

# Performance of predictive models of survival in horses undergoing emergency exploratory laparotomy for colic

Rebecca C. Bishop DVM, MS  |

Santiago D. Gutierrez-Nibeyro DVM, MS, DACVS-LA, DACVSMR-EQ |

Matthew C. Stewart BVSc, MVet, PhD, DACVSMR-EQ |

Annette M. McCoy DVM, MS, PhD, DACVS-LA 

Department of Veterinary Clinical  
Medicine, University of Illinois, Urbana,  
Illinois, USA

## Correspondence

Rebecca C. Bishop, University of Illinois  
College of Veterinary Medicine, 1008 West  
Hazelwood Drive, Urbana IL 61802, USA.  
Email: [rb17@illinois.edu](mailto:rb17@illinois.edu)

## Abstract

**Objective:** To evaluate previously published predictive survival models in a population of horses undergoing colic surgery in the midwestern United States.

**Study design:** Retrospective cohort study; single referral hospital.

**Animals:** A total of 260 horses met the inclusion criteria.

**Methods:** Medical records of horses undergoing surgical treatment for colic were reviewed. Previously published models were applied to cohort data to predict outcome. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy for prediction of short-term survival were calculated.

**Results:** Single-variable and multivariable models performed similarly for prediction of survival, with a mean 79% sensitivity (range: 44%–94%), 48% specificity (range: 22%–83%), 63% PPV (range: 56%–72%), 73% NPV (range: 60%–83%), and 64% accuracy (range: 59%–72%). Blood lactate  $\leq 6$  mmol/l and the colic severity score (CSS) were highly sensitive for prediction of survival; however, both had poor specificity.

**Conclusion:** Single-variable and multivariable predictive models did not perform as well for prediction of survival in the study cohort compared to original reports, suggesting that population-specific factors contribute to patient survival.

**Clinical significance:** Predictive models of survival developed in one population may be less reliable when used to predict outcome in horses undergoing colic surgery from an independent population. Additional model testing and refinement using data from multiple surgical centers could be considered to improve prediction of outcome for horses undergoing laparotomy for treatment of colic.

## 1 | INTRODUCTION

The decision to treat or euthanize a horse with severe colic is significantly influenced by the clinical judgment of the attending clinician. This subjective evaluation could result in pre- or intraoperative euthanasia of horses that may have survived if surgical intervention was pursued.<sup>1</sup> Possible reasons for premature destruction may include misconceptions regarding (1) lesion-specific prognosis; (2) the impact of surgery on future athletic performance; or (3) the effect of age or concurrent disease on recovery. Applying predictive models could augment clinical judgment by providing more objective criteria to help decide if a positive outcome is likely with surgical treatment. However, a predictive model developed in one horse population may not perform well in another horse population and the use of unvalidated predictive models in clinical practice is not recommended.<sup>2</sup>

Single variables used as predictors of prognosis in clinical practice are extrapolated from reports of similar case series or retrospective clinical studies. Abnormal values for multiple cardiovascular parameters (heart rate, capillary refill time, packed cell volume) have consistently been associated with perioperative survival.<sup>3–8</sup> Hematological parameters used to evaluate the risk of mortality include anion gap, L-lactate, ionized calcium, ionized magnesium, creatinine, and glucose.<sup>9–16</sup> Evaluation of peritoneal fluid lactate concentrations and appearance are also suggested to be useful to predict survival of horses with specific types of colic.<sup>6,8,17–21</sup> The advantage of single variable predictors is that they can be rapidly assessed in the course of clinical evaluation and do not require reference to algorithms or complex calculations. However, evaluation of a single variable cannot capture the entire clinical picture of a case.

Several prognostic models have been developed based on multivariable regression modeling of specific colic populations.<sup>22,23</sup> The colic severity score (CSS) uses numerical scores (0 to 4) to each of four categories (pulse, peritoneal total protein, blood L-lactate, and mucous membrane appearance) based on semi-quantitative cutoffs with a total score >7 strongly predictive of death.<sup>24</sup> Grulke et al. developed a shock score that also uses numerical scores (1 to 4) and quantitative cutoffs for several categories, including heart rate, respiratory rate, packed cell volume, among other variables.<sup>25</sup> More recently, McConachie et al. developed the equine multiorgan dysfunction syndrome for surgical gastrointestinal disease (MODS SGI) scoring system to predict survival in horses with acute surgical colic.<sup>26</sup> This model uses scores (0 to 3) based on qualitative descriptions or quantitative cut-offs for parameters of cardiovascular, renal, hepatic, respiratory, musculoskeletal, coagulation,

and gastrointestinal status that may not readily available in the emergency clinical setting.<sup>26</sup>

Although some prognostic models were developed for specific populations,<sup>2,7,11,12</sup> most have been developed using data from horses with a wide variety of underlying conditions that were treated either surgically or medically. Additionally, few models have been validated by testing in an independent population of horses. The authors have observed that in practice, clinicians often apply predictive models or single variable cutoff points without consideration of whether the patient being evaluated fits the specific inclusion criteria used for development of the relevant model. This is particularly true for patients that may require surgery, as there are few predictive models developed specifically for horses requiring surgical intervention for treatment of colic.

The objective of this study, therefore, was to evaluate previously published predictive survival models in a population of horses undergoing colic surgery in the mid-western United States. We hypothesized that the previously reported survival models would not perform as well in this study cohort as in the original populations. We further hypothesized that multivariable models would perform better than single variable models for predicting survival.

## 2 | MATERIALS AND METHODS

### 2.1 | Study design

Medical records for all horses that underwent exploratory laparotomy at the University of Illinois Veterinary Teaching Hospital between January 1, 2009 and December 31, 2020 were reviewed. Horses that underwent exploratory laparotomy for reasons unrelated to abdominal pain (e.g., traumatic body wall rupture), foals <6 months of age, and cases in which euthanasia was elected by owner despite recommendation for treatment were excluded. For horses undergoing multiple laparotomies during the same hospitalization, only the initial surgery was included in the analysis.

Admission data included signalment, duration of colic signs prior to presentation, physical examination findings (heart rate, respiratory rate, rectal temperature), and time between admission and surgery. Mucous membrane color and capillary refill time (CRT) were recorded. Mucous membrane appearance was considered abnormal if CRT >3 seconds or if color was described as hyperemic, muddy, or petechiated. Time of admission was recorded; admission between 5:00 p.m. and 8:00 a.m., or on a weekend or holiday was considered outside of normal working hours.

