

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

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

Background: Objective risk stratification models are used routinely in human critical care medicine. Applications include quantitative and objective delineation of illness severity for patients enrolled in clinical research, performance benchmarking, and protocol development for triage and therapeutic management.

Objective: To develop an accurate, validated, and user-friendly model to stratify illness severity by mortality risk in hospitalized dogs.

Animals: Eight hundred and ten consecutive intensive care unit (ICU) admissions of dogs at a veterinary teaching hospital.

Methods: Prospective census cohort study. Data on 55 management, physiological, and biochemical variables were collected within 24 hours of admission. Data were randomly divided, with 598 patient records used for logistic regression model construction and 212 for model validation.

Results: Patient mortality was 18.4%. Ten-variable and 5-variable models were developed to provide both a high-performance model and model maximizing accessibility, while maintaining good performance. The 10-variable model contained creatinine, WBC count, albumin, SpO₂, total bilirubin, mentation score, respiratory rate, age, lactate, and presence of free fluid in a body cavity. Area under the receiver operator characteristic (AUROC) on the construction data set was 0.93, and on the validation data set was 0.91. The 5-variable model contained glucose, albumin, mentation score, platelet count, and lactate. AUROC on the construction data set was 0.87, and on the validation data set was 0.85.

Conclusions and Clinical Importance: Two models are presented that enable allocation of an accurate and user-friendly illness severity index for dogs admitted to an ICU. These models operate independent of primary diagnosis, and have been independently validated.

Key words: Clinical epidemiology; Critical care; Dog; Illness severity; Mortality risk model; Statistical modeling.

Objective risk stratification models are used routinely in human critical care medicine for estimates of illness severity.^{1,2} Scoring systems for illness severity typically are based on a number of clinical variables that predict mortality risk, and provide an objective basis for patient triage and risk stratification for scientific purposes. In the context of clinical research, the ability to quantify disease severity across treatment groups allows the benefit of therapy to be determined with greater certainty. Scoring systems should be based on objective criteria, and be accurate and easy to use.¹ The predictive accuracy of the system or model should be demonstrated in a sample of patients independent of that used to develop the model.²

Several diagnosis-specific scoring systems have been developed recently.^{3–5} A diagnosis-independent severity score has several advantages. These include application to disease groups for which a diagnosis-specific scoring system is not available, application to patients early in the hospital stay when a diagnosis has not yet been made,

Abbreviations:

| | |
|--------|---|
| APACHE | acute physiology and chronic health evaluation |
| AUROC | area under the receiver operator characteristic |
| ICU | intensive care unit |
| SPI | survival prediction index |
| SPI2 | survival prediction index 2 |

and application to patients with several disease processes. King et al^{6,7} developed a diagnosis-independent survival prediction index (SPI) for dogs in an intensive care unit (ICU) setting. Despite appropriate study methodology, model performance on an independent validation sample was suboptimal (area under the receiver operator characteristic [AUROC] reduced model survival prediction index 2 [SPI2] = 0.68 on validation dataset), and the manipulations required for manual score calculation were relatively complex.

The aim of the study reported here was to develop and independently validate an easy-to-use model that accurately predicted the probability of death, thus providing an objective measure of illness severity. The model was intended to operate independent of primary diagnosis, and was designed to be based on readily available and objective clinical and laboratory variables.

Materials and Methods

This was a single center prospective cohort study conducted at a teaching hospital serving a predominantly referral patient population. Over the year 2008, 79% of emergency appointments were referral and 21% were first opinion assessments. Of the patients seen, 99.8% were from within the province, and 82% of ICU admissions were dogs. ICU admission occurred for hospitalized

From Emergency and Critical Care Department (Hayes, Mathews), Internal Medicine (Kruth), Imaging (Nykamp), Clinical Studies (Boston), and Epidemiology (Poljak, Dewey), University of Guelph, Guelph, Ontario, Canada, and Intensive Care Medicine, University of Sydney (Doig). This work was done at the University of Guelph Veterinary Teaching Hospital. The paper was presented at IVECCS Chicago, 2009.

Corresponding author: Galina Hayes, VTH, University of Guelph, 50 Stone Rd, Guelph, Ontario, Canada N1G 2W1; e-mail: galina@uoguelph.ca.

Submitted December 10, 2009; Revised February 21, 2010; Accepted April 7, 2010.

Copyright © 2010 by the American College of Veterinary Internal Medicine

10.1111/j.1939-1676.2010.0552.x

patients requiring IV access for continual fluid or medication administration, or close monitoring for any reason.

The study population was 825 consecutive client-owned dogs admitted to the ICU over the study period. Fifteen dogs were excluded because of missing outcome information or age < 4 months, leaving a study population of 810 dogs.

Data were collected over a 6-month period by 4 trained investigators, according to a predetermined written protocol of precise rules and definitions. The protocol detailed the process of data point selection and subsequent spreadsheet entry coding for each variable. Interobserver agreement on data points collected was assessed by κ analysis. κ analysis consisted of having all 4 data collectors collect all variables on the same 4 patients, totalling 224 data points. The degree to which the same data values were identified by all collectors then was assessed.

Approval from the hospital ethics committee was obtained, and given the noninterventional study design, informed owner consent was waived.

Variables for the model were selected a priori based on an anticipated relationship with mortality from existing canine and human models, primary literature, and expert clinical opinion.^{3,6,7} These variables are detailed in Table 1.

All variables were assessed over the first 24 hours following ICU admission with the exception of mentation score. Mentation score was assessed at admission to gain the best possible assessment of true patient baseline status. This was carried out to gain the best possible assessment of true patient baseline status before the administration of analgesics or sedation. For the remaining variables, when several measurements for the same variable were available within the initial 24-hour period, the measurement deviating most from the mid-point of the normal range was selected for model entry. No testing or data collection beyond that obtained for clinical management was performed for the purposes of the study.

Variables were collected as continuous, binary, or ordinal. All physiologic and laboratory variables were recorded as continuous wherever possible. All present/absent variables were recorded as binary. Ordinal variables included nutritional status, EKG findings, and obesity score. Nutritional status was assessed using the categories normal appetite, anorexic, and tube fed. EKG findings were assessed using the categories sinus rhythm only, abnormalities present, and abnormalities present with antiarrhythmics instituted.

Table 1. Predictor variables collected from canine patients for development of the Canine Acute Patient Physiologic and Laboratory Evaluation (APPLE) scores.

| Subcategory | Variables |
|---------------|---|
| Management | Emergency versus scheduled admission, admission service (medical versus surgical) |
| Therapeutic | Oxygen support, indwelling urinary catheter, mechanical ventilation, vasoactive drug treatment, mode of nutritional support |
| Historical | Admission for management of acute/chronic problem, current or prior diagnosis of neoplastic disease process |
| Physiological | Sex, age, heart rate, respiratory rate, mean arterial blood pressure, systolic blood pressure, temperature, EKG findings, mentation score, body weight, obesity score, urine output, presence of free fluid in a body cavity, primary diagnostic category, SpO ₂ |
| Laboratory | PvCO ₂ , PvO ₂ , HCO ₃ , BE, pH, PaO ₂ , albumin, Na, K, PCV/TS, glucose, lactate, iCa, creatinine, ACT, PT, PTT, platelet count, WBC count, % bands, CK, urea, cholesterol, total bilirubin, fibrinogen |

Obesity score was subjectively graded between 1 and 5, where 5 was severely obese, 3 was normal body condition, and 1 was markedly underweight. Urine output was recorded in mL/kg/h. Blood pressures were assessed by indirect^a or direct^b methods. Collection methodology for the variables retained in the final models is detailed in Tables 2 and 3.

The variables neoplastic diagnosis and activated clotting time (ACT) were excluded from the model build. The model goal of independence of patient primary diagnosis was contravened by the variable neoplastic diagnosis, which might also have decreased model transferability because of limitations in accessing cytology results within 24 hours of admission in many centers. The ACT variable was also withdrawn because of limited external test availability and differing test methodologies likely to compromise the validity of future score calculation.

