%)	<0.001
Cancer					
Primary disease	118 (19.6%)	1364 (29.9%)	1331 (26.2%)	172 (27.4%)	<0.001
Metastatic disease	33 (5.5%)	507 (11.1%)	790 (15.6%)	126 (20.1%)	
Comorbidity indices					
Charlson comorbidity index, mean (SD)	0.83 (1.79)	1.59 (2.37)	2.28 (2.70)	3.25 (2.78)	<0.001
Revised cardiac risk index, mean (SD)	0.21 (0.42)	0.36 (0.57)	0.88 (0.88)	1.61 (1.13)	<0.001
Outcomes					
Postoperative myocardial injury	0 (0.0%)	19 (0.4%)	56 (1.1%)	28 (4.5%)	<0.001
30 day in-hospital mortality	0 (0.0%)	11 (0.2%)	25 (0.5%)	15 (2.4%)	<0.001
Hospital length of stay (mean, SD)	3.0 (2.7)	4.0 (5.3)	6.0 (8.9)	8.0 (11.2)	<0.001

In a secondary analysis, ASA-PS ratings in the operating theatre had higher correlations with age (Spearman's ρ, 0.28; CI, 0.26–0.29), Charlson comorbidity index (Spearman's ρ, 0.28; CI, 0.27–0.30), RCRI (Spearman's ρ, 0.42; CI, 0.41–0.44), and hospital length of stay (Spearman's ρ, 0.20; CI, 0.19–0.22). Ratings in the operating theatre had moderate ability to predict mortality (AUC 0.74; CI, 0.68–0.80) and myocardial injury (AUC 0.75; CI, 0.71–0.79). Compared with clinic ratings, ratings in the operating theatre differed significantly with respect to predicting myocardial injury (P=0.01) but not mortality (P=0.17).

Discussion

Given the ubiquity of the ASA-PS scale in clinical practice, it is important to define its reliability and validity. In this large single-institution study, the ASA-PS scale had moderate inter-rater reliability, despite its inherent subjectivity. Furthermore, it demonstrated validity as a measure of preoperative health status, showing expected patterns of association with patient characteristics and postoperative outcomes.

Poor reliability has been among the largest criticisms of the ASA-PS scale.[17–21] For example, a previous study relying on

Table 3 Agreement between ASA-PS ratings in the preoperative assessment clinic *vs* operating theatre

ASA-PS rating in the operating theatre	ASA-PS rating assigned in the preoperative assessment clinic			
	ASA I (*n*=602)	ASA II (*n*=4562)	ASA III (*n*=5073)	ASA IV (*n*=627)
ASA I (*n*=515)	285 (47.3%)	201 (4.4%)	28 (0.6%)	1 (0.2%)
ASA II (*n*=3905)	264 (43.9%)	2814 (61.7%)	807 (15.9%)	20 (3.2%)
ASA III (*n*=5689)	52 (8.6%)	1497 (32.8%)	3857 (76.0%)	283 (45.1%)
ASA IV (*n*=755)	1 (0.2%)	50 (1.1%)	381 (7.5%)	323 (51.5%)
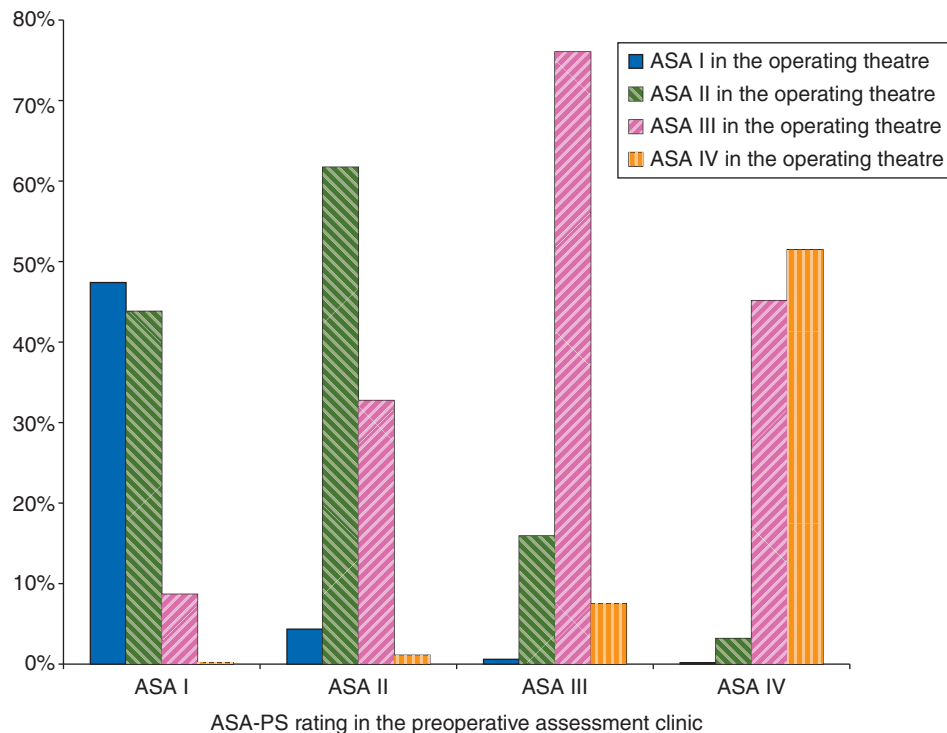
Fig 1 Distribution of ASA-PS ratings in the operating theatre, within strata defined by ASA-PS rating in the preoperative assessment clinic.