Preoperative hematological values included packed cell volume (PCV) and total solids (TS). Venous blood gas analysis was performed with a point of care blood analyzer (pHOx Ultra, Nova Biomedical, Waltham, Massachusetts) and included pH, pCO<sub>2</sub>, pO<sub>2</sub>, oxygen saturation, hematocrit, hemoglobin, bicarbonate, base excess, oxygen content, sodium, potassium, chloride, ionized calcium, ionized magnesium, glucose, lactate, blood urea nitrogen (BUN), creatinine, total CO<sub>2</sub>, anion gap, and osmolality. Results of abdominocentesis were recorded when available and scored as normal appearance, serosanguinous, turbid, or no fluid obtained. Analyses performed on abdominal fluid included lactate (Lactate Scout, EKF Diagnostics, Cardiff, Wales, UK) and TS.

Intraoperative data retrieved from the surgery reports and anesthesia records included duration of general anesthesia and surgery, surgical diagnosis, and location of the lesion. Hematological variables from the first intraoperative blood sample were recorded including arterial blood gas values as for venous blood gas (pHOx Ultra, Nova Biomedical, Waltham, Massachusetts), PCV and TS. Arterial blood was collected from an indwelling catheter (facial or transverse facial artery) and maintained in a closed system until analysis. Surgery performed between 5:00 p.m. and 8:00 a.m., or on a weekend or holiday was considered outside of normal working hours.

Postoperative colic was defined as an episode of abdominal pain requiring sedation or other therapeutic intervention (e.g., nasogastric intubation or additional analgesic medication) after the horse had recovered and been returned to its stall. Presence of postoperative colic (yes/no), treatment, diagnosis, and relaparotomy were recorded as appropriate. Short-term survival was defined as survival to hospital discharge, and horses were classified as survivors or nonsurvivors (euthanized or died). The number of days of hospitalization following surgery was recorded. For nonsurvivors, the reason for euthanasia was recorded and categorized as: recommended due to grave prognosis, postoperative complications, or natural death.

## 2.2 | Predictive models

PubMed was searched for studies containing predictive models for horses with colic (using the following terms: “equine” “colic” and “prediction OR prognostic”) published between 1986 and 2020. This search yielded 137 publications, 100 of which had full text available for review. Models were included if the study population was clearly described and limited to referral practice, and if the variables and models were described in sufficient detail to successfully reproduce the model. Eight single variable

cutoffs<sup>17,27–31</sup> and four multivariable models<sup>23–25,32</sup> met these criteria (Table 1). Models were applied to the study cohort to determine predicted outcome (survival or nonsurvival to hospital discharge). Individuals with missing data relevant to each model were excluded from model analysis. Two multivariable models included measurements that were not available in the retrospective medical records. First, systolic arterial pressure is assessed in determination of the Shock Score but was not measured in our cohort before induction of anesthesia.<sup>25</sup> However, because the Shock Score is determined by the highest score out of six parameters, the predicted outcome in the study cohort was calculated on the basis of the five available parameters.<sup>25</sup> Second, the linear model by Thoenes et al. includes a categorical pain coefficient (none/mild, moderate, and severe),<sup>32</sup> which was not specifically recorded in our cohort. To calculate predicted outcome for this model in the study cohort, heart rate was used to approximate pain (moderate if HR < 80 bpm, severe if HR ≥ 80 bpm; it was assumed that colic surgery would not be performed in horses displaying no or mild pain).

## 2.3 | Statistical analysis

All analyses were conducted in R version 4.1.2<sup>33</sup> within the RStudio environment, version 1.3.1093<sup>34</sup> with two-sided tests of hypotheses;  $p < .05$  was the criterion for statistical significance. For categorical variables, descriptive statistics were reported as frequency count (percentage of total). For continuous variables, normality was assessed by Shapiro–Wilk test and visual evaluation of histogram and QQ plots. Descriptive statistics for continuous variables were reported as mean and standard deviation if normally distributed, and median and interquartile range (IQR; lower quartile, upper quartile) if they were not normally distributed.

Diagnostic test performance indices were calculated for all cases in the study cohort with each single-variable and multivariable model using the confusionMatrix function of R package caret, version 6.0.90.<sup>35</sup> These included sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy, calculated by comparing the predicted to actual outcome with survival considered the “positive” result and nonsurvival the “negative”. These performance indices were compared between single and multivariable models, as well as between models developed from a mixed population (medical and surgical) or surgical population, by Student's t-test using the R package stats, version 4.1.2.<sup>33</sup> All plots were generated using the R package ggpubr, version 0.4.0.<sup>36</sup>

**TABLE 1** Summary of diagnostic criteria and thresholds for single variable and multivariable predictive models as applied to the study cohort

Model	Criteria considered for study cohort	Performance and population in original report	
Single variable models			
Johnston et al. <sup>17</sup>	Peripheral lactate at admission <6 mmol/l predicted to survive	Se 84%, Sp 83% (large colon volvulus)	Large colon volvulus
Radcliffe et al. <sup>27</sup>	Peripheral lactate at admission <8 mmol/l predicted to survive	No predictive value reported	Surgical colic
Orsini et al. <sup>28</sup>	PCV at admission ≤43% predicted to survive	Serum lactate and PCV combined predictive value of 94%	Medical and surgical colic
Orsini et al. <sup>28</sup>	PCV at admission ≤50% predicted to survive		
Puotunen-Reinert et al. <sup>29</sup>	Heart rate at admission ≤60 bpm predicted to survive	No predictive values reported	Medical and surgical colic
Puotunen-Reinert et al. <sup>29</sup>	Heart rate at admission ≤80 bpm predicted to survive	No predictive values reported	Medical and surgical colic
Delesalle et al. <sup>30</sup>	Peritoneal lactate at admission ≤6 mmol/l predicted to survive	Probability of death if lactate >6 is 29% without strangulating obstruction, 52% with strangulating obstruction	Medical and surgical colic
McCoy et al. <sup>31</sup>	Intraoperative lactate <5 mmol/l predicted to survive	2.5x higher relative risk of complications if lactate >5 mmol/l (arterial lactate in recovery)	Surgical colic
Multivariable Models			
Grulke et al. <sup>25</sup>	Shock Score: Horses with shock score of 3 (meeting one or more of the following criteria) predicted to die: heart rate ≥ 80 bpm; respiratory rate ≥ 35 bpm; PCV ≥55%; blood lactate ≥8.3 mmol/dl; blood urea nitrogen ≥55 mEq/dl	Odds ratio of death 10.8 for shock score 3 vs. shock score 1	Surgical colic
Thoefner et al. <sup>32</sup>	Linear model including pain coefficient, PCV, capillary refill time, and rectal temperature. Cutoff for survival set at probability of death ≤0.86	PPV and NPV 87%	Medical and surgical colic
Reeves et al. <sup>23</sup>	Linear model including age, sex, capillary refill time, PCV and heart rate. Cutoff for survival set at probability of death ≤0.6	Se 70, Sp 89, PPV 80, NPV 82	Medical and surgical colic
Furr et al. <sup>24</sup>	Colic severity score: Summed scores based on four criteria: heart rate, peritoneal total protein, blood lactate, and mucous membrane appearance. Horses with total score <7 are predicted to survive	Development: Se 52.9, Sp 90.1, PPV 58.1, NPV 88.1 Validation: Se 66.7, Sp 100, PPV 100, NPV 91.8	Medical and surgical colic

Abbreviations: NPV, negative predictive value; PPV, positive predictive value; Se, sensitivity; Sp, specificity.