Outcome was recorded as survival status at hospital discharge. For animals that were euthanized, the primary reason for euthanasia was determined after discussion with the primary clinician. Likely because of the complex decision making involved in the euthanasia process, difficulty determining the primary reason was sometimes experienced. Consequently, euthanasia for any reason and natural death were allotted equivalent nonsurvival status in the primary analysis.

Before model construction, a formal survey of faculty and residents at the teaching hospital was performed, evaluating perceptions of illness severity scores and preferred score presentation format. Results from this survey were used to determine model presentation format, and are detailed in Appendix 1.

Statistical Methods

Descriptive statistics were assessed as means \pm standard deviation for data that was normally distributed, and as medians and interquartile range for data that was not normally distributed. Normality was tested with the Shapiro-Wilk test. Associations among categorical data were tested with Fisher's exact test. Associations among continuous data were tested with Student's *t*-test for data that were normally distributed and the Mann-Whitney test for data that were not normally distributed.

Score Development

The cases making up the dataset were randomly split into a model construction or training cohort of 598 dogs (approximately 75% of the group) and a validation cohort of 212 dogs. The random division was performed by computer software assignment of a random number to each dog. The dogs then were ranked by the assigned random number, with the 598 highest dogs retained as the training data set. A 75:25 split was selected to maximize the information available from the training group, while providing a validation cohort of reasonable size.

A priori goals of the model build were to produce a highly predictive model with a format and variable number tailored to maximize end-user uptake.

Predictor variables for which values were missing for over 30% of patients were dropped. Missing physiologic or laboratory values

Table 2. Variable measurement methodology.

| Variable | Measurement Methodology |
|---|---|
| Glucose, albumin, creatinine, total bilirubin | Hitachi 911 automated chemistry analyzer |
| Lactate | ABL 800 Flex, Radiometer Medical Aps, Bronshoj, Denmark |
| SpO ₂ | Hand-held pulse oximeter, Nonin 8500 digital |
| Platelet count, WBC count | Advia 120 hematology analyzer |

Table 3. Body cavity fluid score and mentation score calculation.

| Mentation Score: Assessed at Admission before Sedation/Analgesic Administration | Fluid Score (Ultrasonographic Evaluation, as Assessed by FAST or TFAST Technique) ¹⁸ |
|---|---|
| 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. Two or more of abdominal, thoracic, and pericardial free fluid identified |
| 3. Unable to stand, responsive | |
| 4. Unable to stand, unresponsive | |

for the remaining variables were replaced with normal values or mean values where the term normal did not apply (eg, age, body weight).

The relationship between continuous variables and the binary mortality outcome was assessed using locally weighted scatter plot smoothes (LOWESS).^{8,9} 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 data having x values within a suitable interval about x , known as the bandwidth, are taken. A linear regression 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 .^{10,11} 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 and assist ease of interpretation from a clinical interest perspective (eg, the smoothed plot of observed % mortality against PCV) or on the logit scale. A selection of the mortality functions derived for continuous variables by this method is shown in Figure 1.

Nonlinearity suggested by graphical analysis on the logit scale was confirmed by identifying power terms significantly associated with mortality when entered in a univariable model. Nonlinear variables were categorized. The continuous variable was subdivided into categories (eg, category 1 for temperature > 39.5 , category 2 for temperature ≥ 38.5 , < 39.5), and each patient assigned to the appropriate category for that variable, as previously described.¹² Category cut-points were selected after graphical analysis to reflect the low, medium, and high-risk groups for the variable. When subsequently entered into a logistic regression model, this allowed each category of the variable concerned to receive the appropriate risk coefficient independent of the adjoining categories, thus relaxing the regression linearity assumption. After categorization or not, 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 category associated with the lowest mortality risk was used as the referent group. Variables were put forward for consideration in the multivariable model if they achieved significance at $P < .2$ in the univariable analysis.

A stepwise backward elimination multivariable logistic regression procedure was performed to further eliminate variables with poor explanatory power in a multivariable context, with a cut-off for retention of $P < .05$ by the likelihood ratio test. Collinearity was assessed by monitoring standard errors. After the backwards elimination procedure, the variables dropped were re-entered one by one back into the model and again assessed for significance. The variables selected at the end of this process then were then alternately

dropped from the multivariable logistic regression model in a manual build process. Each model was calculated in the construction or training data set, and then assessed for AUROC performance and Hosmer-Lemeshow C statistic calibration in both the construction and validation data sets. Bayesian Information Criteria also were assessed in the construction data set to allow comparison of non-nested models.

Two models were selected to optimize the balance between variable parsimony and model performance in both construction and validation data sets. The models were checked by graphical examination of deviance residuals, leverage, and the Pregibon Delta beta measure.

Further cross validation was performed to assess both the potential impact of euthanasia bias, and stability of the models on different patient groups. The discrimination and calibration of the SPI2 score also was assessed in an external validation procedure in both the construction and validation cohorts.

The best models were converted to integer scores as follows. The desired maximal score values for each model were selected arbitrarily as 80 for the 10-variable model and 50 for the 5-variable model. The lowest risk categories for each variable were set as the referent categories. For each model, the coefficient assigned to the highest risk category for each variable in the model was summed. The desired maximal score for the model was divided by this value to obtain a constant multiplier. All of the coefficients for each category of each variable in the model then were multiplied by this value, and the resulting numbers rounded to the nearest integer to obtain the integer scores for each category of each variable in the model. The referent categories received a score of 0. Conversion to an integer score was done to facilitate the presentation of convenient, transparent, and user-friendly models that could be manually calculated. Models were presented in both Imperial and Systeme International (SI) units. Logistic regression was used to relate the score to a predicted probability of mortality.

Finally, the extent to which the rounding-off procedure influenced the discriminative power of the model was evaluated by testing the discrimination and calibration of each integer score. This was done by calculating each score for each patient, and entering each score (APPLE_{fast} and APPLE_{full}) into univariable logistic regression analysis against the mortality outcome. Hosmer-Lemeshow C statistics and AUROC values were calculated and compared between both the 10-variable and 5-variable multivariable models and the univariable integer score models. All analyses were performed in a commercially available statistical software program.^c

Results

κ Analysis

Interobserver agreement over collected data points was assessed as excellent with a κ statistic of 0.82. When present, discrepancies typically arose in selecting between primary and ancillary diagnoses, and in the selection of a single value when multiple values of a continuous variable were available.

Patient Population Characteristics and Univariable Analysis

Case type and mortality characteristics of the patient population are shown in Table 4. For 810 ICU admissions, mortality risk was 18.4% ($n = 149$), with 96% ($n = 143$) of deaths occurring by euthanasia. Of euthanasia deaths, 66% ($n = 98$) occurred in association with poor current health status, and 14% ($n = 21$) in association with diagnosis of a terminal disease. Time to death was

