Title

2024 RECOVER Guidelines: Methods, evidence identification, evaluation, and consensus process for development of treatment recommendations

Journal

Journal of Veterinary Emergency and Critical Care, 34(S1)

ISSN

1534-6935

Authors

Fletcher, Daniel J

Boller, Manuel

Burkitt-Creedon, Jamie M

et al.

Publication Date

2024-07-01

DOI

10.1111/vec.13388

Received: 22 March 2024 | Accepted: 25 April 2024

DOI: 10.1111/vec.13388

SPECIAL ARTICLE

Journal of Veterinary Emergency and Critical Care | WILEY

2024 RECOVER Guidelines: Methods, evidence identification, evaluation, and consensus process for development of treatment recommendations

Daniel J. Fletcher PhD, DVM, DACVECC[1] | Manuel Boller Dr med vet, MTR, DACVECC[2,3] | Jamie M. Burkitt-Creedon DVM, DACVECC[4] | Erik Fausak MSLIS, RVT[5] | Megan G. Van Noord MSIS[5] | Kim Mears MLIS, AHIP[6] | Kate Hopper BVSc, PhD, DACVECC[4] | Steven E. Epstein DVM, DACVECC[4]

[1]Department of Clinical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, New York, USA

[2]VCA Canada Central Victoria Veterinary Hospital, Victoria, British Columbia, Canada

[3]Faculty of Veterinary Medicine, Department of Veterinary Clinical and Diagnostic Sciences, University of Calgary, Calgary, Alberta, Canada

[4]Department of Surgical and Radiological Sciences, School of Veterinary Medicine, University of California, Davis, Davis, California, USA

[5]University Library, University of California, Davis, Davis, California, USA

[6]Robertson Library, University of Prince Edward Island, Charlottetown, Prince Edward Island, Canada

Correspondence
Daniel J. Fletcher, Department of Clinical Sciences, College of Veterinary Medicine, Cornell University, DCS Box 31, 930 Campus Rd., Ithaca, NY 14853, USA.
Email: djf42@cornell.edu

Funding information
Boehringer Ingelheim Animal Health; Zoetis Animal Health

Abstract

Objective: To describe the methodology used by the Reassessment Campaign on Veterinary Resuscitation (RECOVER) to re-evaluate the scientific evidence relevant to CPR in small and large animals, to newborn resuscitation, and to first aid and to formulate the respective consensus-based clinical guidelines.

Design: This report describes the evidence-to-guidelines process employed by RECOVER that is based on the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach and includes Information Specialist-driven systematic literature search, evidence evaluation conducted by more than 200 veterinary professionals, and provision of clinical guidelines in the domains of Preparedness and Prevention, Basic Life Support, Advanced Life Support, Post-cardiac Arrest Care, Newborn Resuscitation, First Aid, and Large Animal CPR.

Setting: Transdisciplinary, international collaboration in academia, referral practice, and general practice.

Results: For this update to the RECOVER 2012 CPR guidelines, we answered 135 Population, Intervention, Comparator, and Outcome (PICO) questions with the help of a

Abbreviations: ALS, Advanced Life Support; BLS, Basic Life Support; CI, confidence interval; CT, clinical trial; DC, Domain Chair; EE, Evidence Evaluator; EEST, Evidence Evaluation Summary Table; EPW, Evidence Profile Worksheet; ERT, Evidence Review Table; GRADE, Grading of Recommendations Assessment, Development, and Evaluation; ILCOR, International Liaison Committee on Resuscitation; IS, Information Specialist; PICO, Population, Intervention, Comparator, and Outcome; RECOVER, Reassessment Campaign on Veterinary Resuscitation; RoB, risk of bias; ROSC, return of spontaneous circulation; RR, relative risk.

Daniel J. Fletcher, Manuel Boller, and Jamie M. Burkitt-Creedon contributed equally to this work.

team of Domain Chairs, Information Specialists, and more than 200 Evidence Evaluators. Most primary contributors were veterinary specialists or veterinary technician specialists. The RECOVER 2024 Guidelines represent the first veterinary application of the GRADE approach to clinical guideline development. We employed an iterative process that follows a predefined sequence of steps designed to reduce bias of Evidence Evaluators and to increase the repeatability of the quality of evidence assessments and ultimately the treatment recommendations. The process also allowed numerous important knowledge gaps to emerge that form the foundation for prioritizing research efforts in veterinary resuscitation science.

Conclusions: Large collaborative, volunteer-based development of evidence- and consensus-based clinical guidelines is challenging and complex but feasible. The experience gained will help refine the process for future veterinary guidelines initiatives.

KEYWORDS

cardiopulmonary resuscitation, clinical trials, consensus guidelines, critical care, development and evaluation, evidence-based medicine, GRADE, grading of recommendations assessment

1 | INTRODUCTION

In 2012, the Reassessment Campaign on Veterinary Resuscitation (RECOVER) initiative, a collaboration between the American College of Veterinary Emergency and Critical Care and the Veterinary Emergency and Critical Care Society (VECCS), released the first evidence-based consensus guidelines on veterinary CPR (2012 RECOVER CPR Guidelines).[1] These guidelines were developed using an approach similar to that used by the International Liaison Committee on Resuscitation (ILCOR), an international collaboration that evaluates the evidence on human CPR and generates a consensus on science based on the current resuscitation literature.[2] In the years since the publication of the 2012 RECOVER CPR Guidelines, a more robust approach to evidence-based clinical guidelines development has been adopted by ILCOR and many other medical guidelines organizations: the Grading of Recommendations Assessment, Development, and Evaluation (GRADE).[3] The RECOVER 2024 Guidelines represent the first veterinary application of this new approach to clinical guideline development. For this process, we developed a multistep, iterative process that starts with asking relevant clinical questions and ends with the provision of clinical treatment recommendations for Prevention and Preparedness, Basic Life Support (BLS), Advanced Life Support (ALS), Monitoring, Post-cardiac Arrest Care, First Aid, Newborn Resuscitation, and Large Animal Resuscitation. An overview of this process is presented in Figure 1.