### 3 | RESULTS

#### 3.1 | Study population

A total of 260 horses (age range: 6 months to 32 years, mean: 10.8 ± 6.9 years) met the inclusion criteria. The cohort included 103 (39.6%) mares, 34 (13.1%) stallions,

and 123 (47.3%) geldings. Breed distribution was representative of the patient population seen for emergency care at the University of Illinois, with 35 unique breeds (Table S1) categorized into eleven groups: stock horses (104 [40%]), gaited horses (23 [8.9%]), Thoroughbred (22 [8.5%]), Standardbred (20 [7.8%]), Arabian (20 [7.7%]), warmblood (20 [7.7%]), draft (14 [5.4%]), mixed breed

**TABLE 2** Summary of admission data as mean  $\pm$  SD or median (quartile 1, quartile 3) with number of observations (N) for each variable, grouped by survivors ( $N = 156$ ) and nonsurvivors ( $N = 104$ )

Variable	N	Survivors	N	Nonsurvivors
Rectal temperature ( $^{\circ}$ F)	124	99.5 $\pm$ 1.4	77	99.5 $\pm$ 2.1
Heart rate (bpm)	148	57 $\pm$ 17.3	99	72 $\pm$ 20.5
Respiratory rate (bpm)	128	24 (19.5, 32)	89	28 (20, 40)
Peritoneal lactate (mmol/l)	45	3.2 (2.1, 7.1)	41	9.9 (6.4, 16)
Peritoneal total solids (g/dl)	51	2.6 $\pm$ 1.2	36	3.6 $\pm$ 1.6

(13 [5%]), Morgan (8 [3.1%]), pony (8 [3.1%]), and miniature horse (8 [3.1%]). Mean bodyweight was 1063  $\pm$  263 pounds. Admissions occurred outside of normal working hours for 159/260 (61.2%) cases.

### 3.2 | Admission data

Quantitative admission data are summarized in Table 2. Mucous membrane appearance was abnormal in 42/260 (16.2%) horses. Abdominocentesis was performed in 167/260 (64.2%) horses, and fluid was obtained in 151 horses. Peritoneal fluid was described in all 151 horses when it was obtained: 86 (57%) serosanguinous, 46 (30.5%) normal, 16 (10.6%) turbid or cloudy, two (1.3%) bloody, and one enterocentesis (0.7%). Venous blood gas results are summarized in Table S2.

### 3.3 | Surgical and anesthetic data

Median anesthesia time was 148 min (IQR 110, 195) and median surgical time was 105 min (IQR 70, 155). Surgical diagnoses are summarized in Table 3. No definitive diagnosis was recorded in the medical records of one horse which was euthanized for intractable colic 10 days postoperatively.

### 3.4 | Postoperative colic

Sixty-five of the 188 (34.5%) horses that recovered from anesthesia exhibited colic signs postoperatively. Of these, 33/65 (50.8%) responded to medical treatment, 16/65 (24.6%) underwent repeat exploratory laparotomy, and 16/65 (24.6%) were euthanized without a second exploratory. Ten of the 16 horses that underwent repeat laparotomy (62.5%) were euthanized; diagnoses included ileus ( $n = 10$ ), ischemia–reperfusion injury ( $n = 8$ ), anastomotic stricture ( $n = 2$ ), severe adhesions ( $n = 2$ ), and rupture of the right dorsal colon. The remaining 6/16 (37.5%) horses survived to hospital discharge. Surgical diagnoses in horses that survived after repeat laparotomy were anastomosis site complications ( $n = 3$ ), gas

distension, ileus, and devitalization/reperfusion injury. One of the six horses had a third laparotomy during the same hospitalization at which time adhesions were identified as the cause of colic signs. That individual ultimately survived to discharge.

### 3.5 | Outcomes

Duration of follow up was limited to the time of discharge from the hospital (median 8 days [IQR 6,10]). Of the 260 horses undergoing general anesthesia, 188 (72.3%) survived to anesthetic recovery. Of those, 156 (83%) survived to hospital discharge. Overall short-term survival rate was 60% (156/260).

Of 104 nonsurviving horses, 72 horses were euthanized intraoperatively based upon recommendation of the attending veterinarian due to a grave prognosis. Thirty-two horses were euthanized due to complications that developed after surgery that were unresponsive to medical and/or surgical therapy. Reasons for euthanasia are summarized in Table 4. Lesions were categorized as unresectable due to the length involved in 24 horses, and due to the location of affected viscus in 20 horses. Irreparable lesions included: extensive mesenteric rent, irreducible ileocecal intussusception, irreducible jejunal intussusception, and a large mass adhered to multiple organs. Two horses were euthanized in the recovery box as a result of inability to stand; one due to hypovolemic shock and secondary renal failure, and one in which a cause was not identified.

### 3.6 | Performance of predictive models

A summary of the evaluated models, the populations from which they were generated and their performance (as reported in the original publication) is found Table 1. The Shock Score<sup>25</sup> and the linear model by Thoenner et al.<sup>32</sup> included measurements that were unavailable in the retrospective medical records. Predicted outcome for these two models were calculated with minor modifications as described in the methods.

