

The Feline Acute Patient Physiologic and Laboratory Evaluation (Feline APPLE) Score: A Severity of Illness Stratification System for Hospitalized Cats

G. Hayes, K. Mathews, G. Doig, S. Kruth, S. Boston, S. Nykamp, Z. Poljak, and C. Dewey

Background: Scores allowing objective stratification of illness severity are available for dogs and horses, but not cats. Validated illness severity scores facilitate the risk-adjusted analysis of results in clinical research, and also have applications in triage and therapeutic protocols.

Objective: To develop and validate an accurate, user-friendly score to stratify illness severity in hospitalized cats.

Animals: Six hundred cats admitted consecutively to a teaching hospital intensive care unit.

Methods: This observational cohort study enrolled all cats admitted over a 32-month period. Data on interventional, physiological, and biochemical variables were collected over 24 hours after admission. Patient mortality outcome at hospital discharge was recorded. After random division, 450 cats were used for logistic regression model construction, and data from 150 cats for validation.

Results: Patient mortality was 25.8%. Five- and 8-variable scores were developed. The 8-variable score contained mentation score, temperature, mean arterial pressure (MAP), lactate, PCV, urea, chloride, and body cavity fluid score. Area under the receiver operator characteristic curve (AUROC) on the construction cohort was 0.91 (95% CI, 0.87–0.94), and 0.88 (95% CI, 0.84–0.96) on the validation cohort. The 5-variable score contained mentation score, temperature, MAP, lactate, and PCV. AUROC on the construction cohort was 0.83 (95% CI, 0.79–0.86), and 0.76 (95% CI, 0.72–0.84) on the validation cohort.

Conclusions and Clinical Importance: Two scores are presented enabling allocation of an accurate and user-friendly illness severity measure to hospitalized cats. Scores are calculated from data obtained over the 1st 24 hours after admission, and are diagnosis-independent. The 8-variable score predicts outcome significantly better than does the 5-variable score.

Key words: Cats; health outcome assessment; Illness severity score; Intensive care unit.

Studies of new management and therapeutic protocols in veterinary medicine frequently are of an observational design. The experienced clinician recognizes that morbidity and mortality outcomes depend not only on the therapy or management protocol under study, but also on the patient's underlying health status and physiologic reserve. Randomization can be applied to minimize disparity in illness severity between treatment groups¹; however, observational studies require robust and valid methods of risk adjustment to decrease bias and improve interpretability and transferability of results.² Illness severity scores have evolved to meet this need, and typically provide a measure of mortality risk.

The ideal score should reflect illness severity accurately, use objective measures that are accessible to most clinicians, and be simple to calculate.^{2,3} Scores that function independent of primary diagnosis maximize prospective applicability and decrease the risk of introducing bias in a risk-adjusted analysis by exclusion of patient groups.⁴ Scores can use patient data collected on the 1st intensive care unit (ICU) day or data collected

Abbreviations:

AUROC	area under the receiver operator characteristic curve
ICU	intensive care unit
IV	intravenous
MAP	mean arterial pressure

every day throughout the patient hospital stay. Repetitive scores can become difficult to interpret when length of hospital stay varies across treatment groups.⁵ Typically the most abnormal values of physiologic variables collected over the defined time period are selected for entry into score development or calculation, because they provide a convenient, predictive, and unbiased summary measure.^{6–9} Diagnosis-independent illness severity scores are available for humans, dogs, and horses.^{10–13} No such scores are available for cats.

The objective of this study was to construct a validated and robust mortality risk prediction score for hospitalized cats. We wish to develop a score with a high degree of predictive accuracy that could be presented in an accessible format allowing manual calculation. This then would allow assignment of an objective measure reflecting severity of illness to any cat enrolled in clinical research.

Materials and Methods

This was a single center cohort study conducted at a teaching hospital serving a predominantly referral patient population. All cats admitted to the ICU were enrolled over a 32-month period, from January 2007 until August 2009. The minimum criteria for admission to the ICU were the need for ongoing intravenous (IV) access for continuous administration of IV fluids or medication or the need for close observation. Patients were subsequently excluded from the study if inconsistencies in patient identification were identified.

From the Veterinary Teaching Hospital (Hayes, Mathews, Kruth, Nykamp, Boston) and the Department of Population Medicine, University of Guelph, Guelph, ON, Canada (Dewey, Poljak); and the Department of Intensive Care Medicine, Northern Clinical School, Royal North Shore Hospital, University of Sydney, Sydney, Australia (Doig). This work was done at the University of Guelph Veterinary Teaching Hospital.

Corresponding author: Dr G. Hayes, VTH, University of Guelph, 50 Stone Road, Guelph, ON, Canada N1G 2W1; e-mail galina@uoguelph.ca.

Submitted March 15, 2010; Revised June 10, 2010; Accepted October 7, 2010.

Copyright © 2010 by the American College of Veterinary Internal Medicine

10.1111/j.1939-1676.2010.0648.x

Data were collected by 4 investigators according to a written protocol. Approval from the hospital ethics committee was obtained and informed owner consent waived because of the noninterventive nature of the study design. Potential predictive variables ($n = 41$) were selected a priori based on an anticipated relationship with mortality risk from existing canine and human scores, primary literature, and expert clinical opinion. Some variables, reported previously as predictive, were not collected because of concerns that they would limit future accessibility or applicability of the score or be liable to measurement error because of lack of measurement standardization among centers.

The variables assessed are presented in Table 1, and the measurement methodology in Table 2. Body cavity fluid scores and mentation scores were assigned as detailed in Table 3. All variables were assessed over the 1st 24 hours after ICU admission, with the exception of "mentation score," which was assessed at admission. For the remaining variables, the most abnormal values observed over the 24 hours after admission were recorded. No additional diagnostic testing beyond that required for clinical decision making was performed for the purposes of the study. For the blood pressure variable, blood pressure was measured by both direct and indirect manometry and Doppler flow signals. For Doppler techniques, mean arterial pressure (MAP) was recorded as the cuff pressure associated with the 1st return of an audible signal on cuff deflation.

At hospital discharge, outcome was recorded as mortality status. For cats that were euthanized, the reasons for euthanasia were reviewed in the written medical record and discussed with the primary clinician if clarification was required. Euthanized patients were subcategorized as euthanasia performed to relieve suffering in the face of anticipated imminent death, euthanasia performed due to current patient morbidity combined with diagnosis of a disease likely to be ultimately terminal, and euthanasia performed in the face of clinical illness and financial pressure. Euthanasia for any reason and non-euthanasia death were allotted equivalent status in the primary model build. For the validation process, the predictive performance of the score was evaluated for a group restricted to include noneuthanasia death only and then also for noneuthanasia death and various euthanasia subsets.

Table 1. Variables selected a priori for assessment of association with mortality risk.

Variable Type	Variable
Physical examination	Age, mentation score , EKG score , body cavity fluid score , obesity score, bodyweight , rectal temperature , heart rate , respiratory rate , MAP , urine output, SpO_2
Benchtop laboratory results	Glucose , lactate , PCV , TP , PvCO_2 , HCO₃ , pH , Na , Cl , K , iCa , ACT , PT , aPTT
Full laboratory results	Creatinine , urea , albumin , total bilirubin , WBC count , platelet count , CK , cholesterol
Historical information	Chronicity
Intervention	Provision of oxygen support , pressor support , mechanical ventilation support , nutrition score , urinary catheter placement, transfusion support

MAP, mean arterial pressure; iCa, ionized calcium; ACT, activated clotting time; PT, prothrombin time.

The variables in bold are those that were entered into the multivariable logistic regression procedure after elimination of variables with a univariable mortality association of $P > .2$ or with $> 30\%$ observations missing.

Table 2. Methods and measurement methodology for variables included in building the APPLE score multivariable models.

