Journal of Veterinary Internal Medicine, 2026, 40(1), aalaf062
https://doi.org/10.1093/jvimsj/aalaf062
Systematic Review | Small Animal Internal Medicine Neurology

A systematic review of acute pain scoring instruments and their measurement properties in cats and dogs

Jungyoon Lee[1,*] and Paulo V. Steagall[1,2,3]

[1]Department of Veterinary Clinical Sciences, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong, China
[2]Department of Clinical Sciences, Université de Montréal, St-Hyacinthe, Quebec, Canada
[3]Centre of Animal Health and Welfare, City University of Hong Kong, Hong Kong, China
*Corresponding author: Paulo V. Steagall, Department of Veterinary Clinical Sciences, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, 31 To Yuen Street, Tat Chee Avenue, Kowloon Tong, Hong Kong, China (pmortens@cityu.edu.hk).

Abstract

Background: **Assessment of acute pain can involve the use of pain scoring instruments.**

Hypothesis/Objectives: **To evaluate the measurement properties of instruments scoring acute pain in cats and dogs according to the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) for Outcome Measurement Instruments (OMIs).**

Animals: **This study did not include live animals.**

Methods: **Five bibliographic databases were searched without restrictions on date. Inclusion criteria were original studies reporting the development or validation of instruments assessing acute pain for cats and dogs and their measurement properties. Exclusion criteria comprised studies reporting chronic pain, no or non-ordinal scoring systems, and studies/instruments not in English. Two investigators reviewed study titles, abstracts, and full texts. The COSMIN Risk of Bias checklist was used to evaluate the following measurement properties: content validity, internal consistency, reliability, measurement error, criterion validity, construct validity, and responsiveness.**

Results: **A total of 25 studies met the inclusion criteria, with 15 instruments assessing acute pain in cats and dogs. The UNESP-Botucatu multidimensional pain scale, its short form, and the Feline Grimace Scale demonstrated the highest quality of evidence and findings, with appropriate criterion and construct validity, reliability, and responsiveness. In general, instruments for dogs lacked development studies, thorough reporting and showed weaker evidence and lower quality of measurement properties than the ones in cats.**

Conclusions and clinical significance: **The quality of evidence and gaps of knowledge across various instruments evaluating acute pain were identified. The measurement properties of instruments in cats are currently superior to those in dogs.**

Keywords pain measurement, pain scales, feline, canine, COSMIN, PRISMA

Abbreviations CMPS, Glasgow composite measure pain scale; CSU-FAPS, Colorado State University feline acute pain scale; COSMIN, COnsensus-based Standards for the selection of health status Measurement INstruments; FGS, feline grimace scale; NRS, numerical rating scale; OMIs, outcome measurement instruments; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; SDS, simple descriptive scale; MCPS, UNESP-Botucatu Multidimensional pain scale; UFEPS, UNESP-Botucatu feline pain scale; UMPS, University of Melbourne pain scale; VAS, visual analog scale

Introduction

Pain is a multidimensional and individual experience with sensory and emotional components that negatively affect animal welfare.[1] Pain changes behavioral expressions, affects quality of life, and leads to suffering, fear, and anxiety.[2] Since animals cannot self-report pain, scoring instruments are pivotal for pain assessment, and they should be able to discriminate between painful and non-painful individuals in a repeatable manner that is effective across various settings and users. These instruments are helpful in guiding clinical decision-making (ie, the need for analgesics).[3] Acute pain usually arises from tissue damage and is self-limiting, whereas chronic pain is generally defined as ongoing pain longer than 3 months and can be a disease itself.[4,5] Accurate recognition of pain is crucial in veterinary medicine; pain scoring instruments and their measurement properties have been investigated in several species.[6,7]

Clinical acute pain assessment in cats and dogs relies on the observation of behavioral changes, such as body posture, vocalization, interaction with the observer/environment, attention to the painful area, response to touch/palpation, and facial expressions.[1,8] Indeed, most of these behaviors are evaluated as part

Received: July 16, 2025. **Revised:** December 9, 2025. **Accepted:** December 9, 2025

of instruments assessing acute pain in small animals. However, a systematic review of the measurement properties of these scales evaluating acute pain has not been reported yet. Measurement or psychometric properties, refer to the quality aspects of the outcome measurement instrument (OMI; ie, pain scoring instrument), including validity, reliability, and responsiveness.[9] The following measurement properties are evaluated when using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines[6,10]: (1) content validity; (2) internal consistency; (3) reliability (inter- and intra-reliability); (4) measurement error; (5) criterion validity; (6) construct validity and (7) responsiveness. These properties are essential aspects of an OMI.