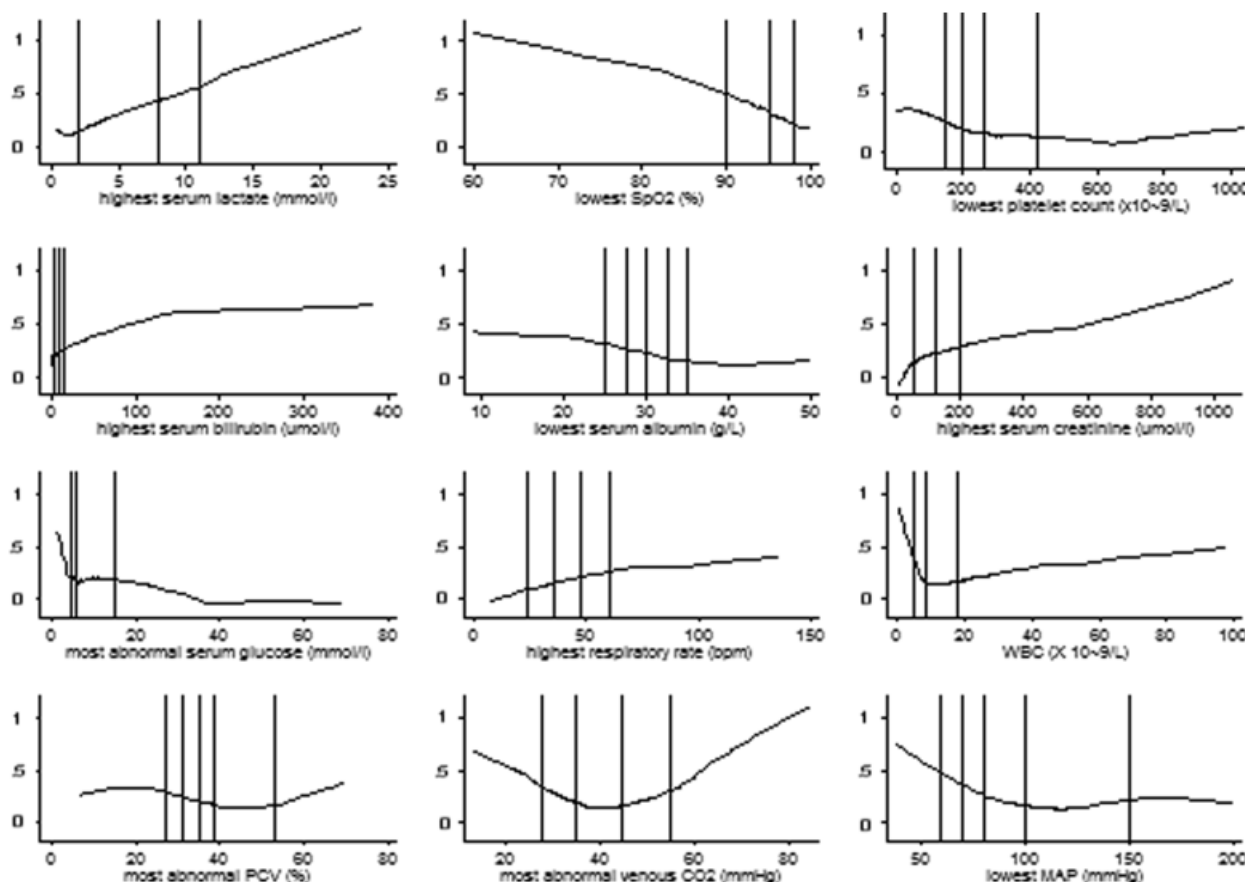


Fig 1. Locally weighted scatterplot smooth (LOWESS) plots of risk of hospital mortality against each continuous variable. Running mean smooth with bandwidth = 0.8. Vertical lines show the limits of the categories applied to the continuous variables before entry into the logistic regression models. All variables showed a statistically significant association with mortality risk ($P < .05$, likelihood ratio test) in both univariable and multivariable analysis.

biphasic in distribution, with peaks at 1 and 3 days. Median ICU stay was 2 days (range 1–20). ICU stay was significantly shorter in the nonsurvivors (median, 1 day; $P < .001$) compared with the survivors. There was no significant difference in mortality risk between the construction and validation patient groups.

Mortality risk was significantly lower in patients hospitalized because of trauma compared with the general patient population ($P = .001$). Patients that failed to eat spontaneously over the first 24 hours of admission had a higher mortality risk than those that ate (OR, 3.02; 95% CI, 1.98–4.59; $P < .001$). Manifestation of EKG abnormalities in the first 24 hours of admission also was associated with a significantly higher mortality risk (OR, 3.78; 95% CI, 2.33–6.11; $P < .001$). In a subgroup of patients ($n = 78$) for which ACT, PT, and PTT data were available, ACT was found to have a linear association in the logit with mortality risk which reached statistical significance. Each 10-second prolongation of ACT was associated with an estimated 20% increase in mortality odds (OR 1.20; 95% CI, 1.02–1.42; $P = .005$). The associations between PT and PTT and mortality risk failed to reach statistical significance in the same group.

Oncologic patients constituted 16.9% of the total patient population. The highest mortality risk occurred for patients on the medical oncology service.

Model Build

Figure 1 shows the LOWESS plots of the risk of hospital mortality against each of 12 continuous variables measured within the first 24 hours of admission. Vertical lines show the limits of the categories subsequently entered into the logistic models.

The variables dropped and retained at each stage of the model build are shown in Figure 2. The variables for coagulation time data, ionized calcium, CK, cholesterol, PaO_2 , and urine output were dropped because of lack of availability for $> 70\%$ of patients.

Thirty-one variables were entered into a multivariable backward stepwise elimination model build after univariable analysis. No variables were dropped after assessment for collinearity. After multivariable backward stepwise elimination and re-entry, 20 variables remained ($P < .05$) and entered the manual build.

Two models ultimately were selected to satisfy the goals of providing both a high-performance model and a model maximizing parsimony while maintaining good performance. The 10-variable model (APPLE_{full}) contained creatinine, WBC count, albumin, SpO_2 as detected by pulse oximetry, total bilirubin, mentation score (see Table 3), respiratory rate (bpm), age (years), lactate, and presence of free fluid in a body cavity as

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

| | No. | Median (IQR) or % | Hospital Mortality, % |
|---|-----|-------------------|-----------------------|
| Age (years) | 810 | 6.5 (3–10) | |
| Sex | | | |
| Male | 462 | 57 | 18.6 |
| Female | 348 | 43 | 16.7 |
| Surgical status | | | |
| Surgical | 292 | 36 | 8.6 ^a |
| Nonsurgical | 518 | 64 | 23.4 ^a |
| Source of admission | | | |
| Referral | 575 | 71 | 17.9 |
| Primary care | 235 | 29 | 18.6 |
| Admission type | | | |
| Emergency | | 89 | 19.2 |
| Planned admission | | 11 | 12.2 |
| Length of hospital stay (days) | | 3 (1–4) | |
| Primary reason for admission | | | |
| GI/pancreatitis | 107 | 13.2 | 14.0 |
| Trauma | 87 | 10.7 | 5.7 ^a |
| Spinal/peripheral neuropathy | 87 | 10.7 | 8.0 ^a |
| Medical oncology | 83 | 10.2 | 47.0 ^a |
| Immune mediated disease | 66 | 8.1 | 22.7 |
| Surgical oncology | 54 | 6.7 | 20.4 |
| Respiratory | 53 | 6.5 | 24.5 |
| Intracranial disease | 49 | 6.0 | 20.4 |
| Renal/urinary tract | 47 | 5.8 | 25.5 |
| Cardiac | 40 | 4.9 | 20.0 |
| Hepatic | 34 | 4.2 | 11.7 |
| Sepsis | 32 | 4.0 | 15.6 |
| Ophthalmological | 24 | 3.0 | 0.0 ^a |
| Toxicity | 17 | 2.1 | 5.9 |
| Endocrine disorder | 16 | 2.0 | 6.2 |
| Miscellaneous | 8 | 1.0 | 37.5 |
| Reproductive | 6 | 0.7 | 0.0 |
| Vasoactive drug treatment during first 24 hours | 19 | 2.3 | 57.9 ^a |
| Mechanical ventilation during first 24 hours | 11 | 1.4 | 81.8 ^a |
| Mortality-total deaths all cause | 149 | 18.4 | |
| Type of death | | | |
| Natural death | 6 | 4.0 | |
| Euthanasia due to | | | |
| Financial constraints | 24 | 16.1 | |
| Terminal disease diagnosis | 21 | 14.1 | |
| Current severity of illness | 98 | 65.8 | |
| Exclusions due to age/unavailable records | 15 | | |

^aMortality difference statistically significant at $P < .05$ using a two-tailed χ^2 test of difference in proportion tested against opposing category or population as a whole.

detected by ultrasonographic screening. The AUROC on the construction cohort was 0.93 (95% CI, 0.90–0.95) and on the validation cohort was 0.91 (95% CI, 0.89–0.96). There was no significant lack of calibration in either cohort (Hosmer-Lemeshow $\chi^2_8 = 5.14$ and $\chi^2_{10} = 7.12$, $P = .74$ and $.71$ on construction and validation cohorts, respectively). The 5-variable model (APPLE_{fast}) contained glucose, albumin, mentation score, platelet count, and lactate. The AUROC on the construction cohort was 0.86 (95% CI, 0.82–0.90) and on the validation cohort was 0.85 (95% CI, 0.80–0.92), again with no significant lack of calibration in either cohort (Hosmer-Lemeshow $\chi^2_8 = 8.51$ and $\chi^2_{10} = 8.25$, $P = .38$ and $.60$).