2 | EVIDENCE EVALUATION PROCESS

2.1 | Organizational structure

The RECOVER 2024 evidence evaluation process was overseen by 3 Co-Chairs, who are veterinary emergency and critical care specialists. A series of 8 domains of interest were identified by the Co-Chairs: Prevention and Preparedness, BLS, ALS, Monitoring, Post-cardiac Arrest Care, First Aid, Newborn Resuscitation, and Large Animal Resuscitation. For each domain, 2–3 Domain Chairs (DCs) were appointed to lead the domain. In addition, each domain was assigned 1–2 Information Specialists (ISs) who worked with the DCs to complete all literature searches relevant to the questions of interest within that domain. Individual questions of interest in each domain were then assigned to 1–2 Evidence Evaluators (EEs), who read each identified relevant paper. They answered a series of standard questions to evaluate the quality of each paper and the certainty of its findings as related to the question of interest. Project milestone tracking was managed using a commercial web-based relational database system.[a] The database contained tables tracking information about each domain's questions, the EEs, and each EE's question assignment, including tracking details about each stage of the evidence evaluation process such as due dates and dates of completion. Communication with participants, including reminder emails and progress tracking, was semiautomated using commercial web-based automation software.[b]

2.2 | Selecting the PICO questions

All questions were written in the standardized PICO format: (P) patient population, (I) intervention, (C) comparison intervention, and (O) outcome(s) of interest.[4] For each domain, questions from the 2012 RECOVER CPR Guidelines were evaluated along with questions from the ILCOR PICO question list relevant to the domain.[c] Finally, additional veterinary-specific PICO questions were generated by the DCs and the Co-Chairs. The DCs and Co-Chairs ranked all candidate PICO questions using a 3-tier system—(1) critical question, (2) high-priority question, and (3) lower priority question—and evidence evaluation resources were allocated to prioritize tier 1 and tier 2 questions.

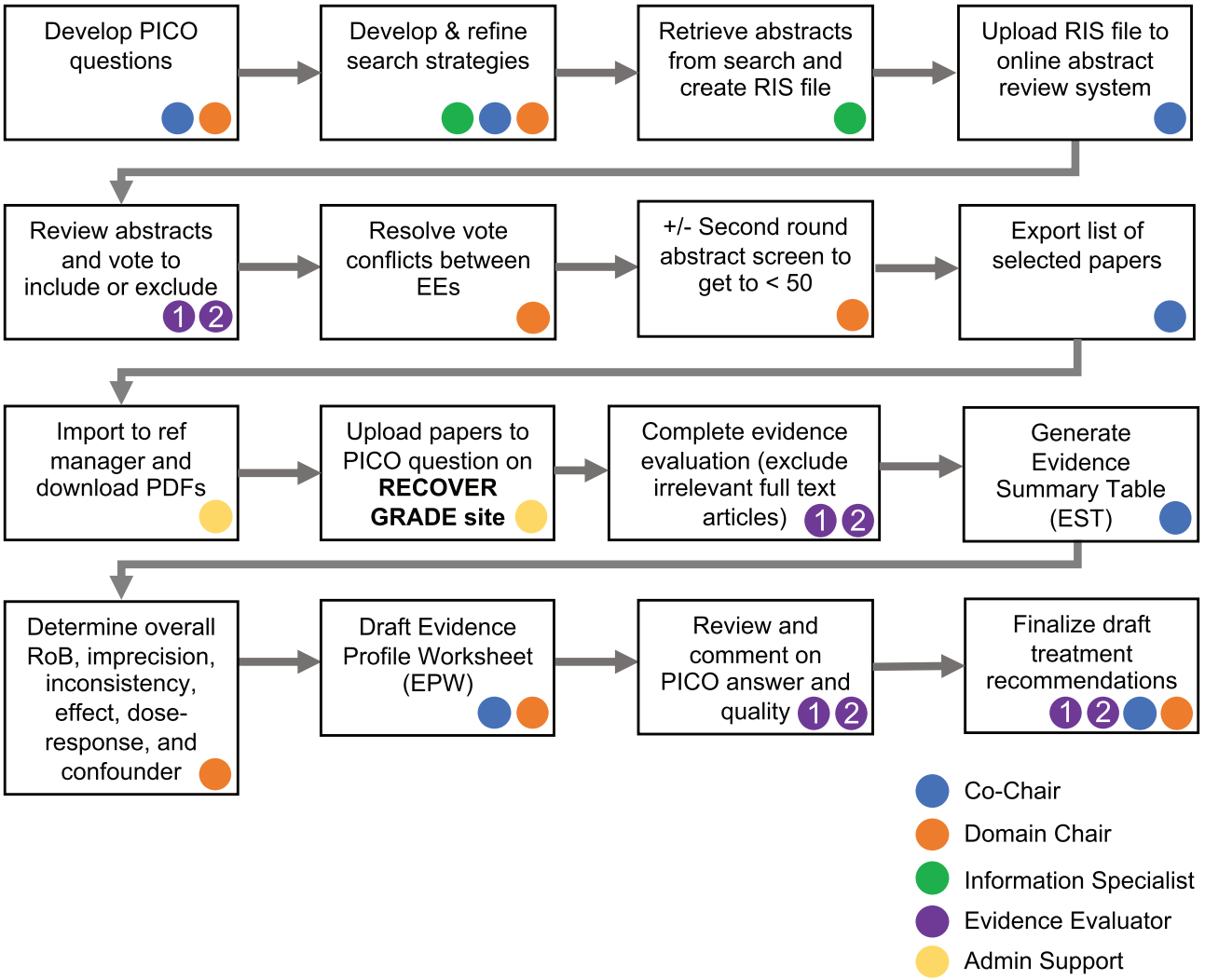

FIGURE 1 Evidence Evaluation process overview for the 2024 RECOVER CPR Guidelines. EE, Evidence Evaluator; EPW, Evidence Profile Worksheet; EST, evidence summary table; GRADE, Grading of Recommendations Assessment, Development, and Evaluation; PICO, Population, Intervention, Comparator, and Outcome; RECOVER, Reassessment Campaign on Veterinary Resuscitation; RoB, risk of bias.

DCs also generated a list of potential outcomes for each PICO question and rated these outcomes according to clinical relevance. For example, for many CPR-related PICO questions, the outcomes of improved surrogate markers of perfusion, return of spontaneous circulation (ROSC), survival to hospital discharge, and survival with favorable neurologic outcome were identified as clinically relevant. However, for most of these questions, the order of priority assigned was (1) survival with good neurologic outcome, then (2) survival to discharge, with ROSC and surrogate markers of perfusion assigned lower priority in terms of clinical relevance.

2.3 | Searching the literature

IS involvement in systematic reviews is associated with higher quality reporting of search strategies and lower risk of selection bias.[5,6] For this reason, a team of ISs with experience in conducting systematic reviews and searches in veterinary medicine and agricultural science was recruited to develop, document, and execute database-specific search strategies. ISs were previously used for developing ILCOR search strategies for human CPR guidelines, but inclusion of IS teams in the RECOVER guidelines process began with RECOVER 2024.[7]

The Co-Chairs appointed a lead IS (EF) to communicate with all ISs and develop search and documentation workflows (in collaboration with other ISs). While ideas and strategies were adapted from ILCOR and early versions of the literature search extension of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA-S), many unique veterinary challenges made this workflow novel.[7,8] At least 2 ISs were assigned to each of the 8 subject domains to work with veterinary content experts (the DCs). Library school graduate students and floating ISs also supported domains as needed. DCs provided PICOs, ranging from 10 to 30 per domain, to be searched by IS teams. Anticipating that these would become living reviews, searches were conducted between 2019 and 2022.