Systematic reviews might compile data from various studies developing or evaluating OMIs; it allows for an evidence-based selection of an OMI to measure a construct of interest (ie, pain) in the target population.[10,11] In this case, gaps in knowledge about instruments assessing acute pain in cats and dogs might be identified, providing a current state of evidence and future perspectives for research studies. This systematic review aimed to evaluate the evidence of the measurement properties of instruments scoring acute pain in cats and dogs using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) and the COSMIN guidelines.

Materials and methods

This systematic review is reported according to the 2024 PRISMA for OMIs and COSMIN guidelines version 2.0, encompassing the COSMIN Risk of Bias checklist version 3.0.[9,11]

Databases and search terms

Five bibliographic databases (MEDLINE via EBSCOHost, PubMed, Web of Science, CAB Abstracts, and SCOPUS) were searched to identify peer-reviewed studies without any limitations to the publication period. The search terms were refined and tested using PubMed and the Medical Subject Heading terms, allowing consistency in indexing scientific articles. The search terms were based on previous studies,[6] supported by a librarian, as follows: [("PAIN MEASUREMENT∗") or ("PAIN ASSESSMENT") or ("PAIN SCOR∗") or ("PAIN SCALE") or ("PAIN INDEX") or ("PAIN INDICES") or ("PAIN RATING∗") or ("PAIN INTENSI∗") or ("PAIN THRESHOLD") or ("PAIN QUANTIFICATION") or ("PAIN EVALUAT∗")] and [(CAT or CATS or FELINE∗) or (DOG or DOGS or CANINE∗)]. The search was last updated in February 2025. The field of construct of interest (ie, pain) and type of instrument (ie, pain measurement) were limited to the title and abstract of the studies. The population of interest (ie, cat or dog) was limited to the title of the studies.

Eligibility criteria

The inclusion criteria consisted of original studies that reported the development or validation of instruments for evaluating acute pain in dogs and cats, as well as their measurement properties. Any source of acute pain was eligible. Exclusion criteria consisted of studies reporting the development and assessment of measurement properties of instruments scoring chronic pain (ie, osteoarthritic or neuropathic pain) or reports and instruments not published in English. Studies reporting indicators of pain without a scoring system or non-ordinal pain assessment variables were not included. Studies reporting the use of pain scoring instruments to measure constructs other than pain (eg, sedation, animal welfare, or quality of life), nociceptive testing alone, or surveys about knowledge and opinions regarding pain assessment in dogs and cats were also excluded. Studies reporting dichotomous agreement were excluded, as dichotomous agreement values are not considered a measurement property when evaluated according to the COSMIN guidelines.[9]

Literature search

Study titles and abstracts were initially screened for eligibility by the 2 investigators of the study. One investigator was a specialist by the American College of Veterinary Anesthesia and Analgesia with experience in systematic reviews (PVS), and the other was a final-year veterinary student (JL) who had been taught a 4-credit course in anesthesia and pain management. Full texts were then selected without masking of authorship. The references were exported into EndNote (version 20) and Covidence (an online platform integrated with the Cochrane's review production toolkit, designed to streamline systematic reviews; www.covidence.org) for duplicate removal. The 2 investigators independently reviewed the full articles to confirm eligibility criteria. "Snowball" methods, searching references of eligible articles and electronic citation tracking, were used to maximize the retrieval of relevant studies.