When the models were cross validated on the validation cohort after censoring of patients with neoplastic

diagnoses, the models showed good stability with AUROCs increasing slightly to 0.95 for the APPLE_{full} model and 0.87 for the APPLE_{fast} model, with retention of calibration.

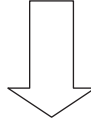
Model discrimination also was retained for both models as the patients included in each category of euthanasia were successively censored from the validation cohort. The best model discrimination for both models ultimately was shown for the validation cohort when patient deaths were restricted to those occurring by natural death only (ie, all euthanasia cases censored). The discrimination and calibration of the SPI2 score also were assessed in an independent external validation procedure. The SPI2 performed well, with an AUROC of 0.82 on the combined construction and validation cohorts and good calibration (Hosmer-Lemeshow

Variables collected:

Emergency vs scheduled admission, admission service (medical vs surgical), oxygen support, indwelling urinary catheter, mechanical ventilation, vasoactive drug treatment, mode of nutrition, admission for management of acute/ chronic problem, current or prior diagnosis of neoplastic disease process, sex, age, heart rate, respiratory rate, mean arterial blood pressure, systolic blood pressure, temperature, EKG score, mentation score, body weight, obesity score, urine output, presence of free fluid in a body cavity, primary diagnostic category, SpO₂, PvCO₂, PvO₂, HCO₃, BE, pH, PaO₂, albumin, Na, K, PCV/TS, glucose, lactate, iCa, creatinine, ACT, PT, PTT, platelet count, WBC count, %bands, CK, urea, cholesterol, total bilirubin, fibrinogen

Variables dropped as <70% data availability:

urine output, iCa, ACT, PT, PTT, CK, cholesterol, fibrinogen
PaO₂



Variables remaining after variables with <70% data availability dropped:

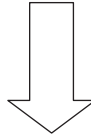
Emergency vs scheduled admission, admission service (medical vs surgical), oxygen support, indwelling urinary catheter, mechanical ventilation, vasoactive drug treatment, mode of nutrition, admission for management of acute/ chronic problem, current or prior diagnosis of neoplastic disease process, sex, age, heart rate, respiratory rate, mean arterial blood pressure, systolic blood pressure, temperature, EKG score, mentation score, body weight, obesity score, presence of free fluid in a body cavity, primary diagnostic category, SpO₂, PvCO₂, PvO₂, HCO₃, BE, pH, albumin, Na, K, PCV/TS, glucose, lactate, creatinine, platelet count, WBC count, %bands, urea, total bilirubin,

Variables dropped as contravening model goals:

Neoplastic diagnosis, primary diagnostic category

Variables dropped as P>0.2 in univariable analysis:

emergency admission, body condition score, sex, body weight, indwelling urinary catheter, %bands, urea
vasoactive drug treatment



Variables remaining after variables with P>0.2 dropped:

Admission service (medical vs surgical), oxygen support, mechanical ventilation, mode of nutrition, admission for management of acute/ chronic problem, age, heart rate, respiratory rate, mean arterial blood pressure, systolic blood pressure, temperature, EKG score, mentation score, presence of free fluid in a body cavity, SpO₂, PvCO₂, PvO₂, HCO₃, BE, pH, albumin, Na, K, PCV/TS, glucose, lactate, creatinine, platelet count, WBC count, total bilirubin,

Variables dropped in stepwise backwards

elimination with forward re-entry if P=>0.05:

Admission service, mechanical ventilation,
acute/chronic, heart rate, systolic bp, temperature,
PvO₂, BE, pH, PCV, TS



Variables remaining after multivariable stepwise elimination model:

Oxygen support, mode of nutritional support, age, respiratory rate, mean arterial blood pressure, EKG findings, mentation score, presence of free fluid in a body cavity, SpO₂, PvCO₂, HCO₃, albumin, Na, K, glucose, lactate, , creatinine, platelet count, WBC count, total bilirubin,

Manual build:



Variables remaining in final models:

age, respiratory rate, mentation score,
presence of free fluid, SpO₂, albumin,
glucose, lactate, creatinine, platelet count,
WBC count, total bilirubin

Fig 2. Variables dropped and retained at each stage of the model build.

$P = .49$). Performance characteristics of the models are summarized in Table 5 and cross-validation results are shown in Table 6.

The 2 models were used to develop 2 objectively weighted multivariate prognostic scores, the canine Acute

Patient Physiologic and Laboratory Evaluation (APPLE_{full} and APPLE_{fast}) scores ranging from 0 to 80 and 0 to 50, respectively. The algorithms for score calculation are shown in Figures 3 and 4. The central cell on each table represents the range of values for which 0 points

Table 5. Performance characteristics of APPLE models and SPI2 score on study construction (n = 598) and validation (n = 212) populations.

| | APPLE _{full} Model | APPLE _{fast} Model | SPI2 Score |
|---|-----------------------------|-----------------------------|----------------------|
| AUROC | | | |
| Construction cohort | 0.93 | 0.87 | 0.83 |
| Validation cohort | 0.91 | 0.84 | 0.79 |
| Postconversion to integer score | 0.92 | 0.86 | |
| On combined construction and validation cohort | | | 0.82 |
| Hosmer-Lemeshow <i>C</i> statistic | | | |
| Construction cohort χ^2_8 | 5.14, <i>P</i> = .74 | 8.51, <i>P</i> = .38 | 9.90, <i>P</i> = .27 |
| Validation cohort χ^2_{10} | 7.12, <i>P</i> = .71 | 8.25, <i>P</i> = .60 | 9.06, <i>P</i> = .34 |
| Post conversion to integer score | 8.72, <i>P</i> = .37 | 10.65, <i>P</i> = .22 | |
| On combined construction and validation cohort | | | 7.43, <i>P</i> = .49 |
| Score sensitivity for predicting mortality when scores > 40/80 (APPLE _{full}) and > 25/50 (APPLE _{fast}) are taken to predict death | 40.9% | 67.1% | |
| Score specificity (at > 40/80 and > 25/50 cut-offs) | 98.3% | 85.3% | |
| Score sensitivity for predicting mortality when scores > 30/80 (APPLE _{full}) and > 22/50 (APPLE _{fast}) are taken to predict death | 81.2% | 73.8% | |
| Score specificity (at > 30/80 and > 22/50 cut-offs) | 89.4% | 80.0% | |

would be assigned for the variable. The cells to either side of the central cell show the appropriate score for the relevant variable range. The final score for the patient is achieved by summing the scores for each variable. Patient score then can be correlated with mortality risk, if required either by the graphs in Figure 5 or the equations listed below. Figure 5 depicts the relationship between the APPLE scores and predicted probability of mortality. The equations below describe the relationship between each score and the predicted mortality risk *P*, where *R* is logit *P*. There was minimal change in AUROC values and no loss of calibration after conversion of the models to integer score. Results of score discrimination and calibration are shown in Table 5. The score sensitivity (proportion of patients that died that were predicted to die) and specificity (proportion of patients that did not die that were predicted not to die) at various cut-points also are shown in Table 5. The sensitivity and specificity results corresponding to the midpoint of each score and the score values optimizing the sum of sensitivity and specificity were selected for display. It should be emphasized that these are population-averaged results. The confidence intervals for mortality prediction surrounding each cut-point are wide, as shown in Figure 4. The APPLE scores in SI units are detailed in Appendix 2, Figures A1 and A2.