Variable	Measurement Methodology
Albumin, creatinine, total bilirubin, urea	Hitachi 911 automated chemistry analyzer, Hoffmann-La Roche Limited, Mississauga, Ontario, Canada
Glucose, lactate, Na, Cl, HCO ₃ , pH, K	ABL 800 Flex, Radiometer Medical Aps, Bronshøj, Denmark
Platelet count, WBC count	Advia 120 hematology analyzer, Siemens Healthcare Diagnostics, Deerfield, IL
Body cavity fluid score	Ultrasound assessment (FAS technique, see Table 3) with a Sonosite Titan SonoSite Canada Inc., Markham, Ontario, Canada. Assessments were performed as a routine component of triage by both interns and residents without specialized training in u/s. If medical history and physical exam did not trigger u/s assessment, the u/s score was recorded as 0
Mentation score	Clinical assessment as detailed in Table 3, applied at admission
PCV	Manual read from heparinized hematocrit tubes after centrifugation at 30,000 rpm for 3 minutes
TP	Refractometric measurement from total solids using heparinized plasma taken from hematocrit tubes spun at 30,000 rpm for 3 minutes and read on a temperature compensated hand-held refractometer
EKG score	Clinical score applied after visual assessment of EKG at admission or over the 1st 24 hours of ICU stay, where 1 = normal, 2 = abnormalities present but no specific pharmacotherapy applied, 3 = abnormalities present and pharmacotherapy applied
Rectal temperature	Most abnormal rectal temperature taken with digital thermometer
MAP	Most abnormal MAP recorded with either direct bp monitoring, or an automated indirect manometer Cardell 9401 (Thames Medical, Worthing, West Sussex, UK), or a hand held Doppler probe (Parks 811-b, Parks Medical Electronics Sales, Inc., Las Vegas, NV) with manual cuff inflation-first returning signal noise recorded as MAP ¹⁴
Nutrition score	0 = ate within 1st 24 hours of admission, 1 = anorexic, 2 = fed via an e-tube or PEG tube 3 = parenteral nutrition
Chronicity variable	Symptomatic for > 4 weeks with clinical signs associated with primary reason for admission

APPLE, Acute Patient Physiologic and Laboratory Evaluation; ICU, intensive care unit; MAP, mean arterial pressure.

Statistical Methods

Descriptive statistics were reported as mean and standard deviation where data were normally distributed, and a median and interquartile range where data were not normally distributed. Normality of data was tested with the Shapiro-Wilks test. Associations between mortality and categorical data were tested with Fisher's

exact test. Associations between mortality and continuous data were tested with the Student *t*-test where normally distributed, and the Mann-Whitney test where not normally distributed. Linearity assumptions were assessed and addressed as described below.

Score Development Methodology

The study population was randomly divided into a score construction or “training” cohort (3/4 of patients) and a score validation cohort (1/4 of patients).

Putative variables for which values were missing for over 30% of patients were excluded from further analysis. Consistent with established score development methodology, where variables had fewer than 30% missing values, the missing values were replaced with normal values where that made biological sense or mean values where normal did not apply. The variables “age” and “body weight” are instances where normal did not apply.^{7,8} Categorical and binary variables then were assessed for association with mortality as described above (Table 3).

The relationship between continuous variables and the binary mortality outcome was assessed by locally weighted scatter plot smoothing (LOWESS).^{15–17} This method overcomes the problem of determining the functional relationship of a continuous variable with an outcome when the outcome is binary. To obtain the smoothed value of *Y* at *X* = *x*, all the data having *x* values within a suitable interval about *x*, known as the bandwidth, are taken. A regression line is fitted to all of these points, with the points closer to the value of *x* being weighted in their contribution to the smoothed *y* value. The predicted value from this regression at *X* = *x* is taken as the estimate of *E*(*Y*|*X* = *x*). Smoothed values for *E*(*Y*) are obtained for each observed value of *X*. We selected a conservative bandwidth of 0.8 to minimize “wriggle” in the resulting functions. LOWESS lines can be fitted both on the original scale of *Y*, which assists ease of interpretation from a clinical interest perspective (showing for example the smoothed plot of observed percentage mortality against PCV) or on the logit scale.

After assessment of the linearity assumption on the logit scale of *Y*, continuous variables not fulfilling this assumption were categorized. The continuous variable was subdivided, for example, category 1 for temperature > 39.5°C, category 2 for temperature 38.5–39.5°C, category 3 for temperature 37.5–38.4°C, and each

patient then was assigned to the appropriate category for the variable. When subsequently entered into a logistic regression model, this allowed each category to receive the appropriate risk coefficient independent of the adjoining categories, thus bypassing the linearity assumption.¹⁸ We selected the points for divisions within each category divisors by percentiles of risk. For instance, for temperature, the LOWESS curve indicated a risk range of approximately 20–60% between the lowest and highest observed values. The range of 38.5–39.5°C appeared to represent the lowest risk category for that variable. Categories of 37.5–38.5°C and > 39.5°C were selected as representing the next highest risk categories, then 36.0–37.0°C, and finally < 36.0°C. After categorizing variables as indicated by the logit scale assessments, putative variables were entered into univariable logistic regression models to obtain a measure of statistical association between the variable and mortality outcome by the likelihood ratio test. The referent categories for each variable were set as those with the lowest mortality risk. Variables not associated with nonsurvival in the univariate analyses (*P* > .2) were dropped at this stage.

The remaining variables then were entered into a stepwise backward elimination procedure in the construction cohort to further eliminate variables with poor explanatory power in a multivariable context. The cut-off for retention was set at *P* < .05 by the likelihood ratio test. Bootstrapping with a sample size of *n* = 225 and 100 repetitions per iteration also was performed at each stage of the elimination procedure to attempt to improve the stability of the variable selection process given the relatively small sample size available. Variable collinearity was assessed by a symmetric matrix inversion routine, with variables dropped based on numerical accuracy assessments. After the backwards elimination procedure, the variables dropped were re-entered one by one back into the model and assessed for significance. Each variable then was assessed in a manual build process.

For the manual build, each model was developed in the construction or “training” cohort, and then assessed for area under the receiver operator characteristic curve (AUROC) performance and Hosmer-Lemeshow *C* statistic calibration in first the construction and then the validation cohort. Degrees of freedom were adjusted appropriately. Bayesian Information Criteria (BIC) for the non-nested models also were assessed in the construction cohort. An absolute difference in BIC of > 6 was taken as strong evidence of model improvement. Before final presentation, adjoining categories within variables were collapsed if the coefficient assigned to the category failed to achieve statistical significance compared with the referent category, which was again set as the category for each variable with the lowest risk. The models were checked by graphical examination of deviance residuals, leverage, and the Pregibon delta beta measure. Finally, the model discrimination and calibration characteristics also were assessed in subsets of the validation cohort that varied by cause of death.

The models were converted to an integer score as follows: The referent categories for each variable had been previously reset to the lowest risk category identified for each variable in the multivariable context. This had the effect of making all the coefficients for each category of each variable positive. Two desired maximal score values were empirically selected as 80 for the 8-variable model and 50 for the 5-variable model. For each model, the coefficient assigned to the category achieving the greatest magnitude (and thus reflecting the highest risk) for each variable was identified and summed across all the variables in the model. This value then was used as the divisor for the maximum score selected for the model. A multiplier was thus identified that when applied to the coefficient of the highest risk category for each variable would result in the sum of the maximum score achievable for the model equalling the desired maximum score. All the coefficients for each category of each variable then were multiplied by the appropriate multiplier for each model, and rounded to the nearest integer. These then constituted the integer scores assigned to each category in each variable. To check that the rounding process

Table 3. Algorithm for assignment of the body cavity fluid score and mentation score used for feline APPLE score calculation.