hypothetical case scenarios found only fair inter-rater agreement (κ 0.21–0.4).[19] Another study found moderate inter-rater agreement (κ 0.53) when instead comparing ratings by the responsible anaesthesiologist *vs* a different blinded anaesthesiologist reviewing the same medical record.[22] In contrast to previous work, our study evaluated the inter-rater reliability of the ASA-PS scale in 'real-world' clinical practice. Since we compared ASA-PS ratings performed by two anaesthesiologists involved in the clinical care of the same patient, both raters had the opportunity to interview, physically examine, and participate in clinical decision-making. This increased degree of clinical engagement may have explained, in part, the higher observed inter-rater reliability. This degree of inter-rater agreement is remarkable for a subjective rating scale, with 67% of patients being assigned the same ASA-PS score, and almost 99% being assigned scores within one ASA-PS class of each other.

Despite the increased degree of inter-rater reliability in our present study, the ICC (0.61) and weighted κ (0.61) still decreased below the minimum of 0.90 recommended by McHorney and Tarlov.[28] The absence of high inter-rater reliability is also not surprising. There is inherent subjectivity to differentiating between patients with 'mild systemic disease', 'severe systemic disease', and 'severe systemic disease that is a constant threat to life', especially in the absence of a 'moderate systemic disease' category or further standardized information to help define the current existing categories.

We identified several factors associated with inter-rater disagreement, namely age, surgery, hypertension, malignancy, and comorbidity burden. Age has been previously noted as a source of disagreement in ASA-PS ratings,[20] especially since there are no guidelines on how patients' age should be considered when assigning ASA-PS scores. Nonetheless, the association between age and inter-rater disagreement in our study

Table 4 Characteristics of categories defined by level of inter-rater agreement for ASA-PS rating. ENT, ear–nose–throat; SD, standard deviation. *Defined as preoperative dialysis requirement or preoperative creatinine concentration exceeding 176 μmol litre^{-1} (2.0 mg dl^{-1})

