Comparative Medicine
Copyright 2023
by the American Association for Laboratory Animal Science

Vol 73, No 6
December 2023
Pages 446–460

Original Research

A Novel Scoring System for Humane Endpoints in Mice with Cecal Ligation and Puncture-Induced Sepsis

Lindsey T Ferguson,[1,*] Ammar A Rashied,[4] Zhe Liang,[2] Tetsuya Yumoto,[2,5] Jerome C Anyalebechi,[2] David A Swift,[2] Marina S Hernandes,[3] Robert T Krafty,[4] Craig M Coopersmith,[2] and Vanessa K Lee[1]

Animal-based research is essential to the study of sepsis pathophysiology, diagnostics, and therapeutics. However, animal models of sepsis are often associated with high mortality because of the difficulty in predicting imminent death based on premortem assessment of the animals. The use of validated visual scoring would allow researchers to systematically identify humane endpoints but visual approaches require high interobserver agreement for accurate results. The objective of this study was to establish a scoring system for mice undergoing cecal ligation and puncture (CLP)-induced sepsis based on 3 visual parameters: respiratory status, activity and response to stimulus (ASR), and eye appearance, with scores ranging from 0 to 3. In the first study, we evaluated interobserver agreement. Veterinary and investigative staff assessed 283 mice with CLP and had substantial to near-perfect agreement for all 3 parameters as evaluated using weighted Cohen κ statistic. The second study assessed the ability of the scoring system and temperature to predict death. The scoring system and subcutaneous transponders were used to monitor C57BL/6J mice (n = 80, male and female) until death or for 7 days after CLP. Results showed that the scoring system discriminates between surviving (n = 26) and nonsurviving (n = 54) septic mice. The scoring system was accurate in predicting death, with an AUC of 0.8997. The sensitivity and specificity of the ASR parameter were 96% and 92%, respectively, and for the eye parameter were 94% and 73%. A sum of the ASR and eye scores that was 5 or more was also predictive of death. Temperature was a quantitative predictor, with sensitivity and specificity of 93% and 92%, respectively. This scoring system refines the CLP model by allowing identification of humane endpoints and avoidance of spontaneous death.

Abbreviations and Acronyms: ASR, activity and response to stimulus; CLP, cecal ligation and puncture; MGS, mouse grimace score; M-CASS, mouse clinical assessment score for sepsis; modified M-CASS, modified mouse clinical assessment score for sepsis; MSS, murine sepsis score; ROC, receiver operating characteristic

DOI: 10.30802/AALAS-CM-22-000124

Introduction

Sepsis is a life-threatening syndrome characterized by organ dysfunction due to a dysregulated host response to infection.[78] Sepsis is an important clinical syndrome in human medicine, with an incidence of 49 million cases globally and a contribution of 20% to all global deaths in 2017.[73] While consensus statements on the clinical care of human patients have improved the diagnosis and treatment of sepsis,[16,17] preclinical research remains necessary to elucidate the pathophysiology of sepsis, evaluate diagnostics, and test novel therapeutics. Mice are the most common animal used for sepsis-related research, with sepsis most commonly induced by the surgical procedure of cecal ligation and puncture (CLP),[45] which is considered by many to be the best available model of polymicrobial sepsis.[12]

Sepsis models in animals generally induce acute severe progressive disease, and evaluation of experimental treatment often relies on mortality rates. Consequently, sepsis research has prompted animal welfare concerns about the use of death as an experimental endpoint.[46,62-65,91] The *Guide for the Care and Use of Laboratory Animals* states that "the use of humane endpoints contributes to refinement by providing an alternative to experimental endpoints that result in unrelieved pain and distress, including death."[31] Humane endpoint criteria that guide euthanasia should be animal model-specific and allow for prompt decision-making by both veterinarians and investigators.[31,64] The establishment of humane endpoints is both a refinement and an alternative,[60,75] which should be reported as recommended by the *ARRIVE Guidelines 2.0* and can contribute to reproducible and rigorous animal research.[70] Specific to sepsis research, proposals have been made to establish consistent monitoring schemes and euthanasia criteria for animals, but no consensus has been reached.[26,45,46,63,65,66,80,91] Ideal surrogate humane endpoints in murine CLP-induced sepsis research would have high predictive capacity for death,

Submitted: 02 Dec 2022. Revision requested: 26 Jan 2023. Accepted: 14 Mar 2023.
[1]Division of Animal Resources, Emory University, Atlanta, Georgia; [2]Department of Surgery and Emory Critical Care Center, Emory University School of Medicine and Emory Healthcare, Atlanta, Georgia; [3]Department of Medicine, Emory University School of Medicine, Atlanta, Georgia; [4]Department of Biostatistics and Bioinformatics, Emory University Rollins School of Public Health, Atlanta, Georgia; and [5]Department of Emergency, Critical Care and Disaster Medicine, Okayama University Graduate School of Medicine, Kita-ku, Okayama, Japan.
**Corresponding author. Email: Lindsey.Ferguson@nationwidechildrens.org*

thus creating endpoints that meet scientific aims and avoid unrelieved animal pain and distress.

Quantitative measurements have the benefit of objectivity as surrogate endpoints for sepsis models, but temperature is currently the only such parameter that has shown promise as a useful tool. Previous studies have measured circulating inflammatory cytokines, reactive metabolites, tissue enzymes, and cardiopulmonary function as predictors of sepsis outcomes, but these measurements are invasive and affect the animals, costs, and personnel time, which makes them impractical for routine use.[28,30,44,72,77,87] Body temperature is less invasive than blood sampling and has been studied as a surrogate endpoint in various murine infectious and inflammatory disease models.[1,19,25,34,61,83,86] The studies specific to sepsis models suggest that temperature monitoring is a promising method for humane endpoint identification; however, the limited studies relied on thermometry methods that are unreliable and induce acute stress from restraint.[23,36,41,42,48,52,55,86] Newer noncontact, infrared (IR) thermometry has promise but requires manual restraint and produces variable measurements of skin surface temperature.[54,85] Subcutaneous temperature microchips are less invasive than rectal thermometry, biotelemetry, and IR but they have not been previously evaluated in a murine model of CLP-induced sepsis.[34] While temperature measurement has advantages, it requires specialized equipment and whether it can be generalized among mice of housed under various conditions is unknown.

An observational scoring system for surrogate endpoints as an adjunct or replacement for temperature monitoring could help overcome the limitations of objective parameters. Observational parameters are powerful tools for cage-side welfare assessments and for identifying humane endpoints by using parameters that are specific to the model, are limited to those found to be most predictive of adverse outcomes and can be generalized across different observers.[7] However, to date, observational methods have not been generally accepted as adequate surrogates for predicting death in the CLP model. The mouse grimace scale (MGS) has not been validated for use in mouse models of sepsis.[37] Although the murine sepsis score (MSS) and the mouse clinical assessment score for sepsis (M-CASS) have been established, neither of these studies used the CLP sepsis model.[30,77] Subsequent studies evaluated versions of these scoring systems in a mouse CLP model, but neither these nor the original versions of these scoring systems were studied as a marker for imminent death, so the true ability to predict mortality is unknown.[48,77,81]

Our goal was to create an observational scoring system that could aid in the identification of early humane endpoints in mice with CLP. We hypothesized that our scoring system would predict death in this model. We also evaluated the sensitivity and specificity of scoring system parameters and total scores to determine specific scores that would be predictive of death. Furthermore, we hypothesized that our scoring system would provide a high level of interobserver agreement. We also hypothesized that subcutaneous temperatures would differ between surviving and nonsurviving septic mice and evaluated the sensitivity and specificity of temperature to establish a temperature threshold for the quantitative prediction of death.

Materials and Methods

Mice. All mice were housed in an AAALAC International accredited facility in compliance with the *Guide for the Care and Use of Laboratory Animals*. All animal procedures were reviewed and approved by Emory University's Institutional Animal Care and Use Committee.

C57BL/6J mice (n = 42 males, n = 42 females, JAX stock number 000664) were obtained from The Jackson Laboratory (Bar Harbor, ME) for the study aim of assessing the ability of the scoring system to predict death. All mice were purchased at 6 wk of age and were used for the study by 8 wk of age. After their arrival at our facility, mice were grouped by sex, housed in individually ventilated cages (IVC; Super Mouse 750 ventilated rack, Lab Products, Seaford, DE), and allowed to acclimate for at least 72 h prior to experimental manipulation. Mice were housed with the same cage mates in groups of 4 to 5 mice per cage until all mice in the cage had been used experimentally.

Mice were housed on 1/8-in. corncob bedding (Bed-o'-Cobs, The Andersons, Maumee, OH) in polysulfone IVC caging (cage bottom number 75031, Lab Products, Seaford, DE), with 2 × 2 in cotton squares (Ancare, Bellmore, NY) for enrichment. Mice were fed irradiated pelleted rodent chow (LabDiet 5053, LabDiet, St. Louis, MO) and received reverse-osmosis–filtered and UV-light–treated water through an automated watering system. For the duration of the experiment, the conditions in the rooms housing study mice were set at 72 °F (22 °C; measured range of 71.6 to 74.5 °F [22 to 24 °C]), 50% humidity (measured range of 42 to 66%), 10 air changes hourly, and on a 12:12 light:dark cycle with lights on at 0700 and off at 1900. Intracage air changes were set as 35 per hour.

The facility colony health surveillance program conducted environmental health monitoring quarterly with sentinel free soiled bedding samples sent to third-party commercial laboratories for PCR analysis. Excluded agents were Sendai Virus, mouse hepatitis virus, mouse minute virus, mouse parvovirus 1 and 2, Theiler Virus (GDVII or TMEV), epizootic diarrhea of infant mice, lymphocytic choriomeningitis virus, polyoma virus, mouse cytomegalovirus, Ectromelia Virus, K Virus, *Mycoplasma pulmonis*, mouse adenovirus 1 and 2, pneumonia virus of mice, Reovirus, Hantaan virus, lactate dehydrogenase elevating virus, mouse thymic virus, *Filobacterium rodentium*, *Encephalitozoon cuniculi*, fur mites (*Myobia*, *Myocoptes*, *and Radfordia* spp.), and pinworms (*Aspiculuris* and *Syphacia* spp.).

The few animal health concerns that occurred over the course of the study happened either before or at the time of surgery. Specifically, 3 mice required euthanasia due to dehydration from drinking valve malfunction and one mouse acutely died during recovery from CLP. No other clinical unexpected abnormalities were detected during the study.

Experimental design. The study was designed to use 2 different groups of mice to address 2 goals. The first goal was to assess the interobserver agreement of our scoring system when used by personnel in a sepsis research laboratory. The second goal was to evaluate the predictive capacity of the scoring system for the death of mice after the induction of polymicrobial sepsis by CLP.

An a priori power analysis was used to determine that a sample size of 174 observations was needed to achieve a κ statistic of 0.90 with 80% or more power and a significance level of 0.05 by using a two-sided Z-test in the κ module of PASS 15 Power Analysis and Sample Size Software (NCSS LLC, Kaysville, Utah). A power analysis was also used to determine sample sizes needed to validate the scoring system with the CLP model and to predict death, which was conducted with PASS 15 and two-sided unequal-variance t-tests and based on data from a published study.[48] The time of death can vary in the CLP model, but for an approximate mortality level of 40% over a 7-d timeline, a priori power analysis determined a total sample size of 65 mice was needed to achieve 80% power to

detect temperature differences of 4.5 °C with a significance level of 0.05. A final sample size of 42 male and 42 female C57BL/6J mice was selected to account for varying survival rates between male and female septic mice and for animal losses unrelated to sepsis.[15,90] Control groups were not used, as the survival rate for sham laparotomy mice has been repeatedly documented to be 100% and the observer is inherently blind at the time of scoring as to whether a mouse will survive.[43,58]

To evaluate the scoring system for interobserver agreement (goal 1), the scores from 5 researchers were compared with the scores from a single veterinarian. The 5 researchers worked in the same sepsis research laboratory and had experience ranging from one to over 10 y duration with the CLP model. Observers scored mice that underwent CLP surgery in ongoing 7-d post-CLP survival curve experiments through individual paired veterinarian-researcher monitoring sessions. The paired veterinarian-researcher monitoring sessions occurred on different days and at different times. The study of the scoring system's interobserver agreement was noninvasive and was superimposed on to ongoing studies in the same sepsis research laboratory. The pool of mice consisted of various strains on a C57BL/6 background (n = 283) and were assigned to different experiments assessing for the effect of sepsis and comorbidities on sepsis pathophysiology.