Mentation score: assessed at admission before sedation/analgesic administration	Body cavity fluid score (ultrasonographic evaluation, as assessed by FAST or TFAST technique ^a)
0 = normal	0 = no abdominal, thoracic or pericardial free fluid identified
1 = able to stand unassisted, responsive but dull	1 = abdominal OR thoracic OR pericardial free fluid identified
2 = can stand only when assisted, responsive but dull	2 = 2 or more of abdominal, thoracic, or pericardial free fluid identified
3 = unable to stand, responsive	
4 = unable to stand, unresponsive	

APPLE, Acute Patient Physiologic and Laboratory Evaluation.

^aBoysen, Rozanski et al. Evaluation of a FAST protocol to detect free abdominal fluid. JAVMA 2004;225:1198–1204.

Table 4. Population characteristics and outcome for feline intensive care unit admissions.

	Number	% of Group	Hospital Mortality %	<i>P</i> Value for Test of Mortality Difference between Case Group and Overall Hospital Mortality
Gender				
Male	339	56.5	26.0	
Female	261	43.5	25.7	
Source of admission				
Referral		74.1		
Primary care		25.9		
Admission type				
Emergency		28.8	29.2	
Planned admission		71.2	24.0	
Primary reason for admission				
GI/pancreatitis ^a	83	13.8	16.9	.045
Trauma	73	12.2	19.2	.165
Spinal/peripheral neuropathy	7	1.2	14.3	.484
Medical oncology ^a	70	11.7	50.0	.000
Immune mediated disease	20	3.3	15.0	.261
Surgical oncology	28	4.7	17.9	.327
Respiratory	56	9.3	26.8	.858
Intracranial disease	18	3.0	44.4	.068
Renal/urinary tract	74	12.3	21.6	.377
Cardiac	45	7.5	26.7	.895
Hepatic ^a	28	4.7	42.9	.035
Sepsis	23	3.8	39.1	.138
Ophthalmological	4	0.7	25.0	.971
Toxicity	12	2.0	16.7	.466
Endocrine disorder	8	1.3	0.0	.117
Diabetic	35	5.8	14.3	.109
Reproductive	3	0.5	0.0	.306
Miscellaneous	13	2.2	15.4	.389
Mortality-total deaths all cause	155	25.8		
Type of death				
Euthanasia due to				
Current severity of illness	82	52.9		
Terminal disease diagnosis	29	18.7		
Financial constraints	17	11.0		
Noneuthanasia death	27	17.4		
Exclusions due to age/unavailable records	6			

^aMortality difference statistically significant at $P < .05$ by a 2-tailed chi-square test of difference in proportion tested against population as a whole.

had not resulted in significant loss of information, the score for each model for each patient was calculated, and AUROC discrimination and calibration statistics were evaluated by checking a univariable logistic regression model based on each of the scores.

All analyses were performed in Stata version 10.1.^a Statistical significance was set at $P < .05$ unless otherwise stated.

Results

Study Population Characteristics

The study population consisted of 606 cats consecutively admitted to the ICU over the 32-month study period. Six patients subsequently were excluded because of inconsistencies in identification and consequently missing or uncertain outcome information. Cats constituted 18% of total ICU admissions over the study period. Death occurred in 155 cats from the study population of 600 cats. Over the study period, 74% of cats admitted to the ICU were admitted after referral appointments, whereas 26% of admissions followed 1st opinion assessments. Emer-

gency or unplanned admissions constituted 29% of the admissions, whereas the remainder of admissions followed scheduled appointments.

The primary diagnostic categories and mortality characteristics of the study population are shown in Table 4. The 3 most common reasons for ICU admission in this group of patients included GI disease or pancreatitis, renal or urinary tract disease, or trauma. When mortality rates within the groups were tested against overall hospital mortality, patients with GI disease or pancreatitis had significantly lower mortality (16.9%, $P = .045$) and patients admitted with a medical oncological problem had significantly higher mortality (50.0%, $P < .001$) compared with the patient group as a whole (25.8%).

Of the 155 deaths, 35 deaths (22.6%) occurred within the medical oncology case group. For deaths within this group, 2 cats died, and 19 were euthanized to relieve suffering in the face of anticipated imminent death. The remaining 14 cats euthanized in this case group were judged to be clinically stable, but were euthanized at the

owner's request because of perceived poor overall quality of life and diagnosis of a terminal disease. The overall mortality rate for the study population ($n = 600$) was 25.8% ($n = 155$), of which 17.4% ($n = 27$) of cats died and 71.6% ($n = 111$) were euthanized either in association with anticipated imminent death, or in the face of clinical morbidity and diagnosis of terminal disease. For 11.0% of deaths ($n = 17$) the primary clinician listed financial constraints as a primary motivating feature in the euthanasia. The presenting complaints for these 17 cats were as follows: persistent seizures, magnetic resonance imaging declined ($n = 3$); traumatic diaphragmatic hernia, surgery declined ($n = 2$); multiple pelvic fractures, surgery declined ($n = 4$); renal failure, hospitalization declined ($n = 2$); abdominal effusion and mass effect, biopsy declined ($n = 1$); recurrent feline lower urinary tract disease and urethral obstruction, catheterization declined ($n = 2$); suspect hepatic failure, diagnostic work-up declined ($n = 1$); septic peritonitis, surgery declined ($n = 1$); granular lymphocytic lymphoma, chemotherapy declined ($n = 1$).

Median time to death was 3 days, with a range of 0.5–12 days. ICU stay was on average 1 day shorter for non-survivors compared with survivors (median of 2 days versus 3 days, $P < .001$). There was no difference in mortality risk between construction and validation patient groups (26.0 versus 25.3%, $P = .87$). The mortality risk associations for the putative variables, together with the data availability, are shown in Table 5. The AUROC characteristic for each variable as a univariable mortality predictor also is shown.

Although many variables had a statistically significant association with mortality, the 4 variables with the greatest discriminant capacity as assessed by AUROC analysis were urea (0.78), chloride (0.77), body temperature (0.71), and lactate (0.70). No significant association was identified between chloride corrected for sodium and mortality. Cholesterol, ionized calcium (iCa), and activated clotting time (ACT) were also associated with mortality, although for ACT and iCa the availability of data limited entry into the multivariable model build. The exploratory LOWESS curves for these variables are shown in Figure 1. For a subgroup of cats ($n = 139$) with ACT in the range 90–150 seconds, each 10-second increase in ACT was associated with an estimated 50% increase in mortality odds (OR, 1.48; 95% CI, 1.09–2.00; $P = .01$). For a subgroup of cats ($n = 163$) with an iCa < 1.2 mmol/L, each decrease in iCa of 0.1 mmol/L was associated with an estimated 40% increase in mortality odds (OR, 1.39; 95% CI, 1.06–1.83; $P = .02$). For a subgroup of cats ($n = 25$) with a cholesterol < 2.3 mmol/L, each decrease in cholesterol of 0.1 mmol/L was associated with an estimated 30% increase in mortality odds (OR, 1.35; 95% CI, 1.02–1.80; $P = .04$).

Details of Score Development

The coagulation time variables (ACT, prothrombin time, and partial thromboplastin time), and the variables “urine output (in mL/kg/h)” and “SpO₂ (%)” were dropped because of lack of available data on $> 30\%$ of

patients. The mortality functions derived for selected putative continuous variables by the LOWESS method are shown in Figure 2.

After categorization of variables demonstrating lack of linearity in the logit, a total of 37 variables were assessed for univariable association with mortality. Nine variables demonstrated a univariable lack of association with mortality risk ($P > .2$) and were dropped. Twenty-eight variables were entered into the stepwise backwards elimination procedure. Ten variables were selected at the end of this process (bicarbonate, temperature, urea, body cavity fluid score, chloride, mentation score, PCV, creatinine, MAP, and lactate). The variables “bicarbonate” and “creatinine” were removed in the manual build to improve calibration of the model in the validation cohort, with minimal loss in performance, resulting in the final 8-variable model. A 5-variable model, consisting of the variables from the group that are typically the most accessible by “benchtop” methods (temperature, mentation score, PCV, MAP, and lactate), also was evaluated.