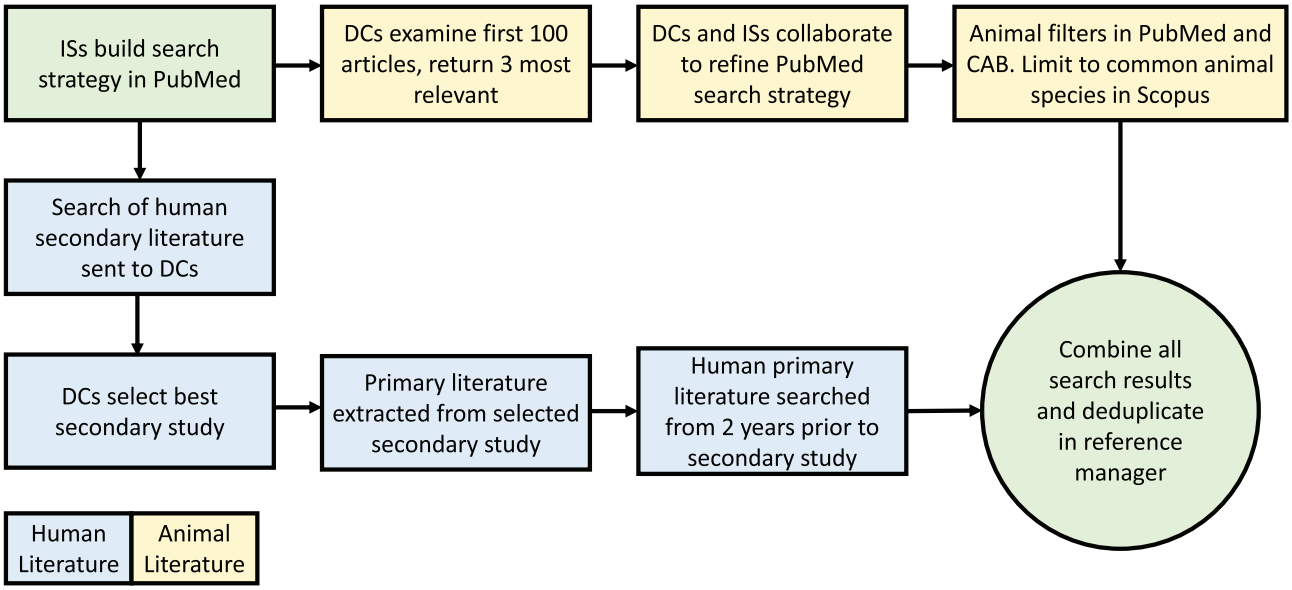

FIGURE 2 An overview of the literature search process used for each PICO question in the RECOVER 2024 Guidelines. CAB, Centre for Agriculture and Biosciences; OSF, Open Science Framework; PICO, Population, Intervention, Comparator, and Outcome; DC, Domain Chair; IS, Information Specialist.

For most PICOs, a 2-pronged approach was used to find both human-based studies relevant to the question and animal-based studies. First, for human-based studies, an initial search of secondary literature (eg, existing guidelines, systematic reviews) was conducted in Medline (PubMed). DCs reviewed the results of these searches and identified relevant records. These records were used to collect primary human studies as follows: primary literature included in the secondary studies was collected, and a second search of human-only literature from 2 years prior to the publication of the most recent secondary study was screened for relevant studies that were more recent. Second, for animal-based studies, a search string of common clinical veterinary and research animal species was developed to limit searches to primary, animal-only literature. This search string was used with PICO searches developed collaboratively with DCs to search 3 databases: Medline via PubMed, Centre for Agriculture and Biosciences (CAB) Abstracts via CAB Direct, and Scopus. Filters were applied to the primary literature searches to exclude non-English language publications, case reports, reviews, meta-analyses, letters to the editor, and other nonanalytical literature. The results of these searches were deduplicated across databases using Zotero[d] and then screened by the DCs. For some PICOs, the above process identified little to no human or animal literature. For these PICOs, searches were run unconstrained for date or species, and all results from the 3 databases were screened. The workflow for the process is outlined in Figure 2.

Search strategies and results are documented in detailed worksheets, and peer review of search strategies occurred using modified Peer Review of Electronic Search Strategy Guidelines and informal meetings.[9] All search worksheets, which include exact search dates, search strings, and the numbers of search results, and peer review documents are available in Open Science Framework.[e]

2.4 | Recruiting and training EEs

EEs were recruited using email announcements to membership lists from the following organizations: the Veterinary Emergency and Critical Care Society, the American College of the Veterinary Emergency and Critical Care, the European Veterinary Emergency and Critical Care Society, the European College of Veterinary Emergency and Critical Care, the Latin American Veterinary Emergency and Critical Care Society, the Academy of Veterinary Emergency & Critical Care Technicians and Nurses, the American College of Veterinary Anesthesia and Analgesia, and the American College of Veterinary Internal Medicine. Additional EEs were recruited via social media posts and through word of mouth. Minimal requirements for participation included a veterinary doctoral or veterinary technology degree, but most veterinarians and veterinary technicians were board-certified veterinary specialists or veterinary technician specialists.

Once recruited, EEs received training via an online training system, which included video recordings of presentations on the evidence evaluation approach, the logistics of the evidence evaluation process, and links to the evidence evaluation manual.[f]

2.5 | Evidence evaluation

The evidence evaluation process was based on the GRADE approach to evidence-based clinical guideline development, first described in 2000 and currently in use by over 110 organizations in 19 countries around the world, including ILCOR.[3] This robust approach provides a structured framework for critically evaluating the literature relevant to a specific PICO question. The steps in the evidence evaluation process are summarized in Figure 1.

Both EEs assigned to a PICO question completed an initial screening of the abstracts for all studies identified in the literature searches provided by the ISs. These abstracts were uploaded to an online study screening tool, and each EE voted independently whether to consider the study in the full evidence evaluation[g]. Once both EEs voted, 1 of the DCs for the domain reviewed and resolved any conflicts between the EEs. For most PICO questions, this yielded a manageable number of candidate studies for evidence evaluation. However, when more than 50 relevant abstracts were identified, the DC completed a secondary screening to remove studies that appeared to bear the least relevance to the PICO question. The remaining articles were considered candidate studies for evidence evaluation and were loaded into a web-based system for EE review.

Each EE performed a full text screening of all candidate studies. Studies that were not primary literature (eg, systematic reviews, meta-analyses) were rejected. For PICO questions for which a large amount of literature was available, EEs rejected studies with substantial indirectness to the PICO question. All studies that remained after this full text screening were included for full evidence evaluation. It should be noted that this sometimes led to the 2 EEs assigned to a single PICO question rejecting different studies; a study underwent evidence evaluation if a single EE included it.

