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

Lesley G. King, MVB; James S. Wohl, DVM; Ann Marie Manning, DVM; Susan G. Hackner, DVM; Marc R. Raffe, DVM; Greg Maislin, MS, MA

Objective—To prospectively evaluate a survival prediction index (SPI) in dogs admitted to intensive care units (ICU) and to generate and test an improved SPI (ie, SPI2).

Sample Population—Medical records of 624 critically ill dogs admitted to an ICU.

Procedure—Data were collected from dogs within 24 hours after admission to an ICU. Variables recorded reflected function of vital organ systems, severity of underlying physiologic derangement, and extent of physiologic reserve; outcome was defined as dogs that survived or did not survive until 30 days after admission to the ICU. Probabilities of survival were calculated, using an established model (SPI). We then performed another logistic regression analysis, thereby reestimating the variables to create the new SPI2. Cross-validation of the models obtained was performed by randomly assigning the total sample of 624 dogs into an estimation group of 499 dogs and validation group of 125 dogs.

Results—Testing of SPI resulted in an area under the curve (AUC) of 0.723. Testing of SPI2 revealed an AUC of 0.773. A backwards-elimination procedure was used to create a model containing fewer variables, and variables were sequentially eliminated. The AUC for the reduced model of SPI2 was 0.76, indicating little loss in predictive accuracy.

Conclusions and Clinical Relevance—The new SPI2 objectively stratified clinical patients into groups according to severity of disease. This index could provide an important tool for clinical research. (*Am J Vet Res* 2001;62:948–954)

Research is necessary to advance our understanding of pathophysiologic characteristics and management of diseases. Although animals can be important for use in models of naturally developing diseases that affect humans and domestic animals, their use has certain disadvantages. Experimentally induced disease may not be identical to that which develops naturally, and it is often homogeneous, compared with the heterogeneity of naturally developing diseases. Clinically affected animals with naturally developing disease usually represent a mixture of ages and breeds, and they may represent various stages of progression or severity of disease. They may have received a variety of treatments, and they may have multiple concurrent diseases in addition to the primary disease.

Although use of animals in a controlled research setting can provide vital information, we will never be able to replace the information obtained from clinical studies for testing various management techniques or evaluating new treatments. Conclusions from clinical studies, however, often are limited by an inability to define a homogeneous patient population. For example, the results of a study testing a new drug for dogs with immune-mediated hemolytic anemia may be difficult to interpret unless it can be verified that the dogs receiving the new treatment all have the same type and severity of hemolysis, the same degree of systemic illness, and, therefore, the same risk of mortality as the dogs receiving the standard (traditional) treatment. This problem can be partially addressed by developing a method for scoring severity of disease that attempts to place a numeric value on the degree of illness in the patients included in a clinical study. Ideally, such indices would be independent of diagnosis and, therefore, applicable to any patient and any disease. In addition, an ideal prediction index would use readily available information that could be collected early during the hospitalization period (ie, prior to initiation of treatment [a clinical study]). When it can be documented that the severity of illness initially is similar in 2 groups that then receive differing treatments, the results of the clinical study are considered more reliable. Numeric scoring of the severity of disease is analogous to risk stratification for prediction of survival.

A **survival prediction index (SPI)** has been developed.[1] That index was created by use of data from a group of 200 dogs admitted to the **intensive care unit (ICU)** at the Veterinary Hospital of the University of Pennsylvania. The SPI was calculated, using clinical variables collected by a single person within the first 24 hours after a dog was admitted to the ICU. One objective of the multiple-center study reported here was to prospectively test the accuracy of the SPI in a large cohort of dogs admitted to ICU at various facilities in an attempt to include critically ill dogs from sev-

Received Dec 13, 1999.
Accepted Jun 19, 2000.
From the Section of Critical Care, Department of Clinical Studies, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA 19104-6010 (King); the Department of Small Animal Surgery and Medicine, School of Veterinary Medicine, Auburn University, Auburn, AL 36849 (Wohl); Angell Memorial Animal Hospital, 350 S Huntingdon Ave, Boston, MA 02130 (Manning); VCA Veterinary Referral Associates Inc, 15021 Dufief Mill Rd, Gaithersburg, MD 20878 (Hackner); Department of Veterinary Clinical Medicine, University of Illinois, Urbana, IL 61802 (Raffe); and Biomedical Statistical Consulting, 1357 Garden Rd, Wynnewood, PA 19096 (Maislin).
Supported by the Bernice Barbour Foundation, Hackensack, NJ.

eral locations and to use data collected by numerous personnel. A second objective was to use the data to generate a new **improved SPI (SPI2)**. Finally, using estimation and validation samples, we intended to test the predictive accuracy of SPI2 for future use.