All the evaluated single variable cutoffs and multivariable models performed similarly for prediction of

**TABLE 3** Intraoperative diagnoses, organized by primary intestinal segment affected. Percentage column refers to total cases (260) for intestinal segment, and number of cases within the intestinal segment for specific diagnosis (i.e., pedunculated lipoma, 34 of 112 small intestine cases)

Diagnosis	Number	Percent
Small intestine	<b>112</b>	<b>43.1%</b>
Pedunculated lipoma	34	30.4%
Anterior enteritis	17	15.2%
Mesenteric rent	11	9.8%
Epiploic foramen entrapment	10	8.9%
Volvulus	7	6.3%
Small intestinal strangulation of unidentified cause	6	5.4%
Adhesions	6	5.4%
Omental strangulation	5	4.5%
Gastrosplenic ligament entrapment	3	2.7%
Intussusception	3	2.7%
<i>Two jejunojejunal, 1 ileocecal</i>		
Impaction	3	2.7%
Diaphragmatic hernia	2	1.8%
Indirect scrotal hernia	2	1.8%
Infarction	2	1.8%
Pyloric thickening	1	0.9%
Large colon	<b>104</b>	<b>40%</b>
Volvulus	40	38.5%
Displacement	30	28.8%
<i>Seventeen right dorsal, 10 left dorsal, 3 not specified</i>		
<i>Twelve of 30 had concurrent impaction</i>		
Impaction	25	24%
Enterolith	6	5.8%
Colitis	3	2.9%
Small colon	<b>31</b>	<b>11.9%</b>
Fecalith/enterolith	17	54.8%
Impaction	11	35.5%
Pedunculated lipoma	1	3.2%
Nephrosplenic entrapment	1	3.2%
Volvulus	1	3.2%
Cecum	<b>4</b>	<b>1.5%</b>
Impaction	2	50%
Displacement	1	25%
Incarceration in diaphragmatic hernia	1	25%
Stomach	<b>4</b>	<b>1.5%</b>
Gastric rupture	4	100%
Extra-gastrointestinal	<b>4</b>	<b>1.5%</b>

**TABLE 3** (Continued)

Diagnosis	Number	Percent
Suspected renal failure and disseminated intravascular coagulation	1	
Enlargement of kidney and multiple lymph nodes	1	
Hemoabdomen secondary to unidentified mass	1	
Bile duct rupture secondary to cholelithiasis	1	

**TABLE 4** Case outcomes and reasons for euthanasia

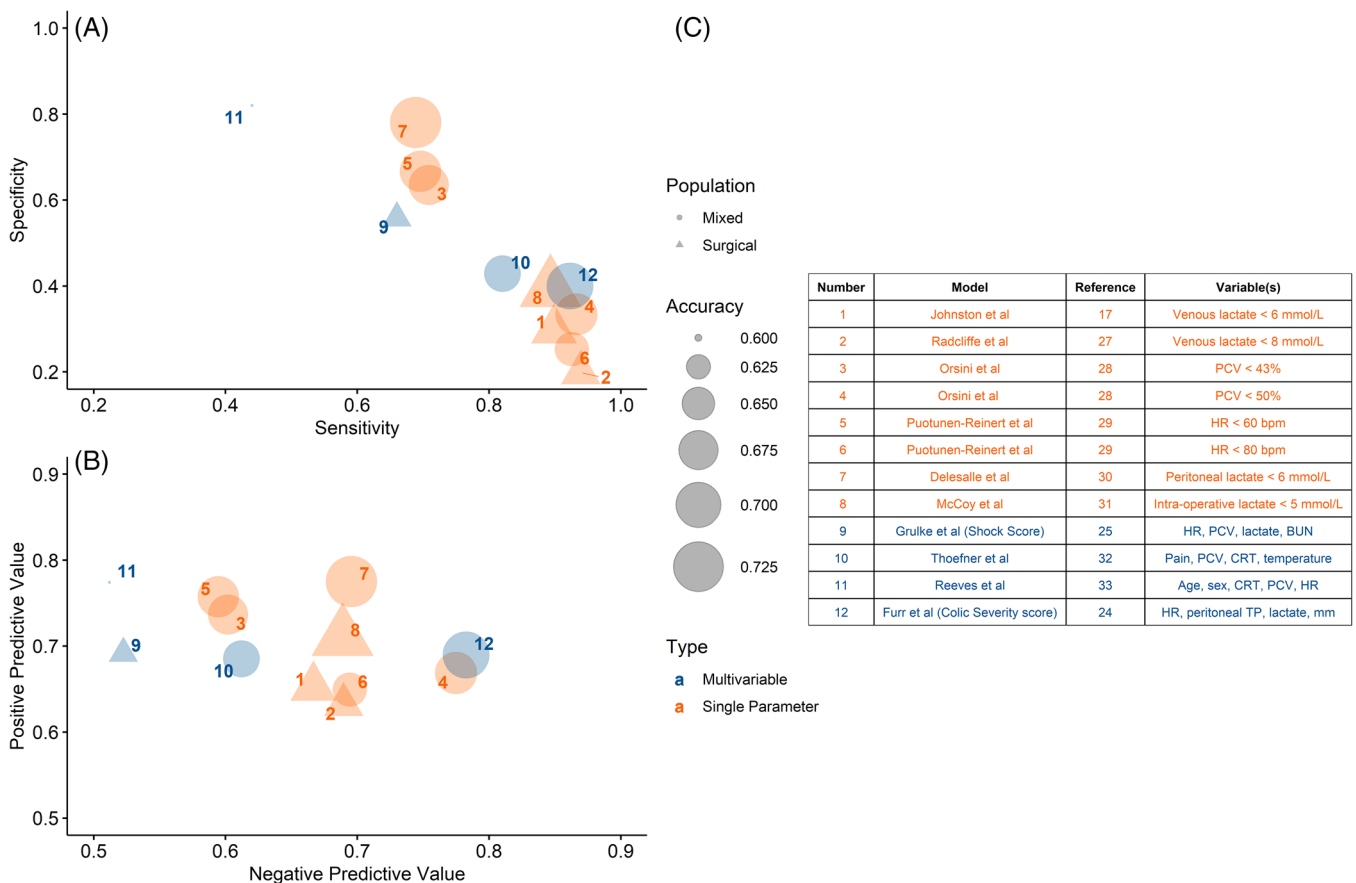
Outcome	Number	Percent
Survival to discharge	<b>156</b>	<b>60%</b>
Intraoperative euthanasia	<b>72</b>	<b>27.7%</b>
Unresectable lesion	44	61.1%
Rupture of intestinal viscus	18	25%
Irreparable lesion	4	5.6%
Excessive adhesions	3	4.2%
Peritonitis	3	4.2%
Postoperative euthanasia	<b>32</b>	<b>12.3%</b>
Postoperative colic	23	71.9%
Endotoxemia and deterioration in face of treatment	3	9.4%
Inability to stand in recovery	2	6.3%
Laminitis	2	6.3%
Severe intra-abdominal hemorrhage	1	3.1%
Disseminated intravascular coagulation	1	3.1%

outcome in the study cohort, with a mean 79% sensitivity (range 44%–94%), 48% specificity (20%–82%), 70% PPV (63%–78%), 65% NPV (51%–78%), and 67% accuracy (60%–73%) (Table 5). There was no difference between single variable cutoffs and multivariable models (all  $p > .3$ ), or between models derived from a mixed medical and surgical population versus those derived from a surgical population, (all  $p > .1$ ) for any measures of diagnostic performance in the study cohort (Figure 1, Table 6). Single variable cutoffs of blood lactate (6 or 8 mmol/l), PCV (50%), and heart rate (80 bpm) were highly sensitive for prediction of survival, but poorly specific, with better NPV than PPV. The most accurate single variable predictor was abdominal lactate (cutoff of lactate  $\leq 6$  mmol/l predictive of survival), at 73% accuracy. The CSS<sup>24</sup> was the most sensitive of the multivariable predictive models