Equations to calculate mortality probability (*P*):

- (1) APPLE_{fast} score: where $P_1 = \exp(R_1)/(1 + \exp[R_1])$

$$R_1 = (0.249 \times \text{APPLE}_{\text{fast}}) - 7.020$$
- (2) APPLE_{full} score: where $P_2 = \exp(R_2)/(1 + \exp[R_2])$

$$R_2 = (0.237 \times \text{APPLE}_{\text{full}}) - 8.294$$

Discussion

Patients admitted to veterinary hospitals that require continuous IV access or close monitoring constitute a population with substantial emotional and financial investment on the part of their owners. Early prediction of high mortality risk allows triage to an environment where optimization of volume resuscitation, gas exchange, nutrition, and treatment of both the primary problem and all associated complications can be performed effectively. Thus, early stratification of illness severity has important implications for management and timely intervention. These patients are the focus of substantial clinical research by many specialties in veterinary medicine today. The ability to objectively categorize

Table 6. Cross validation results of canine APPLE_{full} and APPLE_{fast} scores trained on the construction population (n = 598).

| Patient Population | AUROC for APPLE _{full} | Hosmer-Lemeshow <i>C</i> Statistic for APPLE _{full} | AUROC for APPLE _{fast} | Hosmer-Lemeshow <i>C</i> Statistic for APPLE _{fast} |
|---|---------------------------------|--|---------------------------------|--|
| Validation group with "neoplasia" patients excluded (n = 174) | 0.95 | 11.65, <i>P</i> = .31 | 0.87 | 16.80, <i>P</i> = .09 |
| Validation group with financial euthanasias excluded (n = 206) | 0.91 | 6.07, <i>P</i> = .81 | 0.84 | 14.01, <i>P</i> = .17 |
| Validation group with financial and terminal disease euthanasias excluded (n = 199) | 0.93 | 6.41, <i>P</i> = .78 | 0.86 | 12.58, <i>P</i> = .25 |
| Validation group with all euthanasias excluded (n = 174) | 0.94 | 26.08, <i>P</i> = .003 | 0.89 | 31.89, <i>P</i> < .001 |

APPLE_{full} Score US units

| | | | | | | |
|-----------|--------------|--------------|-----------------------------------|----------------|----------------|------------|
| | | | creatinine (mg/dL) 0-0.62 | 1 0.63-1.35 | 8 1.36-2.26 | 9 >2.26 |
| | | | 9 <5.1 | 2 5.1-8.5 | 3 8.6-18 | |
| 6 <2.6 | 7 2.6-3.0 | 9 3.1-3.2 | albumin (g/dL) 3.3-3.5 | 2 >3.5 | | |
| 10 <90 | 4 90-94 | 1 95-97 | SpO ₂ (%) 98-100 | | | |
| | | | total bilirubin (mg/dL) 0-0.23 | 6 0.24-0.46 | 4 0.47-0.93 | 3 >0.93 |
| | | | mentation score 0 | 5 1 | 7 2 | 8 3 |
| | | | respiratory rate (bpm) <25 | 3 25-36 | 5 37-48 | 6 49-60 |
| | | | age (years) 0-2 | 6 3-5 | 8 6-8 | 7 >8 |
| 3 2 | 4 1 | | fluid score 0 | | | |
| | | | lactate (mg/dL) <18.0 | 2 18.0-71.2 | 3 71.3-90.1 | 6 >90.1 |

Footnote: See Table 3 for calculation of 'fluid score' and 'mentation score'. A value of zero is ascribed for each parameter in the central zone. 'Mentation score' is collected at admission, for all others utilise the most abnormal value identified over the 24 hour period following admission. If history and physical exam fail to prompt assessment of SpO₂ or fluid score, assign zero.

Fig 3. Canine 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 10 parameters listed, with a maximum potential score of 80.

patients by illness severity assists effective clinical research. Treatment groups with substantial differences in illness severity call into question findings regarding treatment efficacy. Reporting illness severity in objective and transferable terms lends context to case reports and case series. Some treatments may only be effective or appropriate in the context of severe patient compromise, which can be formally identified and recorded using a severity score. For observational studies utilizing regression modeling to identify predictors of specific outcomes, entering illness severity into a multivariable model can allow a risk-adjusted analysis to be performed.

Model discrimination refers to the ability of the model to accurately differentiate those animals that will live from those that will die. Model calibration refers to the ability of a model to accurately predict the appropriate number of deaths at each level of risk across a

population. Model discrimination can be assessed by the AUROC characteristic. An AUROC of 1.0 implies perfect performance, whereas an AUROC of 0.5 implies a model with no better discrimination than a coin flip. Model calibration is assessed by the Hosmer-Lemeshow statistic, with a *P* value > .05 implying acceptable calibration.

Two objective illness severity scores were developed based on a study cohort of all ICU admissions over a 6-month period. In the study, hospital ICU admission was required for any patient requiring continuous IV access or close monitoring. The study population constituted a broad case mix with primary clinicians from many services including internal medicine, critical care, surgery, cardiology, oncology, and others. The APPLE scores reflect the severity of derangement of normal physiology identified by abnormalities in clinical and laboratory

APPLE_{fast} Score US units

| | | | | | | |
|-----------|--------------|--------------|---|-------------------------|----------------|-------------|
| 7 <84 | 8 84-102 | 9 103-164 | 10 165-273 | glucose (mg/dL) >273 | | |
| 8 <2.6 | 7 2.6-3.0 | 6 3.1-3.2 | albumin (g/dL) 3.3-3.5 | 2 >3.5 | | |
| | | | lactate (mg/dL) <18.0 | 4 18.0-72.1 | 8 72.2-90.1 | 12 >90.1 |
| 5 <151 | 6 151-200 | 3 201-260 | platelet count (x10 ⁹ /L) 261-420,000 | 1 >420 | | |
| | | | mentation score 0 | 4 1 | 6 2 | 7 3 |
| | | | | | | 14 4 |

Footnote: See Table 3 for calculation of 'mentation score'. A value of zero is ascribed for each parameter in the central zone. 'Mentation score' is collected at admission, for all others utilise the most abnormal value identified over the 24 hour period following admission.

Fig 4. Canine Acute Patient Physiologic and Laboratory Evaluation (APPLE_{fast}) score: Calculated by summing the value in the upper left corner of the appropriate cell for each of the 5 parameters listed, with a maximum potential score of 50.