Materials and Methods

Study population—Data were collected for dogs admitted to the small animal ICU at 4 locations (location 1, University of Pennsylvania; location 2, Tufts University; location 3, University of Minnesota; location 4, VCA Veterinary Referral Associates, Gaithersburg, Md). Critically ill dogs with naturally developing conditions requiring medical or surgical treatment were eligible for inclusion in the study. Dogs that had been treated by a referring veterinarian, at an emergency clinic, or in the ward area of a participating hospital prior to admission to the ICU were eligible for inclusion. Dogs that were euthanatized because they were judged to be in the terminal stages of an incurable disease also were eligible for inclusion. Dogs were excluded when data collection posed a risk to the animal, they were euthanatized purely on the basis of financial reasons, or data collection was incomplete.

Data collection—All data were collected within 24 hours after each dog was admitted to an ICU. Variables were chosen on the basis of the established SPI[1]; variables were selected to reflect function of vital organ systems, severity of underlying physiologic derangement, and extent of physiologic reserve.

Variables such as WBC count and concentrations of creatinine, albumin, bilirubin, and bicarbonate usually were measured only once during the 24-hour period. Some variables were measured serially during routine monitoring in the 24-hour period (eg, rectal temperature, heart rate, respiratory rate, **mean arterial pressure [MAP]**, **oxygen saturation [Sao_2]**, PCV, and concentrations of **total solids [TS]** and glucose). When serial measurements of a variable were available, the most abnormal value (ie, the value that had the greatest deviation from the reference range) was used in the analysis, except for dogs after surgery, because the initial rectal temperature usually was low and then increased to a plateau during the first few hours. In these dogs, the plateau temperature was used for the analysis rather than the initial rectal temperature at time of admission to an ICU.

Dogs were categorized with regard to neurologic status on the basis of whether they had objective evidence of diseases of the brain or spinal cord. Dogs also were categorized on the basis of medical versus surgical admissions. Surgical admissions could be admitted to an ICU before or after surgery. Dogs were categorized on the basis of the chronic nature of their illness. Chronic disease was defined as a dog having failure of a major organ persisting for > 1 month (eg, chronic renal failure, chronic hepatic insufficiency, or chronic congestive heart failure), systemic neoplasia (eg, lymphosarcoma or hemangiosarcoma), any form of immunosuppression before admission to an ICU (eg, chemotherapy, treatment with immunosuppressive drugs, hyperadrenocorticism, or diabetes mellitus), or a congenital disease that could lead to compromised organ function (eg, brachycephalic airway conformation, immune deficiency, portosystemic shunt, or cardiovascular anomalies).

Outcome—Outcome was defined as dogs that survived or did not survive until 30 days after admission to an ICU. Outcome was determined by examination of a dog 30 days after admission to an ICU that was conducted by the investigators or a referring veterinarian. Outcome also was determined for some dogs by a telephone conversation with the owners.

Statistical analysis—Categoric and continuous variables were summarized and tested for significant differences among the 4 locations. For categoric variables, differences among locations were tested by use of the χ^2 test, whereas the Kruskal-Wallis test was used for variables measured on an interval scale. Significance was set at a value of $P < 0.05$.

There were 2 main approaches to the multivariable analysis. First, we calculated probabilities of survival in this multiple-center group of dogs, using the previously established SPI.[1] The logit *P* value was calculated for this group of dogs, using the following equation:

$$\text{logit } P = -19.9676 + (0.1537 \bullet \text{rectal temperature}) + (0.00742 \bullet \text{MAP}) - (0.0144 \bullet \text{body weight}) - (0.00164 \bullet \text{heart rate}) + (0.00423 \bullet \text{respiratory rate}) + (0.0668 \bullet Sao_2) + (0.0123 \bullet \text{bicarbonate}) - (1.2567 \bullet \text{neurologic status}) - (0.8375 \bullet \text{creatinine}) + (0.0350 \bullet \text{PCV}) - (0.0403 \bullet \text{WBC}) - (0.2732 \bullet \text{albumin}) + (0.5082 \bullet \text{TS}) - (0.1163 \bullet \text{age}) + (0.00280 \bullet \text{glucose}) - (0.2114 \bullet \text{bilirubin}) - (0.9294 \bullet \text{medical vs surgical status}) - (0.0796 \bullet \text{chronic vs acute status})$$

where medical vs surgical status was a dichotomous variable (medical = 1, surgical = 0), and chronic status was a dichotomous variable (chronic = 5, acute = 0).

The SPI value was calculated for this group of dogs, using the following logistic equation:

$$P(\text{SPI}) = \frac{e^{\text{logit } P}}{(1 + e^{\text{logit } P})}$$

Next, we used the data from this group of dogs to perform another logistic regression analysis, thereby reestimating the variables to create a new full model. Cross-validation of the models was performed by randomly assigning the total sample of 624 dogs into an estimation group of 499 dogs and validation group of 125 dogs. In addition, when appropriate, a second validation group was created, using the data reported for the original study (a group of 200 dogs at location 1).[1] A backwards-elimination procedure was used to sequentially eliminate variables that had a minimal incremental impact on the accuracy of the new full model. We sequentially eliminated any variable that had a value of $P > 0.10$. Therefore, we obtained an SPI2 with fewer variables while retaining a comparable amount of discrimination for predicting outcome. Cross-sample validation was used to assess the degree to which stepwise removal of variables reduced potential generalizability.