	Lower ASA-PS rating in operating theatre [$n=1340$ (12.3%)]	No change in ASA-PS rating [$n=7279$ (67.0%)]	Higher ASA-PS rating in operating theatre [$n=2245$ (20.7%)]	P-value
Patient characteristics				
Age (yr), mean (range)	56.3 (18–94)	59.1 (18–95)	59.9 (18–99)	<0.001
Male sex	681 (12.6%)	3612 (66.9%)	1110 (20.5%)	0.69
Surgical service				
ENT surgery	261 (19.2%)	803 (58.9%)	299 (21.9%)	
General surgery	206 (11.4%)	1238 (68.4%)	367 (20.3%)	
Gynaecology	144 (16.4%)	485 (55.2%)	249 (28.4%)	
Neurosurgery	74 (8.3%)	676 (75.6%)	144 (16.1%)	
Ophthalmology	1 (25.0%)	2 (50.0%)	1 (25.0%)	
Orthopaedic surgery	244 (9.2%)	2063 (77.5%)	355 (13.3%)	<0.001
Plastic surgery	43 (14.7%)	172 (59.5%)	74 (25.6%)	
Thoracic surgery	129 (13.2%)	583 (59.7%)	264 (27.1%)	
Urology	202 (12.4%)	1049 (64.4%)	379 (23.3%)	
Vascular surgery	36 (10.1%)	208 (58.3%)	113 (31.7%)	
Comorbid disease				
Coronary artery disease	177 (11.5%)	1077 (69.8%)	288 (18.7%)	0.04
Congestive heart failure	28 (13.4%)	141 (67.5%)	40 (19.1%)	0.82
Peripheral vascular disease	24 (8.2%)	176 (59.9%)	94 (32.0%)	<0.001
Cerebrovascular disease	70 (12.4%)	393 (69.6%)	101 (17.9%)	0.25
Hypertension	537 (11.3%)	3277 (69.2%)	921 (19.5%)	<0.001
Diabetes	200 (11.7%)	1189 (69.3%)	327 (19.1%)	0.09
Renal insufficiency*	18 (10.6%)	116 (68.2%)	36 (21.2%)	0.78
Chronic obstructive pulmonary disease	68 (10.1%)	461 (68.6%)	143 (21.3%)	0.20
Asthma	160 (14.4%)	760 (68.5%)	189 (17.0%)	0.002
Rheumatic disease	23 (18.0%)	86 (67.2%)	19 (14.8%)	0.07
Peptic ulcer disease	10 (16.4%)	43 (70.5%)	8 (13.1%)	0.27
Liver disease	32 (11.6%)	178 (64.3%)	67 (24.2%)	0.82
Malignancy				
Primary disease	398 (13.3%)	1821 (61.0%)	766 (25.7%)	<0.001
Metastatic disease	207 (14.2%)	881 (60.5%)	368 (25.3%)	
Comorbidity indices				
Charlson comorbidity index, mean (SD)	2.14 (2.68)	1.82 (2.50)	2.31 (2.72)	<0.001
Revised cardiac risk index, mean (SD)	0.64 (0.84)	0.67 (0.85)	0.68 (0.83)	0.30

should be viewed cautiously since its statistical significance was not strong. Additionally, the association did not follow a logical pattern, such as increasing inter-rater disagreement at the extremes of age. Surgical procedure has also previously been identified as a source of inter-rater disagreement.[17] [18] For example, Haynes and Lawler[18] found that anaesthesiologists assigned patients undergoing minor surgical procedures to lower ASA-PS classes than would be otherwise expected, even when the patients had serious medical disease. The influence of surgical procedure on inter-rater disagreement is likely driven by misunderstanding of the ASA-PS classification system, which was developed to measure preoperative health status, not operative risk. Indeed, in his original paper, Saklad[1] stated that the ASA-PS grade had 'no relation to the operative procedure, the ability of the surgeon or anesthetist, nor the

type of anesthesia the patient will receive'. Nonetheless, many anaesthesia providers still consider the ASA-PS scale an anaesthetic risk predictor.[17] The association of specific comorbidities with inter-rater disagreement in our study has some consistency with previous research.[22] In addition, our results suggest that clinicians are less likely to agree on how some medical conditions (e.g. cancer) impact on preoperative physical status, but more likely to agree on the impact of the total burden of comorbidity.

In addition to evaluating the reliability of the ASA-PS scale, we assessed its construct, concurrent, and predictive validity.[29] The scale showed construct validity, based on fair correlation with patient age, and an increased burden of comorbidities in patients with higher ASA-PS scores (Table 2). Our findings confirm previous work, such as a single-centre study showing

Table 5 Adjusted association of patient characteristics with disagreement in ASA-PS class rating. ENT, ear–nose–throat. *Defined as preoperative dialysis requirement or preoperative creatinine concentration exceeding 176 μmol litre^{-1} (2.0 mg dl^{-1})