Risk Stratification Scores

The performance characteristics of the final models are presented in Table 6. The feline Acute Patient Physiologic and Laboratory Evaluation (APPLE_{full}) model had excellent performance characteristics for discrimination, with AUROCs of 0.91 (95% asymptotic CI, 0.87–0.94) on the training cohort and 0.88 (95% asymptotic CI, 0.84–0.96) on the validation cohort. The feline APPLE_{fast} model had good discrimination characteristics, with an AUROC of 0.83 (95% asymptotic CI, 0.79–0.86) on the training cohort and an AUROC of 0.76 (95% asymptotic CI, 0.72–0.84) on the validation cohort. Calibration was good for both models in both cohorts.

Discrimination characteristics also were assessed for each model in the validation data set after sequential censoring of patients to which the various categories of euthanasia applied, and finally with the deaths within the validation data set restricted only to patients that arrested spontaneously. Discrimination improved slightly for both models when patients undergoing financially driven euthanasia were excluded, improved again when patients undergoing financially driven and terminal disease euthanasia were excluded, and improved markedly when patient deaths were restricted solely to patients experiencing spontaneous arrest. Calibration was retained in all validation subsets, except when deaths were restricted to patients experiencing spontaneous arrest, at which time the overall mortality rate was significantly lower than that predicted by the models, with fewer observed deaths in the lower risk categories.

The models showed minimal loss of performance after conversion to integer score format, despite the associated rounding error, with the APPLE_{full} score AUROC = 0.90 (asymptotic 95% CI, 0.87–0.93) and APPLE_{fast} score AUROC = 0.82 (asymptotic 95% CI, 0.76–0.84) on the training data set. Calibration also remained good with Hosmer-Lemeshow C statistic P values of .22 and .82, respectively. The final risk models, presented after conversion to integer scores, are shown in Figures 3 and 4.

Table 5. Univariable association between putative variables and mortality in cats admitted to the ICU of a referral hospital.

Variable	Number of Cases with Information Available	Cats that Lived Mean±SD or Median (IQR) or %	Cats that Died, Mean±SD or Median (IQR) or %	P Value	AUROC	Overall Mean±SD or Median (IQR) or %
ICU stay ^a	600	3 (2–4)	2 (1–4)	<.001		2 (1–4)
Age	596	8 (3–12)	8 (4–12)	.394	0.54	8 (3–12)
Chronic dx ^a	600	42%	28%	.002	0.57	38.3%
Nutritional status	600				0.53	
Normal		62.2%	67.1%	.275		
Anorexic		26.1%	24.5%	.695		
Feeding tube		11.4%	7.7%	.195		
Parenteral		0.2%	0.6%	.436		
Urinary catheter placement	600	10.6%	11.0%	.888	0.50	10.7%
Oxygen support ^a	600	19.6%	40.6%	.000	0.60	25%
Pressor support ^a	600	1.1%	7.1%	.000	0.53	2.7%
Transfusion support	600				0.51	
FFP		4.5%	3.2%	.485		4.2%
PRBC/whole blood		5.4%	4.5%	.662		5.2%
Any product		9.9%	7.7%	.417		9.3%
Mechanical ventilation sup- port ^a	600	0.2%	7.7%	.000	0.54	2.2%
Body weight (kg)	600	4.8±2.2	5.0±3.1	.254	0.55	4.9±2.4
Body temperature ^a (°C)	569	38.5 (37.6–39.2)	37.4 (36.4–38.9)	.000	0.71	38.4 (37.2–39.1)
Heart rate ^a (bpm)	587	200±36	189±42	.002	0.57	198±38
Respiratory rate ^a (bpm)	579	44±19	48±23	.090	0.57	45±20
MAP ^a (mmHg)	444	109±27	95±29	.000	0.68	105±28
Urine output (mL/kg/h)	42	2.1 (1.0–3.1)	1.8 (0.8–2.7)	.504		1.9 (1.0–3.0)
Body condition score (1–5)	533	2.9±0.9	2.8±1.0	.65	0.53	2.9±1.0
Mentation score ^a (0–4)	587	0 (0–1)	1 (0–2)	.000	0.67	0 (0–1)
EKG score ^a	561				0.52	
0		95.5%	90.8%	.037		94.1%
1		3.6%	7.1%	.082		4.6%
2		1.0%	2.1%	.008		1.3%
Body cavity fluid score ^a	590				0.59	
0		86.9%	68.2%	.000		82.2%
1		10.9%	26.4%	.000		14.7%
2		1.4%	5.4%	.006		2.4%
3		0.9%	0%	.247		0.7%
SpO ₂ (pulse oximeter)	31	98% (96–99)	97% (94–99)	.477		98% (94–99)
PvCO ₂ (mmHg)	546	41.5 (37.5–46)	40.7 (35.7–46.3)	.491	0.57	41.3 (37.1–46.1)
HCO ₃ (mmol/L) ^a	546	19.7±4.4	18.3±5.5	.001	0.61	19.3±4.7
pH ^a	546	7.31 (7.26–7.34)	7.27 (7.19–7.34)	.000	0.66	7.30 (7.24–7.34)
Na (mmol/L) ^a	546	153±5.2	150±7.7	.000	0.62	152±6.1
Cl (mmol/L) ^a	546	120±6.3	117±8.9	.000	0.77	119±7.2
Corrected Cl (mmol/L)	546	118±5.3	117±6.7	.322	0.51	118±5.7
K (mmol/L) ^a	546	3.7 (3.4–4.1)	3.8 (3.3–4.6)	.000	0.62	3.8 (3.4–4.2)
Glucose (mmol/L) ^a	548	8.2 (6.3–10.7)	7.8 (5.9–10.7)	.002	0.61	8 (6.2–10.7)
Lactate (mmol/L) ^a	546	1.9 (1.3–2.7)	2.3 (1.5–4.3)	.000	0.70	2.0 (1.3–2.9)
iCa (mmol/L) ^a	401	1.26 (1.20–1.32)	1.22 (1.14–1.41)	.001	0.67	1.25 (1.18–1.32)
PCV (%)	546	31±10.1	31±9.9	.180	0.56	31±10.0
TP (g/dL) ^a	532	7.1±1.24	6.9±1.59	.000	0.63	7.1±1.34
ACT (seconds) ^a	187	116 (105–134)	128 (116–172)	.001	0.65	119 (106–147)
Creatinine (μmol/L)	510	111 (80–155)	107 (80–179)	.113	0.18	110 (80–163)
CK	510	460 (210–1876)	529 (269–1818)	.716	0.51	472 (221–1834)
Urea (mmol/L) ^a	510	8.9 (6.9–15.3)	9.8 (6.1–18.7)	.000	0.78	9.1 (6.8–15.9)
Albumin (g/L) ^a	510	33±5.9	31±8.2	.027	0.59	33 (6.5)
Total bilirubin (μmol/L) ^a	510	3 (2–5)	3 (2–8)	.030	0.58	3 (2–6)
Cholesterol ^a	510	3.6 (2.9–4.7)	4.0 (3.3–5.1)	.046	0.18	3.8 (3.0–4.9)
WBC count	462	10.9 (7.4–16.0)	11.6 (7.7–14.9)	.071	0.58	11.3 (7.5–15.8)
Platelet count	462	231 (152–338)	232 (146–377)	.075	0.57	232 (150–349)

ICU, intensive care unit; AUROC, area under the receiver operator characteristic curve; IQR, interquartile range; MAP, mean arterial pressure; ACT, activated clotting time.

^aVariables associated with mortality risk with $P < .05$ in univariable analysis.