Two model-building strategies were implemented. The first strategy involved extensive individual analysis for each prognostic variable by which the most predictive functional relationship with the log odds of survival (linear, quadratic, step-function, or categoric) was identified and used in the stepwise algorithm. The second strategy simply assumed linear associations with log odds for all interval prognostic variables. Although the first strategy produced a higher value for **area under the curve (AUC)** in the estimation sample, dramatic shrinkage on cross-validation required us to abandon this strategy (data not shown).

Receiver operating characteristic (ROC) curves were plotted to document the tradeoff between true-positive (sensitivity) versus false-positive (1 – specificity) rates as a function of varying the predictive cut-point. The AUC for the ROC curve was computed to produce summary indices of predictive discrimination,[2] using a formula equivalent to the following: (Somer's D rank correlation coefficient/2) + 0.5. Magnitude of the AUC represents the probability that a randomly selected survivor has a larger predicted probability of survival than a randomly selected nonsurvivor. Therefore, the AUC is a summary measure of predictive value across all possible cut-points of the prediction index. Determination of a specific cut-point on which to base prediction of a specific dog's survival depends on the survival rate and the ratio of

costs associated with false-positive predictions to costs associated with false-negative predictions.[3] The cut-point that minimizes total prediction errors is optimal when costs are assumed to be equal. Assuming that the costs of errors were equal in the study reported here, the optimal cut-point was derived from the ROC curve, using the ratio of nonsurvival rates to survival rates. The optimal cut-point was located at the point where the slope of the line tangent to the curve was equal to this ratio.

Potential generalizability was assessed by computing the proportional reduction in the AUC discrimination index.[4] This value, when computed in the validation sample by use of an equation derived in the estimation sample, was compared with its value obtained from the estimation sample. Thus, predictive shrinkage was computed as follows:

$$\left(\frac{AUC_{estimation} - AUC_{validation}}{AUC_{estimation}}\right) \times 100$$

Hosmer-Lemeshow goodness-of-fit tests[5] were constructed for logistic regression models. The null hypothesis for this test was that the model adequately fit the true survival probabilities. It was constructed by categorizing the predicted probabilities into deciles and comparing the observed and expected number of survivors and nonsurvivors within each decile. When the model fits, the test statistic has an approximate χ^2 distribution with 8 degrees of freedom.

Standardized odds ratios with 95% **confidence intervals (CI)** were computed to assess the clinical relevance of regression coefficients in the reduced-model SPI2 following the backwards-elimination procedure. Standardized odds ratios estimate relative changes in probability of survival associated with changes of 1 SD in each variable.

As a final analysis, predictive values for dogs that were medical or surgical admissions were compared. This was an attempt to determine the potential applicability of this test to differing patient populations.

Results

Patient population—Data were collected from 624 dogs within 24 hours after admission to an ICU. Data were collected by several people, including the investigators, veterinary technicians, and veterinary students. There were 336 dogs from location 1, 147 from location 2, 16 from location 3, and 125 from location 4. There were 286 (47.4%) female dogs, 227 of which were spayed, and 328 male dogs, 177 of which were neutered. There were 109 (17.5%) mixed-breed dogs, and 1 or more of 82 purebred dogs, with German Shepherd Dogs, Golden Retrievers, Labrador Retrievers, and Cocker Spaniels represented most frequently.

Overall survival rate was 61.1% (381/624 dogs alive at 30 days). We did not detect a significant difference in survival rate among locations (**Table 1**). Of the 624 dogs, 407 (65.2%) were medical admissions, and 217 (34.8%) were surgical admissions. There were 404 (64.7%) dogs admitted with acute disease and 220 (35.3%) admitted with chronic disease. There were 544 (87.2%) dogs without evidence of neurologic disease, but 80 (12.8%) had objective evidence of disease in the brain or spinal cord. Descriptive statistics for continuous variables were summarized (**Table 2**).

Use of SPI to predict outcome—The SPI originally was derived by use of data collected at location 1 from 200 dogs; the AUC of SPI was 0.89. When the SPI was tested on data collected from the new (estimation) group of dogs (n = 499), the AUC decreased to 0.723. The SPI then was tested, using only data from the estimation sample for location 1 (274 dogs), to determine whether location significantly influenced results. Using this group of dogs, the AUC remained 0.723.

Table 1—Categoric variables for dogs admitted to intensive care units (ICU), on the basis of location of data collection

	Location				
Variable	1 (n = 336)	2 (n = 147)	3 (n = 16)	4 (n = 125)	All (n = 624)
Survival at 30 days after admission	217 (64.6)	84 (57.1)	12 (75.0)	68 (54.4)	381 (61.1)
Medical vs surgical admission*	194 (57.7)	113 (76.9)	8 (50.0)	92 (73.6)	407 (65.2)
Chronic disease*	200 (59.5)	115 (78.2)	10 (62.5)	81 (64.7)	404 (64.7)
Neurologic disease*	29 (8.6)	29 (19.7)	2 (12.5)	20 (16.0)	80 (12.8)

Values in parentheses are percentages.
*Values differ significantly ($P < 0.05$) among locations (χ^2 test).