Before the experimental scoring of mice, the veterinarian trained the researchers by using an interactive PowerPoint presentation to score various videos and images in a 30-min didactic session. Videos and images showed C57BL/6 background, post-CLP mice from the same sepsis laboratory, and accounted for 15 min of the didactic training session. In addition, independent hands-on 15-min training sessions were conducted between the veterinarian and individual researchers before the initial experimental scoring session. For each experimental scoring session, the veterinarian gave individual researchers a packet that contained a colored printout of scoring system (Figure 1), a timer for counting breaths (Digital Countdown Timer, Traceable Products, Webster, TX), a laboratory composition notebook, and ink for the permanent entry of scores into the notebook. The researcher and veterinarian then independently scored mice based on "Scoring System Development and Description." The veterinarian and researcher did not discuss mice or scores and were blind to each other's entries. Combining the scores awarded across all the individual researcher-veterinarian monitoring sessions, 849 parameter scores were recorded by observing 283 mice with CLP-induced sepsis.

For validation of the scoring system to predict death (goal 2), mice (C57BL/6J, n = 39 females, n = 41 males) underwent CLP surgery on day 0 and were monitored through day 7 or until death occurred. The aim of goal 2, evaluating the ability of the scoring system to predict death, was performed as a standalone project, not as an add-on to ongoing experiments in the sepsis research laboratory. The research laboratory used a standardized approach to perform CLP, and therefore surgery and perioperative care were identical for mice used to accomplish both goals of the study. CLP surgeries were performed in 3 groups of mice, and all surgeries were completed within a one-month period. The same veterinarian performed all surgeries for each group on one standard business day and also performed subsequent observations for this portion of the study.

After mice underwent CLP surgery, the veterinarian monitored them at 0700, 1100, 1500, and 1900 every day for 7 d and recorded scores and body temperatures; temperatures were obtained by using subcutaneous transponders as described in the "Scoring System Development and Description" section. The monitoring intervals were selected to allow consistent data collection every 4 h during the facility's light phase and to evaluate the ability of the scoring system to predict death at times when mice would typically be monitored by research and veterinary staff. The veterinarian also documented the presence or absence of a nest during the 24 h after CLP surgery. Nesting material was consistently provided in all mouse cages, and cages were not scheduled for routine cage change during the 7-d experiment, thus minimizing the effect of husbandry parameters on the study. Body weights were measured after the completion of monitoring and temperature measurements at 0700 every day. Notes were made if atypical behavior and clinical signs were detected at any time point. At 1900 on day 7 after CLP, a final temperature was measured in surviving mice by using a rectal

Parameter	Score	Parameter description
Respiratory status	0	Normal rhythm, rapid mouse respiration, with RR > 120 (too fast to count)
	1	RR > 120, but with periods of labored breathing
	2	Slow respirations (RR= 61-120)
	3	Respiratory rate less than or equal to 60
Activity, response to stimulus (ASR)	0	Mouse is performing any prior to stimulus: Climbing, running, fighting. Normal spontaneous walking. Mouse displays normal reaction in response to stimulus.
	1	Slightly suppressed activity prior to stimulus. Spontaneous walking but slower than normal. Mouse moves forward in response to stimulus but slower than normal.
	2	Moderately suppressed activity. No spontaneous walking observed prior to stimulus. Mouse moves at least two steps forward in response to stimulus but much slower than normal.
	3	No activity. No spontaneous walking observed prior to stimulus. Mouse moves less than 2 steps forward in response to touch. Mouse may experience tremors in response to touch.
Eyes	0	Open
	1	Partially closed
	2	Half-closed
	3	Mostly or completely closed

Figure 1. Scoring system for prediction of death in mice with CLP-induced sepsis.

thermometry probe. Euthanasia was then performed by using CO_2 asphyxiation.

CLP surgery. Sepsis was created surgically by CLP. All procedures were conducted in a designated procedure space of the animal research facility. Each mouse was placed in an induction chamber with 100% oxygen and 3% isoflurane at a flow rate of 1 L/min. After losing the righting reflex, the mouse was removed from the induction chamber and moved to the preparation table. A nose cone and 2% isoflurane were used maintain anesthesia for the remainder of the procedure. All mice received buprenorphine hydrochloride (0.1 mg/kg, Par Pharmaceutical, Rochester, MI) as preemptive analgesia prior to the skin incision. The mouse's lower left abdominal quadrant was clipped to expose a 2 × 2–cm square of skin. A surgical scrub consisting of 3 applications of combined chlorohexidine and isopropyl alcohol swabs (Prevantics, PDI, Woodcliff Lake, NJ) was applied to the surgical area. Thermal support was provided by a circulating water heating blanket (38 °C; model TP700T, Gaymar Industries, Orchard Park, NY). Mice were draped with ethylene oxide sterilized (Sterrad NX, Advanced Sterilization Products, Irvine, CA) transparent cling film (Press'n Seal, The Glad Products Company, Oakland, CA). A left paramedian incision was made through the skin and abdominal muscle, and the cecum was exteriorized. The cecum was ligated 1 cm from its base with autoclaved 4-0 silk (SUT-15 to 2, Roboz Surgical Instruments, Gaithersburg, MD). Halfway between the distal aspect of the cecum and the silk ligature, the cecum was then punctured completely with a 25-gauge needle (25G × 1 in. hypodermic needle, Excel, Redondo Beach, CA). To check for patency, a small amount of cecal contents was extruded from the puncture site. The intestine was then placed back into the abdomen. The abdominal body wall was closed with 4-0 polyglactin suture (Coated Vicryl Polyglactin 910, Ethicon, Somerville, NJ) in a single cruciate pattern and the skin incision was approximated with surgical adhesive (Webglue, Patterson Veterinary, Greeley, CO). The surgery required less than 10 min and the length for cecal ligation was selected to produce a mortality rate of approximately 50%.[74]

Prior to recovery from anesthesia, mice were implanted between the scapulae with sterile subcutaneous temperature radio frequency identification (RFID) transponders (Temperature programmable microchip, product UCT-2112, UID, Lake Villa, IL) by using a handheld microchip injector (Microchip injector with ejector, product UPGI-Q, UID, Lake Villa, IL). A postoperative temperature was recorded using a handheld RFID reader (URH-1HP handheld reader, UID, Lake Villa, IL). Mice received unique ear tag identification (La Pias mouse ear tags, Braintree Scientific, Braintree, MA). For cages housing 5 mice, the fifth mouse did not receive an ear tag and was identified with a unique number. To mimic the clinical care of human patients with sepsis,[17] all mice received 1 mL of warmed sterile saline subcutaneously for fluid replacement and subcutaneous broad-spectrum antibiotics (ceftriaxone 50 mg/kg; Hospira, Lake Forest, IL and metronidazole 35 mg/kg; Baxter Healthcare, Deerfield, IL) postoperatively. At 0700 and 1900 for 48 h after CLP,[33] all mice continued to receive subcutaneous ceftriaxone and metronidazole after the recording of scoring system observations and body weight measurements. A heating pad was placed under the recovery cage to avoid hypothermia. Mice were monitored continuously until recovery from anesthesia and then every 15 min until they were ambulatory. After recovery from anesthesia, male and female mice were returned to their original groups for the remainder of the experiment.

Scoring system development and description. We performed a pilot study to refine the scoring system descriptions and methodology in order to optimize interobserver agreement. In the pilot, 6 members of the veterinary staff at the research institution used the scoring system. Participating veterinary staff members had varying years of experience with research mice. A faculty veterinarian, resident veterinarians, and veterinary technicians were paired with the primary author to observe a total of 50 CLP mice assigned to various 7-d survival experiments on different days and times over the course of 1 mo. Mice used in these experiments were removed from the study when moribund and thus were not monitored until spontaneous death. During the pilot study, observers scored mice on the 3 parameters of respiratory status, activity and stimulus response (ASR), and eye appearance. The scoring system was refined based on feedback that would increase interobserver agreement to create a final version (Figure 1). Adjustments to the final version of the scoring system included the use of exact respiratory rates for given scores, removal of agonal breaths as a characteristic for a respiratory score of 3, and characterization of ASR scores of 2 and 3 as failure of the mouse to make any spontaneous movement and to move only in response to the stimulus.

The final scoring system (Figure 1) was designed to include key physiologic and behavioral parameters that we predicted would be simple enough to produce high interobserver agreement and accurately differentiate severely affected mice in the CLP model, based on previous research, our pilot study, and the expertise of sepsis researchers at our institution.[30,48,77] Our scoring system involves 3 components: respiratory status, ASR, and eyes. For the respiratory parameter, labored breathing was defined as the display of increased abdominal effort or an inconsistent rate, which created a flutter-type appearance to the sides of the thorax. For the ASR parameter, a tail pull was selected as the stimulus because it minimized animal handling and was piloted to be more reliable than pushing the mouse's dorsum with the observer's finger, as the push elicited a downward rather than forward motion. For the eye parameter, in cases of unilateral changes, the eye that was the most open was evaluated and scored.

All mice in a cage had undergone CLP surgery on the same day, and they remained in the home cage for the duration of the study. The veterinarian first removed the cage from the rack and placed it in a running biosafety cabinet. The cage lid and the food hopper were removed and observers then scored the respiratory status of all mice in the cage by observing thoracic wall movements for 30 s. Mice were identified during the noninvasive monitoring by cage number and by unique colored ear tags described in the "CLP Surgery" section. Nesting and moist chow materials were then temporarily removed from the cage and mice were observed for spontaneous activity. To evaluate the stimulus response, individual mice were subjected to a gentle pull on the base of the tail. The tail pull stimulus was brief and consistently applied by the veterinarian in all monitoring sessions, even when paired with observers from the research staff. If the mouse reacted by moving backward, a second stimulating tail pull was elicited to observe for forward motion. After scoring the stimulus response, the eyes of each mouse were assessed for degrees of openness or orbital tightening. Lastly, mouse temperature were measured in the home cage by waving the URH-1HP reader (UID Devices) over individual mice or through the cage wall to detect a signal. Temperature values were recorded in °C. To avoid background signal from cohoused mice, the observer's hand was lightly placed over other mice as a barrier. At 1900 on day 7 after CLP, rectal

thermometry was used to measure temperature for comparison with transponder values. The rectal probe (RET-3 mouse rectal probe, Braintree Scientific, Braintree, MA) was coated with a small amount of sterile lubricant and then inserted to a marked length of 20 mm from its distal tip. The rectal probe was connected to a monitoring device (Microtherma 2 thermometer, Thermoworks, Lindon, UT) and temperature was recorded in °C. The transponder and rectal thermometry comparison was limited to mice that had survived for 7 days after CLP to avoid the potential confounding factor of frequent restraint for rectal probe insertion on the behavioral parameters. Any additional necessary restraint, handling, or injections that were capable of inducing an acute stress response were performed after scoring.

Statistical analysis. For the evaluation of interobserver agreement of the 3 observational parameters, researcher observations were consolidated and compared against the trained veterinarian's observations using weighted Cohen κ statistic. All 283 mice from the paired researcher-veterinarian monitoring sessions were included in the statistical analysis as we had no predetermined exclusion criteria. The statistical significance of κ was analyzed by threshold $\alpha = 0.05$. Weighted Cohen κ statistic values between 0.01 and 0.20 are considered indicative of slight interobserver agreement, between 0.21 and 0.40 indicate fair agreement, between 0.41 and 0.60 indicate moderate agreement, between 0.61 and 0.81 indicate substantial agreement, and κ values higher than 0.81 indicate near-perfect agreement.[51] All analyses for interobserver agreement were performed using SAS 9.4 (SAS Institute, Cary, NC).