The GRADE process for each candidate study consisted of (1) determining which of the outcomes of interest were directly addressed in the study, (2) identifying the study type, (3) answering questions focused on assessing the risk of bias (RoB) in the study, (4) evaluating the degree of indirectness in the study related to the PICO question, (5) determining the degree of imprecision in the study results, (6) identifying any aspects of the study that might increase the relative strength of the study results, and (7) writing a brief statement describing the treatment recommendation(s) related to the PICO question that could be made based on the results of the study. The EEs repeated this process for every study they included for evaluation; details of the evidence evaluation process follow below. All 7 steps of this process and the initial full text screening were facilitated using an online evidence evaluation system developed specifically for the RECOVER 2024 Guidelines process.

2.5.1 | Determining the outcome(s) addressed in the study

EEs were asked to identify which of their PICO question's assigned outcomes were addressed in the study being evaluated. All outcomes deemed relevant to the PICO question were listed, and the EEs were asked to select which were addressed either directly or indirectly in the study being evaluated.

2.5.2 | Identifying the study type

Only primary literature was evaluated, and each study was assigned by the EE to 1 of 3 types: clinical trial (CT), experimental animal study, or observational study. This was noted for each study in the online evidence evaluation system by each EE. Subsequent questions asked by the system depended on study type as identified by the EE.

CTs carry the highest level of evidence in the GRADE system. These are prospective studies that include recruitment of clinical patients (humans or animals) and in which the investigator intervenes to prevent or treat a condition or disease. The intervention in the study should coincide with the "I" in the PICO question for which the study is being evaluated. In the majority of CTs, the effect of the intervention in the study group is compared to a control group (the "C" in the PICO question) that does not receive the intervention. Ideally, study subjects are allocated to these groups in a random fashion; such studies are randomized, controlled CTs. Proper randomization increases the quality of the evidence, and absent or poor/inadequate randomization decreases it. CTs were assigned an initial high quality of evidence.

Experimental animal studies involve the use of experimental or laboratory animals. They are different than CTs in animals because in experimental studies, disease states are induced, while in CTs, disease states occur naturally. Because the diseases and conditions are induced, the evidence generated by these types of studies is considered of lower quality than that generated by CTs. In the RECOVER GRADE process, experimental animal studies were assigned an initial moderate quality of evidence.

Observational studies include recruitment of clinical subjects after an intervention or exposure in which the investigator does not decide whether the patient received the exposure/intervention. Group assignment is determined outside of the investigator's control by means such as chance, client choice, or clinician decision. A cause-and-effect relationship between intervention/exposure and outcome cannot be determined in observational studies. These types of studies generally provide the lowest level of evidence for clinical guidelines. Since observed patients are assigned to groups after an intervention/exposure has occurred, the decisions that led to the intervention are made based on clinical judgment rather than a predetermined experimental approach. There are 4 main types of observational studies: cohort studies, cross-sectional studies, case–control studies, and case reports/case series. In the RECOVER GRADE process, observational studies were assigned an initial low quality of evidence.

2.5.3 | Assessing the study for RoB

RoB assessment is a structured way to assess study limitations. In the GRADE process, the RoB assessment is conducted for each study. Bias introduces systematic error into research studies and can occur during study design, data collection, or analysis of the results of the study. The RoBs to be addressed vary with study design, and the RECOVER online GRADE system asked the EEs only the questions relevant to the type of study being assessed. For each question, the EE could choose "yes," "no," or "unsure" regarding the type of bias. If the answer was "yes" or "unsure," the EE then entered a brief comment into the system describing the bias or the reason they could not determine whether bias was present, which occurred most commonly because

TABLE 1 Risk of bias questions. The online Grading of Recommendations Assessment, Development, and Evaluation ("GRADE") system asked Evidence Evaluators questions relevant to the type of study being assessed during the evidence evaluation process for 2024 RECOVER CPR Guidelines development.

Study type	Bias type	Question(s)
Clinical trial	Selection bias: generation	Was the method used to generate the allocation sequence described in sufficient detail to allow an assessment of whether it should produce comparable groups?
	Selection bias: concealment	Was the method used to conceal the allocation sequence described in sufficient detail to determine whether intervention allocations could have been foreseen in advance of, or during, enrollment?
	Performance bias: participants/pet owners	Were measures used to blind study participants/pet owners and personnel from knowledge of which intervention a participant received? Was the intended blinding effective?
	Detection bias: assessors	Were measures used to blind outcome assessors from knowledge of which intervention a participant received? Was the intended blinding effective?
	Attrition bias	Were the outcome data complete for each main outcome, including attrition and exclusions from the analysis?
	Reporting bias	Did the study report appropriate outcomes (ie, to avoid selective outcome reporting)?
	Other bias	Was the study otherwise free of important sources of bias not already reported previously?
Observational study	Selection bias: enrollment	Were appropriate eligibility criteria developed and applied to both the cohort of interest and the comparison cohort?
	Selection bias: confounding	Was confounding adequately controlled for?
	Detection bias: exposure and outcome	Was measurement of exposure and outcome appropriate and consistently applied to both the cohort of interest and the comparison cohort?
	Attrition bias	Was follow-up complete?
	Other bias	Was the study otherwise free of important sources of bias not already reported previously?
Experimental study	Selection bias: generation	Was the allocation sequence adequately generated and applied?
	Selection bias: baseline characteristics	Were the groups similar at baseline or were they adjusted for confounders in the analysis?
	Selection bias: concealment	Was the allocation adequately concealed?
	Performance bias: random housing	Were the animals randomly housed during the experiment?
	Performance bias: blinding	Were the caregivers and/or investigators blinded from knowledge which intervention each animal received during the experiment?
	Detection bias: random outcome assessment	Were animals selected at random for outcome assessment?
	Detection bias: assessor blinding	Was the outcome assessor blinded?
	Attrition bias	Were incomplete outcome data adequately addressed?
	Reporting bias	Are reports of the study free of selective outcome reporting?
	Other bias	Was the study otherwise free of important sources of bias not already reported previously?

the methods or results did not explain the approach in enough detail to be sure. A list of each of the RoB questions assessed for each type of study appears in Table 1.

The GRADE manual provided to all EEs contained detailed definitions of each of the RoB questions as well as a prompt written in the form of a question for the EE to answer about the presence of the bias, a series of signaling questions to help convey the range of ways the type of bias could be present, and specific examples of the type of bias. EEs were instructed to reach out to the DCs for their domain if they had questions about a specific study.

RoB questions for CTs and observational studies have been standardized through the GRADE process and were used unaltered in the RECOVER 2024 Guidelines development process.[10] The GRADE process does not include standardized RoB questions for experimental

TABLE 2 Indirectness questions. Signaling questions asked of Evidence Evaluators during the evidence evaluation process for 2024 RECOVER CPR Guidelines development.