**TABLE 5** Diagnostic performance of single variable and multivariable predictive models for predicting outcome (survival to hospital discharge) in the study cohort. Unless otherwise specified, variables were assessed at hospital admission

Model	Variable (S)	Sensitivity	Specificity	PPV	NPV	Accuracy	N
Single variable models							
Johnston et al. <sup>17</sup>	Venous lactate <6 mmol/l	90%	30%	65%	67%	65%	249
Radcliffe et al. <sup>27</sup>	Venous lactate <8 mmol/l	94%	20%	63%	69%	64%	249
Orsini et al. <sup>28</sup>	PCV < 43%	71%	63%	74%	60%	68%	227
Orsini et al. <sup>28</sup>	PCV < 50%	93%	33%	67%	78%	69%	227
Puotunen-Reinert et al. <sup>29</sup>	HR < 60 bpm	70%	67%	76%	59%	68%	247
Puotunen-Reinert et al. <sup>29</sup>	HR < 80 bpm	93%	25%	65%	69%	66%	247
Delesalle et al. <sup>30</sup>	Peritoneal lactate <6 mmol/l	69%	78%	78%	70%	73%	86
McCoy et al. <sup>31</sup>	Intraoperative lactate <5 mmol/l	89%	39%	71%	69%	70%	210
Multivariable models							
Grulke et al. <sup>25</sup> (Shock Score)	HR, PCV, lactate, BUN	66%	56%	69%	52%	62%	260
Thoefner et al. <sup>32</sup>	Pain, PCV, CRT, temperature	82%	43%	69%	61%	66%	176
Reeves et al. <sup>23</sup>	Age, sex, CRT, PCV, HR	44%	82%	77%	51%	60%	187
Furr et al. <sup>24</sup> (colic severity score)	HR, peritoneal TP, lactate, mm	92%	40%	69%	78%	71%	110

Abbreviations: BUN, blood urea nitrogen; CRT, capillary refill time; HR, heart rate; mm, mucous membrane appearance; N, number of cases included; NPV, negative predictive value; PCV, packed cell volume; PPV, positive predictive value.

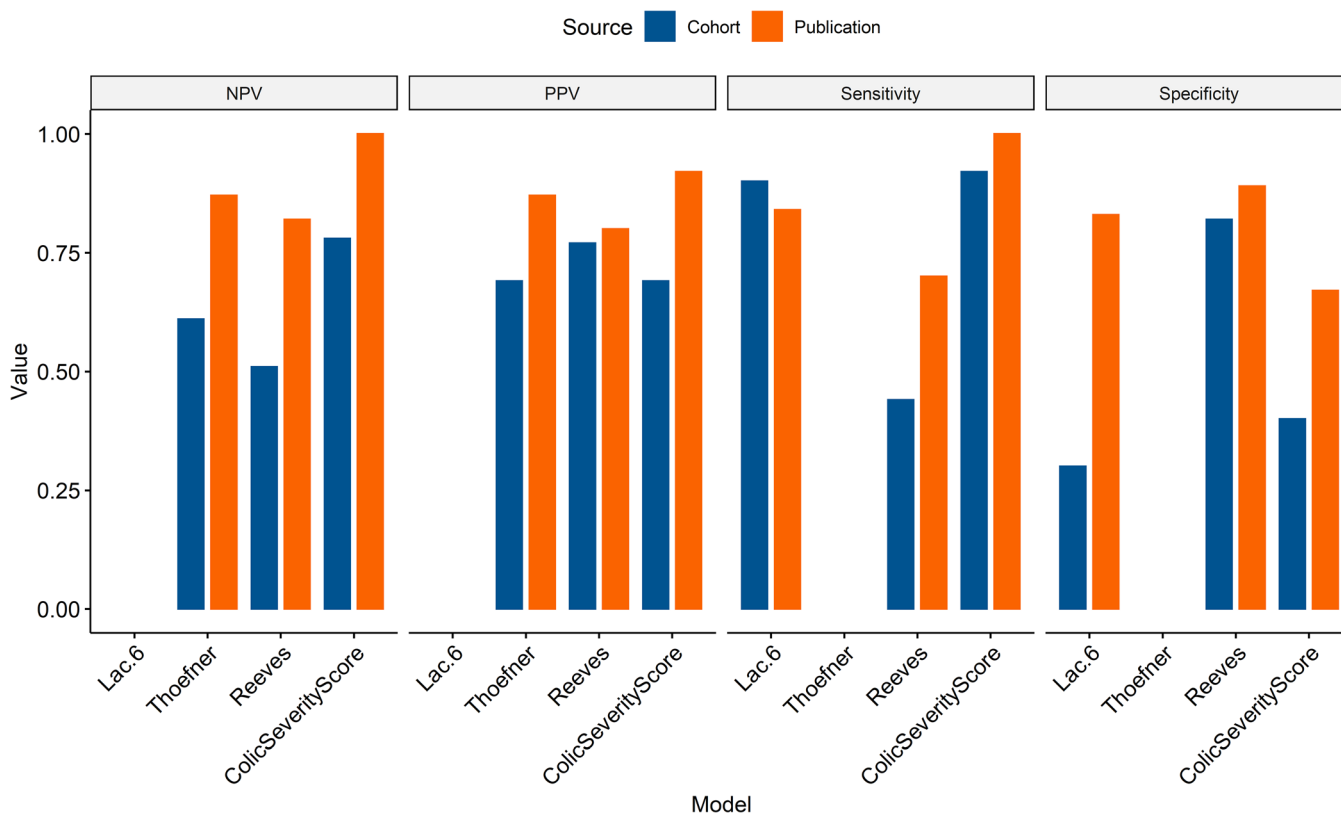


**FIGURE 1** Diagnostic performance of all tested models plotted by sensitivity and specificity (A) and positive predictive value (PPV)/negative predictive value (NPV) (B). Dot size reflects model accuracy, color is model type (single variable in orange, multivariable in blue), and shape is original study population (circle is mixed surgical and medical colic, triangle is surgical colic)