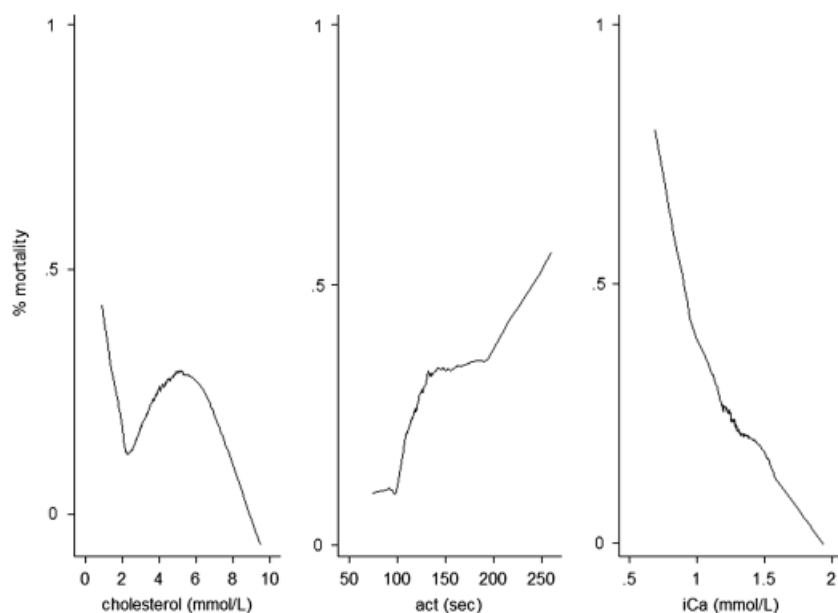


Fig 1. LOWESS curves to show the relationship between cholesterol, ACT, and iCa and mortality risk for cats admitted to the ICU of a referral hospital. LOWESS, locally weighted scatter plot smoothing; ACT, activated clotting time; ICU, intensive care unit; iCa, ionized calcium.

The scores in conventional (US) units are shown in Appendix 1, as Figures A1 and A2. Body cavity fluid scores and mentation scores were assigned as detailed in Table 3.

The central cell in each figure represents the range of values for the variable for which 0 points would be as-

signed. The cells to either side show the appropriate score for the corresponding range of the variable. The final score for the patient is achieved by summing the scores for each variable. Reflecting the underlying multivariable logistic model, no variables can be omitted

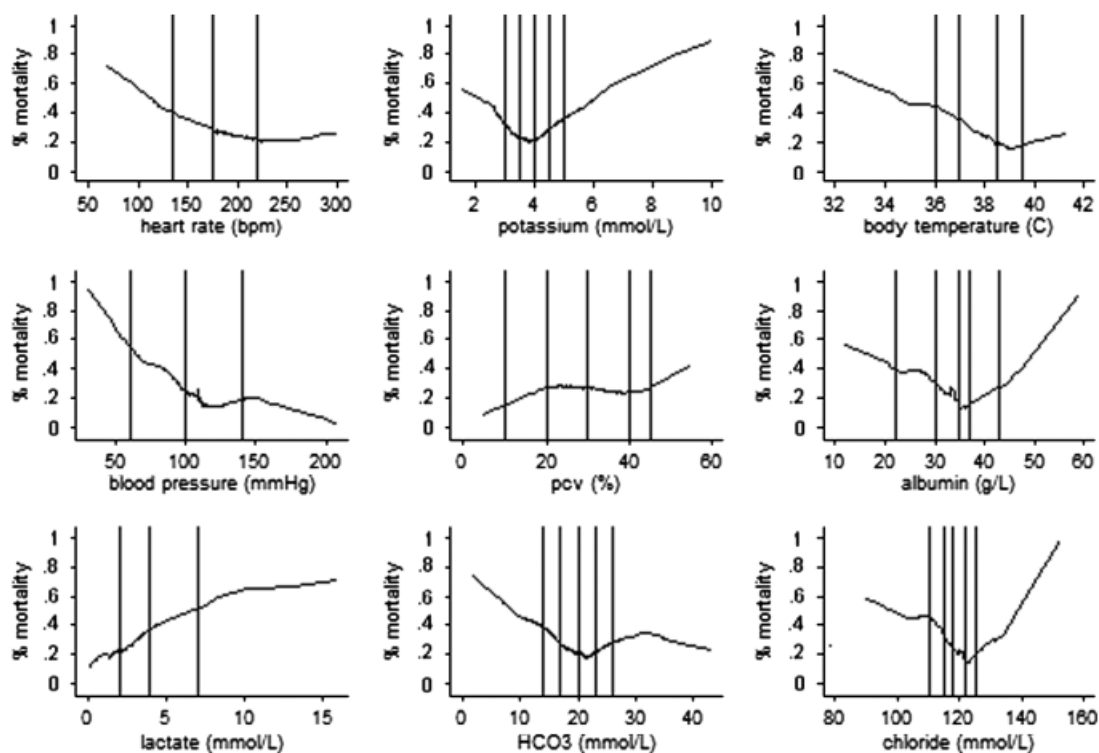


Fig 2. LOWESS curves to show the relationships between mortality risk and selected continuous variables in a population of cats admitted to the ICU of a referral hospital. The vertical lines mark the category limits subsequently applied when the variables were categorized, before entry into a logistic regression model. LOWESS, locally weighted scatter plot smoothing; ICU, intensive care unit.

Table 6. Performance characteristics of the feline APPLE models.

Data Set Restrictions	AUROC for AP- PLE _{full} Model	Hosmer-Lemeshow <i>C</i> Statistic	AUROC for AP- PLE _{fast} Model	Hosmer-Lemeshow <i>C</i> Statistic
Construction data set	0.91	$\chi^2_8 = 9.90, P = .27$	0.83	$\chi^2_8 = 0.72, P = .99$
Validation data set	0.88	$\chi^2_{10} = 14.88, P = .14$	0.76	$\chi^2_{10} = 9.93, P = .45$
Validation data set, financial euthanasia censored	0.89	$\chi^2_{10} = 15.22, P = .12$	0.78	$\chi^2_{10} = 14.48, P = .15$
Validation data set, financial and terminal disease euthanasia censored	0.90	$\chi^2_{10} = 15.22, P = .12$	0.78	$\chi^2_{10} = 14.79, P = .14$
Validation data set, all euthanasia censored	0.93	$\chi^2_{10} = 22.64, P = .01$	0.80	$\chi^2_{10} = 41.48, P \leq .01$

AUROC, area under the receiver operator characteristic curve; APPLE, Acute Patient Physiologic and Laboratory Evaluation.

when calculating the score for a patient because the score assignments for the remaining variables would change, resulting in inaccuracy. The relationship between each variable and mortality risk is dependent on the other variables within the model. The integer scores can be converted to mortality risk probabilities (range, 0–1) by the equations shown in Figure 5, where P is the mortality risk probability and R is logit (P). Figure 6 depicts the relationship between the feline APPLE scores and the predicted probability of mortality. An example of score calculation is shown in Figure 7.

Discussion

The objective of this study was to construct and validate a robust mortality risk prediction score for hospitalized cats. This would facilitate assignment of an objective score reflecting severity of illness to any cat enrolled in clinical research. The scores we developed had a high degree of predictive accuracy when assessed in a group of patients independent to those used for the model build, with AUROCs of 0.88 and 0.76, and no loss of calibration. The scores are transparent and easy to

calculate manually. The scores are based on variables that are inexpensive to measure and readily available. The full 8-variable score performed better than the fast 5-variable score based on the AUROC 95% confidence intervals.