Table 2—Descriptive statistics for continuous variables in data collected from 624 dogs at 4 locations within 24 hours after admission to an ICU

Variable	Mean	SD	Minimum	Maximum
Age (y)*	7.4	4.2	0.2	17.0
Body weight (kg)	25.1	15.1	1.4	87.0
Rectal temperature (F)	101.4	2.19	90.9	107.0
Heart rate (bpm)*	134.8	39.2	43.0	264.0
Respiratory rate (rpm)*	50.4	25.0	10.0	180.0
MAP (mm Hg)*	90.9	25.7	34.0	185.0
Sao_2 (%)	92.4	5.8	53.0	99.8
PCV (%)*	32.6	12.9	6.0	74.0
Total solids (g/dl)*	5.3	1.6	0.6	12.0
Glucose (mg/dl)*	114.6	97.0	15.0	1,193.0
WBC (× 1,000 cells/μl)	19.3	11.4	0.5	82.0
Creatinine (mg/dl)*	1.74	2.74	0.2	37.0
Albumin (g/dl)*	2.75	0.8	0.7	6.2
Bilirubin (mg/dl)	0.89	2.54	0.0	29.0
Bicarbonate (mmol/L)*	20.1	5.6	2.4	56.6

*Values differ significantly ($P < 0.05$) among locations.
bpm = Beats per minute. rpm = Respirations per minute. MAP = Mean arterial blood pressure. SaO_2 = Oxygen saturation.
To convert Fahrenheit to Celsius, subtract 32 and multiply the difference by 5/9.

Calculation of SPI2—Results of the new multivariable logistic regression analyses were calculated, using the estimation sample of 499 dogs (**Table 3**). To facilitate future application of the equation, we used a weighted average of the prevalence of survival for each of the 4 locations to adjust the intercept rather than a set of location-specific intercepts. The weighted intercept was computed as follows:

$$\text{Adjusted intercept} = \text{model intercept} + \log([1 - p1]/p1) + \log(p2/[1 - p2])$$

where p1 was the baseline prevalence for the reference location (ie, location 1), and p2 was the baseline prevalence for the weighted average of all locations combined. For the study reported here, value of p1 was 0.6715 (274 dogs at location 1), and value of p2 was 0.635 (499 dogs). Thus, for the SPI2, the new adjusted intercept was computed as follows: $-7.50 = -7.34 + \log([1 - 0.6715]/0.6715) + \log(0.635/[1 - 0.635])$.

Table 3—Results of logistic regression for the estimation sample of 499 dogs, using full and reduced models

Variable	Full modelFeIF Full Model			Reduced-model SP12*			OR for variables in the reduced model			
	Beta	SE	*P*	Beta	SE	*P*	Change in variable (± 1 SD)	OR	95% Confidence limits Lower	Upper
Intercept	−7.5100	5.4692	—	0.3273	0.5770	—	—	—	—	—
Temperature	0.0598	0.0501	0.232	—	—	—	—	—	—	—
MAP	0.0081	0.0043	0.059	0.0108	0.0041	0.008	26.1699	1.328	1.077	1.638
Body weight	−0.0069	0.0077	0.37	—	—	—	—	—	—	—
Heart rate	−0.0044	0.0031	0.151	—	—	—	—	—	—	—
Respiratory rate	−0.0098	0.0044	0.026	−0.0102	0.0041	0.014	−24.7648	1.288	1.054	1.574
Sa_{O_2}	0.0237	0.0180	0.188	—	—	—	—	—	—	—
Bicarbonate	0.0342	0.0200	0.088	—	—	—	—	—	—	—
Neurologic status	−0.2101	0.3457	0.543	—	—	—	—	—	—	—
Creatinine	−0.2074	0.0519	0.0001	−0.2183	0.0506	0.0001	−2.9343	1.897	1.418	2.538
PCV	0.0107	0.0095	0.262	0.0164	0.0086	0.055	12.8209	1.234	0.996	1.530
WBC	−0.0101	0.0094	0.282	—	—	—	—	—	—	—
Albumin	0.3201	0.1575	0.042	0.3553	0.1417	0.012	0.8021	1.330	1.064	1.661
Total solids	0.0823	0.0917	0.369	—	—	—	—	—	—	—
Age	−0.1146	0.0277	0.0001	−0.1184	0.0257	0.0001	−4.1977	1.644	1.330	2.031
Glucose	−0.0001	0.0012	0.924	—	—	—	—	—	—	—
Bilirubin	−0.0379	0.0426	0.374	—	—	—	—	—	—	—
Acute vs chronic	−0.2259	0.2390	0.3445	—	—	—	—	—	—	—
Medical vs surgical admission	−0.8249	0.2529	0.001	−0.8069	0.2336	0.001	—	—	—	—

*Reduced-model of the new improved survival prediction index (SPI2) created by a backward-elimination procedure, eliminating variables with $P > 0.10$. OR = Odds ratio. — = Not determined.