**TABLE 6** Diagnostic performance of predictive models grouped by model type (single variable vs. multivariable) and original study population (mixed surgical and medical colic vs. surgical colic only). Presented as mean  $\pm$  standard deviation (SD); *p*-value represents result of Student's *t*-test between the two groups

	Sensitivity		Specificity		PPV		NPV		Accuracy	
	Mean $\pm$ SD	<i>p</i> -value	Mean $\pm$ SD	<i>p</i> -value	Mean $\pm$ SD	<i>p</i> -value	Mean $\pm$ SD	<i>p</i> -value	Mean $\pm$ SD	<i>p</i> -value
Model type										
Single variable	84 $\pm$ 12%	0.333	44 $\pm$ 22%	0.413	70 $\pm$ 5%	0.674	68 $\pm$ 6%	0.362	68 $\pm$ 3%	0.302
Multivariable	71 $\pm$ 21%		55 $\pm$ 19%		71 $\pm$ 4%		61 $\pm$ 13%		65 $\pm$ 5%	
Original study population										
Mixed	77 $\pm$ 17%	0.382	54 $\pm$ 21%	0.134	72 $\pm$ 5%	0.102	66 $\pm$ 9%	0.759	68 $\pm$ 4%	0.372
Surgical	85 $\pm$ 13%		36 $\pm$ 15%		67 $\pm$ 4%		64 $\pm$ 8%		65 $\pm$ 4%	

Abbreviations: NPV, negative predictive value; PPV, positive predictive value; SD, standard deviation.



**FIGURE 2** Diagnostic performance of tested models in the study population (cohort, blue) compared to reported values from original publication when available (publication, orange). Blank space indicates that the parameter was not included in the original publication. PPV, positive predictive value; NPV, negative predictive value

(sensitivity 92%, NPV 78%), but had poor specificity (40%).

Measures of diagnostic performance were only found in the original reports for four models. Lactate cutoff of 6 mmol/l had higher sensitivity in our study cohort (90%) compared to the original report (84%), but much lower

specificity (30% vs. 83%).<sup>17</sup> This model was originally described in horses specifically treated for large colon volvulus. When applied to the cases of the study cohort with a diagnosis of large colon volvulus ( $n = 40$ ), the model still did not perform well (sensitivity 75%, specificity 46%, PPV 48%, NPV 73%, accuracy 60%). For multivariable models, all



measures of diagnostic performance were lower in the study cohort than reported in the original publications (Figure 2).

## 4 | DISCUSSION

All tested models underperformed the original reports when used to predict outcome in this independent study cohort, supporting the hypothesis that the models would not perform as well when applied to a different population. Interestingly, diagnostic performance for the previously reported models was similar between single variable cutoffs and multivariable models when applied to the study cohort, therefore we reject the hypothesis that multivariable models would be more accurate. These findings suggest that while general associations between clinical variables and outcome may be broadly applicable across populations, population- and facility-specific factors still play an important role in predicting survival in horses undergoing colic surgery.

The great variety of lesions underlying in signs of colic and the associated pathological processes cause a wide spectrum of patient morbidity, presenting an inherent challenge in prediction of survival for this heterogeneous horse population. A primary goal of this analysis was to determine the prognostic accuracy for survival using information available prior to placing a horse under general anesthesia and making a definitive intraoperative diagnosis. Additional intraoperative variables, such as performance of a resection and anastomosis, have been associated with survival to discharge in other selected populations.<sup>21</sup> We elected to exclude these variables because of their bias towards the portion of the cohort that recovered from anesthesia, while we were interested in prognostication for the entire group.

Performance indices for the models were calculated considering survival as the positive outcome. In this context, “sensitivity” is the proportion of horses correctly identified as survivors, and “specificity” is the proportion of horses correctly identified as nonsurvivors. Most of the single variable cutoff values assessed were moderately to highly sensitive and poorly specific. Clinically, this means that single value cutoffs are best at predicting nonsurvival when the horse has a high value (that is, above the cutoff). A PCV threshold of 50%<sup>28</sup> and the CSS (a summed score based on heart rate, peritoneal total protein, blood lactate, and mucous membrane appearance)<sup>24</sup> had the highest NPV of all models considered, indicating they were best for predicting nonsurvival in the study population. The multivariable model by Reeves et al. (a linear model incorporating age, sex, capillary refill time, PCV and heart rate)<sup>23</sup> and peritoneal lactate cutoff of 6 mmol/l ( $\leq 6$  mmol/l predictive of survival) had the highest PPV.

Overall accuracy was similar for single variable and multivariable models in the study cohort. Peritoneal lactate cutoff of 6 mmol/l,<sup>30</sup> was the most accurate predictor of outcome at 73%. This means that, at best, nearly a third of patients would be incorrectly predicted to either survive or not survive according to the models.

The roles of pre-emptive euthanasia and economic factors in colic survival present an additional complication when evaluating outcomes. It can be argued that all horses euthanized intraoperatively should be excluded from analysis, as those horses were not given the opportunity to survive, and a subset may have survived if further treatment was pursued. However, this approach does not account for cases in which the lesion identified at surgery was incompatible with life. These inclusion criteria are important to consider in evaluation of predictive models, as the predictive value and accuracy of a diagnostic test are affected by disease prevalence, in addition to the test's sensitivity and specificity. The overall survival rate in the study population was lower than some contemporary reports which focus only on horses that recovered from general anesthesia.<sup>21,37,38</sup> However, it is critical to note that the study cohort also includes horses that were euthanized intraoperatively due to the nature of their lesion, and survival rate is comparable to other reports where all surgical cases are included.<sup>37,39</sup> This highlights that inclusion criteria of a study can alter the consequent survival rate and offer a more (or less) optimistic outlook than is appropriate for the broader surgical population.

The cutoff value for a predictive model can be selected to optimize the sensitivity or specificity of the model. The cutoffs for mathematical predictive models were selected to maximize the sum of specificity and sensitivity.<sup>23,32</sup> However, the severity of false negative and false positive errors is not equivalent in the case of life-or-death clinical decisions. If a false negative error (predicting a horse will die that could otherwise live) is considered the more grievous error (because the owner would be prompted to euthanize the horse), it may be more desirable to select a cutoff value that maximizes sensitivity, while accepting lower specificity. While peritoneal lactate cutoff of 6 mmol/l<sup>30</sup> had the greatest overall accuracy for prediction of outcome, the CSS,<sup>24</sup> venous lactate cutoff of 8 mmol/l,<sup>27</sup> and admission heart rate cutoff of 80 bpm<sup>29</sup> may be the most applicable for a clinical setting due to their high sensitivity, meaning that horses with values below the cutoff are correctly identified as survivors.