Use of scoring system for prediction of death. Data were analyzed to determine whether the scoring system could predict death after CLP in mice. The dataset included 80 mice, with 28 monitoring sessions and observation points possible for each mouse. Three parameters—respiratory status, ASR and eyes—were scored as 0 to 3. A fourth factor, temperature, was also included and analyzed. The monitoring session in which a mouse was found dead was defined as a censoring event, as no scores could be assigned to the mice that died prior to the end of the 7-d period after CLP. After excluding censoring events, retrospective data analysis for the prediction of mortality encompassed 1,336 observations for each parameter and temperature. The full dataset used sensitivity and specificity over the 7-d observation period and directly preceding a censoring event to detect parameters that performed well in predicting death. Sensitivity and specificity for individual parameters, total parameter scores, combined ASR and eye scores, and temperature were calculated as follows, sensitivity was defined as equal to true positives/(true positives + false negatives) and specificity as equal to true negatives/(true negatives + false positives). True positives were defined as mice that died after receiving a score of 3 for a parameter, whereas false positives were defined as surviving mice that had received a score of 3 in a parameter. True negatives were defined as mice that survived without receiving a score of 3 in a parameter, whereas false negatives were defined as mice that died without receiving a score of 3 in a parameter.

The sensitivity and specificity determinations for temperature used data collected in Celsius format. The median temperature for all mice was 32 °C. Therefore, the target of interest for temperature was a subcutaneous temperature below 32 °C and its potential use as a cut-off to quantitatively predict death. This temperature was cross tabulated against an outcome of death or survival to calculate sensitivity and specificity. For the sensitivity and specificity, true positives were defined as mice that died after receiving a last recorded subcutaneous transponder temperature of less than 32 °C, whereas false positives were defined to as mice that survived to the experimental endpoint but received a last recorded subcutaneous transponder temperature of less than 32 °C. True negatives were defined as mice that were alive at the experimental endpoint and received a last recorded subcutaneous transponder temperature greater than or equal to 32 °C, whereas false negatives were defined as mice that died after receiving a last recorded subcutaneous transponder temperature greater than or equal to 32 °C. We also evaluated temperatures of surviving and nonsurviving mice longitudinally over the course of the 7-d period after CLP; thus, the mean temperatures and standard deviation for all mice in the surviving and nonsurviving subgroups were calculated across all 28 monitoring sessions.

A time course analysis was also conducted to determine whether pivotal parameter scores rose or fell during the 24-h period preceding death and how these scores changed as compared with the overall experimental timeline. To do this, mice were grouped into 2 subsets: survivors and nonsurvivors. Each scoring system parameter was analyzed at the last 3 time points, with the fourth time point being either death or the experimental endpoint of day 7 after CLP. The 3 parameters and the temperature data were averaged across all mice in the 2 subgroups for the antepenultimate (2 monitoring sessions before last), penultimate (monitoring session before last), and last recorded time points and were evaluated for significant differences. A total score was created by adding the scores of all 3 parameters (respiratory, ASR, and eyes). The total score was then averaged across all mice in the 2 subgroups by time point for analysis. To determine whether the scoring system could be refined to 2 parameters, combined scores for ASR and eyes were generated by adding the scores from the 2 parameters. The combined ASR and eye scores were averaged across all mice in the 2 subgroups by time point for analysis. Data are presented as mean ± SD. Differences between the surviving and nonsurviving subgroups for the average of individual parameter scores, the average of total scores, the average of combined ASR and eye scores, and the average subcutaneous temperatures across the final 3 time points prior to euthanasia or death were evaluated for statistical significance by using Mann–Whitney U and Wilcoxon signed rank tests. Analyses were performed using Python 3.10.0 (Python Software Foundation, Wilmington, DE).

The highest total score and highest combined ASR-eye score given to a surviving mouse throughout the course of the experiment and the final scores given to a nonsurviving mouse prior to its death were also evaluated. The sensitivity and specificity of total score was calculated by defining true positives as nonsurviving mice that received total scores greater than 6, whereas false positives were defined as surviving mice that received total scores greater than 6. True negatives were defined as surviving mice that received a total score less than or equal to 6, whereas false negatives were nonsurviving mice that received total scores of less than or equal to 6. Combined ASR-eye scores range from 0 to 6. Therefore, threshold scores of 5 and 6 were compared. The sensitivity and specificity of the combined ASR-eye scores was calculated by defining true positives as nonsurviving mice that received combined scores greater than or equal to 5, whereas false positives were defined as surviving mice that received a combined score greater than or equal to 5. True negatives were defined as surviving mice that received a combined score of less than 5, while false negatives were nonsurviving mice that received a combined score less than 5. All analyses for evaluating scoring and temperature

as predictors of death were performed using Python 3.10.0 (Python Software Foundation, Wilmington, DE) and Prism 9.2 (GraphPad Software, La Jolla, CA).

Validation of scoring system as a diagnostic for prediction of death. Receiver operating characteristic (ROC) curves show the true positive rate against the false positive rate. ROC curves diagnose the ability of a binary classifier to correctly classify observations. The true positive rate (TPR) is defined as the proportion of observations that were correctly predicted as positive out of all positive observations. The false positive rate (FPR) is defined as the proportion of observations that were incorrectly predicted as positive out of all negative observations. The 45-degree diagonal of a ROC curve indicates random chance classification; movement of the curve away from the diagonal toward the top left corner of the graph indicates better predictions. For the purposes of this study, ROC curves were built based on a logistic regression model. The binary outcome was death, with 0 indicating living throughout the study and 1 indicating death at some time point during the study. The first logistic regression model had 3 input features: respiratory status, ASR, and eyes. The second logistic regression model had 2 input features: ASR and eyes.

An area under the curve (AUC) was calculated for each ROC curve to summarize the overall diagnostic accuracy of the scoring system, which was the ability of the scoring thresholds to predict the death of CLP mice. An AUC of 0.5 indicates no discrimination and diagnostic accuracy by random chance, a value of 0.7 to 0.8 is considered acceptable for prediction, 0.8 to 0.9 is considered excellent for prediction, and lastly, an AUC greater than 0.9 is considered outstanding for diagnostic accuracy.[49,93]

Results

Interobserver agreement. The interobserver agreement between research personnel and the veterinarian was calculated for observations obtained from 283 CLP mice on the parameters of respiration, ASR, and eyes (Table 1). The interobserver agreement qualifies as near-perfect agreement for all 3 parameters. The 95% confidence interval for all 3 parameters indicated consistency among the observers, which included all participating research personnel and the veterinarian.

Prediction of death. ***Mortality and clinical signs.*** Of the 80 mice that underwent CLP, 54 (68%) died within 7-d and were defined as nonsurvivors (Figure 2). The 26 mice that reached the 7-d experimental endpoint were defined as survivors. In addition to the clinical signs that were included in the scoring system, mice rarely displayed seizure activity, ataxia with the ability to move forward in response to tail pull stimulus, or recumbency (lateral or dorsal) that required brief handling to position the mouse in a righted posture to test the response to a tail pull stimulus. Mice engaged in minimal nesting behavior in the immediate 24 h period after CLP. Only one cage of mice integrated the clean cotton square into a nest within the first 24-h after surgery. The cotton square in the other cages (n = 17 cages) was not integrated into a nest during the first 24-h after CLP and was thus deconstructed by the observing veterinarian and left in the cage.

Scoring system. For each parameter of the scoring system, 1,336 scores were obtained from the 80 mice over the course of the experiment. For corresponding time points, parameter scores summed to produce 1,336 total scores and 1,336 combined ASR-eye scores. The averages of the last 3 recorded scores for each individual parameter, total scores, and combined ASR-eye scores were calculated for the surviving and nonsurviving subgroups (Figure 3).

The average scores for each of the 3 parameters were significantly lower across the final 3 time points prior to euthanasia for surviving mice as compared with those of nonsurviving mice prior to death ($P < 0.0001$, Figure 3A). The highest mean scores of surviving mice through euthanasia on day 7 after CLP were 1.0 ± 0.9 for respiratory, 1.1 ± 0.9 for ASR, and 1.3 ± 0.9 for eyes. With regard to time course, the highest mean scores for surviving mice at the final 3 time points before euthanasia were 0.4 ± 0.6 for respiratory, 0.6 ± 0.9 for ASR, and 0.5 ± 1.0 for eyes, as compared with scores for nonsurviving mice of 2.1 ± 0.6 for respiratory, 2.9 ± 0.4 for ASR, and 2.7 ± 0.5 for eyes. The respiratory, ASR, and eye scores for surviving mice contributed average total scores of 1.3 ± 2.1, 1.5 ± 2.3, and 1.3 ± 2.4, respectively, at the antepenultimate, penultimate and final sessions prior to euthanasia (Figure 3B). The highest average total score for all 3 parameters for surviving mice at any time during the experiment was 3.4 ± 2.5. Overall, there was a positive relationship between average total score and death of the animal. All nonsurviving mice approached an average maximal total score of near 7 in the final 3 monitoring sessions before death (Figure 3B). Specifically, the average total scores for nonsurviving mice were 7.0 ± 1.2, 7.4 ± 1.0, and 7.5 ± 0.8, respectively, for antepenultimate, penultimate, and final sessions prior to death. The average individual parameter scores and the average total scores of the surviving and nonsurviving mice were significantly different for the final 3 time points before euthanasia or death, ($P < 0.0001$).

Sensitivity and specificity for individual parameters were calculated for obtaining a score of 3 both at any time during the study and immediately before death (Table 2). The respiratory parameter demonstrated lower sensitivity and specificity for all mice across the experiment and prior to death, as compared to

Table 1. Interobserver scoring agreement of research personnel and veterinarian for all monitoring sessions

Monitoring parameter	Weighted Cohen κ statistic	95% CI
Respiratory	0.8161	(0.76 to 0.87)
Activity-stimulus	0.8964	(0.86 to 0.93)
Eyes	0.8072	(0.76 to 0.85)

Interobserver agreement (weighted Cohen Kappa statistic) results are statistically defined as slight agreement (0.01–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.81), and near perfect (> 0.81).[51] n = 283 mice monitored for agreement when researchers compared with veterinarian.

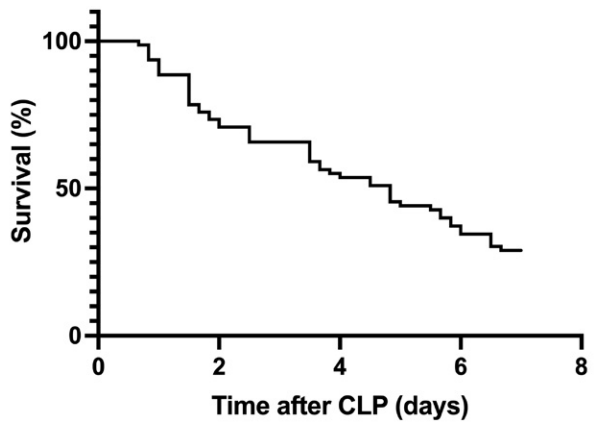

Figure 2. Kaplan-Meier curve for overall survival of mice (n = 80) over 7 d after CLP surgery.