The *P* values reported here represent the significance for each variable after stepwise elimination of nonsignificant variables. Standardized odds ratios provide the proportional change in odds of survival associated with a change of 1 SD in each variable while holding values for all other variables constant, which allows assessment of the clinical relevance of the coefficient for each variable. Area under the curve (AUC) for the full-model SPI2 was 0.773, and AUC for the reduced-model SPI2 was 0.758.

See Table 2 for remainder of key.

In the new (log-linear) model of SPI2, AUC obtained for the estimation sample was 0.773. To ascertain whether a model containing fewer variables could adequately predict survival, a backwards elimination procedure was used. Variables deleted included rectal temperature, body weight, heart rate, Sa_{O_2}, bicarbonate concentration, neurologic status, WBC count, concentrations of total solids, glucose, and bilirubin, and chronic vs acute status (Table 3). The AUC of the estimation sample for this reduced-model SPI2 was 0.76, indicating that there was not an appreciable loss of accuracy.

For the reduced-model SPI2, the new adjusted intercept was computed as follows:

$$0.3273 = 0.4885 + \log([1 - 0.6715]/0.6715) + \log(0.635/[1 - 0.635]).$$

The logit *P* value for the reduced-model SPI2 was calculated as follows:

Logit *P* = 0.3273 + (0.0108 • MAP) – (0.0102 • respiratory rate) – (0.2183 • creatinine) + (0.0164 • PCV) + (0.3553 • albumin) – (0.1184 • age) – (0.8069 • medical vs surgical status)

where medical vs surgical status was a dichotomous variable (medical = 1, surgical = 0), and chronic status was a dichotomous variable (chronic = 5, acute = 0). The adjusted intercept subsequently could be used at any location. However, for increased accuracy, a new intercept for any sample at a new location with a different baseline prevalence of survival could be computed as follows:

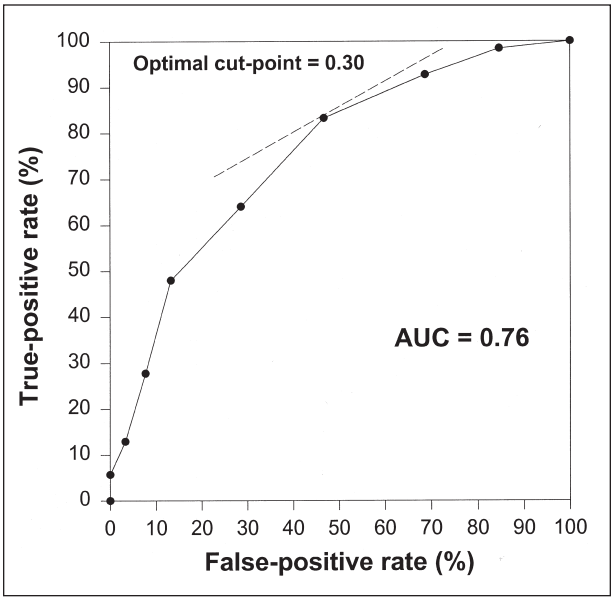

Figure 1—Receiver operating characteristic (ROC) curve for the reduced model of the new survival prediction index (SPI2) used on data for 499 dogs from 4 locations. Magnitude of the area under the curve (AUC) represents the probability that a randomly selected survivor has a larger predicted probability of survival than a randomly selected nonsurvivor. Therefore, the AUC is a measure of predictive value across all possible cut-points of the prediction index. The optimal cut-point was derived from the ROC curve, using the ratio of nonsurvival rate to survival rate, and is located at the point where the slope of the line tangent to the curve is equal to this ratio.

$$\text{New intercept} = 0.3273 + \log([1 - 0.635]/0.635) + \log(p3/[1 - p3])$$

Table 4—Results of AUC for the estimation sample of 499 dogs, validation sample of 125 dogs, and original sample of 200 dogs*

	Estimation sample	Validation sample		Original sample	
Calculation	AUC	AUC	Shrinkage (%)	AUC	Shrinkage (%)
Original model SPI[1]	0.72†	0.65	24	0.89	NA
Full-model SPI2	0.77	0.72	6.5	0.75	2.6
Reduced-model SPI2‡	0.76	0.68	10.5	0.71	6.6

*The estimation sample of 499 dogs and validation sample of 125 dogs were randomly allocated from 624 dogs from 4 locations from which data were collected. Data for the original sample of 200 dogs were collected only from location 1 and has been reported elsewhere.[1] †Represents 19.1% shrinkage when the original survival prediction index (SPI) calculation was applied to the new sample of 499 dogs. ‡Reduced-model SPI2 created by a backward-elimination procedure, eliminating variables with $P > 0.10$.

NA = Not applicable.