Of the models tested, measures of diagnostic accuracy were reported in the original publications for three multivariable models and one single variable predictor (Table 1). The models performed better in the original

populations than in the current study population, apart from venous lactate cutoff of 6 mmol/l, which had a sensitivity of 90% in the current study, compared to 84% originally reported (Figure 2).<sup>17</sup> This is unsurprising, as it is generally expected that a model will perform best in the population from which it is developed. Validation of a predictive model in multiple populations is recommended before making assumptions about its general applicability for use in clinical decision-making.<sup>2</sup> Without such external validation, clinicians should exercise caution when applying predictive models to clinical cases.

Three of the multivariable models tested here were originally developed in general colic populations.<sup>23,24,32</sup> As our study population included only colic patients requiring surgical intervention, this probably explains why these multivariable models did not perform as well in our study cohort when compared to the respective original populations. Further, some single variable cutoffs were derived from studies focused on specific lesions (i.e., large colon volvulus)<sup>17</sup> or based on specific time points (intraoperative and recovery/postoperative arterial lactate)<sup>31</sup> and may not be applicable beyond those specific situations. Notably, when applied to horses in the study cohort with large colon volvulus, the model developed specifically for that lesion still did not perform as well as in the original report. Further, there was no statistically significant difference identified when measures of diagnostic performance were compared between the models from a mixed colic population and those from surgical colic cases. This suggests that even the models developed from a more similar patient population did not perform well when applied to a different cohort.

Models originally designed to predict nonsurvival, rather than survival, such as those reported by Reeves et al.<sup>23</sup> and Thoenner et al.,<sup>32</sup> could have had improved performance if a different cutoff value was used for survival. However, clinical utility of these two specific models is limited by the required computations. In contrast, the CSS,<sup>24</sup> which is the most accurate multivariable model tested, is easily applied using information rapidly acquired from a routine colic evaluation.

#### 4.1 | Limitations

Inherent limitations of a retrospective study include the quality of available medical record data, lack of randomly assigned treatment groups, and clinician-dependent differences in case management. The effect of clinician bias is impossible to eliminate, as case management and euthanasia recommendations are affected by surgeon assessment, experience, and judgment. Duration of follow-up in this study was limited to the time of

discharge from hospital. Inclusion of longer-term follow up may have captured additional post-operative complications and mortality, such as incisional infections and recurrent colic. Prediction of long-term survival is arguably of more interest to the horse owner; however, this was beyond the scope of the current study, which focused on short-term outcomes. Horses that were euthanized rather than undergo exploratory laparotomy or were euthanized intraoperatively due to financial considerations were excluded. Results of this study reflect cases from one institution, and (as demonstrated by this study) the performance of these models may differ when applied to other populations.

#### 4.2 | Conclusions

Single-variable and multivariable predictive models did not perform as well for prediction of outcome in the study cohort compared to their original reports. These findings suggest that while general associations between clinical variables and outcome may be broadly applicable across populations, population-specific factors still play an important role in survival. Not surprisingly, patient outcomes and performance of prognostic models in this cohort consisting only of surgical cases appears to be less favorable than previous reports that include both of medical and surgical colic cases. The best-performing single variable cutoffs and multivariable models were highly sensitive for prediction of survival, but poorly specific.

#### 4.3 | Clinical significance

While no single variable was sufficient to accurately predict the outcome of a horse requiring exploratory laparotomy for colic, our findings support the continued use of parameters commonly considered in clinical assessment of patient status, such as admission heart rate, packed cell volume, and blood lactate. Predictive models of survival developed in one population are highly likely to be less reliable when used to predict outcome in horses undergoing colic surgery in an independent population. Additional model testing and refinement using data from multiple surgical centers could be considered to improve prediction of outcome for horses undergoing laparotomy for treatment of colic.

#### AUTHOR CONTRIBUTIONS

R.C. Bishop: Study initiation and design, data collection, data analysis and interpretation, drafting of the manuscript, approved the final version of the manuscript. S.D. Gutierrez-

Nibeyro: Study design, manuscript revision, approved the final version of the manuscript. M.C. Stewart: Study design, manuscript revision, approved the final version of the manuscript. A.M. McCoy: Study design, data analysis and interpretation, manuscript revision, approved the final version of the manuscript.

## ACKNOWLEDGMENTS

The authors would like to thank Dr. Alex Lipka for assistance with statistical analysis.

## FUNDING INFORMATION

There was no funding specific to this work.

## CONFLICT OF INTEREST

There was no funding specific to this work. The authors declare no conflict of interest related to this report.