PICO question component	Signaling question
"P," Patient population	Are there clinically relevant differences between the experimental population described in the "P" portion of the PICO question and the experimental population in this paper? If so, please answer "yes" and explain in the comments box.
"I," Intervention	Are there clinically relevant differences between the intervention described in the "I" portion of the PICO question and the intervention studied in this paper? If so, please answer "yes" and explain in the comments box.
"C," Comparator	Are there clinically relevant differences between the comparison described in the "C" portion of the PICO question and the comparison or control intervention in this paper? If so, please answer "yes" and explain in the comments box.
"O," Outcome	Is there any clinically relevant difference between the outcome(s) evaluated in this paper and this specific outcome from the "O" portion of the PICO question? If there is more than 1 outcome evaluated in the paper, answer this question regarding the outcome most similar to this "O" from the PICO question. If there is a difference, please answer "yes" and explain in the comments box.

Note: Separate indirectness questions were targeted at each of the 4 parts of the PICO question and assessed for every study.
Abbreviation: PICO, Population, Intervention, Comparator, and Outcome.

animal studies, but an adaptation of the GRADE RoB questions has been developed and validated by the Systematic Review Centre for Laboratory animal Experimentation (SYRCLE), and these questions were used in the RECOVER 2024 Guidelines process.[11]

2.5.4 | Evaluating the study for indirectness

Directness of a study refers to the degree to which study design elements align with each part of the PICO question. For example, studies in people relevant to a RECOVER 2024 PICO question are inherently indirect in the patient population (P) being studied, and the strength of the evidence provided by such a study would be reduced due to this indirectness.[12] Indirectness in the comparison (C) portion of the PICO question may occur when studies do not directly compare the intervention (I) of interest to the specific comparator described in the PICO question. While the evidence provided by such a study is not entirely direct, it may still inform an answer to the PICO question, but with less certainty.

Separate indirectness questions were targeted at each of the 4 parts of the PICO question and assessed for every study. Each question asked whether there were clinically relevant differences between the aspect of the PICO question being queried in the study and that aspect of the RECOVER PICO question itself. The P, I, and C questions were asked only once per study, and the GRADE system then looped through each individual outcome identified by the EE and asked an indirectness question for each of these outcomes separately. In cases where the EE believed that indirectness was present, they were asked to enter a brief description of the nature of the indirectness into the system. The signaling question asked for each PICO question element is listed in Table 2.

2.5.5 | Evaluating the study for imprecision

Any study in which a small number of subjects is included or a small number of events occurred will have effect estimates with wide 95% confidence intervals (CIs). The results from these studies are therefore considered to be imprecise, which is interpreted as lower quality of evidence. In general, the GRADE approach recommends that studies be downgraded for imprecision when the clinical interpretation would differ if the actual value fell at the higher end of the 95% CI than at the lower end of the 95% CI.[13] The online system looped through all outcomes and asked the imprecision questions for each outcome for which the EE believed relevant data were available in the study. Different criteria to identify imprecision were recommended to EEs for dichotomous outcomes (eg, ROSC) than for continuous outcomes (eg, coronary perfusion pressure).

For dichotomous outcomes, EEs were asked to enter the total number of subjects in each group as well as the number of subjects in each group with the outcome of interest. The relative risk (RR) and CI were calculated by the system from the data entered. Based on the RR and CI, EEs were asked if they had concern that there was imprecision in the result for the outcome based on the CI. If it was very wide, and especially if it contained or came close to including 1.0, EEs were instructed to note that imprecision existed for the outcome.

For continuous and other nondichotomous outcomes, EEs were instructed to consider the statistical results (eg, means, medians, proportions) and decide (1) if the differences between the groups were of sufficient magnitude and (2) if the degree of variability within each group (eg, standard deviation, range) was small enough for a clinically relevant difference between the groups. If the differences reached statistical significance, and the magnitude of effect was large enough and the degree of variability was small enough to be considered clinically significant, EEs were instructed to note no evidence of imprecision.

2.5.6 | Upgrading study results

The final step in the evidence evaluation process for each study was to identify any factors that might lead to upgrading the quality of evidence assigned to an individual study for a specific outcome. Three factors

were considered: (1) a large magnitude of effect, (2) a dose–response effect, and (3) confounders that would likely decrease the chance of identifying an effect (and yet an effect was still found).[14] The GRADE system looped through all outcomes identified by the EE for the study and asked questions regarding the 3 factors for each outcome.

For dichotomous outcomes, variables for which the RR was very large (>2.0) would indicate a large magnitude of effect. For continuous variables, a large and clinically relevant absolute difference in the variable between groups would similarly indicate a large magnitude of effect. This was considered independent of the CIs or standard deviations/ranges of the point estimates of the effects since this was examined by the imprecision metrics. When a large magnitude of effect was noted, the quality of evidence for this outcome in this study was upgraded.

When multiple doses of an intervention were evaluated in a study and a clear dose–response effect was noted in the statistical analysis, the quality of evidence was upgraded because a clear dose–response relationship suggests an effect of the intervention that is independent of the actual effect size.

Finally, the EE was asked to look for any obvious confounders in the study that would have reduced the likelihood of finding an effect of the intervention. In some studies, the characteristics of the intervention or comparator group would likely result in an underestimate of an apparent treatment effect. If, for instance, only sicker patients receive an experimental intervention or exposure, yet they still fared better, it is likely that the actual intervention or exposure effect was even larger than the data suggested. In these cases, the quality of evidence for the study was upgraded for the outcome.

2.5.7 | Treatment recommendation summary

Following evidence evaluation for each study, the EE wrote a brief summary of the treatment recommendation for the PICO question that could be derived from the study being evaluated for each of the outcomes addressed in the study. The EE entered a brief description of the rationale for the treatment recommendation (or lack of recommendation) based on data from the study for each outcome.

2.6 | Evidence Evaluation Summary Tables

Once an EE completed the evidence evaluation process for the relevant outcomes in all relevant studies, the RECOVER GRADE system generated a spreadsheet with an Evidence Evaluation Summary Table (EEST) for each outcome. The EEST was a condensed view of all of the data collected during the evidence evaluation process, with 1 row for each study evaluated, grouped by the study type. Columns were generated for each of the RoB questions, the raw data entered by the EE for dichotomous outcomes, and the EE's interpretation of the degree of indirectness for each of the 4 aspects of the PICO question as well as imprecision and factors that might lead to an upgrade in the quality of evidence. Individual cells in each column were color-coded to reflect the seriousness of any issue that might cause the study to be downgraded in quality, with red cells denoting the presence of an issue, yellow cells denoting that the EE could not determine from the manuscript if the issue was present, and green cells indicating that the issue was not likely present. Hovering over each red or yellow cell would reveal the comment the EE had entered into the system describing the nature and severity of the issue. This allowed the EE to develop a subjective overview of the severity of issues in each category across the various studies.

The EEs used this information and any additional notes they collected during the evidence evaluation process to write a 3-part structured summary of the answer to the PICO question based on the evidence in all the studies they had read. The summary contained (1) a consensus on science, (2) treatment recommendations, and (3) knowledge gaps.