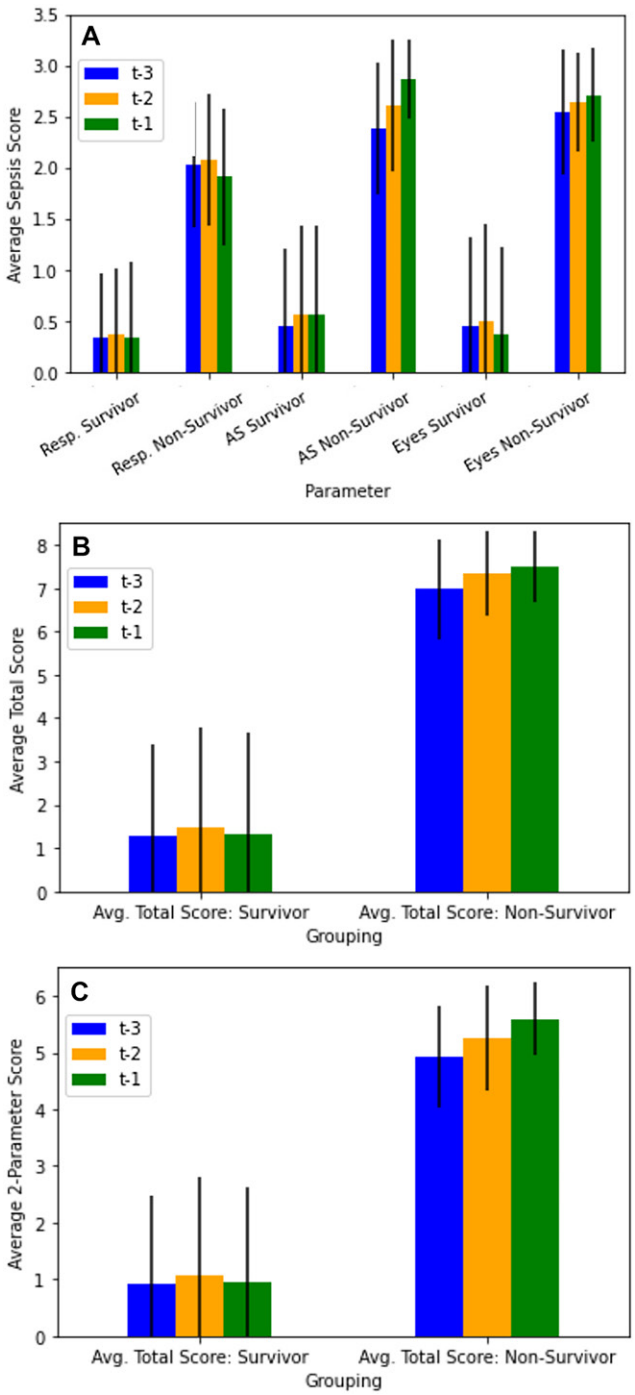

Figure 3. Average recorded scores and standard deviations for the last 3 monitoring sessions of surviving and nonsurviving mice for respiratory, ASR (activity and response to stimulus), and eye parameters. t-3, t-2 and t-1 are respectively the antepenultimate (third from last), penultimate (second from last) and final (last) recorded monitoring sessions. Across each time point, scores for each parameter were averaged for all surviving (n = 26) and nonsurviving mice (n = 54). (A) Average scores for the respiratory, ASR, and eye parameters. (B) Average total recorded scores for surviving and nonsurviving mice, determined by summing scores for the respiratory, ASR and eye parameters. (C) Average combined ASR-eye scores for nonsurviving and surviving mice, determined by summing ASR and eye scores and excluding respiratory scores.

the sensitivity and specificity of the ASR and eyes parameters. Respiratory scores of 1 and 2 were awarded more immediately before death compared to scores of 3, which included 43 mice that received a respiratory score of 1 and 2 and 11 mice that received a score of 3 at the monitoring session directly before death. Sensitivity and specificity for total parameter scores were calculated for obtaining a total score greater than 6 immediately before death, while the sensitivity and specificity for combined ASR-eye scores were calculated for obtaining a score of greater than or equal to 5, excluding the less sensitive and specific respiratory parameter as an input. The sensitivity and specificity for a total score greater than 6 corresponding to the death of a mouse at the next monitoring session was 96% and 92% respectively (Table 3). The sensitivity and specificity for a combined ASR-eye score of at least 5 corresponding to death of a mouse at the next monitoring session was also 96% and 92% respectively (Table 3). Five mice were excluded from these evaluations because they died within the 24-h period after CLP and before receiving scores or temperatures from at least 3 monitoring sessions; however, these mice were included in the sensitivity and specificity calculations that did not require at least 3 measurements.

Over the course of all monitoring sessions, 41 mice received a score of 3 in the respiratory category and 36 were found dead, 54 mice received a score of 3 in the ASR category and 52 were found dead, and 58 mice received a score of 3 for the eyes category and 51 were found dead; these animals were defined as true positives. True negatives were comprised of surviving mice that did not receive a score of 3 in the 3 parameters, specifically 21 mice that did not receive a score of 3 in the respiratory category and survived, 24 mice that did not receive a score of 3 in the ASR category and survived, and 19 mice that did not receive a score of 3 for the eyes category and survived (Table 4). Two surviving mice received scores of 3 at their final monitoring session before euthanasia. One of these mice had a score of 3 for both ASR and eyes, while the other mouse had scores of 3 for all parameters. One previously mentioned mouse had scores of 3 in all parameters (total score of 9) and survived until the last monitoring session on day 7. With regard to scores awarded to nonsurviving mice at the last monitoring session before death, 1 mouse had a total score of 5, 1 mouse had a total score of 6, 26 mice had a total score of 7, 20 had a score of 8, and 6 had a score of 9 (Table 5). If euthanasia had been performed at the moment a mouse received a total score of greater than or equal to 6 for all 3 parameters, pain and distress would have been avoided for 53 mice (98%) of the mice that died spontaneously in the current study (Table 5). Performing euthanasia at a total score of greater than or equal to 7 would have captured one fewer mouse, avoiding pain and distress for 52 mice (96%) that died spontaneously in the current study (Table 5).

Diagnostic accuracy of the scoring system. ROC analysis was performed based on the true positive and false positive rates for the 3 parameters combined against the outcome of death of a mouse during the study in order to determine the accuracy of our scoring system in predicting the death of septic mice. For the 3 inputs of respiratory, ASR, and eyes, the AUC generated from ROC analysis was 0.907 (Figure 4A).

Two-parameter scoring system. Because of its low sensitivity to predict death the respiratory parameter was excluded from additional statistical analysis. For the combined ASR-eye scores, nonsurviving mice had higher average scores in the final 3 monitoring sessions before death than did the surviving mice. Specifically, combined ASR-eye scores for nonsurviving mice were 4.9 ± 0.9, 5.3 ± 0.9, and 5.6 ± 0.6, respectively, at the antepenultimate, penultimate and final sessions before death (Figure 3C). The combined ASR-eye scores of surviving mice

Table 2. Sensitivity and specificity for correspondence of a score of 3 with the death of mouse after CLP

Monitoring parameter	Sensitivity (score of 3 at any time)	Specificity (score of 3 at any time)	Sensitivity (score of 3 before death)	Specificity (score of 3 before death)
Respiratory	70%	81%	20%	96%
Activity and stimulus response	96%	92%	87%	92%
Eyes	94%	73%	74%	96%

True positive = mouse that received a score of 3 and died.
False positive = mouse that received a score of 3 and survived.

Table 3. Criteria used to determine sensitivity and specificity of both total scores and combined activity-stimulus and eye scores for prediction of death by the time of the next monitoring session

Classifications	Number of mice	Outcome
True positive	52	Died
False positive	2	Survived
True negative	24	Survived
False negative	2	Died

Although the same number of mice were sorted into the same groups for sensitivity and specificity calculations for both metrics, cutoffs for total scores and scores from the 2-parameter system differ and are described in legend.
True positives for total score = nonsurviving mouse with a total score greater than 6.
True negatives for total score = surviving mouse with a total score less than or equal to 6.
True positives for ASR-eye score = nonsurviving mouse with a combined ASR-eye score greater than or equal to 5.
True negatives for ASR-eye score = surviving mouse with a combined ASR-eye score less than 5.

Table 4. Criteria used to determine sensitivity and specificity of each scoring system parameter for the prediction of death of mice after CLP at any time during the experimental timeline.

Monitoring parameter	True positives	True negatives
Respiratory	36	21
Activity-stimulus response	52	24
Eyes	51	19

True positive = mouse that received a score of 3 and died.
True negative = mouse that did not receive a score of 3 and survived.

were 0.9 ± 1.6, 1.1 ± 1.7, and 1.0 ± 1.7, respectively, at the antepenultimate, penultimate and final sessions prior to euthanasia (Figure 3C). The highest average combined ASR-eye score for surviving mice at any time during the study was 2.4 ± 1.7. The combined ASR-eye scores of surviving and nonsurviving mice were significantly different ($P < 0.0001$).

For combined ASR-eye scores given at the last monitoring session before death, one mouse received a score of 3, one a score of 4, 19 mice received scores of 5, and 33 received scores of 6 (Table 5). If the combined ASR and eye scores and a threshold score of 5 had been used, euthanasia could have been performed for 52 mice (96%) that died spontaneously in the current study, thereby avoiding an additional period of potential pain and distress (Table 5). Using a score of 6 as the threshold would have avoided pain and distress for only 33 mice (61%) that died spontaneously in the current study (Table 5). The sensitivity and specificity of an ASR-eye score of greater than or equal to 5 corresponding to death of a mouse by the time of the next monitoring session was 96% and 92% respectively.

Over the course of the study, mice that survived had significantly lower total and ASR-eye scores than did mice that died ($P < 0.0001$). However, mice that survived had relatively few high scores during the study. Specifically, 6 surviving mice had total scores greater than 6 during the study and 5 mice had ASR-eye scores of greater than or equal to 5 (Table 5). Total scores greater than 6 were given during the first 48-h after CLP for 4 out of 6 surviving mice, while the combined ASR-eye scores greater than or equal to 5 were given during the first 48-h after CLP for 3 of 5 surviving mice.

Diagnostic accuracy of 2-parameter scoring system. ROC analysis was repeated as described above with exclusion of the respiratory parameter and use of the combined ASR-eye parameters. ROC analysis of the 2-parameter scoring system resulted in an AUC of 0.8997 (Figure 4B).

Temperature. A total of 1,336 temperature measurements were obtained using the subcutaneous RFID-transponder temperature. Mouse temperatures ranged from 32 °C to 36 °C during anesthetic recovery from CLP surgery. The average temperatures during the last 3 monitoring sessions before the death of nonsurviving mice were 29.0 ± 1.6 °C, 29.0 ± 1.6 °C, and 29.0 ± 1.8 °C, with the third value being the last one recorded before death (Figure 5). The average temperatures over the course of the last 3 monitoring sessions before euthanasia of mice that survived until the endpoint were 35.9 ± 2.9 °C, 35.9 ± 2.7 °C, and 36.0 ± 2.7 °C, with the third value being the last temperature average recorded before euthanasia (Figure 5). The lowest average subcutaneous temperature recorded for surviving mice throughout the experiment was 34.2 ± 3.1 °C, which rose toward 36.0 °C; in contrast, nonsurviving mice averaged lower subcutaneous temperatures throughout the study (Figure 6). The temperatures of surviving and nonsurviving mice were significantly different ($P < 0.0001$) across all monitoring sessions. All 54 nonsurviving mice had a subcutaneous temperature below 32 °C at some point during the 7 d after surgery. Eleven of 26 surviving mice also had subcutaneous temperature below 32 °C at some point during the monitoring scheme; the low temperatures for 10 of these mice were recorded during the first 48-h after CLP. In the time course analysis, the lowest average temperature recorded across the last 3 monitoring sessions were 35.9 ± 2.7 °C for surviving mice and 29.0 ± 1.6 °C for nonsurviving mice.

At the time of the monitoring session closest to death, 52 mice had subcutaneous temperatures less than 32 °C, and 50 of those mice were found dead. These 50 mice were defined as true positives, and the other 2 were defined as false positives. True negatives comprised 24 surviving mice that developed subcutaneous temperatures greater than or equal to 32 °C at the final monitoring session before euthanasia, while 4 mice were considered false negatives due to being found dead despite having subcutaneous temperatures greater than or equal to 32 °C at the previous monitoring session. The sensitivity and specificity

Table 5. Distribution for the number of septic mice and the total scores across the respiratory, activity and response to stimulus (ASR), and eyes parameters and the combined ASR-eye scores at the last monitoring session (monitoring session before death) for nonsurviving mice, and the highest total scores and highest combined ASR-eye scores received during the study for surviving mice

Score	Number of nonsurviving mice that received this total score before death (last recorded score)	Number of nonsurviving mice that received this ASR-eye score before death (last recorded score)	Number of surviving mice that received this total score as highest score during study	Number of surviving mice that received this ASR-eye score as highest score during study
0	0	0	2	1
1	0	0	0	2
2	0	0	2	5
3	0	1	2	6
4	0	1	6	7
5	1	19	3	3
6	1	33	5	2
7	26	n/a*	4	n/a*
8	20	n/a*	0	n/a*
9	6	n/a*	2	n/a*

Total nonsurviving mice, n = 54.
Total surviving mice, n = 26, * denotes that scores >6 were not possible for the combination of activity and eyes alone because the maximal sum of the 2 scores is 6.