## ORCID

Rebecca C. Bishop  <https://orcid.org/0000-0002-9660-732X>

Annette M. McCoy  <https://orcid.org/0000-0003-4088-6902>

## REFERENCES

- Freeman DE. Fifty years of colic surgery. *Equine Vet J.* 2018;50:423-435.
- Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ.* 2009;338:b605.
- Proudman CJ, Edwards GB, Barnes J, French NP. Modelling long-term survival of horses following surgery for large intestinal disease. *Equine Vet J.* 2005;37:366-370.
- Proudman CJ, Dugdale AH, Senior JM, et al. Pre-operative and anaesthesia-related risk factors for mortality in equine colic cases. *Vet J.* 2006;171:89-97.
- Mair TS, Smith LJ. Survival and complication rates in 300 horses undergoing surgical treatment of colic. Part 1: short-term survival following a single laparotomy. *Equine Vet J.* 2005;37:296-302.
- Garcia-Seco E, Wilson DA, Kramer J, et al. Prevalence and risk factors associated with outcome of surgical removal of pedunculated lipomas in horses: 102 cases (1987-2002). *J Am Vet Med Assoc.* 2005;226:1529-1537.
- van der Linden MA, Laffont CM. Sloet van Oldruitenborgh-Oosterbaan MM: prognosis in equine medical and surgical colic. *J Vet Intern Med.* 2003;17:343-348.
- Hackett ES, Embertson RM, Hopper SA, Woodie JB, Ruggles AJ. Duration of disease influences survival to discharge of thoroughbred mares with surgically treated large colon volvulus. *Equine Vet J.* 2015;47:650-654.
- Hinchcliff KW, Rush BR, Farris JW. Evaluation of plasma catecholamine and serum cortisol concentrations in horses with colic. *J Am Vet Med Assoc.* 2005;227:276-280.
- Bristol DG. The anion gap as a prognostic indicator in horses with abdominal pain. - abstract - Europe PMC. *J Am Vet Med Assoc.* 1982;181:63-65.
- Garcia-Lopez JM, Provost PJ, Rush JE, Zicker SC, Burmaster H, Freeman LM. Prevalence and prognostic importance of hypomagnesemia and hypocalcemia in horses that have colic surgery. *Am J Vet Res.* 2001;62:7-12.
- Johansson AM, Gardner SY, Jones SL, Fuquay LR, Reagan VH, Levine JF. Hypomagnesemia in hospitalized horses. *J Vet Intern Med.* 2003;17:860-867.
- Delesalle C, Dewulf J, Lefebvre RA, Schuurkes JAJ, van Vlierbergen B, Deprez P. Use of plasma ionized calcium levels and Ca<sup>2+</sup>-substitution response patterns as prognostic parameters for ileus and survival in colic horses. *Vet Q.* 2005;27:157-172.
- Groover ES, Woolums AR, Cole DJ, LeRoy BE. Risk factors associated with renal insufficiency in horses with primary gastrointestinal disease: 26 cases (2000-2003). *J Am Vet Med Assoc.* 2006;228:572-577.
- Hollis AR, Boston RC, Corley KT. Blood glucose in horses with acute abdominal disease. *J Vet Intern Med.* 2007;21:1099-1103.
- Hassel DM, Hill AE, Rorabeck RA. Association between hyperglycemia and survival in 228 horses with acute gastrointestinal disease. *J Vet Intern Med.* 2009;23:1261-1265.
- Johnston K, Holcombe SJ, Hauptman JG. Plasma lactate as a predictor of colonic viability and survival after 360 degrees volvulus of the ascending colon in horses. *Vet Surg.* 2007;36:563-567.
- Bergren AL, Credille BC, Epstein KL, Giguère S. Retrospective comparison of Gastrosplenic entrapment of the small intestine to other strangulating small intestinal lesions in adult horses. *Vet Surg.* 2015;44:535-539.
- Barton MH, Collatos C. Tumor necrosis factor and interleukin-6 activity and endotoxin concentration in peritoneal fluid and blood of horses with acute abdominal disease. *J Vet Intern Med.* 1999;13:457-464.
- Sauze MN, Cebra CK, Tornquist SJ. The diagnostic and prognostic value of alkaline phosphatase activity in serum and peritoneal fluid from horses with acute colic. *J Vet Intern Med.* 2004;18:564-567.
- Espinosa P, Le Jeune SS, Cenani A, et al. Investigation of perioperative and anesthetic variables affecting short-term survival of horses with small intestinal strangulating lesions. *Vet Surg.* 2017;46:345-353.
- Pascoe PJ, Ducharme NG, Ducharme GR, Lumsden JH. A computer-derived protocol using recursive partitioning to aid in estimating prognosis of horses with abdominal-pain in referral hospitals. *Can J Vet Res.* 1990;54:373-378.
- Reeves MJ, Curtis CR, Salman MD, Hilbert BJ. Prognosis in equine colic patients using multivariable analysis. *Can J Vet Res.* 1989;53:87-94.
- Furr MO, Lessard P, White NA 2nd. Development of a colic severity score for predicting the outcome of equine colic. *Vet Surg.* 1995;24:97-101.
- Grulke S, Olle E, Detilleux J, Gangl M, Caudron I, Sertejn D. Determination of a gravity and shock score for prognosis in equine surgical colic. *J Vet Med A Physiol Pathol Clin Med.* 2001;48:465-473.
- McConachie E, Giguere S, Barton MH. Scoring system for multiple organ dysfunction in adult horses with acute surgical gastrointestinal disease. *J Vet Intern Med.* 2016;30:1276-1283.
- Radcliffe RM, Divers TJ, Fletcher DJ, Mohammed H, Kraus MS. Evaluation of L-lactate and cardiac troponin I in horses undergoing emergency abdominal surgery. *J Vet Emerg Crit Care (San Antonio).* 2012;22:313-319.

28. Orsini JA, Elser AH, Galligan DT, Donawick WJ, Kronfeld DS. Prognostic index for acute abdominal crisis (colic) in horses. *Am J Vet Res.* 1988;49:1969-1971.
29. Puotunen-Reinert A. Study of variables commonly used in examination of equine colic cases to assess prognostic value. *Equine Vet J.* 1986;18:275-277.
30. Delesalle C, Dewulf J, Lefebvre RA, et al. Determination of lactate concentrations in blood plasma and peritoneal fluid in horses with colic by an Accusport analyzer. *J Vet Intern Med.* 2007;21:293-301.
31. McCoy AM, Hackett ES, Wagner AE, et al. Pulmonary gas exchange and plasma lactate in horses with gastrointestinal disease undergoing emergency exploratory laparotomy: a comparison with an elective surgery horse population. *Vet Surg.* 2011;40:601-609.
32. Thoenfer MB, Ersboll AK, Hesselholt M. Prognostic indicators in a Danish hospital-based population of colic horses. *Equine Vet J Suppl.* 2000;32:11-18.
33. R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, 2021 (<https://www.R-project.org/>)
34. RStudio Team: RStudio: Integrated Development for R, RStudio, PBC, 2020 <http://www.rstudio.com/>
35. Kuhn M: caret: Classification and Regression Training, R package version 6.0-90, 2021 <https://CRAN.R-project.org/package=caret>
36. Kassambara A: ggpubr: 'ggplot2' Based Publication ready Plots, R package version 0.4.0, 2020 <https://CRAN.R-project.org/package=ggpubr>
37. Kilcoyne I, Dechant JE, Nieto JE. Comparison of clinical findings and short-term survival between horses with intestinal entrapment in the gastrosplenic ligament and horses with intestinal entrapment in the epiploic foramen. *J Am Vet Med Assoc.* 2016;249:660-667.
38. Whyard JM, Brounts SH. Complications and survival in horses with surgically confirmed right dorsal displacement of the large colon. *Can Vet J.* 2019;60:381-385.
39. van Bergen T, Haspeslagh M, Wiemer P, et al. Surgical treatment of epiploic foramen entrapment in 142 horses (2008-2016). *Vet Surg.* 2019;48:291-298.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Bishop RC, Gutierrez-Nibeyro SD, Stewart MC, McCoy AM. Performance of predictive models of survival in horses undergoing emergency exploratory laparotomy for colic. *Veterinary Surgery.* 2022; 51(6):891-902. doi:10.1111/vsu.13839