See Table 3 for remainder of key.

where p3 is the prevalence of survival at the new location. Then the value for SPI2 for an animal could be obtained, using the same exponential equation described previously for calculation of SPI from logit *P*.

Use of ROC analysis to assess predictive value— The ROC curve for the reduced-model SPI2 was plotted (**Fig 1**). The optimal cut-point was derived from the ROC curve, using the ratio of nonsurvival rate to the weighted average for survival rate; values for the study reported here were 0.365 and 0.635, respectively. Use of this tangent line yielded a cut-point of approximately 0.30. Estimated positive- and negative-predictive values for this optimal cut-point were 75.6% (95% CI, 71.1 to 80.1%) and 64.7% (95% CI, 57.0 to 72.3%), respectively. Estimated sensitivity and specificity at this cut-point were 83.3% (95% CI, 79.2 to 87.4%) and 53.3% (95% CI, 46.0 to 60.5%), respectively.

Assessment of generalizability for SPI and reduced-model SPI2—An AUC value of 0.89 for SPI was calculated for the data from the original 200 dogs.[1] When SPI was prospectively validated by use of data from the 499 dogs used to estimate reduced-model SPI2, AUC declined to 0.72, representing shrinkage of 19.1%. When SPI was validated by use of data from the 125 dogs used to validate the reduced-model SPI2, AUC decreased from 0.89 to 0.65, representing shrinkage of 24%.

Compared with the reduction in the AUC for SPI, the AUC for the full-model SPI2 had less shrinkage (**Table 4**). The AUC decreased from 0.77 to 0.72, a proportional decrease of 6.5%. The AUC for the reduced-model SPI2 had shrinkage of 10.5%.

As a final validation analysis, we then applied SPI2 to the sample of 200 dogs used to develop the original SPI (Table 4). The AUC values were larger than those computed by use of the validation sample of 125 dogs from the 4 locations. Shrinkage was 2.6% for the full-model SPI2 and 6.6% for the reduced-model SPI2.

For the full-model SPI2, the Hosmer-Lemeshow goodness-of-fit χ^2 value was 5.17 (8 degrees of freedom). The null hypothesis of an adequate fit of the model to the actual survival probabilities was not rejected ($P = 0.740$). A similar finding was observed for the reduced-model SPI2, with $\chi^2 = 6.60$ (8 degrees of freedom, $P = 0.580$). Thus, these 2 models appeared to produce predicted probabilities that adequately fit the true survival probabilities.

Assessment of clinical relevance of the reduced-model SPI2—Standardized odds ratios with 95% CI were generated to characterize the clinical relevance of the regression coefficients in the reduced model (Table 3). Standardized odds ratios provide the proportional change in probability of survival associated with a change of 1 SD in each variable, allowing assessment of the clinical relevance of the coefficient for each variable. Analysis of the results revealed that our findings for all variables made sense clinically: the odds of survival increased with an increase in MAP, PCV, and albumin concentration and with a decrease in respiration rate, creatinine concentration, and age. For example, when all other variables were held constant, an increase of 1 SD in MAP (ie, increase of 26 mm Hg) improved the odds of survival by a factor of 1.328, and a decrease in creatinine concentration of 1 SD (ie, 2.9 mg/dl) improved the odds of survival by a factor of 1.897.

Comparison of predictive value for medical and surgical admissions—We investigated the predictive accuracy of the reduced-model SPI2 for medical admissions, using the estimation and validation samples. When applied to this subsample, the AUC was 0.75. When applied to surgical admissions, the AUC was 0.70. Thus, it appears that the relative predictive value was higher for dogs in an ICU with a medical admission. We then examined results of separate stepwise elimination procedures, on the basis of admission, to determine whether the importance of specific variables differed by type of admission. When we reestimated the reduced-model SPI2 for medical admissions, variables with a value of $P < 0.10$ included rectal temperature, MAP, heart rate, respiratory rate, bicarbonate concentration, creatinine concentration, PCV, albumin concentration, and age (data not shown). In contrast, only PCV, albumin concentration, and age had a value of $P < 0.10$ for surgical admissions.

Discussion

By deriving and validating a method for objectively stratifying clinical patients into groups on the basis of severity of disease, the study reported here provides

researchers with a vitally important tool for clinical research. When treatments or outcomes are being studied in clinical patients, we now have a numeric tool to quantify the severity of disease, thereby allowing researchers to statistically document that groups of patients are similar with regard to their likelihood of short-term survival prior to treatment.