In the consensus on science, the EEs summarized the body of evidence that addressed the PICO across all outcomes considered, prioritizing the evidence for the most critical outcomes. They were instructed to reference specific studies they believed most strongly supported their conclusion and to provide a brief summary of any data from those studies that helped them reach the conclusion. The generic format for the statement was "For the outcome of Z (eg, survival to hospital discharge), we have identified X# type studies in species 1 that showed benefit. In addition, we found X# type studies in species 2 enrolling # patients showing no benefit." EEs were next instructed to provide a statement on their treatment recommendation based on evaluation of the evidence, and their assessment of the strength of the evidence (strong, weak, none). Finally, the EEs were asked to provide a statement on important gaps in knowledge that hindered their ability to make a more confident treatment recommendation.

Once both EEs assigned to a PICO question had completed this task, a consolidated EEST was generated containing data collected by both EEs along with their structured summaries.

3 | CLINICAL GUIDELINE DEVELOPMENT

3.1 | Evidence Profile Worksheets

The consolidated EESTs were sent back to the DCs and the Co-Chairs. One member of that group was assigned primary responsibility (the main author) for drafting a standardized Evidence Profile Worksheet (EPW) for each PICO question. This format was based on the GRADE Evidence Profile, which is a structured and detailed assessment of the quality of evidence for each outcome of a PICO question.[15] The EPW contained the PICO question, a prioritized list of outcomes, the names of the DCs, EEs, and Co-Chairs who participated in the evidence evaluation for the PICO question, and an accounting of any conflicts of interest for any of the participants. The main author completed a standardized Evidence Review Table (ERT) for the PICO question, which summarized for each outcome the quality factors for the various types of studies that informed the ultimate treatment recommendation for the PICO question.

The main author completed a 5-step process to generate the treatment recommendation(s) associated with the PICO question: (1) prioritize outcomes for the PICO, (2) review the structured summaries written by the EEs, (3) review the consolidated EEST with input from both EEs, (4) assign an initial quality of evidence based on the data available for the highest priority outcome for which evidence was available and then adjust that metric based on the evidence quality data collected by the EEs, and finally (5) formulate treatment recommendation(s) for or against the intervention and assign a strength of the recommendation and complete a structured summary for the PICO question.

3.1.1 | Outcome measure prioritization

Each PICO question was initially developed by the DCs and Co-Chairs with up to 7 specific outcome measures. Before examining the evidence collected by the EEs, the EPW main author evaluated all 7 of these outcomes and developed a rank order list of the outcomes from most critical to least critical.

3.1.2 | Structured summary reviews

The main author reviewed the structured summaries submitted by the EEs to get an overview of the data extracted and the conclusions drawn by the EEs. Discrepancies between EEs were noted, and the extracted data from the evidence evaluation process were examined to resolve them.

3.1.3 | Review of the consolidated EEST

The EEST was examined by the main author, starting with the most critical outcome identified. If no evidence was available for the highest priority outcome, the main author sequentially moved down the list of outcomes until an outcome with substantial evidence was found, and the highest quality evidence for that outcome was reviewed. This review included evaluation of the treatment recommendation summaries for these studies written by the EEs as well as review of the primary manuscripts if needed. Links to each manuscript were included in the EEST.

If a compelling, well-supported treatment recommendation could be written for a PICO question based on a specific higher priority outcome, the main author did a cursory review of the data for lower priority outcomes to determine whether any data were present that might temper the treatment recommendation. The result of this step was an initial treatment recommendation for or against the intervention being examined in the PICO question.

3.1.4 | Assigning an overall quality of evidence to the treatment recommendation

Once the highest priority outcome with evidence that informed a treatment recommendation was identified, the main author filled in an ERT for this highest priority outcome on the EPW. The ERT summarized the studies available to support the treatment recommendation for each outcome and the positive and negative quality metrics identified by the EEs. The main author then assigned an initial overall quality of evidence to the treatment recommendation, which was based on the highest quality evidence found for the most critical outcome for which evidence was available. From this starting point, the overall evidence quality was adjusted up or down using the GRADE metrics extracted by the EEs described above. An example ERT for PICO question BLS-07 is shown in Figure 3.

Figure 4 shows an example in which the best evidence comes from randomized CTs (which start as "high" quality evidence) and is downgraded for RoB, indirectness, and imprecision, and then upgraded for large treatment effect to ultimately yield an overall "low" quality of evidence rating.

3.1.5 | Complete first draft of EPW

The primary author used the EEST and the summaries written by the EEs for the PICO question to construct a 5-part structured summary for the PICO question: (1) introduction, (2) consensus on science, (3) treatment recommendation(s), (4) justification for the treatment recommendation(s), and (5) a list of high-priority knowledge gaps.

The EPW introduction explained the rationale for investigating the PICO question and a description of the importance of the question in veterinary CPR. The consensus on science summarized the studies used to formulate the treatment recommendation(s) and justified the ultimate quality of evidence assigned. The treatment recommendation(s) that were generated from the evidence evaluation were enumerated next. A brief justification for each treatment recommendation, explaining the rationale, the quality of evidence, and the ultimate strength of the recommendation, was provided. The strength of the recommendation was based on both the quality of evidence and the perceived risk:benefit ratio of the intervention. In some cases, despite a low quality of evidence (or no evidence), strong treatment recommendations were made because of the importance of an intervention and the perceived favorable risk:benefit ratio. The main author explained the rationale for the recommendation and the strength of the recommendation explicitly in the justification section. Finally, the main author listed important knowledge gaps that remained after the evidence evaluation for the PICO question to provide a roadmap for high-priority future research.

Once all 5 sections of the EPW were completed by the main author, they were reviewed by the writing group, which consisted of the DCs and Co-Chairs. Final edits based on that process were made, and the first draft of the EPW was finalized.

3.2 | The CPR algorithm

Once EPWs for all PICO questions were completed, the Co-Chairs and DCs developed a draft CPR algorithm summarizing the important aspects of small animal CPR. BLS, ALS, and Monitoring priorities were

Study Design		Reduced Quality Factors 0 = not serious, - = serious, - - = very serious				Positive Quality Factors 0 = none, + = one, ++ = multiple			Overall Quality High, moderate, low, very low, none
No of studies	Study Type	RoB	Indirectness	Imprecision	Inconsistency	Large Effect	Dose-Response	Confounder	
Outcome: Favorable neurologic outcome									
2	EXP	-	-	-	0	0	0	+	Very low
3	OBS	- -	-	0	0	0	0	0	Very low
Outcome: Survival to discharge									
4	OBS	- -	- -	-	0	0	0	0	Very low
Outcome: ROSC									
3	EX	-	0	-	0	0	0	0	Very low
5	OBS	- -	-	-	0	0	0	0	Very low

FIGURE 3 Evidence Review Table. Each Evidence Review Table summarized the studies available to support the treatment recommendation for each outcome and the positive and negative quality metrics identified by the Evidence Evaluators.