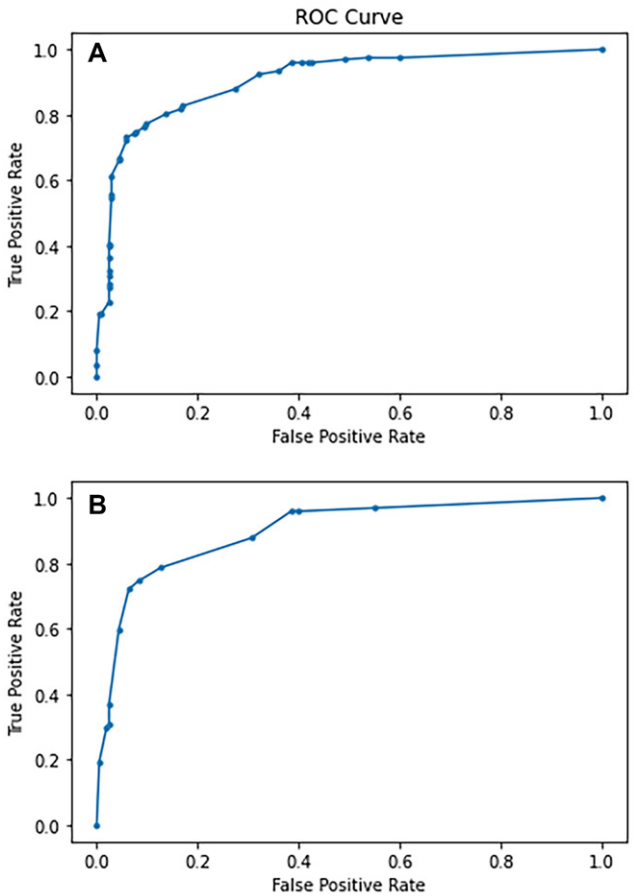

Figure 4. Receiver operating characteristic (ROC) curve plot analyses generated from logistic regression with the inputs of the true positive rate (Sensitivity) and false positive rate (1 - Specificity). (A) ROC curve plot analysis for the respiratory, ASR, and eye parameters against the outcome of death for septic mice. The AUC was 0.907. (B) ROC curve plot analysis generated for combined ASR and eye parameters for the outcome of death for septic mice. The AUC was 0.8997. ROC, receiver operating characteristics; AUC, area under the curve.

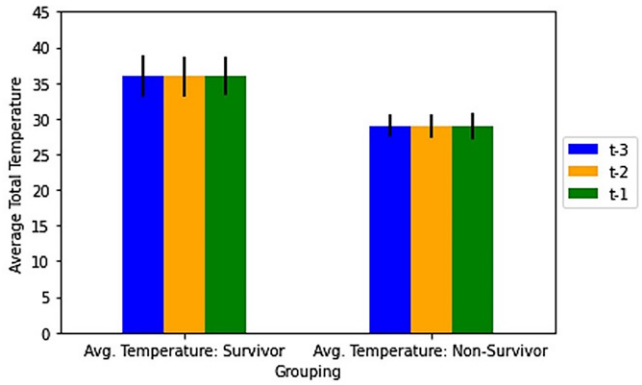

Figure 5. Average recorded temperatures and standard deviations from last 3 monitoring sessions for nonsurviving (n = 54) and surviving mice (n = 26). t-3, t-2 and t-1 are respectively penultimate (third from last) antepenultimate (second from last) and final (last) recorded monitoring session. For each time point, temperatures were averaged across all mice in the nonsurviving and surviving subgroups.

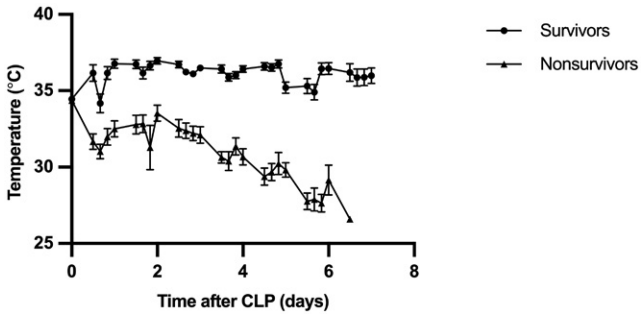

Figure 6. Average recorded temperatures and standard deviations for surviving (n = 26) and nonsurviving (n = 54) mice after CLP. For each session, the recorded temperatures were averaged across all mice in the nonsurviving and surviving subgroups, which were significantly different (P < 0.0001). The total number of monitoring sessions, which occurred 4 times per day for 7 d after surgery was n = 28, with 0 representing the day of CLP surgery. No mice died after the 25th monitoring session.

of temperatures below a threshold of 32.0 °C corresponding to death of a mouse were 93% and 92% respectively. The mean temperatures for surviving mice at the time of euthanasia were 36.0 ± 2.6 °C by subcutaneous RFID microchips and 35.0 ± 2.9 °C by rectal thermometry ($P = 0.2060$).

Discussion

Our novel scoring system predicted death with a high rate of sensitivity and specificity, as well as high diagnostic accuracy shown by ROC curve analysis and an AUC of 0.907 for the 3 observed parameters and an AUC of 0.8997 for the combined ASR-eye parameters. Specifically, a combined ASR-eye score of at least 5 had high sensitivity and specificity for identifying septic mice that would die before the next monitoring session. Various personnel used our scoring system with high interobserver agreement, indicating the generalizability of the system to be used in practice to determine humane endpoints for CLP-septic mice. Although our scoring system does not rely on temperature monitoring, we found that temperature also was a highly sensitive and specific quantitative predictor of death when measured using minimally invasive RFID microchips. We identified a threshold temperature of 32 °C that was predictive for death in IVC-housed groups of male and female mice that were maintained under standard conditions after CLP. These findings support our hypotheses that our refined scoring system identifies early humane endpoints for mice with CLP-induced sepsis, is practical to use by trained personnel with an acceptable level of agreement, and that subcutaneous temperature can identify surviving and nonsurviving mice.

The ASR parameter exhibited the highest sensitivity and specificity for indicating that a septic mouse would die at some point during the experiment. This parameter incorporated both the undisturbed and the stimulated movements of a mouse in order to reveal signals for intervention before the mouse reached a moribund state or died spontaneously from CLP-induced sepsis. The moribund condition has been used as a surrogate for death in rodent studies; the condition implies severe debilitation and unresponsiveness that precedes imminent death.[82] However, the moribund condition may not be an adequate humane endpoint, as unresponsive subjects may experience nociception and distress in the moribund state.[82] Our scoring system differs from those used in prior other studies[30,35,48,77]; ours has more descriptive respiratory and activity parameters that may capture clinical signs specifically displayed by mice after CLP, as shown in our pilot study. In addition, the ASR parameter demonstrated a higher interobserver agreement compared with the interobserver agreement for the respiratory and eyes parameters. However, the ability of only the ASR parameter to detect mice that spontaneously die by the time of the next monitoring session demonstrated lower sensitivity and specificity compared with the sensitivity and specificity of the parameter to detect mice that will die at some point during the experimental timeline.

For timely determination of humane endpoints, we recommend interpreting the ASR score in combination with the eye score, rather than interpreting individual parameter scores. Analgesic regimens may mask an animal's clinical condition, activity, or response to stimulus.[27] A previous study initially found differences in the clinical condition of CLP mice given analgesia with buprenorphine HCl as compared with buprenorphine sustained-release for analgesia; clinical condition scores included the assessment of activity and reactivity to handling; however, the activity and reactivity of mice did not differ based on analgesic regimen by 36-h after CLP.[27] The mice in our study received one perioperative dose of buprenorphine HCl for analgesic relief for 6 to 12 h, thus their ASR scores presumably match the effect of sepsis progression on spontaneous and stimulated animal activity with minimal confounding from the analgesic. Observational parameters may require additional validation for sepsis models with regard to specific analgesic regimens, as consensus on the use of analgesic regimens in studies of sepsis in mice is still being debated by experts in the field.[5,8,91]

The sensitivity of the eye parameter was high for the detection of mice that die over the course of the study, while the specificity of the eye parameter was high for the detection of immediate death. The sensitivity of 94% for the eye parameter documented that this feature may generate few false negatives, thus missing fewer cases and predicting septic mice that would die at some point during a study. However, compared to the respiratory and ASR parameters, the specificity of the eye parameter indicated that this parameter may generate more false positive results when used alone. Facial expressions encoded into grimace scores represent a well-established approach to the identification and assessment of pain in mammalian species used in research.[11,37,39,50,59,84] Mice undergoing CLP may experience pain related to the laparotomy required to induce sepsis and or related to disease progression, as pain is often cited as a symptom in human sepsis.[8] The eye parameter in our scoring system differs from the MGS, which focuses on orbital tightening over 4 different appearances. Adaptions of the MGS have been used to assess humane endpoints in several disease models and to evaluate pain in rodents.[39,48]

Our scoring system was validated using live-scoring of eye changes rather than retrospective scoring based on images or videos. Live-scoring is preferred over retrospective scoring to allow prompt humane intervention to prevent unnecessary pain and distress. Scores may differ when comparing live and retrospective scoring.[56] Our system is valuable for conducting cage-side evaluations of eye changes in mice with CLP-induced sepsis. Scores for the eye parameter were assigned after the scores for the respiratory and ASR had been assigned. This scoring order was used because the eyes of CLP mice were observed to resume their pre-stimulus orbital tightness promptly after the gentle tail pull was elicited; similarly, brief tail restraint during cage changes did not evoke significant changes to grimace scores.[57] The high sensitivity of the eye parameter supports the use of the measure to identify progression of sepsis toward imminent death in mice after CLP. Finally, the overall interobserver agreement for the eye parameter was substantial with a weighted Cohen κ-statistic of 0.8072, but observing the eyes of a mouse with a darkly pigmented coat can be difficult and training is important for successful implementation of this scoring system. However, the interobserver reliability of the eye parameter in our scoring system aligns with that reported previously for MGS and other systems.[11,29,37] Although the mice in our study had no diagnosed ophthalmic abnormalities, C57BL/6 and other inbred strains should be screened for microphthalmia and anophthalmia before using a clinical scoring system that includes eye assessment.[67,68]

The respiratory parameter had the lowest sensitivity for predicting mouse mortality and did not contribute significantly to increasing the predictive value, sensitivity, or specificity of the total score. The respiratory parameter in our scoring system was modified from the MSS.[77] Our respiratory parameter uses numeric definitions, whereas MSS and adapted-MSS are qualitative and difficult to evaluate.[77,81] However, our respiratory parameter was of limited use when considering high-value

scores of 3, since we observed more scores of 1 or 2 awarded at the monitoring session directly before death. The high respiratory rates that occurred before death could perhaps be due to altered acid-base status. Severely septic mice develop a mixed respiratory and metabolic acidosis,[92] and compensatory mechanisms might correlate to high respiratory rates and effort. Other possible causes for high respiratory rates include hypoxemia, stress, or antibiotic regimens.[32] Regardless of the cause, our observations complicate the use of respiratory status for predicting death. Further investigation is needed to incorporate respiratory rate changes or oxygen saturation as more specific prognostic factors for septic mice, perhaps mirroring the situation in human medicine.[40]

We found a near-perfect agreement between scores given by the veterinarian and research personnel in the scoring of 283 mice after CLP based on weighted Cohen κ-statistics. Cohen κ is a common and robust measure of interobserver reliability or agreement in veterinary medicine and other fields.[7] We used weighted Cohen κ statistics for the interobserver agreement portion of our study, thereby placing more weight on the degree of disagreement between the scores awarded by research personnel and the veterinarian.[18] To reduce animal numbers, the mice monitored by multiple raters to evaluate interobserver agreement were concurrently assigned to ongoing experiments being performed by the sepsis research laboratory. The experiments were diverse and aimed at elucidating the mechanisms of sepsis in the presence of comorbidities such as altered gut microbiome, cancer, alcohol ingestion, and immune cell alterations. Therefore, the high interobserver agreement for all observational parameters given the number of observers, number of observations, and the diverse background and use of the septic mice we observed supports the general utility of our scoring system.