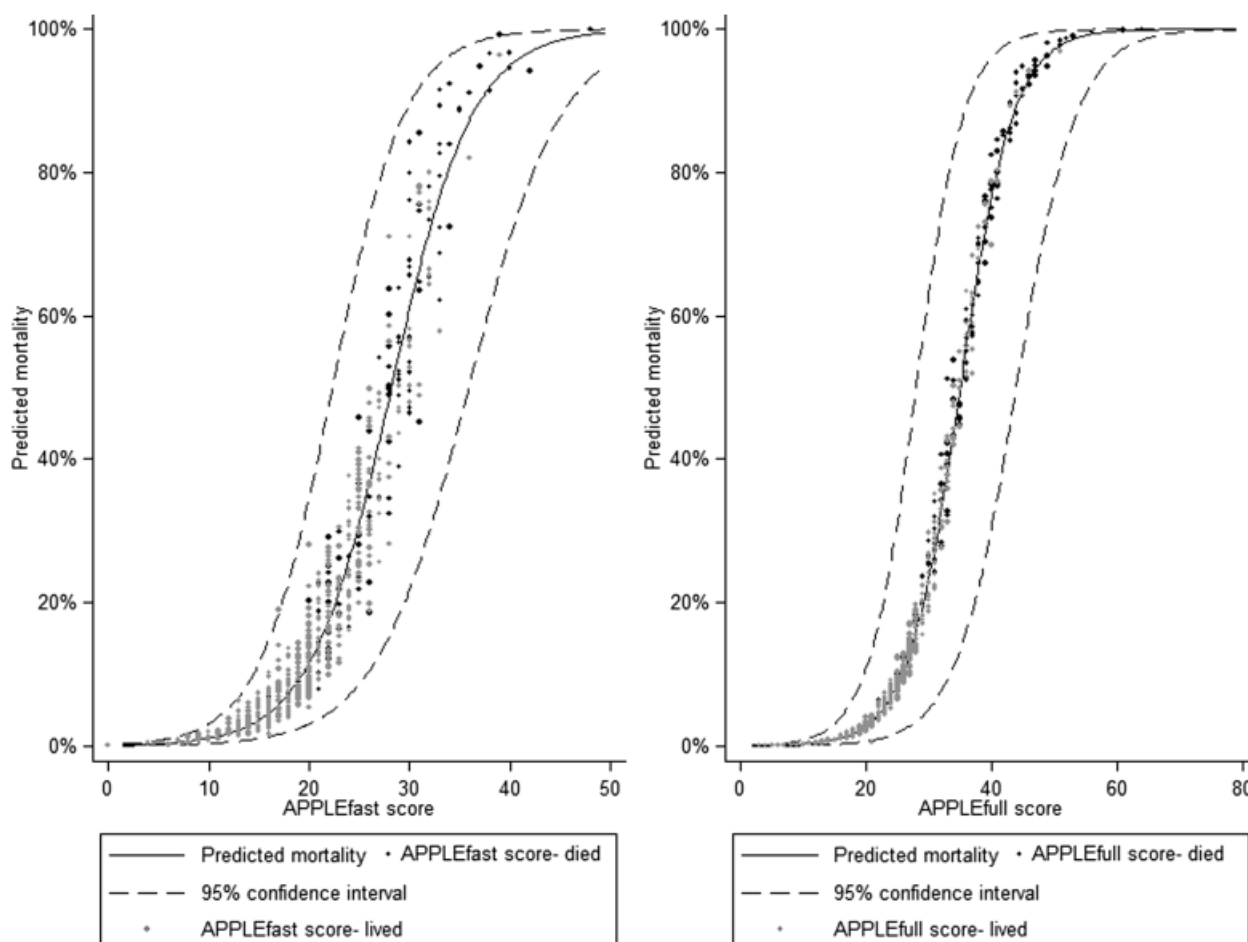


Fig 5. Graphs to show the relationship between Acute Patient Physiologic and Laboratory Evaluation (APPLE) scores and predicted mortality (P) with confidence intervals for population predictions, with underlying scatterplots depicting the study population by mortality group for each score.

variables, and correlate with likelihood of survival to hospital discharge.

Two scores were developed to maximize end-user uptake. The 10-variable score (APPLE_{full}) optimizes predictive accuracy, however, where clinical information or available time is more limited, the 5-variable model (APPLE_{fast}) can be used. However, the user must accept some loss of model discrimination when using the reduced model. An important strength of the models is the simplicity of their interpretation. The scores have been tailored to allow rapid cage-side calculation with the use of simple and objective clinical criteria. A further strength of this study was the validation process. A major concern of model development is how well model performance will be maintained if the model is transferred to a new set of patients either temporally or geographically. Although some loss of model performance is inevitable, a model shown to perform well in a validation cohort is desirable. A model for which no validation procedure has been performed is of serious concern to the end-user. Discrimination characteristics of both the APPLE_{fast} and APPLE_{full} models were excellent in both the construction and validation cohorts. Both models calibrated well.

Patients with neoplastic diagnoses made up a relatively large proportion of our patient group, and had the highest patient group mortality risk. Because of concerns regarding the ability of the model to transfer well to a patient population with a lower proportion of patients with neoplastic disease, model discrimination, and calibration also was assessed in the validation cohort with these patients censored. Model performance remained excellent.

The performance of the SPI2 score was also assessed in an independent external validation procedure and found to be very good, with an AUROC of 0.82 on the combined construction and validation cohorts and good calibration.

A particular challenge in the development of veterinary risk prediction models is the prevalence of euthanasia as the major mortality outcome in veterinary patient populations. This can result in a form of information bias known as euthanasia bias. 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. If the clinician perceives a particular clinical sign to be very negative and this perception is relayed to the client triggering the

euthanasia decision, a mortality association will be created whether this association truly exists or not. Mortality associations may be biased in this way toward a false-positive identification of association where none exists. 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.^{13,14} The approach taken to date in human mortality risk prediction studies is to assume that patients dying after withdrawal of care (eg, discontinuation of mechanical ventilation) ultimately would have died had care been maintained. Risk prediction models then have been calculated with no differentiation between patients that died in the face of maximal care and those that died after withdrawal of care. Despite this issue, these models have offered a reliable tool to risk stratify patients enrolled in clinical research for many years.

We elected to take a slightly different approach. We acknowledged that the purest form of score development would take place in a patient population in which the only patients dying would be those experiencing natural death in the face of maximal intervention. However, this population does not exist in veterinary medicine. The performance of humane euthanasia to relieve unnecessary suffering in moribund animals considered to be awaiting death is common within the culture of veterinary medicine. Exclusion of euthanized patients obviously was not a practical approach for a study of this type. Instead, we elected to categorize the primary reason for euthanasia in each case, and perform a cross-validation of the models while censoring successive euthanasia categories, recording model discrimination at each step.

The final group contained only dogs that experienced natural death. We found that the discrimination of our models increased as euthanasia categories were censored, with AUROCs increasing from 0.91 to 0.94 and 0.84 to 0.89. This suggested that the variables and coefficients assigned in the training data set truly reflected mortality probability rather than euthanasia risk based on false premise. Calibration was lost in the final natural death only group, likely reflecting the higher all-cause mortality in the construction data set relative to mortality restricted to natural deaths only in the validation data set.

Our cross-validation results suggest that the models proposed should provide a reliable assessment of individual illness severity, and truly reflect underlying mortality probabilities. In common with all models, however, they may under- or over predict mortality rates for the group as a whole if applied prospectively to populations with substantially lower or higher mortality rates than those reported for our populations. Because the primary intended use for these models 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.²

Non-nested models, defined as models with differing variables, were constructed as dictated by a variable se-

lection process based on optimization of performance in construction and validation cohorts while imposing a parsimonious approach on the number of variables to be included. The final models were selected on the basis of achieving maximal information for minimal number of variables with minimal shrinkage between the construction and validation data sets and retention of calibration. The smaller model contains variables not present in the larger model, which may at first seem counterintuitive. However, in the multivariable context, the predictive variables retained are not those that are the best predictors in the univariable context, but those that best explain the mortality variance not already explained by the other variables in the model. Equally, the risk coefficients assigned to each category reflect the risk not already explained by the other variables within the model, and may also appear counter-intuitive. Thus, the predictive power of an individual variable within a multivariable model is codependent in direction and magnitude on the other variables present. As an example, when pH is introduced into a model containing lactate, the risk coefficients assigned to pH will explain only the mortality risk not already captured by lactate, and may appear counter intuitive. Equally, if lactate is removed, the predictive power of pH will increase, and the risk coefficients will change. In this way, selection of variables targeted at explanatory power may well result in nonnested models in biological systems.