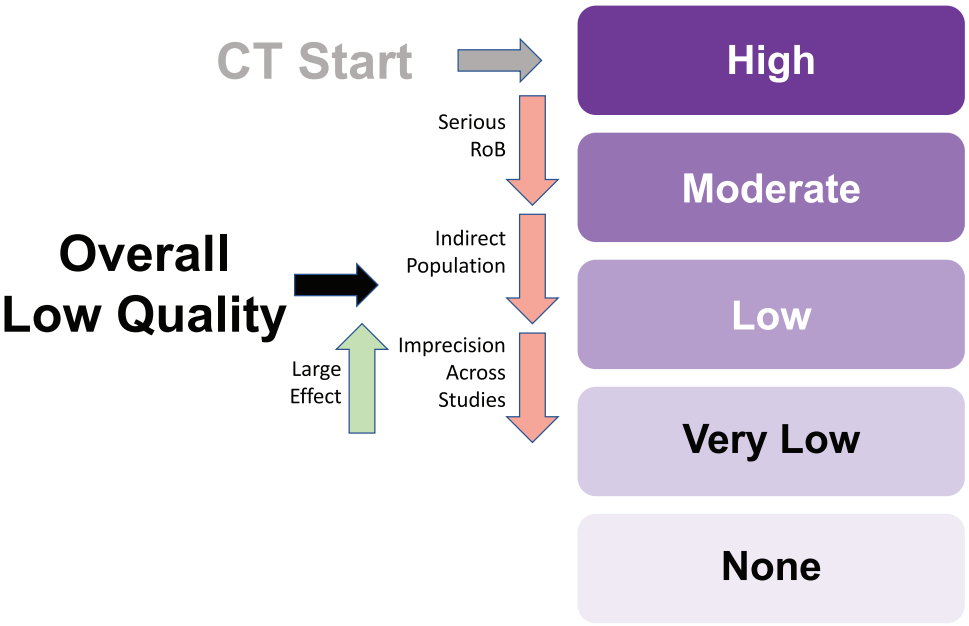

FIGURE 4 Determination of the overall quality of evidence for clinical trials (CTs) for a given outcome. In this example, the initial quality of evidence for a CT started as high, but then was downgraded 3 times: for serious risk of bias (RoB), serious indirectness for deviation in the population, and serious imprecision. Due to the observed large effect on the outcome under scrutiny, the quality of evidence was upgraded by 1 step, resulting in an overall low quality of evidence.

included in the algorithm with the intention that it could act as a cognitive aid in the clinical setting to help rescuers adhere to the clinical guidelines as specified in the treatment recommendations.

3.3 | The consensus process

Once all EPWs and the CPR algorithm were completed, they were posted to an online commenting system for 30 days.[h] Comments from the veterinary profession were solicited via email to the same group email lists used to recruit EEs as well as via social media posts and the RECOVER website. The commenting software tracked all input, and the consolidated comments were screened by the Co-Chairs and DCs, who then edited the EPWs and the CPR algorithm where appropriate.

4 | RESULTS

Seven small animal domains and a total of 135 PICO questions were evaluated. The 5 core domains from RECOVER 2012 were updated with this new process and included BLS (20 PICO questions), ALS (17 PICO questions), Monitoring (13 PICO questions), Prevention and Preparedness (16 PICO questions), and Postcardiac Arrest Care (27 PICO questions). In addition, 2 new small animal domains were investigated: First Aid (14 PICO questions) and Newborn Resuscitation (28 PICO questions).

A total of 227 EEs participated in the RECOVER 2024 process. Of the 135 PICO questions for which evidence evaluation was completed, 14 were evaluated by only 1 EE rather than the planned 2 EEs (Monitoring: 1/13, Postcardiac Arrest Care: 12/27, Prevention and Preparedness: 1/16).

A team of 11 ISs from 6 veterinary colleges in the United States, Canada, and the United Kingdom managed all of the literature searches in collaboration with the DCs.

5 | DISCUSSION

The RECOVER 2024 Guidelines are the first veterinary clinical guidelines developed using the GRADE approach to the authors' knowledge. GRADE is used widely in human clinical guideline processes and is a more robust and reproducible methodology for guideline creation than the process used for RECOVER 2012.[2,3,16,17] In addition, while RECOVER 2012 relied upon EEs to develop the literature searches for their PICO questions, RECOVER 2024 included a team of ISs with expertise in robust literature searches, likely improving the quality and completeness of the searches for each PICO question. Most PICO questions were evaluated by 2 EES, compared to 1 EE in RECOVER 2012.

Although experimental animal studies have not been traditionally included in GRADE analyses for human clinical guidelines, the relevance of these studies, especially those in the target species, was deemed significant and we therefore included them in the RECOVER 2024 evidence evaluation process. In order to maintain the objectivity of evidence evaluation, we used the RoB questions developed by the SYstematic Review Centre for Laboratory animal Experimentation (SYRCLE) to evaluate RoB in the included experimental studies.[11] Experimental animal studies were considered to start at a moderate quality of evidence, acknowledging that although studies in the target species are desirable, experimental animal studies in which disease processes are induced provide weaker evidence than those in animals with naturally occurring disease. Importantly, veterinary CTs and observational studies were categorized as such in the RECOVER 2024 evidence evaluation process, and the experimental animal category was used only for laboratory-based studies in experimental animals.

There were several important limitations of this first use of the GRADE approach in veterinary medicine that were almost exclusively a consequence of the large size of the project. Managing this large group of volunteer EEs and guiding them through a complex evidence evaluation process was challenging. The large number of PICO questions evaluated concurrently contributed to the challenges. Providing adequate training to the large group of EEs with various levels of preexisting expertise in evidence evaluation proved to be a project in its own right and relied upon online content and support from the DCs and Co-Chairs. Human GRADE guideline processes often involve in-person training of EEs and compensated support staff, neither of which was possible in the RECOVER 2024 effort. Better resources to recruit support staff to manage such large guidelines processes would be warranted in future projects. In addition, the COVID 19 pandemic and the associated increased work demands on veterinary professionals further slowed project progress.[18]

EEs did not complete full text reviews of studies before beginning the full evidence evaluation because it would have delayed the process. This meant that for many PICO questions, EEs evaluated different subsets of studies to answer the same PICO question. This also made it challenging to extract summary measures of effect sizes across studies, leading to less robust quantitative summary measures to consider when drafting treatment recommendations. Future GRADE-based guideline processes would likely benefit from this added step to maintain better consistency between EEs.

Finally, the consensus process was conducted remotely using an online commenting system. This allowed a large number of veterinary professionals to have input on the guidelines and voice any concerns; it also made it possible to receive feedback on a very large number of treatment recommendations. However, this approach also limited the opportunity for discussion and debate about treatment recommendations. In-person forums at international meetings or real-time online town halls might provide a more robust consensus process. A Delphi methodology utilizing rounds of surveys with iteratively adapted guideline wording until consensus is reached is a well-established approach to find explicit agreement within a guidelines panel.[19,20] Unfortunately, we lacked the resources to implement this given the size of the project. Overall, we believe that a smaller number of PICO questions would have prevented the limitations we experienced.