	Adjusted odds ratio for disagreement in ASA rating	95% confidence interval	*P*-value
Patient characteristics			
Male sex	1.04	0.95–1.14	0.40
Age			
40 yr or less	Reference category		
41–60 yr	0.84	0.74–0.96	
61–80 yr	0.94	0.82–1.08	0.03
81 yr or more	0.89	0.72–1.10	
Surgical service			
ENT surgery	Reference category		
General surgery	0.66	0.57–0.77	
Gynaecology	1.20	1.00–1.43	
Neurosurgery	0.46	0.38–0.55	
Ophthalmology	1.49	0.21–10.67	<0.001
Orthopaedic surgery	0.42	0.36–0.48	
Plastic surgery	0.99	0.76–1.28	
Thoracic surgery	0.97	0.82–1.15	
Urology	0.88	0.75–0.92	
Vascular surgery	1.13	0.82–1.58	
Comorbid disease			
Coronary artery disease	0.89	0.78–1.00	0.06
Congestive heart failure	1.04	0.76–1.41	0.83
Peripheral vascular disease	1.02	0.72–1.43	0.92
Cerebrovascular disease	0.91	0.75–1.10	0.33
Hypertension	0.86	0.78–0.94	0.001
Diabetes	0.95	0.84–1.07	0.42
Renal insufficiency*	0.91	0.65–1.27	0.58
Chronic obstructive pulmonary disease	0.88	0.74–1.05	0.15
Asthma	1.01	0.88–1.15	0.93
Rheumatic disease	1.25	0.86–1.84	0.25
Peptic ulcer disease	0.92	0.52–1.62	0.77
Liver disease	1.16	0.89–1.50	0.27
Malignancy			
Primary disease	1.20	1.07–1.34	0.002
Metastatic disease	1.19	1.04–1.37	

strong interdependence between ASA-PS ratings and National Surgical Quality Improvement Program clinical risk factors.[7] The ASA-PS scale also exhibited concurrent validity. It was correlated with more 'objective' comorbidity indices such as the Charlson comorbidity index, and RCRI. Notably, the correlation of ASA-PS scores with the Charlson comorbidity index was only slight-to-fair in magnitude. The relatively poor correlation may be explained by the subjectivity of the ASA-PS scale, and differences in how the two scales were developed. The ASA-PS scale was intended to measure preoperative physical status, while the Charlson comorbidity index was developed to measure risks of 1 yr mortality in medical inpatients.[1 23] These findings with respect to correlation with other comorbidity measures are consistent with previous research, such as a single-centre study showing correlation of the ASA-PS scale with the

Neurological, Airway, Respiratory, Cardiovascular, and Other model of risk assessment in children.[12]

Our study confirmed the predictive validity of the ASA-PS scale. Even when assessed well before surgery in an outpatient clinic, the scale had moderate ability to predict postoperative mortality and cardiac complications. By comparison, its correlation with hospital length of stay was relatively weak, likely because hospital length of stay is influenced by many distinct clinical factors, such as surgery type. The ability of the ASA-PS scale to predict adverse outcomes has previously been observed for specific surgeries,[4 5 7 9 10 13] where higher ASA-PS scores were associated with higher mortality rates.[14 15] It has also shown modest ability to predict postoperative cardiac complications,[11 30] and been an important component of models designed to predict postoperative mortality and

morbidity.[6 8 31] This consistent demonstration of moderate predictive validity by the ASA-PS scale, both in our present study and previous research, supports its use as a component of risk-adjustment models for comparing surgical outcomes across hospitals. The ASA-PS score is incorporated into the risk-adjustment model used by the National Surgical Quality Improvement Program to measure the quality of surgical care across US hospitals.[32 33]

Notably, we found that the ASA-PS rating in operating theatres exhibited better validity, based on higher correlations with age, comorbidity scores, and hospital length of stay, and also improved prediction of myocardial injury. In some cases, its superior predictive validity could be explained by changes in a patient's medical status between the clinic visit and subsequent surgery. Nonetheless, such cases should be rare since the cohort only included elective surgeries performed within 30 days of a preoperative clinic visit. The superior predictive validity may also have been due to anaesthesiologists in operating theatres being less 'blinded' to eventual outcomes than those in the clinic. For example, patients assigned to class II in the clinic may have been re-assigned to class III in the operating theatre if they develop severe intraoperative hypotension. Nonetheless, such differences in 'blinding' cannot explain the higher correlation of ASA-PS ratings in the operating theatre with age and comorbidity indices. Thus, our findings indicate that ASA-PS ratings have the greatest validity when assigned by the responsible anaesthesiologist in the operating theatre. The results also highlight potential limitations of using hypothetical case scenarios or reviews of medical records as models for evaluating the psychometric properties of the ASA-PS scale.