The strength of a multivariable model lies in the ability of the model to assess the associations between a variable and mortality while allowing for the concurrent effect of all of the other variables in the model. For this reason, the association between a variable and mortality may differ markedly when the variable is considered in isolation compared with in a multivariable context, and is dependent on the other variables included in the model. This is exemplified by the albumin variable. Evaluation of albumin in univariable form, as in Figure 1, shows the clinically anticipated steadily increasing mortality risk as albumin decreases from 40 to 15 g/L. When the concurrent effects on mortality risk of glucose, lactate, platelet count, and mentation score are considered, as in the APPLE_{fast} score, this association changes slightly with an increased mortality risk now associated with an albumin > 35 g/L compared with an albumin of 33–35 g/L. In this context, the albumin categories now are modeling the mortality risk not already captured through glucose, lactate, mentation score, and platelet count. The risk associations for the albumin < 33 g/L categories change slightly again between the APPLE_{fast} and APPLE_{full} models, when the concurrent effects of several additional variables on mortality are taken into account. Although the scores assigned in a multivariable model may not be clinically intuitive, they reflect the mortality risk findings of the dataset. Clinically intuitive scoring should not be anticipated in a multivariable context. In addition, the lowest risk categories for each variable do not necessarily correspond to the normal range in a multivariable context.

In line with previous studies in veterinary and human medicine (SPI1, SPI2, acute physiology and chronic

health evaluation [APACHE], mortality probability model, ICNARC, severe acute physiology score),^{2,6,7,15} we elected to select the most abnormal values of variables observed over the 24-hour period for model entry. There were several reasons for this approach. First, the use of the most abnormal rather than the point-of-admission values made allowance for the difference between patients that continue to deteriorate despite treatment compared with those that stabilize, and were felt to be likely more explanatory. Secondly, as discussed previously, the confidence intervals associated with any individual score prediction were wide, and we hoped to discourage in any way possible the practice of recommending euthanasia at admission based on a patient score. The scores have been developed and validated on 24-hour admission period values only, and should not be applied prospectively to values restricted to the admission period or values at any other time period in the hospital stay.

The APPLE scores were developed from the coefficients assigned to the various categories of the variables by the multivariable logistic models. These coefficients were in turn driven by the associations between variables and outcome in a multivariable context averaged across the patient population. This resulted in score values that may differ from those that would be anticipated in the more clinically familiar univariable context. For example, in the APPLE_{fast} model, the platelet count assigned the highest score was the 150–200,000/ μ L category, whereas in the APPLE_{full} model an albumin of 31–32 g/L was associated with a greater mortality risk than an albumin of <26 g/L when all other variables in the model were taken into account. The strength of a multivariable model lies in the ability of the model to assess the associations between a variable and mortality while allowing for the concurrent effect of all of the other variables in the model. For this reason, the association between a variable and mortality may differ markedly when the variable is considered in isolation compared with in a multivariable context, and is dependent on the other variables included in the model. This is exemplified by the albumin variable. Evaluation of albumin in univariable form, as in Figure 1, shows the clinically anticipated steadily increasing mortality risk as albumin decreases from 40 g/L to 15 g/L. When the concurrent effects on mortality risk of glucose, lactate, platelet count and mentation score are considered, as in the APPLE_{fast} score, this association changes slightly with an increased mortality risk now associated with an albumin >35 g/L compared with an albumin of 33–35 g/L. In this context, the albumin categories now are modelling the mortality risk not already captured through glucose, lactate, mentation score, and platelet count. The risk associations for the albumin <33 g/L categories change slightly again between the CAP_{fast} and CAP_{full} models, when the concurrent effects of several additional variables on mortality are taken into account. Although the scores assigned in a multivariable model may not be clinically intuitive, they reflect the mortality risk findings of the dataset. Clinically intuitive scoring should not be anticipated in a multivariable context. The lowest risk

categories for each variable do not necessarily correspond to the normal range in a multivariable context.

A surprising finding of this study was the prevalence of stress hyperglycemia among dogs admitted to ICU, and the associations between blood glucose concentration and mortality risk in both a univariable and multivariable context. The upper limit of normal blood glucose in the dog is variously reported as 6.1–7.9 mmol/L depending on the methodology and laboratory. When diabetics were excluded from the population, 19% of the remaining admissions demonstrated a blood glucose concentration above 7.9 mmol/L, and 7% of nondiabetic admissions demonstrated a blood glucose concentration above 10.0 mmol/L within 24 hours of admission. When the entire patient population including diabetics was divided into those exhibiting hyperglycemia (defined as blood glucose >7.9 mmol/L) and normoglycemic animals, hyperglycemia was associated with an overall increased mortality risk (OR, 1.67; $P = .01$). This relationship was further accentuated when diabetics were excluded from the population (OR, 1.90; $P = .004$), suggesting a protective effect of diabetic status. The protective effect of diabetes in this context may reflect the success of insulin therapy in managing the physiologic derangements associated with a diabetic crisis, whereas the increased mortality risk association in nondiabetics may be because of upregulation of the adrenocortical axis and secondary insulin resistance proportional to the severity of the primary disease process. However, when the association between glucose and mortality was evaluated while controlling for overall severity of illness by the APPLE_{full} score as a severity indicator, the association between stress hyperglycemia and increased mortality was lost (OR=1.02, $P = .95$). This suggests a role for stress hyperglycemia as an epiphenomenon rather than playing a truly causal role in mortality risk.

An association between prolongation of ACT and increased mortality risk was identified in this study. In a patient subgroup for which full coagulation data were available, each 10-second lengthening of ACT was associated with an estimated 20% increase in mortality odds (OR, 1.20; 95% CI 1.02–1.42; $P = .005$). PT and PTT failed to attain a statistically significant association with mortality risk within the same subgroup. ACT is a readily available and inexpensive bench-top test that has been shown to have strong correlation with both increased C-reactive protein concentrations and decreased antithrombin concentrations in dogs admitted to an ICU.¹⁶ The association between increased mortality risk and lengthening ACT likely reflects the morbidity of systemic inflammatory response-associated coagulopathies.

There are several limitations to this study. First, the hospital infrastructure may be atypical, in that all patients requiring IV access were admitted to the ICU. Thus, the study population was not limited to critically ill dogs, but to all dogs requiring IV access or 24-hour observation. Hence, caution must be exercised if the model is applied to an ICU population with different admission criteria. Secondly, score development and validation were based exclusively on data from a single center. This

raises uncertainty regarding the external validity of the model. Validation on an international scale is recommended. Thirdly, the models were developed on relatively small population samples. Model stability improves as a function of sample size, with perfect stability achieved when the entire population is sampled. The scores assigned to some variable categories may change if the models were constructed on a larger population. To provide context, the last version of a widely used human illness severity model, the APACHE IV, was constructed on a sample of 131,618 patients from 104 ICUs.¹⁴ Fourthly, because of concerns with sample size and model stability when used prospectively, diagnosis specific variables were not used. However, primary diagnosis can be a highly determinative variable for mortality risk. The inclusion of diagnosis specific coefficients could be a useful and highly predictive feature to consider for future model modification.

Finally, a major concern of the model developers was that the models might be inappropriately used to direct euthanasia decisions. Veterinarians are frequently in the position of being called upon by owners to predict patient survival as a factor in the owner's decision of whether or not to pursue treatment. A study in human medicine comparing clinical judgment with a score system (APACHE II) concluded that although clinical judgment outperformed the model for predicting mortality for individual patients, ultimately no method was reliable for survival prediction.¹⁷ Euthanasia decisions are complex and highly contextual, and should remain the collective responsibility of the individual clinician and owner. At most, a severity score can provide an objective adjunct to the informed but subjective opinion of the clinician. Using these tools as part of a decision-making process is reasonable and prudent. Using these tools to dictate individual patient decisions is not appropriate. This is further highlighted by reviewing the performance characteristics of the model. The APPLE_{fast} score has a specificity of 85% when a score above 25 is taken to predict death, implying that when applied to a similar population, 85% of animals that live would in fact be predicted to live by the score. Thus, the false positive fraction (1-Sp) is 15%, implying that 15% of patients predicted to die by the model would in fact live, and score-driven euthanasia would terminate 15% of patients unnecessarily. The strength of illness severity models is in providing a formal and objective patient risk stratification system. This can then be used to strengthen clinical trial design and analyses in veterinary medicine.