To streamline our scoring system, we also evaluated ASR and eyes and excluded respiratory status, which did not contribute greatly to the predictive capacity of the scoring system and was the most complicated to assess. Obtaining a respiratory rate relies on accurate counts within a specific time period. We used the visual counting of breaths in our current study because it did not require specialized equipment, could be performed using routine housing and cageside observations, and could be performed by multiple raters observing awake mice.[24] Respiratory physical examinations are routine in human and veterinary medicine; however, respiratory rates and clinical signs vary substantially for many species in diverse clinical settings.[4,6,14,38] To our knowledge, the reliability of respiratory signs and rates has not been evaluated for cageside assessment in research rodents but anecdotally, training personnel for evaluation of respiration was difficult during our pilot studies. The sensitivity and specificity and the AUC determined from ROC curve analysis for the ASR and eye parameters support the high diagnostic accuracy of our scoring system without including the respiratory parameter. Because lung injury is variably associated with CLP-induced sepsis, we recommend the use of a threshold score for the combined ASR-eye scores of greater than or equal to 5 for the intervention of euthanasia. Delaying euthanasia until a mouse reaches a score of 6 for these 2 parameters would allow many mice to reach a threshold for euthanasia while being inaccurately identified as not progressing to spontaneous death.

The results confirmed our hypothesis that temperatures would be significantly different between surviving and nonsurviving septic mice. Specifically, we determined that temperatures below 32 °C could be a useful humane endpoint for CLP mice housed in conditions similar to those we used. The temperatures of nonsurviving mice across the last 3 recorded values averaged between 28 and 29 °C; however, four of the mice that died had subcutaneous temperatures equal to or slightly higher than 32 °C at their last recorded monitoring session. Nonetheless, these 4 mice had total scores and combined ASR-eye scores that met the thresholds of greater than 6 and 5, respectively. A subcutaneous temperature of less than 32 °C was strongly associated with the death of a CLP mouse by the time of the next monitoring session. The implanted transponders used in this study allowed quantitative data collection, without interfering with the behavioral observations or requiring additional manipulation as would be necessary with rectal temperatures. We validated the temperatures derived from the subcutaneous transponders with those obtained from rectal thermometry, and recommend the use of transponders as a replacement for rectal thermometry due to their noninvasiveness and agreement with rectal temperatures. Rectal thermometry and the required tail restraint can produce stress-induced hyperthermia, and repeated insertion of rectal probes alters body temperatures and induces locomotor activation.[2,3] For surviving mice, the majority of subcutaneous temperatures observed fell within reported reference ranges for normal mice (around 36 °C during the light phase),[22,71] with an exception being one time point at which the average temperature for surviving mice was 34.1 °C but after which temperatures increased to an average of 36 °C.

Our findings with regard to temperature are consistent with previous research on mouse sepsis models in general and with CLP specifically, as the studies indicate that mice become hypothermic during sepsis proportional to the severity of infection.[23,44,81] However, in sepsis research, nonsurviving mice can develop hypothermia ranging from 25 to 37 °C, with the variability attributed to the method used to induce sepsis, housing density, environmental parameters, and temperature monitoring modality and methods.[9,23,36,42,44,48,52,55,66,81] In mice, thermoregulatory processes and core temperature can be influenced by ambient conditions, housing density, husbandry practices, handling, strain, sex, size, and age.[20,21,53,76] The temperatures derived from the subcutaneous transponders cannot be directly compared with thermometry methods used in other sepsis studies. Temperatures might be higher in our study than in others because our measurements were obtained from unanesthetized mice and possibly because of upregulation of nonshivering thermogenesis in interscapular brown adipose tissue.[21,54] An effect of nonshivering thermogenesis is not likely because the rectal and interscapular temperatures were not significantly different; however, we only measured rectal temperature in surviving mice, so we do not know if a difference occurred at lower body temperatures. The majority of the studies that use mouse CLP do not report housing practices; however, social housing is now standard for mice. Mice in our study were socially housed and able to show typical murine huddling behavior for thermoregulation. A study that proposed a temperature cutoff of less than 30 °C measured temperatures by using rectal thermometry under anesthesia.[48] Use of a lower temperature cutoff in our study would have lowered the sensitivity of subcutaneous temperature as a predictor for death and would have neglected to identify 9 mice that would have died before their next assessment. Mice that undergo CLP can be hypothermic in the acute postoperative period and return to baseline temperatures as long as 32 h after surgery;[44,48,79] in our study, 10 of 11 mice that survived for 7 d had one subcutaneous temperature of less than 32 °C during the 48 h after CLP. The frequency of hypothermic events, duration of sustained

hypothermia, and timing of hypothermia relative to CLP are opportunities for further elucidating the value of temperature in predicting death in mice with CLP-induced sepsis. While overall our ASR-eye scores had substantial ability to predict death, temperature provides a quick, quantitative evaluation of a mouse disease state and subcutaneous temperatures below 32 °C could be a distinct humane endpoint for CLP-induced sepsis in mice.

A limitation of our study may be that our scoring system could be further developed, particularly with regard to the ASR and respiratory parameters. We observed additional clinical signs over the course of the study in severely septic mice, including seizures, recumbency, and ataxia. The neurologic clinical signs might indicate the progression of intraabdominal polymicrobial sepsis to septic encephalopathy,[89] while ataxia could be viewed as a type of abnormal ambulation that might manifest as a mouse progresses toward the moribund state. Because only a few mice exhibited these clinical signs during the frequent observations of this study and these mice received scores of three in the ASR parameter, further investigation is warranted to validate these clinical signs as consistent with the mouse CLP phenotype and to evaluate their correlation to disease severity before incorporating them into a scoring system. The variable outcome of CLP-induced lung injury and the technique required for the calculation of respiratory rates make the respiratory parameter relatively complicated to use. Further studies might use our proposed 2-parameter scoring system, ASR and eyes, for humane endpoint monitoring in different sepsis models, with different mouse strains and ages, and in different laboratories while assessing its validity for predicting death. Continued reassessment of interobserver agreement, including research, veterinary and training personnel, could also further validate our system. Novel automated systems and continuous animal monitoring can also utilize our scoring system and further studies may identify additional trends over experimental timelines different from the 7-d period utilized in our study.[13,88] Increased translatability to the clinical time course of human sepsis might be accomplished with longer postoperative follow-up or chronic models of sepsis.[91]

The thresholds identified in our study optimized sensitivity and specificity, but the possibility of false positives still warrants consideration. Our scoring system can be used to determine early humane endpoints in mice with CLP-induced sepsis. Regulatory provisions that discourage the use of death as an experimental endpoint exist in many countries.[10,47,69,91] Because our scoring system offers a means to reliably prompt euthanasia of septic mice that would otherwise experience spontaneous death, the use of our scoring system could increase international and inter-laboratory consistency in research that uses CLP-induced sepsis in mice.

In conclusion, our study describes an approach to identify humane endpoints that can be used routinely with reasonable effort and does not markedly deviate from the projected time of an animal's spontaneous death. Our brief, cageside observation-based scoring system provides surrogate humane endpoints that can be used to trigger timely euthanasia and avoid death as an endpoint in mouse studies of CLP-induced sepsis. The use of standardized humane endpoints is essential for the rigorous conduct of research and for meeting high standards for ethical and humane animal use. Based on our observations, the combined ASR-eye score at a threshold of 5 can replace death as an endpoint and reduce animal pain and distress. Subcutaneous RFID-transponders are a quick, noninvasive thermometry method that does not affect the observation of behavioral parameters for group-housed mice after CLP. We found that a subcutaneous temperature of 32 °C can be used as a quantitative threshold together with the scoring system. These thresholds can be incorporated into any monitoring scheme, although the standard frequency at our academic institution for acute sepsis models is at least twice daily monitoring. Our scoring system has high sensitivity and specificity to predict the death of mice with CLP-induced sepsis and permit the collection of quality samples following euthanasia, has high interobserver agreement, and has several advantages over the M-CASS, MSS, modified-MSS, and the adapted-MSS scoring systems. Our scoring system refines the sepsis model of mouse CLP and contributes to improved welfare of research animals. Lastly, future studies would be useful to validate our scoring system for other mouse strains and ages, sepsis models, and environments.

Acknowledgments

We thank Crystal Gergye, DACLAM for her contributions to early versions of our scoring system and literature review. We thank Wai Hanson, DACLAM, Leeza Birdwell, DACLAM, Melissa Crowe, ALAT, and Cassandra Anthony, RVT for their assistance in scoring mice for our pilot study. We acknowledge Limeng Wan, MSPH, and Reneé Moore, PhD for their initial statistical support and power analysis. For their participation and surgical assistance, we thank Gregory Sousa, VMD, PhD and Ashley Varnadoe, BA. For expert technological assistance and support, we thank Gregory Kable. The authors thank the Emory University Division of Animal Resources for dedicated husbandry and veterinary services. Funding for this project was generously provided by the Division of Animal Resources at Emory University, School of Medicine. Experiments within Dr. Craig Coopersmith's laboratory are funded by grants from the National Institute of General Medical Sciences under award numbers R01GM072808, R01GM104323, R01AA027396, and training grant T32GM095442. Finally, we would like to acknowledge our mice and their contributions to biomedical research and advancements in laboratory animal welfare.

References

1. **Adamson TW, Diaz-Arevalo D, Gonzalez TM, Liu X, Kalkum M.** 2013. Hypothermic endpoint for an intranasal invasive pulmonary aspergillosis mouse model. Comp Med **63**:477–481.
2. **Adriaan Bouwknecht J, Olivier B, Paylor RE.** 2007. The stress-induced hyperthermia paradigm as a physiological animal model for anxiety: A review of pharmacological and genetic studies in the mouse. Neurosci Biobehav Rev **31**:41–59. https://doi.org/10.1016/j.neubiorev.2006.02.002.
3. **Bae DD, Brown PL, Kiyatkin EA.** 2007. Procedure of rectal temperature measurement affects brain, muscle, skin, and body temperatures and modulates the effects of intravenous cocaine. Brain Res **1154**:61–70. https://doi.org/10.1016/j.brainres.2007.03.078.
4. **Berman J, Francoz D, Abdallah A, Dufour S, Buczinski S.** 2021. Evaluation of inter-rater agreement of the clinical signs used to diagnose bovine respiratory disease in individually housed veal calves. J Dairy Sci **104**:12053–12065. https://doi.org/10.3168/jds.2021-20503.
5. **Boehm CA, Nemzek JA.** 2021. Analgesia and Humane Endpoints for Rodents in Sepsis Research, p 221–229. In: Walker WE, editors. Sepsis: Methods and protocols. New York (NY): Springer US.
6. **Brabrand M, Hallas P, Folkestad L, Lautrup-Larsen CH, Brodersen JB.** 2018. Measurement of respiratory rate by multiple raters in a clinical setting is unreliable: A cross-sectional simulation study. J Crit Care **44**:404–406. https://doi.org/10.1016/j.jcrc.2017.12.020.
7. **Buczinski S, Boccardo A, Pravettoni D.** 2021. Clinical scores in veterinary medicine: What are the pitfalls of score construction, reliability, and validation? A general methodological approach applied in cattle. Animals (Basel) **11**:3244. https://doi.org/10.3390/ani11113244.