In another study,[1] we derived an equation for prediction of outcome in critically ill dogs (the SPI). One of the most useful summary assessments of the accuracy of a predictive equation is calculation of the AUC for an ROC curve, which was quite high (0.89) in the initial group of 200 dogs. The first objective of the study reported here was to prospectively test that SPI on data from a larger group of dogs, from multiple locations, and accrued by a number of personnel. When we applied the data from the new group of dogs to the original SPI, we observed that AUC decreased to 0.72. A comparison of the difference in AUC between the original 200 dogs and the new 499 estimation and 125 validation samples comprises the measure of shrinkage for validation analysis (Table 4). Shrinkage estimates of newly estimated models may be expected to be roughly 10%. In this case, the shrinkage was quite high (19.1%). The large AUC value previously reported was positively biased, because it was computed by using the same data that was used to generate the original SPI. Additionally, it is certainly possible that there were differences between the original population of test dogs and those included in this multiple-center study. Although the populations were from similar sources in the 2 studies, changes in medical practice during the 5-year period between the 2 studies may have resulted in slight differences in dogs admitted to an ICU and in differing outcomes for dogs with a specific disease. Involvement of multiple data collectors also may have introduced heterogeneity into the new test population.

Because the findings in this larger group of dogs were probably more representative of the population for which this model was designed, we reevaluated the data to create and validate the modified SPI2. Dogs were randomly allotted into estimation and validation samples. Data from the estimation sample of 499 dogs then were analyzed by use of logistic regression analysis to create a new SPI2. A backward-elimination procedure was performed to remove those variables that made a nonsignificant contribution to the prediction, resulting in a final AUC of 0.76. The predictive accuracy of the reduced-model SPI2 was tested, using the smaller validation group of 125 dogs. Shrinkage for the reduced-model SPI2 was approximately 10% (AUC for validation sample, 0.71), which suggests that the reduced-model SPI2 may be more applicable to new groups of dogs than the original SPI.

Variables included in the model reported here are clinically relevant and routinely monitored in critically ill animals. Because these data can be collected within the first 24 hours after admission to an ICU, severity of disease can be assessed prior to application of the test treatment. In addition, this method potentially can be applied to any critically ill animal, because it is independent of diagnosis. Because it is objective and repeatable, this tool also could be used to compare management within an institution over time or even differences in management among institutions.

To apply the SPI2 prior to initiation of a clinical research study, personnel must simply record the most abnormal values within a 24-hour period for the variables MAP, respiratory rate, serum concentrations of creatinine and albumin, and PCV; age in years; and whether the animal is a medical or surgical admission to an ICU. Numeric values are entered into the linear equation to yield a value for logit *P*, which is equal to the log odds of survival, and then the exponential equation is used to solve for *P*, the predicted probability of survival (ie, SPI2 value). The predicted probability value obtained in this manner is within the range of 0 to 1, with 0 indicating the most severe disease with a high risk of fatality and 1 indicating a disease with a low risk of fatality. Thus, numeric values provide a method for quantifying severity of disease, which then can be statistically compared to document similar or differing severity of disease in the test groups.

The SPI2 implicitly assumes that the log odds of survival are linear with respect to changes in values for clinical variables. Our first modeling strategy exhaustively searched for the optimal functional form (ie, linear, quadratic, step-function, categoric) and used these functional forms to develop a model for the estimation sample. Because shrinkage on cross-validation was sufficiently dramatic, we abandoned this strategy in favor of an assumption of log-linear associations. Shrinkage for the log-linear model was modest. This highlights the necessity of performing a prospective validation to avoid overoptimistic assessments of predictive utility.

Survival rates at the 4 locations ranged from 55 to 75%. The intercept used in the reported equation was adjusted by use of a weighted average of the survival rates for each of the 4 locations. This should provide a reasonable estimate for use at most other potential locations, as long as the population tested has a survival rate in the same range. If the equation is to be used for a patient population with a survival rate that differs substantially, a new intercept should be calculated, using the survival rate of the new location.

The type of statistical analysis we used to create a prediction equation has been reported elsewhere in the veterinary literature.[6-10] For example, this type of data analysis has been used for predicting the need for surgery in horses with colic,[6] prognosticating survival in critically ill foals[7,8] and dogs that were high-risk surgical patients,[9] and, most recently, predicting septicemia in calves with diarrhea.[10] Although these equations have variable predictive accuracy for their respective applications, most are applied to a relatively uniform population of patients. The SPI2 can be applied to any patient, regardless of diagnosis, providing a tool that can be used early during hospitalization for various clinical research studies.

Numerous similar studies have been published for humans, and a number of similar systems are in routine use for clinical research studies of human patients.[11-14] The most commonly used system in human medicine is the **acute physiology, age, chronic health evaluation (APACHE)** scoring system.[11,12] In

studies of human patients, AUC values commonly obtained are as high as 0.90. The numbers of human patients from whom data were collected to develop those equations were much greater than can easily be collected by veterinary researchers. For example, data were accrued from > 17,000 patients for development of the APACHE III scoring system.[12] Including a larger group of animals may make it possible to create another more accurate prediction system.

The APACHE III scoring system is similar to the SPI2, because it includes MAP, respiratory rate, Hct, and concentrations of creatinine and albumin. Some variables that were found to be nonsignificant in the dogs of our report have been included in the APACHE system for human patients, such as heart rate, temperature, partial pressure of oxygen or alveolar-arterial oxygen gradient, WBC count, and concentrations of bilirubin and glucose. Other variables that are included in the APACHE scoring system but that were not tested as part of our study include urine output, blood urea nitrogen concentration, and serum sodium concentration.