The statistical methodology used was crucial to the success of this project, and some areas of the statistical process justify further discussion. We elected to randomly divide our patient group into a construction and validation cohort. This was done to allow validation of the final scores on a patient group independent of that used for construction. Consistent with modeling practice, the number of deaths in the model construction group ($n = 115$) suggested that the final model would be appropriately limited to a maximum of 10 variables to avoid problems with model instability.¹⁵ One of the assumptions of a logistic regression model is that the putative variables have a linear relationship with nonsurvival when nonsurvival is plotted on the logit scale. Therefore variables that are nonlinear must be transformed before model inclusion, or false conclusions about the strength of associations may be drawn. Many clinical variables may be intuitively anticipated to have a nonlinear (typically quadratic) association with mortality risk. For

				Mentation score	4	7	8	9		
				0	1	2	3	4		
	6	4	3	Temperature (C)	1					
	<36.1	36.1-37.0	37.1-38.5	38.6-39.4	>39.4					
9		4		MAP (mmHg)	1					
<61		61-100		101-140	>140					
				lactate (mmol/L)	5	6	9			
				0-1.9	2.0-4.0	4.1-7.0	>7.0			
				PCV(%)	11	16	14	13	17	
				<11	11-20	21-30	31-40	41-45	>45	
11		7		Urea (mmol/L)	12	7	6			
<5.3		5.3-7.6		7.7-8.9	9.0-11.6	11.7-24.9	>24.9			
12		9	11	Chloride (mmol/L)	11	7				
<111		111-115	116-118	119-122	123-125	>125				
				Body cavity fluid score	3	6				
				0	1	2				

Fig 3. Feline Acute Patient Physiologic and Laboratory Evaluation (APPLE_{full}) score: Calculated by summing the value in the upper left corner of the appropriate cell for each of the 8 parameters listed, with a maximum potential score of 80. The central cell in each figure represents the range of values for the variable for which 0 points would be assigned. The cells to either side show the appropriate score for the corresponding range of the variable. The final score for the patient is achieved by summing the scores for each variable. See Table 3 for calculation of “fluid score” and “mentation score.” “Mentation score” is collected at admission, for all others utilize the most abnormal value identified over the 24-hour period following admission. If history and physical exam fail to prompt assessment of fluid score, assign zero. Mean arterial pressure (MAP) reflects the value recorded for MAP if direct or indirect manometry is used, or the pressure corresponding to 1st return of audible signal on cuff deflation if Doppler is used.

				Mentation score	5	6	7	10
				0	1	2	3	4
6	5	4	3	Temperature (C)	1			
<35.6	35.6-36.0	36.1-37.0	37.1-38.5	38.6-39.5	>39.5			
11		7		MAP (mmHg)	1			
<61		61-100		101-140	>140			
				lactate (mmol/L)	6	9	10	
				0-1.9	2.0-4.0	4.1-7.0	>7.0	
				PCV(%)	12	10	9	13
				<16	16-25	26-35	36-45	>45

Fig 4. Feline Acute Patient Physiologic and Laboratory Evaluation (APPLE_{fast}) score: Calculated by summing the value in the upper left corner for each of the 5 parameters listed, with a maximum potential score of 50. The central cell in each figure represents the range of values for the variable for which 0 points would be assigned. The cells to either side show the appropriate score for the corresponding range of the variable. The final score for the patient is achieved by summing the scores for each variable. “Mentation score” is collected at admission (see Table 3), for all others utilize the most abnormal value identified over the 24-hour period following admission. Mean arterial pressure (MAP) reflects the value recorded for MAP if direct or indirect manometry is used, or the pressure corresponding to 1st return of audible signal on cuff deflation if Doppler is used.

instance, body temperature would be expected to be associated with increased mortality risk at extremes of both high and low temperature. This clinical expectation was confirmed in our exploratory data plots (see Fig 1). A number of methods are available for modeling nonlinear variables, including fractional polynomials, linear or cubic splines, or categorization of the continuous variable. After exploration of the various options, we elected to use categorization. Categorization of nonlinear continuous variables is a time-honored approach in medical statistics, and has been used in most of the major risk prediction models used in humans to date.⁶⁻⁸ Although there are several disadvantages with this approach, including the potential loss of information, these are outweighed by the advantages. When the population modeled is large enough, these advantages include flexibility, robustness to spurious risk assignment, and ease of final model interpretation and presentation.

Two important points pertaining to score use deserve emphasis. First, the scores have been constructed and validated on patient data obtained over the initial 24 hours after hospital admission, for patients requiring IV fluid support or close monitoring. As such, the prospective calculation of these scores from patient data limited only to admission, from patient data from days later in the hospital stay, or for outpatients, is not appropriate. The performance of the scores when calculated from data collected over a time period different than that used in this study has not been assessed. Secondly, these scores

are designed to risk stratify populations, not to prognosticate individual patients. The confidence intervals for the scores are wide, as shown in Figure 6, with an APPLE_{full} score of 40 reflecting a mortality risk probability of anywhere from 30 to 50%. Basing a euthanasia recommendation on these scores is likely to result in the euthanasia of an unacceptably large proportion of patients that could in fact have lived, whatever cut point is chosen. Development of guidelines for the appropriate use of the equivalent human scores has resulted in the following broad recommendations. Use of scores for the implementation of withdrawal of care protocols is contraindicated. Use of scores to stratify patients to facilitate triage protocols, for instance scheduling of procedures, is acceptable. Use of scores to assist patient-adjusted clinician performance benchmarking is acceptable, as is use of scores in the context of clinical research. These recommendations appear to apply equally well in veterinary medicine.¹

One of the goals of this study was to develop a score that could achieve a high uptake prospectively across a wide range of treatment centers. In line with this aim, we attempted to select widely available variables for evaluation. We also selected a method of score presentation that facilitated an intuitive and easy approach to patient score calculation, with minimal math. To our knowledge, this particular method of regression model presentation has not been previously reported in veterinary medicine, and we have reported the statistical methods in detail for that reason. Illness severity scores have so far achieved relatively limited acceptance in veterinary medicine, with comments from recent studies indicating that they contain too many variables to be easily calculated, particularly retrospectively.¹⁹ We hoped to circumvent this complaint by presenting a 5-variable score based on only benchtop and clinical variables. All variables were assessed as the “most abnormal” over the 24-hour period after admission, with the exception of mentation score. This was done to gain the best possible assessment of patient baseline status before administration of analgesics or sedation.

A particular challenge in the development of veterinary risk prediction models is the prevalence of euthanasia as the predominant mortality outcome in vet-

Where p =mortality probability and R =logit p

feline APPLE_{full} score: $p_1 = \exp(R_1) / (1 + \exp(R_1))$

$$R_1 = (0.213 \times \text{APPLE}_{\text{full}}) - 9.472$$

feline APPLE_{fast} score: $p_2 = \exp(R_2) / (1 + \exp(R_2))$

$$R_2 = (0.203 \times \text{APPLE}_{\text{fast}}) - 6.142$$

Fig 5. Equations relating the feline Acute Patient Physiologic and Laboratory Evaluation (APPLE) scores to mortality risk probability (P), with a mortality probability of 0–1.