8. **Carpenter KC, Hakenjos JM, Fry CD, Nemzek JA.** 2019. The influence of pain and analgesia in rodent models of sepsis. Comp Med **69**:546–554. https://doi.org/10.30802/AALAS-CM-19-000004.
9. **Carpenter KC, Zhou Y, Hakenjos JM, Fry CD, Nemzek JA.** 2020. Thermoneutral housing temperature improves survival in a murine model of polymicrobial peritonitis. Shock **54**:688–696. https://doi.org/10.1097/SHK.0000000000001551.
10. **Council NHaMR.** 2013. Australian code for the care and use of animals for scientific purposes, 8th ed. Canberra: National Health and Medical Research Council.
11. **Dalla Costa E, Pascuzzo R, Leach MC, Dai F, Lebelt D, Vantini S, Minero M.** 2018. Can grimace scales estimate the pain status in horses and mice? A statistical approach to identify a classifier. PLoS One **13**:e0200339. https://doi.org/10.1371/journal.pone.0200339.
12. **Dejager L, Pinheiro I, Dejonckheere E, Libert C.** 2011. Cecal ligation and puncture: The gold standard model for polymicrobial sepsis? Trends Microbiol **19**:198–208. https://doi.org/10.1016/j.tim.2011.01.001.
13. **Do JP, Defensor EB, Ichim CV, Lim MA, Mechanic JA, Rabe MD, Schaevitz LR.** 2020. Automated and continuous monitoring of animal welfare through digital alerting. Comp Med **70**:313–327. https://doi.org/10.30802/AALAS-CM-19-000090.
14. **Domínguez-Ruiz M, Reinero CR, Vientos-Plotts A, Grobman ME, Silverstein D, Le Boedec K.** 2021. Interclinician agreement on the recognition of selected respiratory clinical signs in dogs and cats with abnormal breathing patterns. Vet J **277**:105760. https://doi.org/10.1016/j.tvjl.2021.105760.
15. **Drechsler S, Weixelbaumer K, Raeven P, Jafarmadar M, Khadem A, van Griensven M, Bahrami S, Osuchowski MF.** 2012. Relationship between age/gender-induced survival changes and the magnitude of inflammatory activation and organ dysfunction in post-traumatic sepsis. PLoS One **7**:e51457. https://doi.org/10.1371/journal.pone.0051457.
16. **Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, Machado FR, McIntyre L, Ostermann M, Prescott HC, Schorr C, Simpson S, Joost Wiersinga W, Alshamsi F, Angus DC, Arabi Y, Azevedo L, Beale R, Beilman G, Belley-Cote E, Burry L, Cecconi M, Centofanti J, Yataco AC, De Waele J, Dellinger RP, Doi K, Du B, Estenssoro E, Ferrer R, Gomersall C, Hodgson C, Moller MH, Iwashyna T, Jacob S, Kleinpell R, Klompas M, Koh Y, Kumar A, Kwizera A, Lobo S, Masur H, McGloughlin S, Mehta S, Mehta Y, Mer M, Nunnally M, Oczkowski S, Osborn T, Papathanassoglou E, Perner A, Puskarich M, Roberts J, Schweickert W, Seckel M, Sevransky J, Sprung CL, Welte T, Zimmerman J, Levy M.** 2021. Executive summary: Surviving sepsis campaign: International guidelines for the management of sepsis and septic shock 2021. Crit Care Med **49**:1974–1982. https://doi.org/10.1097/CCM.0000000000005357.
17. **Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, Machado FR, McIntyre L, Ostermann M, Prescott HC, Schorr C, Simpson S, Wiersinga WJ, Alshamsi F, Angus DC, Arabi Y, Azevedo L, Beale R, Beilman G, Belley-Cote E, Burry L, Cecconi M, Centofanti J, Coz Yataco A, De Waele J, Dellinger RP, Doi K, Du B, Estenssoro E, Ferrer R, Gomersall C, Hodgson C, Hylander Moller M, Iwashyna T, Jacob S, Kleinpell R, Klompas M, Koh Y, Kumar A, Kwizera A, Lobo S, Masur H, McGloughlin S, Mehta S, Mehta Y, Mer M, Nunnally M, Oczkowski S, Osborn T, Papathanassoglou E, Perner A, Puskarich M, Roberts J, Schweickert W, Seckel M, Sevransky J, Sprung CL, Welte T, Zimmerman J, Levy M.** 2021. Surviving sepsis campaign: International guidelines for management of sepsis and septic shock 2021. Crit Care Med **49**:e1063–e1143. https://doi.org/10.1097/CCM.0000000000005337.
18. **Fleiss JL, Cohen J.** 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ Psychol Meas **33**:613–619. https://doi.org/10.1177/001316447303300309.
19. **Gavin HE, Satchell KJF.** 2017. Surface hypothermia predicts murine mortality in the intragastric Vibrio vulnificus infection model. BMC Microbiol **17**:136. https://doi.org/10.1186/s12866-017-1045-z.
20. **Gordon CJ.** 2004. Effect of cage bedding on temperature regulation and metabolism of group-housed female mice. Comp Med **54**:63–68.
21. **Gordon CJ.** 2012. Thermal physiology of laboratory mice: Defining thermoneutrality. J Therm Biol **37**:654–685. https://doi.org/10.1016/j.jtherbio.2012.08.004.
22. **Gordon CJ.** 2017. The mouse thermoregulatory system: Its impact on translating biomedical data to humans. Physiol Behav **179**:55–66. https://doi.org/10.1016/j.physbeh.2017.05.026.
23. **Granger JI, Ratti PL, Datta SC, Raymond RM, Opp MR.** 2013. Sepsis-induced morbidity in mice: effects on body temperature, body weight, cage activity, social behavior and cytokines in brain. Psychoneuroendocrinology **38**:1047–1057. https://doi.org/10.1016/j.psyneuen.2012.10.010.
24. **Grimaud J, Murthy VN.** 2018. How to monitor breathing in laboratory rodents: A review of the current methods. J Neurophysiol **120**:624–632. https://doi.org/10.1152/jn.00708.2017.
25. **Hankenson FC, Ruskoski N, van Saun M, Ying GS, Oh J, Fraser NW.** 2013. Weight loss and reduced body temperature determine humane endpoints in a mouse model of ocular herpesvirus infection. J Am Assoc Lab Anim Sci **52**:277–285.
26. **Hellman J, Bahrami S, Boros M, Chaudry IH, Fritsch G, Gozdzik W, Inoue S, Radermacher P, Singer M, Osuchowski MF, Huber-Lang M.** 2019. Part III: Minimum quality threshold in preclinical sepsis studies (MQTiPSS) for fluid resuscitation and antimicrobial therapy endpoints. Shock **51**:33–43. https://doi.org/10.1097/SHK.0000000000001209.
27. **Herndon NL, Bandyopadhyay S, Hod EA, Prestia KA.** 2016. Sustained-release buprenorphine improves postsurgical clinical condition but does not alter survival or cytokine levels in a murine model of polymicrobial sepsis. Comp Med **66**:455–462.
28. **Hoffman M, Kyriazis ID, Lucchese AM, de Lucia C, Piedepalumbo M, Bauer M, Schulze PC, Bonios MJ, Koch WJ, Drosatos K.** 2019. Myocardial strain and cardiac output are preferable measurements for cardiac dysfunction and can predict mortality in septic mice. J Am Heart Assoc **8**:e012260. https://doi.org/10.1161/JAHA.119.012260.
29. **Hohlbaum K, Corte GM, Humpenöder M, Merle R, Thöne-Reineke C.** 2020. Reliability of the Mouse Grimace Scale in C57BL/6JRj mice. Animals (Basel) **10**:1648. https://doi.org/10.3390/ani10091648.
30. **Huet O, Ramsey D, Miljavec S, Jenney A, Aubron C, Aprico A, Stefanovic N, Balkau B, Head GA, de Haan JB, Chin-Dusting JP.** 2013. Ensuring animal welfare while meeting scientific aims using a murine pneumonia model of septic shock. Shock **39**:488–494. https://doi.org/10.1097/SHK.0b013e3182939831.
31. **Institute for Laboratory Animal Research.** 2011. Guide for the care and use of laboratory animals, 8th ed. Washington (DC): National Academies Press.
32. **Iskander KN, Craciun FL, Stepien DM, Duffy ER, Kim J, Moitra R, Vaickus LJ, Osuchowski MF, Remick DG.** 2013. Cecal ligation and puncture-induced murine sepsis does not cause lung injury. Crit Care Med **41**:159–170. https://doi.org/10.1097/CCM.0b013e3182676322.
33. **Iskander KN, Vaickus M, Duffy ER, Remick DG.** 2016. Shorter duration of post-operative antibiotics for cecal ligation and puncture does not increase inflammation or mortality. PLoS One **11**:e0163005. https://doi.org/10.1371/journal.pone.0163005.
34. **Kort WJ, Hekking-Weijma JM, TenKate MT, Sorm V, VanStrik R.** 1998. A microchip implant system as a method to determine body temperature of terminally ill rats and mice. Lab Anim **32**:260–269. https://doi.org/10.1258/002367798780559329.
35. **Krarup A, Chattopadhyay P, Bhattacharjee AK, Burge JR, Ruble GR.** 1999. Evaluation of surrogate markers of impending death in the galactosamine-sensitized murine model of bacterial endotoxemia. Lab Anim Sci **49**:545–550.
36. **Laitano O, Van Steenbergen D, Mattingly AJ, Garcia CK, Robinson GP, Murray KO, Clanton TL, Nunamaker EA.** 2018. Xiphoid surface temperature predicts mortality in a murine model of septic shock. Shock **50**:226–232. https://doi.org/10.1097/SHK.0000000000001007.