In conclusion, the RECOVER 2024 Guidelines process was the first use of the GRADE approach to develop veterinary clinical guidelines,

with veterinary-specific modifications applied. Although there were limitations due to the sheer size of the project, the process resulted in a more transparent, objective, and rigorous evidence evaluation process than that used for the RECOVER 2012 CPR Guidelines. This shows that the GRADE evidence-to-guidelines process as a more robust approach to clinical guidelines development in veterinary medicine is feasible.

AUTHOR CONTRIBUTIONS

Daniel J. Fletcher, Manuel Boller and Jamie M. Burkitt-Creedon: Conceptualization; data curation; funding acquisition; methodology; writing—review and editing. **Erik Fausak**: Conceptualization; data curation; methodology; project administration; writing—review and editing. **Megan G., Van Noord and Kim Mears**: Data curation; methodology; writing—review and editing. **Kate Hopper and Steven E. Epstein**: Validation; writing—review and editing.

ACKNOWLEDGMENTS

The authors would like to acknowledge the following individuals and organizations, without whom an endeavor of this size would have been impossible. Peter Morley, Vinay Nadkarni, and the International Liaison Committee on Resuscitation provided invaluable mentorship, guidance, and a wealth of experience to inform the RECOVER guidelines process. RECOVER as an organization grew out of the American College of Veterinary Emergency and Critical Care and the Veterinary Emergency and Critical Care Society, and we are grateful for the ongoing support of both organizations as we work to produce guidelines, educational content, and research to improve the care of critically ill and injured animals. We appreciate the tireless efforts of Jamie Holms, who provided ongoing administrative support through the process of developing these guidelines. And finally, Emma Fralin and Samantha Moya, veterinary students at Cornell University, provided support for the extensive reference management needed for this project. Project support was received from Boehringer Ingelheim Animal Health and Zoetis Animal Health, both of which helped to fund the purpose-developed, web-based system used for evidence evaluation.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Daniel J. Fletcher PhD, DVM, DACVECC https://orcid.org/0000-0001-8382-5910

Jamie M. Burkitt-Creedon DVM, DACVECC https://orcid.org/0000-0003-3726-0706

Steven E. Epstein DVM, DACVECC https://orcid.org/0000-0001-6126-6695

ENDNOTES

[a] Airtable [online relational database software]. https://airtable.com.
[b] Make [online automation and software integration system]. https://make.com.
[c] https://www.ilcor.org/documents/continuous-evidence-evaluation-guidance-and-templates.
[d] Zotero [Computer software]. Corporation for Digital Scholarship. https://Zotero.org.
[e] https://osf.io/db2am/, DOI 10.17605/OSF.IO/DB2AM.
[f] https://recoverinitiative.org/ee-instructions/.
[g] Covidence [online computer software]. https://covidence.org.
[h] Now Comment [online commenting platform]. https://nowcomment.com.

REFERENCES

1. Fletcher DJ, Boller M, Brainard BM, et al. RECOVER evidence and knowledge gap analysis on veterinary CPR. Part 7: clinical guidelines: RECOVER clinical guidelines. *J Vet Emerg Crit Care*. 2012;22(Suppl_1):S102-S131.
2. Boller M, Fletcher DJ. RECOVER evidence and knowledge gap analysis on veterinary CPR. Part 1: evidence analysis and consensus process: collaborative path toward small animal CPR guidelines. *J Vet Emerg Crit Care*. 2012;22(Suppl_1):S4-S12.
3. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924-926.
4. Eriksen MB, Frandsen TF. The impact of patient, intervention, comparison, outcome (PICO) as a search strategy tool on literature search quality: a systematic review. *J Med Libr Assoc JMLA*. 2018;106(4):420-431.
5. Aamodt M, Huurdeman H, Strømme H. Librarian co-authored systematic reviews are associated with lower risk of bias compared to systematic reviews with acknowledgement of librarians or no participation by librarians. *Evid Based Libr Inf Pract*. 2019;14(4):103-127.
6. Rethlefsen ML, Farrell AM, Osterhaus Trzasko LC, Brigham TJ. Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. *J Clin Epidemiol*. 2015;68(6):617-626.
7. International Liaison Committee on Resuscitation. The International Liaison Committee on Resuscitation (ILCOR) consensus on science with treatment recommendations for pediatric and neonatal patients: pediatric basic and advanced life support. *Pediatrics*. 2006;117(5):e955-977.
8. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Syst Rev*. 2021;10(1):39.
9. McGowan J, Sampson M, Salzwedel DM, et al. PRESS peer review of electronic search strategies: 2015 guideline statement. *J Clin Epidemiol*. 2016;75:40-46.
10. Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence–study limitations (risk of bias). *J Clin Epidemiol*. 2011;64(4):407-415.
11. Hooijmans CR, Rovers MM, de Vries RB, et al. SYRCLE's risk of bias tool for animal studies. *BMC Med Res Methodol*. 2014;14(1):43.
12. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 8. Rating the quality of evidence–indirectness. *J Clin Epidemiol*. 2011;64(12):1303-1310.
13. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines 6. Rating the quality of evidence–imprecision. *J Clin Epidemiol*. 2011;64(12):1283-1293.
14. Guyatt GH, Oxman AD, Sultan S, et al. GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol*. 2011;64(12):1311-1316.
15. Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2011;64(4):383-394.
16. Zhang Y, Akl EA, Schünemann HJ. Using systematic reviews in guideline development: the GRADE approach. *Res Synth Methods*. 2019;10(3):312-329.
17. Guyatt GH, Oxman AD, Kunz R, et al. Going from evidence to recommendations. *BMJ*. 2008;336(7652):1049-1051.

18. McKee H, Gohar B, Appleby R, et al. High psychosocial work demands, decreased well-being, and perceived well-being needs within veterinary academia during the COVID-19 pandemic. *Front Vet Sci*. 2021;8:746716.
19. Boller M, Fletcher DJ, Brainard BM, et al. Utstein-style guidelines on uniform reporting of in-hospital cardiopulmonary resuscitation in dogs and cats. A RECOVER statement. *J Vet Emerg Crit Care San Antonio Tex 2001*. 2016;26(1):11-34.
20. Nasa P, Jain R, Juneja D. Delphi methodology in healthcare research: how to decide its appropriateness. *World J Methodol*. 2021;11(4):116-129.

How to cite this article: Fletcher DJ, Boller M, Burkitt-Creedon JM, et al. 2024 RECOVER Guidelines: Methods, evidence identification, evaluation, and consensus process for development of treatment recommendations. *J Vet Emerg Crit Care*. 2024;34(Suppl 1):3–15. https://doi.org/10.1111/vec.13388