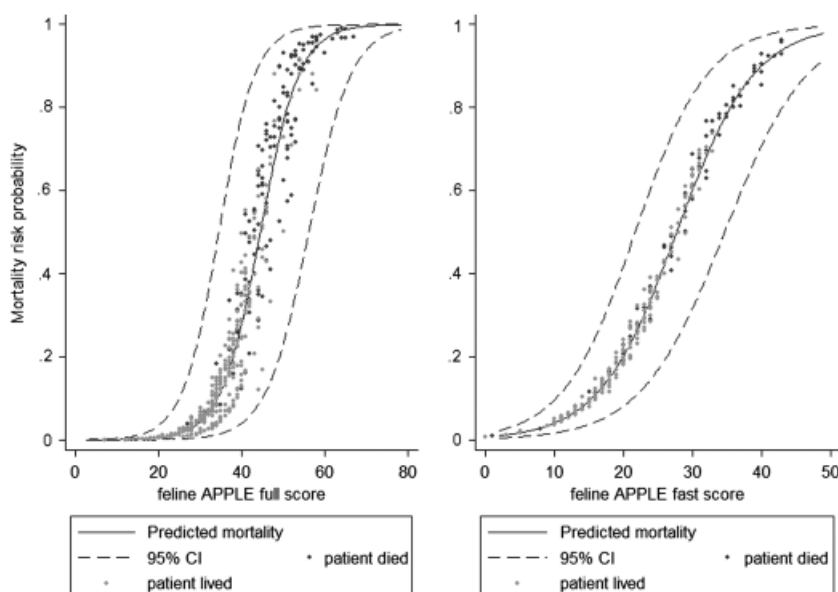


Fig 6. The relationship between the feline Acute Patient Physiologic and Laboratory Evaluation (APPLE) scores and predicted mortality risk probabilities, with 95% CIs. The underlying scatter plots show the scores calculated for the patients in the combined construction and validation data sets, with the dark gray dots representing patients who died, and the lighter gray dots representing patients who lived.

erinary patient populations. In the referral hospital setting, the request for euthanasia of the animal by the client typically is made on the basis of information and opinion received from the clinician. Thus, if the clinician perceives a particular clinical state, for instance, hypotension, as negatively prognostic, and this concern is relayed to the client resulting in a euthanasia decision, then an association between hypotension and death will be created whether this association truly exists or not. Results then may be biased toward a false positive association between a variable and mortality. This quandary is not unique to veterinary medicine. In human neonatal and adult ICUs, between 50 and 90% of deaths occur in association with withdrawal and withholding of care.^{20,21} Administration of opioids and sedatives has been shown to increase after care withdrawal orders.^{20,21} When patients or families are involved in the decision-making process, clinicians have been shown to withhold information on therapies they regard as futile.²² This situation has many parallels with that faced in veterinary medicine with the issue of clinician-informed, client-requested end-of-life care. The approach typically taken in human mortality risk prediction studies is to assume that patients dying after withdrawal of treatment, for instance, discontinuation of mechanical ventilation, would have ultimately died had treatment been maintained. Risk prediction models then have been calculated to date with no differentiation between patients that died in the face of maximal treatment and those that died after withdrawal of treatment. Despite this issue, the human risk prediction models have constituted a robust and highly valuable clinical research tool for many years.

The best known diagnosis-independent illness severity score available in veterinary medicine to date is the Survival Prediction Index.¹¹ For the development of this score, animals that died or were euthanized for any rea-

son were modeled without differentiation, and the prevalence of euthanasia within the group was not reported. We elected to take a slightly different approach. We acknowledged that the purest form of score development would take place on a patient population where the only patients dying would be those experiencing spontaneous death in the face of maximal intervention. However, this population does not exist in medicine. The performance of humane euthanasia to relieve unnecessary suffering in moribund animals considered to be awaiting death is widespread within the culture of veterinary medicine, and appropriately so. To exclude euthanized patients from clinical research risks promoting the ethically unacceptable practice of withholding humane euthanasia for dying animals, as well as generating a study population that does not reflect the target population we work with day to day. Exclusion of euthanized patients obviously was not a practical approach for a study of this type. Instead, we elected to categorize the primary reasons for euthanasia in each case, and perform a cross-validation of the model while censoring successive euthanasia categories, with the final validation reported on a group containing only cats that experienced noneuthanasia death. We found that the discrimination of our score steadily improved as this was done, with AUROCs increasing from 0.88 to 0.93 and 0.76 to 0.80. This suggests that the variables and coefficients assigned in the training data set are truly reflective of mortality probability rather a “euthanasia risk” based on false premises. Calibration was lost in the final “noneuthanasia death only” group, reflecting the higher all-cause mortality in the construction data set relative to mortality restricted to noneuthanasia deaths only in the validation data set. In reflection of the complex decision making involved in euthanasia, difficulty selecting the primary category was occasionally experienced. Re-

‘Rusty’ 6 year old DSH admitted to hospital with acute onset dyspnoea, maintained on oxygen support and IV fluids

Mentation score at admission: Unable to stand, responsive- score 8

Parameter	Most abnormal value detected over the first 24 hours of admission	APPLE _{full} Score
Rectal temperature (°C)	37.0	3
Mean arterial pressure (mmHg) (Doppler)	80	4
Lactate (mmol/L)	3.2	5
PCV (%)	18	11
Urea (mmol/L)	15.2	7
Chloride (mmol/L)	121	0
Body cavity fluid score	Thoracic free fluid present	3

Total APPLE_{full} score= 8+3+4+5+11+7+0+3= 41/80

Conversion to mortality probability (p) using equations in Figure 5:

$$R=(0.213 \times 41) -9.472= -0.739$$

$$p=\exp(-0.739)/(1+\exp(-0.739))=0.32$$

Thus in the range 0 to 1 estimated mortality probability for this patient is 0.32 or 32% risk of non-survival to discharge.



Fig 7. Example of patient Acute Patient Physiologic and Laboratory Evaluation (APPLE_{full}) score calculation and conversion to mortality probability (*P*).

flecting this, euthanasia for any reason and noneuthanasia death were allotted equivalent status in the primary model build, and the most conservative AUROCs for the validation cohort reported in the abstract.

Our cross validation results suggest that the scores proposed should provide a reliable assessment of individual illness severity, and truly reflect underlying mortality probabilities. In common with all scores, however, they may under- or over-predict mortality rates for the group as a whole if applied prospectively to populations with significantly lower or higher mortality rates than those reported for our populations. Because the primary intended use for these scores is to offer an objective measure of baseline illness severity for comparison of groups enrolled in clinical research, we do not regard this as a serious issue for prospective use. Provided the study groups are from the same underlying primary population, loss of calibration is unlikely to bias group comparisons.

The coefficients assigned by the multivariable model to some of the variables had construct validity; for instance, the steadily increasing scores for increasing lactate and decreasing blood pressure make clinical sense. However, the scores assigned in a multivariable model may be very different to those anticipated in the clinically familiar

univariable context. The risk coefficients in a multivariable model reflect the influence of a variable on the outcome after taking into account the simultaneous influence of all the other variables in the model. This key feature may require some additional explanation. As a hypothetical example, if the simultaneous influence of lactate and pH on mortality risk was modeled, it is likely that some, but probably not all, of the influence of high lactate on mortality risk would be mediated through low pH. If both variables were placed in a multivariable model, this would result in the coefficients for lactate reflecting only the direct effect of lactate on mortality risk independent to that mediated through low pH. Thus, as more variables enter the model, some coefficients that may not appear clinically intuitive are likely to arise. This will be particularly true for confounding, intervening, or distorter variables that achieve their predictive utility by refining the model after the impact of the primary predictive variables has occurred.¹⁵ We suspect PCV and chloride may fall into these categories.

There are several areas of weakness in this study. The first is the relatively small number of cats enrolled. Cats do not form as large a group as dogs in our hospital population, and this is reflected in this study. The human predictive models typically are calculated from several

thousand patients from many different hospitals, and this results in a degree of model stability that we cannot hope to replicate here. Despite statistical techniques such as model checking and bootstrapping designed to improve stability as far as possible, considerably higher patient numbers would have been desirable. On a similar note, we were unable to validate our model on a completely independent population from a different hospital or group of hospitals. Ideally, both model construction and validation data sets should be multicenter, again to improve model stability when used prospectively. The 8-variable model showed better predictive performance than the 5-variable model, and as such would be the optimal model to use wherever possible. The 5-variable model has the advantage that it requires only “benchmark” or clinical variables, and therefore may be more accessible in some circumstances.