37. **Langford DJ, Bailey AL, Chanda ML, Clarke SE, Drummond TE, Echols S, Glick S, Ingrao J, Klassen-Ross T, Lacroix-Fralish ML, Matsumiya L, Sorge RE, Sotocinal SG, Tabaka JM, Wong D, van den Maagdenberg AM, Ferrari MD, Craig KD, Mogil JS.** 2010. Coding of facial expressions of pain in the laboratory mouse. Nat Methods **7**:447–449. https://doi.org/10.1038/nmeth.1455.
38. **Latten GHP, Spek M, Muris JWM, Cals JWL, Stassen PM.** 2019. Accuracy and interobserver-agreement of respiratory rate measurements by healthcare professionals, and its effect on the outcomes of clinical prediction/diagnostic rules. PLoS One **14**:e0223155. https://doi.org/10.1371/journal.pone.0223155.
39. **Leach MC, Klaus K, Miller AL, Scotto di Perrotolo M, Sotocinal SG, Flecknell PA.** 2012. The assessment of post-vasectomy pain in mice using behaviour and the Mouse Grimace Scale. PLoS One **7**:e35656. https://doi.org/10.1371/journal.pone.0035656.
40. **Lee CU, Jo YH, Lee JH, Kim J, Park SM, Hwang JE, Lee DK, Park I, Jang DH, Lee SM.** 2021. The index of oxygenation to respiratory rate as a prognostic factor for mortality in sepsis. Am J Emerg Med **45**:426–432. https://doi.org/10.1016/j.ajem.2020.09.052.
41. **Lewis A, Zuckerbraun B, Griepentrog J, Zhang X, Rosengart M.** 2017. Reducing animal use with a biotelemetry-enhanced murine model of sepsis. Sci Rep **7**:6622. https://doi.org/10.1038/s41598-017-05497-5.
42. **Lewis AJ, Yuan D, Zhang X, Angus DC, Rosengart MR, Seymour CW.** 2016. Use of biotelemetry to define physiology-based deterioration thresholds in a murine cecal ligation and puncture model of sepsis. Crit Care Med **44**:e420–e431. https://doi.org/10.1097/CCM.0000000000001615.
43. **Li H, Luo Y-F, Wang Y-S, Xiao Y-L, Cai H-R, Xie C-M.** 2018. Pseudomonas aeruginosa induces cellular senescence in lung tissue at the early stage of two-hit septic mice. Pathog Dis **76**:ftz001.
44. **Li JL, Li G, Jing XZ, Li YF, Ye QY, Jia HH, Liu SH, Li XJ, Li H, Huang R, Zhang Y, Wang H.** 2018. Assessment of clinical sepsis-associated biomarkers in a septic mouse model. J Int Med Res **46**:2410–2422. https://doi.org/10.1177/0300060518764717.
45. **Libert C, Ayala A, Bauer M, Cavaillon JM, Deutschman C, Frostell C, Knapp S, Kozlov AV, Wang P, Osuchowski MF, Remick DG.** 2019. Part II: Minimum quality threshold in preclinical sepsis studies (MQTiPSS) for types of infections and organ dysfunction endpoints. Shock **51**:23–32. https://doi.org/10.1097/SHK.0000000000001242.
46. **Lilley E, Armstrong R, Clark N, Gray P, Hawkins P, Mason K, Lopez-Salesansky N, Stark AK, Jackson SK, Thiemermann C, Nandi M.** 2015. Refinement of animal models of sepsis and septic shock. Shock **43**:304–316. https://doi.org/10.1097/SHK.0000000000000318.
47. **MacArthur Clark JA, Sun D.** 2020. Guidelines for the ethical review of laboratory animal welfare People's Republic of China National Standard GB/T 35892-2018. [Issued 6 February 2018 Effective from 1 September 2018]. Animal Model Exp Med **3**:103–113. https://doi.org/10.1002/ame2.12111.
48. **Mai SHC, Sharma N, Kwong AC, Dwivedi DJ, Khan M, Grin PM, Fox-Robichaud AE, Liaw PC.** 2018. Body temperature and mouse scoring systems as surrogate markers of death in cecal ligation and puncture sepsis. Intensive Care Med Exp **6**:20. https://doi.org/10.1186/s40635-018-0184-3.
49. **Mandrekar JN.** 2010. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol **5**:1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d.
50. **Matsumiya LC, Sorge RE, Sotocinal SG, Tabaka JM, Wieskopf JS, Zaloum A, King OD, Mogil JS.** 2012. Using the Mouse Grimace Scale to reevaluate the efficacy of postoperative analgesics in laboratory mice. J Am Assoc Lab Anim Sci **51**:42–49.
51. **McHugh ML.** 2012. Interrater reliability: The kappa statistic. Biochem Med (Zagreb) **22**:276–282. https://doi.org/10.11613/BM.2012.031.
52. **Mei J, Riedel N, Grittner U, Endres M, Banneke S, Emmrich JV.** 2018. Body temperature measurement in mice during acute illness: Implantable temperature transponder versus surface infrared thermometry. Sci Rep **8**:3526. https://doi.org/10.1038/s41598-018-22020-6.
53. **Meijer MK, Spruijt BM, van Zutphen LF, Baumans V.** 2006. Effect of restraint and injection methods on heart rate and body temperature in mice. Lab Anim **40**:382–391. https://doi.org/10.1258/002367706778476370.
54. **Meyer CW, Ootsuka Y, Romanovsky AA.** 2017. Body temperature measurements for metabolic phenotyping in mice. Front Physiol **8**:520. https://doi.org/10.3389/fphys.2017.00520.
55. **Miao P, Kong Y, Ma Y, Zeng H, Yu Z.** 2013. Hypothermia predicts the prognosis in colon ascendens stent peritonitis mice. J Surg Res **181**:129–135. https://doi.org/10.1016/j.jss.2012.05.078.
56. **Miller AL, Leach MC.** 2015. The Mouse Grimace Scale: A clinically useful tool? PLoS One **10**:e0136000. https://doi.org/10.1371/journal.pone.0136000.
57. **Miller AL, Leach MC.** 2016. The effect of handling method on the mouse grimace scale in two strains of laboratory mice. Lab Anim **50**:305–307. https://doi.org/10.1177/0023677215622144.
58. **Mishra SK, Choudhury S.** 2018. Experimental protocol for cecal ligation and puncture model of polymicrobial sepsis and assessment of vascular functions in mice. Methods Mol Biol **1717**:161–187. https://doi.org/10.1007/978-1-4939-7526-6_14.
59. **Mittal A, Gupta M, Lamarre Y, Jahagirdar B, Gupta K.** 2016. Quantification of pain in sickle mice using facial expressions and body measurements. Blood Cells Mol Dis **57**:58–66. https://doi.org/10.1016/j.bcmd.2015.12.006.
60. **Mohan S, Huneke R.** 2019. The role of IACUCs in responsible animal research. ILAR J **60**:43–49. https://doi.org/10.1093/ilar/ilz016.
61. **Molins CR, Delorey MJ, Young JW, Yockey BM, Belisle JT, Schriefer ME, Petersen JM.** 2012. Use of temperature for standardizing the progression of Francisella tularensis in mice. PLoS One **7**:e45310. https://doi.org/10.1371/journal.pone.0045310.
62. **Morton DB.** 1998. Humane endpoints in animal experimentation for biomedical research: Ethical, legal and practical aspects. Humane Endpoints in Animal Experiments for Biomedical Research. 1–12.
63. **Nandi M, Jackson SK, Macrae D, Shankar-Hari M, Tremoleda JL, Lilley E.** 2020. Rethinking animal models of sepsis - Working towards improved clinical translation whilst integrating the 3Rs. Clin Sci (Lond) **134**:1715–1734. https://doi.org/10.1042/CS20200679.
64. **National Research Council.** 2008. Recognition and alleviation of distress in laboratory animals. Washington (DC): National Academies Press.
65. **Nemzek JA, Hugunin KM, Opp MR.** 2008. Modeling sepsis in the laboratory: Merging sound science with animal well-being. Comp Med **58**:120–128.
66. **Nemzek JA, Xiao HY, Minard AE, Bolgos GL, Remick DG.** 2004. Humane endpoints in shock research. Shock **21**:17–25. https://doi.org/10.1097/01.shk.0000101667.49265.fd.
67. **Nunamaker EA, Anderson RJ, Artwohl JE, Lyubimov AV, Fortman JD.** 2013. Predictive observation-based endpoint criteria for mice receiving total body irradiation. Comp Med **63**:313–322.
68. **Nunamaker EA, Artwohl JE, Anderson RJ, Fortman JD.** 2013. Endpoint refinement for total body irradiation of C57BL/6 mice. Comp Med **63**:22–28.
69. **European Parliament.** 2010. Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. EU Official Journal. L276.
70. **Percie du Sert N, Hurst V, Ahluwalia A, Alam S, Avey MT, Baker M, Browne WJ, Clark A, Cuthill IC, Dirnagl U, Emerson M, Garner P, Holgate ST, Howells DW, Karp NA, Lazic SE, Lidster K, MacCallum CJ, Macleod M, Pearl EJ, Petersen OH, Rawle F, Reynolds P, Rooney K, Sena ES, Silberberg SD, Steckler T, Wurbel H.** 2020. The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. BMC Vet Res **16**:242. https://doi.org/10.1186/s12917-020-02451-y.
71. **Refinetti R.** 2010. The circadian rhythm of body temperature. Front Biosci (Landmark Ed.) **15**:564–594.
72. **Remick DG, Bolgos GR, Siddiqui J, Shin J, Nemzek JA.** 2002. Six at six: Interleukin-6 measured 6 h after the initiation of sepsis predicts mortality over 3 days. Shock **17**:463–467. https://doi.org/10.1097/00024382-200206000-00004.

73. **Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, Colombara DV, Ikuta KS, Kissoon N, Finfer S, Fleischmann-Struzek C, Machado FR, Reinhart KK, Rowan K, Seymour CW, Watson RS, West TE, Marinho F, Hay SI, Lozano R, Lopez AD, Angus DC, Murray CJL, Naghavi M.** 2020. Global, regional, and national sepsis incidence and mortality, 1990-2017: Analysis for the Global Burden of Disease Study. Lancet **395**:200–211. https://doi.org/10.1016/S0140-6736(19)32989-7.
74. **Ruiz S, Vardon-Bounes F, Merlet-Dupuy V, Conil JM, Buleon M, Fourcade O, Tack I, Minville V.** 2016. Sepsis modeling in mice: Ligation length is a major severity factor in cecal ligation and puncture. Intensive Care Med Exp **4**:22. https://doi.org/10.1186/s40635-016-0096-z.
75. **Russell WMS, Burch RL.** 1959. The principles of humane experimental technique. London: Methuen.
76. **Sanchez-Alavez M, Alboni S, Conti B.** 2011. Sex- and age-specific differences in core body temperature of C57Bl/6 mice. Age (Dordr) **33**:89–99. https://doi.org/10.1007/s11357-010-9164-6.
77. **Shrum B, Anantha RV, Xu SX, Donnelly M, Haeryfar SM, McCormick JK, Mele T.** 2014. A robust scoring system to evaluate sepsis severity in an animal model. BMC Res Notes **7**:233. https://doi.org/10.1186/1756-0500-7-233.
78. **Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, Bellomo R, Bernard GR, Chiche JD, Coopersmith CM, Hotchkiss RS, Levy MM, Marshall JC, Martin GS, Opal SM, Rubenfeld GD, van der Poll T, Vincent JL, Angus DC.** 2016. the third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA **315**:801–810. https://doi.org/10.1001/jama.2016.0287.
79. **Stolarski AE, Kim J, Nudel J, Gunn S, Remick DG.** 2022. defining sepsis phenotypes-two murine models of sepsis and machine learning. Shock **57**:268–273. https://doi.org/10.1097/SHK.0000000000001935.
80. **Stortz JA, Raymond SL, Mira JC, Moldawer LL, Mohr AM, Efron PA.** 2017. Murine models of sepsis and trauma: Can we bridge the gap? ILAR J **58**:90–105. https://doi.org/10.1093/ilar/ilx007.
81. **Sulzbacher MM, Sulzbacher LM, Passos FR, Bilibio BLE, de Oliveira K, Althaus WF, Frizzo MN, Ludwig MS, Da Cruz IBM, Heck TG.** 2022. Adapted murine sepsis score: Improving the research in experimental sepsis mouse model. BioMed Res Int **2022**:5700853. https://doi.org/10.1155/2022/5700853.
82. **Toth LA.** 2000. Defining the moribund condition as an experimental endpoint for animal research. ILAR J **41**:72–79. https://doi.org/10.1093/ilar.41.2.72.
83. **Trammell RA, Toth LA.** 2011. Markers for predicting death as an outcome for mice used in infectious disease research. Comp Med **61**:492–498.
84. **Turner PV, Pang DS, Lofgren JL.** 2019. A review of pain assessment methods in laboratory rodents. Comp Med **69**:451–467. https://doi.org/10.30802/AALAS-CM-19-000042.
85. **van der Vinne V, Pothecary CA, Wilcox SL, McKillop LE, Benson LA, Kolpakova J, Tam SKE, Krone LB, Fisk AS, Wilson TS, Yamagata T, Cantley J, Vyazovskiy VV, Peirson SN.** 2020. Continuous and non-invasive thermography of mouse skin accurately describes core body temperature patterns, but not absolute core temperature. Sci Rep **10**:20680. https://doi.org/10.1038/s41598-020-77786-5.
86. **Vlach KD, Boles JW, Stiles BG.** 2000. Telemetric evaluation of body temperature and physical activity as predictors of mortality in a murine model of staphylococcal enterotoxic shock. Comp Med **50**:160–166.
87. **Vyas D, Javadi P, Dipasco PJ, Buchman TG, Hotchkiss RS, Coopersmith CM.** 2005. Early antibiotic administration but not antibody therapy directed against IL-6 improves survival in septic mice predicted to die on basis of high IL-6 levels. Am J Physiol Regul Integr Comp Physiol **289**:R1048–R1053. https://doi.org/10.1152/ajpregu.00312.2005.
88. **Winn CB, Hwang SK, Morin J, Bluette CT, Manickam B, Jiang ZK, Giddabasappa A, Liu CN, Matthews K.** 2021. Automated monitoring of respiratory rate as a novel humane endpoint: A refinement in mouse metastatic lung cancer models. PLoS One **16**:e0257694. https://doi.org/10.1371/journal.pone.0257694.
89. **Yokoo H, Chiba S, Tomita K, Takashina M, Sagara H, Yagisita S, Takano Y, Hattori Y.** 2012. Neurodegenerative evidence in mice brains with cecal ligation and puncture-induced sepsis: Preventive effect of the free radical scavenger edaravone. PLoS One **7**:e51539. https://doi.org/10.1371/journal.pone.0051539.
90. **Zhang M, Montroy J, Sharma R, Fergusson DA, Mendelson AA, Macala KF, Bourque SL, Schlechte JM, Eng MK, McDonald B, Gill SE, Fiest KM, Liaw PC, Fox-Robichaud A, Lalu MM.** 2021. The effects of biological sex on sepsis treatments in animal models: A systematic review and a narrative elaboration on sex- and gender-dependent differences in sepsis. Crit Care Explor **3**:e0433. https://doi.org/10.1097/CCE.0000000000000433.
91. **Zingarelli B, Coopersmith CM, Drechsler S, Efron P, Marshall JC, Moldawer L, Wiersinga WJ, Xiao X, Osuchowski MF, Thiemermann C.** 2019. Part I: Minimum quality threshold in preclinical sepsis studies (MQTiPSS) for study design and humane modeling endpoints. Shock **51**:10–22. https://doi.org/10.1097/SHK.0000000000001243.
92. **Zolfaghari PS, Pinto BB, Dyson A, Singer M.** 2013. The metabolic phenotype of rodent sepsis: Cause for concern? Intensive Care Med Exp **1**:6. https://doi.org/10.1186/2197-425X-1-6.
93. **Zou KH, O'Malley AJ, Mauri L.** 2007. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation **115**:654–657. https://doi.org/10.1161/CIRCULATIONAHA.105.594929.