With the increased prevalence of managed care in the human medical system, there has been considerable discussion about the use of these severity-of-disease scoring systems for prioritization of resources toward patients that have a greater predicted chance of survival. Although such an application may have debatable merit in human medicine from a moral perspective, the high AUC values in the scoring systems for humans suggest that there may be sufficient predictive accuracy to justify their use for this purpose. The AUC of 0.77 reported for the full-model SPI2 provides sufficient predictive accuracy to be a valid tool for statistical analysis but is insufficient to justify its use for decision-making regarding specific dogs. If this equation were used by clinicians to decide which dogs should be treated and which should be euthanatized, decisions would be made that would result in some dogs being euthanatized that would have survived and unsuccessful and costly treatment of other dogs that would not have survived.

Various sources of error exist in the calculation of these predictive equations. The most important of these is the possibility of euthanatizing an animal. Certainly, a source of error was created when decisions were made to euthanatize some of the dogs rather than allowing them to succumb naturally. It is possible that some of the dogs listed as nonsurvivors could, in fact, have survived if treatment had continued. We attempted to minimize this error by excluding dogs that were euthanatized strictly for financial reasons. Relying on the clinical acumen of participating clinicians, we assumed that most euthanatized dogs would have succumbed without treatment. Although it would be ideal to only include dogs that died naturally, this is an unreasonable goal within the standard of practice in modern veterinary medicine. Attempts to create a SPI2 with survival or nonsurvival as the desired end-point must anticipate and accept error inherently incurred by euthanatizing animals.

Additional studies are needed to further validate and refine this system. Experience with similar systems in human medicine has revealed that these systems may be more accurate in certain patient populations. For example, in 1 study,[15] it was documented that the APACHE system is not as accurate in human patients treated for trauma-induced conditions, compared with human patients treated for medical conditions. The authors of that study hypothesized that the prediction for previously healthy humans who had acute trauma would have been more accurate if it had included an anatomic component similar to that suggested in the animal trauma triage scoring system.[16] Analysis of our findings suggest that this also may be true in dogs, because the predictive accuracy was better for medical admissions than for surgical admissions. However, it also should be mentioned that there were approximately twice as many medical admissions in our study, compared with the number of surgical admissions, which may have contributed to these differences.

Additional studies could include analysis of changes in the SPI2 value when it is serially calculated throughout the period of hospitalization. It is possible that this type of analysis could improve the value of the prediction such that it would become clinically useful for decision-making in specific patients. Other studies also should include data collection and creation of predictive equations for use in other species.

References

1. King LG, Stevens MT, Ostro ENS, et al. A model for prediction of survival in critically ill dogs (erattum appears in *J Vet Emerg Crit Care* 1995;5:6). *J Vet Emerg Crit Care* 1994;4:85–99.
2. Harrell FE, Lee KL, Mark DB. Tutorial in biostatistics: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–387.
3. McNeil BJ, Keeler E, Adelstein SJ. Primer on certain elements of medical decision making. *N Engl J Med* 1975;293:211–215.
4. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic curve (ROC). *Radiology* 1982;143:29–36.
5. Hosmer DW, Lemeshow S. Assessing fit. In: Hosmer DW, Lemeshow S, eds. *Applied logistic regression*. New York: John Wiley & Sons, 1989;140-145.
6. Reeves MJ, Curtis CR, Salman MD, et al. Multivariable prediction model for the need for surgery in horses with colic. *Am J Vet Res* 1991;52:1903–1907.
7. Hoffman AM, Staempfli HR, Willan A. Prognostic variables for survival of neonatal foals under intensive care. *J Vet Intern Med* 1992;6:89–95.
8. Furr M, Tinker MK, Edens L. Prognosis for neonatal foals in an intensive care unit. *J Vet Intern Med* 1997;11:183–188.
9. Hardie EM, Jayawickrama J, Duff LC, et al. Prognostic indicators of survival in high-risk canine surgery patients. *J Vet Emerg Crit Care* 1995;5:42–49.
10. Lofstedt J, Dohoo IR, Duizer G. Model to predict septicemia in diarrheic calves. *J Vet Intern Med* 1999;13:81–88.
11. Knaus WA, Draper EA, Wagner DP, et al. APACHE II: a severity of disease classification system. *Crit Care Med* 1985;13: 818–829.
12. Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system: risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991;100:1619–1636.
13. Le Gail JR, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993;270:2957–2963.
14. Lemeshow S, Teres D, Klar J, et al. Mortality probability models (MPM II) based on an international cohort of intensive care unit patients. *JAMA* 1993;270:2478–2486.
15. McAnena OJ, Moore FA, Moore EE, et al. Invalidation of the APACHE II scoring system for patients with acute trauma. *J Trauma* 1992;33:504–506.
16. Rockar RA, Drobatz KJ. Development of a scoring system for the veterinary trauma patient. *J Vet Emerg Crit Care* 1994;4:77–83.