Several study limitations need to be acknowledged. *First,* this was a retrospective cohort study from a single quaternary-care teaching institution, as reflected by the high proportion of ASA III and ASA IV patients. Similar studies at other centres with differing case-mixes are necessary to better generalize our findings. *Secondly,* the cohort only included patients who underwent elective surgery after being assessed in an outpatient preoperative assessment clinic. Thus, the cohort excluded individuals who were assigned ASA-PS class V, class VI, or any emergency modifier ('E') code. Our findings therefore cannot be extrapolated to non-elective surgical procedures. *Thirdly,* anaesthesiologists in operating theatres were not blinded to ASA-PS scores assigned in the clinic, thereby potentially biasing patients' second ASA-PS rating and increasing inter-rater reliability. Nonetheless, this limitation permitted both anaesthesiologists to be able to conduct a face-to-face assessment of patients in a manner consistent with clinical practice. Indeed, the increased inter-rater agreement observed in our present study may be a reflection of anaesthesiologists being able to interview, physically examine, and participate in the medical care of patients, as opposed to reviewing hypothetical case scenarios or blinded medical charts. Furthermore, our sensitivity analysis suggests that these results would only be negated if fully one-third to one-half of all cases of inter-rater agreement were solely attributable to the lack of blinding. We would propose that such a large impact is unlikely. *Fourthly,* there were financial incentives to assign patients in

operating theatres to ASA-PS classes III and IV. Nonetheless, since such incentives would encourage ratings in the operating theatre to disagree with ASA-PS scores assigned in the clinic, this bias would have led to an underestimate of reliability, as evidenced by our sensitivity analysis. The strength of the bias is also likely small, as individual financial premiums would be considerably diluted within a group practice plan of 65 consultant anaesthesiologists. Furthermore, previous research found no systematic differences in ASA-PS scores assigned by anaesthesiologists who used these scores for billing purposes, as opposed to scores assigned by anaesthesiologists who did not.[20]

Conclusions

In a large single-institution cohort study, the ASA-PS scale had moderate inter-rater reliability in clinical practice. The scale also showed validity, based on its correlation with preoperative characteristics and its prediction of postoperative outcomes. Despite the inherent subjectivity of the ASA-PS scale, our findings support its use as a measure of preoperative health status.

Supplementary material

Supplementary material is available at *British Journal of Anaesthesia* online.

Authors' contributions

A.S., S.R.J., G.T., and D.N.W.: conception and design. A.S. and D.N.W.: analysis and interpretation of the data. A.S. and D.N.W.: drafting of the article. All authors: critical revision of the article for important intellectual content. All authors: final approval of the article. W.S.B., G.T., and D.N.W.: provision of study materials or patients. S.R.J. and D.N.W.: statistical expertise. W.S.B. and D.N.W.: obtaining of funding.

W.S.B., G.T., and D.N.W.: administrative, technical, or logistic support.

Declaration of interest

None declared.

Funding

References

1 Saklad M. Grading of patients for surgical procedures. *Anesthesiology* 1941; **2**: 281–4

2 Vogt AW, Henson LC. Unindicated preoperative testing: ASA physical status and financial implications. *J Clin Anesth* 1997; **9**: 437–41

3 Australian Government Department of Health and Aging. *Medicare Benefits Schedule Book*. Commonwealth of Australia, 2005

4 Bjorgul K, Novicoff WM, Saleh KJ. American Society of Anesthesiologist physical status score may be used as a comorbidity index in hip fracture surgery. *J Arthroplasty* 2010; **25**: 134–7

5 Cullen DJ, Apolone G, Greenfield S, Guadagnoli E, Cleary P. ASA Physical Status and age predict morbidity after three surgical procedures. *Ann Surg* 1994; **220**: 3–9

6 Dalton JE, Kurz A, Turan A, Mascha EJ, Sessler DI, Saager L. Development and validation of a risk quantification index for 30-day postoperative mortality and morbidity in noncardiac surgical patients. *Anesthesiology* 2011; **114**: 1336–44