Additional concerns include our use of Doppler instrumentation to assess blood pressure in our patients. We took the point of 1st return of sound to indicate the MAP in our patients. Although this has been shown to be true in anesthetized cats,¹⁴ it has not been validated in awake cats to date. Another potential weakness was the use of ultrasound examination to assess for the presence of free fluid. Variation in equipment sensitivity and operator skill may decrease the repeatability of this variable. Finally, for euthanized patients the primary reason for euthanasia was assigned on the basis of the subjective impression of the primary clinician; owners were not questioned directly. This may have resulted in misclassification error for some patients.

In conclusion, we developed and presented in a user-friendly manner 2 scores that offer a robust measure of illness severity and facilitate objectification of illness severity in clinical research. The scores have been validated on patient groups independent to that used in construction, and are free from euthanasia bias. The 8-variable model offers optimal discrimination, but use is dependent on the availability of a chemistry profile and the use of ultrasound as a routine triage tool. The 5-variable model offers good discrimination, and the variables may be more readily accessible. Both scores calibrate well. Independent multicenter validation is a goal for the future.

Footnote

^a StataCorp LP, College Station, TX

Acknowledgment

The study was supported by a grant from the Ontario Veterinary College Pet Trust.

References

- Higgins TL. Quantifying risk and benchmarking performance in the adult intensive care unit. *J Intensive Care Med* 2007;22:141–156.
- Harrison DA, Parry GJ, Carpenter JR, et al. A new risk prediction model for critical care: The intensive care national audit and research centre (ICNARC) model. *Crit Care Med* 2007;35:1091–1098.
- Hayes G, Mathews K, Kruth S, et al. Illness severity scores in veterinary medicine: What can we learn? *J Vet Intern Med* 2010; 24:457–466.
- Wunsch H, Brady AR, Rowan K. Impact of exclusion criteria on case mix, outcome, and length of stay for the severity of disease scoring methods in common use in critical care. *J Crit Care* 2004;19:67–74.
- Le Gall JR. The use of severity scores in the intensive care unit. *Intensive Care Med* 2005;31:1618–1623.
- Le Gall JR, Loirat P, Alperovitch A, et al. A simplified acute physiology score for ICU patients. *Crit Care Med* 1984;12: 975–977.
- Knaus WA, Wagner DP, Draper EA, et al. The APA CHE III prognostic system: Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991;100:1619–1636.
- Zimmerman JE, Kramer AA, McNair DS, et al. APACHE IV- hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006;34:1297–1310.
- Higgins TL, Teres D, Nathanson B. Outcome prediction in critical care: The mortality probability models. *Curr Opin Crit Care* 2008;14:498–505.
- King LG, Stevens MT, Ostro EN. A model for prediction of survival in critically ill dogs. *J Vet Emerg Crit Care* 1994;4: 85–99.
- King LG, Wohl JS, Manning AM, et al. Evaluation of the survival prediction index as a model of risk stratification for clinical research in dogs admitted to intensive care units at four locations. *Am J Vet Res* 2001;62:948–954.
- Rohrbach BW, Buchanan BR, Drake JM, et al. Use of a multivariable model to estimate the probability of discharge in hospitalised foals that are seven days of age or less. *J Am Vet Med Assoc* 2006;228:1748–1756.
- Hayes G, Mathews K, Doig G, et al. The canine acute patient physiologic and laboratory evaluation (APPLE) score- a severity of illness stratification system for hospitalised dogs. *J Vet Intern Med* 2010;24:1034–1047.
- Caulkett NA, Cantwell, Houston DM. A comparison of indirect blood pressure monitoring techniques in the anesthetized cat. *Vet Surg* 1998;27:370–377.
- Dohoo I, Martin W, Stryhn H. Logistic regression. In: Dohoo I, Martin W, Stryhn H, eds. *Veterinary Epidemiologic Research*. Charlottetown, PEI: AVC Inc; 2003:355–368.
- Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc* 1979;74:829–836.
- Harrell FE. Relaxing linearity assumptions for continuous predictors. In: Harrell FE, ed. *Regression Modeling Strategies*. New York: Springer; 2001:24–25.
- Hosmer DW, Lemeshow S. *Applied Logistic Regression*, 2nd ed. New York: Wiley & Sons; 2000:92–128.
- Weatherston LK, Streeter EM. Evaluation of fresh frozen plasma administration in dogs with pancreatitis: 77 cases (1995–2005). *J Vet Emerg Crit Care* 2009;19:617–622.
- Vincent JL, Berre J, Creteur J. Withholding and withdrawing life prolonging treatment in the intensive care unit. *Chron Respir Dis* 2004;1:115–120.
- Verhagen AA, van der Hoeven MA, van Meerveld RC. Physician medical decision-making at the end of life in newborns: Insight into implementation at two Dutch centers. *Pediatrics* 2007;120:20–28.
- Gedge E, Giacomini M, Cook D. Withholding and withdrawing life support in critical care settings: Ethical issues concerning consent. *J Med Ethics* 2007;33:215–218.

Appendix 1

US Unit Models

				Mentation score 0	4 1	7 2	8 3	9 4			
	6 <36.1	4 36.1-37.0	3 37.1-38.5	Temperature (C) 38.6-39.4	1 >39.4						
9 <61		4 61-100		MAP (mmHg) 101-140	1 >140						
				lactate (mg/dL) 0-17.1	5 17.2-36.0	6 36.1-63.1	9 >63.1				
				PCV(%) <11	11 11-20	16 21-30	14 31-40	13 41-45	17 >45		
11 <14.9		7 14.9-21.3		Urea (mg/dL) 21.4-24.9	12 25.0-32.5	7 32.6-69.8	6 >69.8				
12 <111		9 111-115	11 116-118	Chloride (mEq/L) 119-122	11 123-125	7 >125					
				Body cavity fluid score 0	3 1	6 2					

Fig A1. Feline Acute Patient Physiologic and Laboratory Evaluation (APPLE_{full}) score conventional (US) units: Calculated by summing the value in the upper left corner of the appropriate cell for each of the 8 parameters listed, with a maximum potential score of 80. The central cell in each figure represents the range of values for the variable for which 0 points would be assigned. The cells to either side show the appropriate score for the corresponding range of the variable. The final score for the patient is achieved by summing the scores for each variable. See Table 3 for calculation of “fluid score” and “mentation score.” “Mentation score” is collected at admission, for all others utilize the most abnormal value identified over the 24-hour period following admission. If history and physical exam fail to prompt assessment of fluid score, assign zero. Mean arterial pressure (MAP) reflects the value recorded for MAP if direct or indirect manometry is used, or the pressure corresponding to 1st return of audible signal on cuff deflation if Doppler is used.

				Mentation score 0	5 1	6 2	7 3	10 4		
6 <35.6	5 35.6-36.0	4 36.1-37.0	3 37.1-38.5	Temperature (C) 38.6-39.5	1 >39.5					
11 <61		7 61-100		MAP (mmHg) 101-140	1 >140					
				lactate (mg/dL) 0-17.1	6 17.2-36.0	9 36.1-63.1	10 >63.1			
				PCV(%) <16	12 16-25	10 26-35	9 36-45	13 >45		

Fig A2. Feline Acute Patient Physiologic and Laboratory Evaluation (APPLE_{fast}) score conventional (US) units: Calculated by summing the value in the upper left corner for each of the 5 parameters listed, with a maximum potential score of 50. The central cell in each figure represents the range of values for the variable for which 0 points would be assigned. The cells to either side show the appropriate score for the corresponding range of the variable. The final score for the patient is achieved by summing the scores for each variable. “Mentation score” is collected at admission (see Table 3), for all others utilize the most abnormal value identified over the 24-hour period following admission. Mean arterial pressure (MAP) reflects the value recorded for MAP if direct or indirect manometry is used, or the pressure corresponding to 1st return of audible signal on cuff deflation if Doppler is used.