Conclusions

Our analysis of a database of dogs admitted to an ICU enabled development of 2 new scoring systems based on readily available physiologic and biochemical variables collected within 24 hours of ICU admission. After a survey of clinicians, the scores were presented in a highly accessible format facilitating rapid manual calculation. Performance of both models is higher than that previously reported, and both have been validated on a patient sample independent to that used for model con-

struction. External validation of the models is recommended. Use of a score with good prognostic ability provides the possibility of identifying high-risk patients likely to benefit from aggressive management or new treatment modalities. Of primary importance, however, is the capacity of an accurate and objective stratification system to improve the quality of observational clinical research.

Acknowledgment

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

Footnotes

^a Drager monitor and transducer system, Telford, PA

^b Cardell 9401, Orchard Park, NY

^c Stata version 10.1, StataCorp LP, College Station, TX

References

1. Gunning K, Rowan K. ABC of intensive care: Outcome data and scoring systems. *Br Med J* 1999;319:241–244.
2. LeGall JR. The use of severity scores in the intensive care unit. *Intensive Care Med* 2005;31:1618–1623.
3. Mansfield CS, James FE, Robertson ID. Development of a clinical severity index for dogs with acute pancreatitis. *J Am Vet Med Assoc* 2008;233:936–944.
4. Segev G, Kass PH, Francey T, Cowgill LD. A novel clinical scoring system for outcome prediction in dogs with acute kidney injury managed by hemodialysis. *J Vet Intern Med* 2008;22:301–308.
5. Jergens AE, Schreiner CA, Frank DE, Niyo Y. A scoring index for disease activity in canine inflammatory bowel disease. *J Vet Intern Med* 2003;17:291–297.
6. 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.
7. 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 4 locations. *Am J Vet Res* 2001;62:948–954.
8. 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.
9. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc* 1979;74:829–836.
10. Harrell FE. Relaxing linearity assumptions for continuous predictors. In: Harrell FE, ed. *Regression Modeling Strategies*. New York: Springer; 2001:24–25.
11. Hosmer DW, Lemeshow S. *Applied Logistic Regression*, 2nd ed. New York: John Wiley and Sons Inc; 2000:91–116.
12. Harrison DA, D'Amico G, Singer M. The pancreatitis outcome prediction (POP) Score: A new prognostic index for patients with severe acute pancreatitis. *Crit Care Med* 2007;35:1703–1708.
13. Vincent JL, Berre J, Creteur J. Withholding and withdrawing life prolonging treatment in the intensive care unit. *Chron Respir Dis* 2004;1:115–120.
14. 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;120:20–812.

15. Zimmerman J, Kramer A, McNair D. Acute physiology and chronic health evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006;34:1297–1310.

16. Cheng T, Mathews K, Abrams-Ogg A, et al. Relationship between assays of inflammation and coagulation: A novel interpretation of the canine activated clotting time. *Can J Vet Res* 2009;73:97–102.

17. Meyer AA, Messick WJ, Young R, et al. Prospective comparison of clinical judgment and APACHE II score in predicting outcome in critically ill surgical patients. *J Trauma* 1992;32:747–753.

18. Boysen R, et al. Evaluation of a FAST protocol to detect free abdominal fluid. *JAVMA* 2004;225:1198–1204.

Appendix 1: Survey Conducted to Evaluate Perceptions, Clinical Use, and Preferred Format of Outcome or Illness Scores

Method

A qualitative questionnaire consisting of 11 open questions was administered through face-to-face interviews. Study population was a convenience sample of 23 faculty and residents at a veterinary teaching hospital. The survey was conducted by a single interviewer over a 2-day period. Response rate was 100%. Data analysis was performed in Stata version 10.1 (StataCorp LP, College Station, TX).

Results

The commonest reason given for lack of illness score uptake in veterinary clinical research was low user convenience (48%); scores were perceived as cumbersome to calculate and use. On a similar note, the most desirable score characteristic was ease of use (29%), with score accuracy for predicting outcome and score objectivity the next most requested characteristics (17% for each). The most desirable score format was perceived as one which could be rapidly hand calculated from a table (78%), while the least desirable was one involving calculation from an exponentiated equation (4%). Scores in some form were used by the majority of responders in their day to day clinical practice (74%). Responders were familiar with between 0 and 17 illness scores, with a median of 4. When subjects were polled regarding appropriate illness or outcome score use for individual patients, 65% of subjects felt that scores could be used in conjunction with clinical judgment to appropriately direct therapy/euthanasia, 26% felt they should not be used to direct therapy/euthanasia for the individual but should be applied only to populations, and 9% felt they could be used appropriately as a sole means of directing therapy/euthanasia.

Conclusion

End user convenience is a key factor determining score uptake. Outcome scores are achieving use in clinical practice; however, perceptions of appropriate use vary widely.

Appendix 2: SI unit APPLE models

APPLE_{full} Score SI units

| | | | | | | | |
|-----------|------------|------------|---------------------------------|--------------------------------------|---------------|------------|----------|
| | | | creatinine (umol/l) 0-55 | 1 56-120 | 8 121-200 | 9 >200 | |
| | | | 9 <5.1 | wbc (x10 ⁹ /l) 5.1-8.5 | 2 8.6-18 | 3 >18 | |
| 6 <26 | 7 26-30 | 9 31-32 | albumin (g/l) 33-35 | 2 >35 | | | |
| 10 <90 | 4 90-94 | 1 95-97 | SpO ₂ (%) 98-100 | | | | |
| | | | total bilirubin (umol/l) 0-4 | 6 5-8 | 4 9-16 | 3 >16 | |
| | | | mentation score 0 | 5 1 | 7 2 | 8 3 | 13 4 |
| | | | respiratory rate (bpm) <25 | 3 25-36 | 5 37-48 | 6 49-60 | 5 >60 |
| | | | age (years) 0-2 | 6 3-5 | 8 6-8 | 7 >8 | |
| 3 2 | 4 1 | | fluid score 0 | | | | |
| | | | lactate (mmol/l) 0-1.9 | 2 2.0-7.9 | 3 8.0-11.0 | 6 >11 | |

Footnote: See Table 3 for calculation of 'fluid score' and 'mentation score'. A value of zero is ascribed for each parameter in the central zone. 'Mentation score' is collected at admission, for all others utilise the most abnormal value identified over the 24 hour period following admission. If history and physical exam fail to prompt assessment of SpO₂ or fluid score, assign zero.

Figure A1. Canine 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 10 parameters listed, with a maximum potential score of 80.

APPLE_{fast} Score SI units

| | | | | | | | | |
|------------------|---------------------|---------------------|--------------------------------------|----------------------------------|------------------|------------------|------------------|----------------|
| 7 <4.6 | 8 4.6-5.6 | 9 5.7-9.0 | 10 9.1-15.0 | glucose (mmol/l) >15.0 | | | | |
| 8 <26 | 7 26-30 | 6 31-32 | albumin (g/l) 33-35 | | 2 >35 | | | |
| | | | | lactate (mmol/l) <2 | 4 2-8 | 8 8-10 | 12 >10 | |
| 5 <151 | 6 151-200 | 3 201-260 | platelet count 261-420,000 | | 1 >420 | | | |
| | | | | mentation score 0 | 4 1 | 6 2 | 7 3 | 14 4 |

Footnote: See Table 3 for calculation of 'mentation score'. A value of zero is ascribed for each parameter in the central zone.
'Mentation score' is collected at admission, for all others utilise the most abnormal value identified over the 24 hour period following admission.

Figure A2. Canine Acute Patient Physiologic and Laboratory Evaluation (APPLE_{fast}) score: Calculated by summing the value in the upper left corner of the appropriate cell for each of the 5 parameters listed, with a maximum potential score of 50.