7 Davenport DL, Bowe EA, Henderson WG, Khuri SF, Mentzer RM, Jr. National Surgical Quality Improvement Program (NSQIP) risk factors can be used to validate American Society of Anesthesiologists Physical Status Classification (ASA PS) levels. *Ann Surg* 2006; **243**: 636–41

8 Glance LG, Lustik SJ, Hannan EL, *et al.* The surgical mortality probability model: derivation and validation of a simple risk prediction rule for noncardiac surgery. *Ann Surg* 2012; **255**: 696–702

9 Han K-R, Kim HL, Pantuck AJ, Dorey FJ, Figlin RA, Belldegrun AS. Use of American Society of Anesthesiologists physical status classification to assess perioperative risk in patients undergoing radical nephrectomy for renal cell carcinoma. *Urology* 2004; **63**: 841–6

10 Hightower CE, Riedel BJ, Feig BW, *et al.* A pilot study evaluating predictors of postoperative outcomes after major abdominal surgery: physiological capacity compared with the ASA physical status classification system. *Br J Anaesth* 2010; **104**: 465–71

11 Lee TH, Marcantonio ER, Mangione CM, *et al.* Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. *Circulation* 1999; **100**: 1043–9

12 Malviya S, Voepel-Lewis T, Chiravuri SD, *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

13 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 Acute Care Surg* 2007; **63**: 972–8

14 Vacanti CJ, VanHouten RJ, Hill RC. A statistical analysis of the relationship of physical status to postoperative mortality in 68,388 cases. *Anesth Analg* 1970; **49**: 564–6

15 Wolters U, Wolf T, Stutzer H, Schroder T. ASA classification and perioperative variables as predictors of postoperative outcome. *Br J Anaesth* 1996; **77**: 217–22

16 Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *J Am Med Assoc* 1997; **277**: 488–94

17 Aronson WL, McAuliffe MS, Miller K. Variability in the American Society of Anesthesiologists Physical Status classification scale. *AANA J* 2003; **71**: 265–74

18 Haynes SR, Lawler PG. An assessment of the consistency of ASA physical status classification allocation. *Anaesthesia* 1995; **50**: 195–9

19 Mak PH, Campbell RC, Irwin MG. The ASA Physical Status Classification: Inter-observer consistency. American Society of Anesthesiologists. *Anaesth Intensive Care* 2002; **30**: 633–40

20 Owens WD, Felts JA, Spitznagel EL Jr. ASA physical status classifications: a study of consistency of ratings. *Anesthesiology* 1978; **49**: 239–43

21 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

22 Cuvillon P, Nouvellon E, Marret E, *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

23 Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987; **40**: 373–83

24 Ellenberger C, Tait G, Beattie WS. Chronic beta blockade is associated with a better outcome after elective noncardiac surgery than acute beta blockade: a single-center propensity-matched cohort study. *Anesthesiology* 2011; **114**: 817–23

25 Quan H, Sundararajan V, Halfon P, *et al.* Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* 2005; **43**: 1130–9

26 R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, 2013

27 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–74

28 McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 1995; **4**: 293–307

29 Steiner D, Norman G. *Health Measurement Scales: A Practical Guide to Their Development and Use*, 4th Edn. New York, NY: Oxford University Press, 2008

30 Gilbert K, Larocque BJ, Patrick LT. Prospective evaluation of cardiac risk indices for patients undergoing noncardiac surgery. *Ann Intern Med* 2000; **133**: 356–9

31 Gupta PK, Gupta H, Sundaram A, *et al.* Development and validation of a risk calculator for prediction of cardiac risk after surgery. *Circulation* 2011; **124**: 381–7

32 Birkmeyer JD, Shahian DM, Dimick JB, *et al.* Blueprint for a new American College of Surgeons: National Surgical Quality Improvement Program. *J Am Coll Surg* 2008; **207**: 777–82

33 Hall BL, Hamilton BH, Richards K, Bilimoria KY, Cohen ME, Ko CY. Does surgical quality improve in the American College of Surgeons National Surgical Quality Improvement Program: an evaluation of all participating hospitals. *Ann Surg* 2009; **250**: 363–76

Handling editor: P. S. Myles