Data extraction

The 2 investigators extracted data from selected articles using a predefined data collection sheet (Microsoft Excel, Inc.) according to the COSMIN guidelines.[9] The following information was extracted:

1. Characteristics of the population (species, age, sex, breed, source, and duration of pain).
2. Characteristics of the instrument (name/version, number of items or action units, scoring method such as images, video, or real-time assessment).
3. Setting and purpose for which the instrument is intended: response, intervention (ie, rescue analgesia), age, clinical, or experimental setting.

Assessment of measurement properties

The 2 investigators independently assessed the measurement properties of each instrument via a predesigned Excel spreadsheet. In cases of disagreement, a consensus was reached after discussion between the 2 investigators. All information recorded followed the COSMIN guidelines,[9] which were validated and developed to assess human patients' reported outcomes in systematic reviews. An adapted COSMIN evaluation spreadsheet was used, excluding comprehensibility (ie, the patient's point of view) and interviewing methods,[12] considering the target population of this study (ie, nonverbal animals).

The measurement properties of eligible articles were assessed using the COSMIN Risk of Bias checklist version 3.0, which assesses the methodological quality of studies on the measurement properties of OMIs.[9] The following measurement properties

were evaluated: (1) content validity (assessment of the items of the scale and calculation of content validity index); (2) internal consistency (correlation among the items of the scale); (3) measurement error (including sensitivity, specificity and accuracy); (4) reliability (whether the scores remain consistent across different observers [ie, inter-observer reliability] and over time [ie, intra-observer reliability]); (5) criterion validity (correlation of the proposed instrument with a gold-standard); (6) responsiveness (capability to distinguish changes over time) and (7) construct validity (whether the instrument measures its intended outcome by comparing different known groups, such as pain scores before and after surgery) (Table 1). In veterinary medicine, currently, there is no established gold standard to evaluate pain, whereas in humans, self-reporting pain is possible.[43] According to the COSMIN, reviewers can decide the appropriate OMI for correlation.[9] Thus, a near-equivalent gold standard OMI (ie, validated pain scoring instruments) assessing acute pain within the same species was considered suitable for comparison in this study. Measurement properties were individually assessed for methodological quality as very good (V), adequate (A), doubtful (D), inadequate (I), or not applicable (N). The overall rating of the study's methodological quality was based on the lowest score assigned and the criteria listed in Table 1.[13]

Subsequently, the quality of the findings of each study was rated as sufficient (+), indeterminate (?), or insufficient (−) (Table 2). Afterwards, the overall quality of the findings for an instrument measuring acute pain was summarized as sufficient (+), insufficient (−), indeterminate (?), or inconsistent (+/−). The rating was summarized as inconsistent when the ratings for individual studies were mixed (ie, both sufficient and insufficient). In such cases, explanations for the inconsistencies were explored. It is stated in the COSMIN guidelines that explanations could be found in aspects such as differences in target population within the study for a similar concept (ie, pain measurement), study quality (ie, a study with inadequate methodological quality also presenting insufficient results), or a combination of both.[9] However, if reasons for inconsistency could not be explained, both reviewers could discuss and reach a consensus based on the available evidence. In this case, an overall grading of inconsistent (+/−) was provided, adhering to the COSMIN guidelines.[9]

Following the COSMIN guidelines for systematic reviews,[9] a Grading of Recommendations Assessment, Development and Evaluation (GRADE) was employed to assess evidence quality of the OMI.[14] The GRADE assumes that the quality of evidence is high, then downgrades the quality as risk of bias, inconsistency, imprecision, and indirectness are presented. The final quality of evidence for each measurement property was rated as high, moderate, low, or very low (Table 3).[14] In cases where findings were not reported (ie, no studies available for a specific measurement property), the quality of evidence was not rated for that measurement property.

Results

Overview of studies

The 2020 PRISMA flow diagram provides the identification, screening, and inclusion/exclusion of studies as part of this systematic review (Figure 1). A total of 354 studies were identified; 181 were duplicate records and removed via EndNote20 and Covidence. Title and abstract screening involved 173 manuscripts with an additional 143 records excluded. After full text screening and data extraction, 12 studies were excluded. Using the snowball method, 7 additional studies were retrieved through citation searching, assessed for eligibility, and then included. A total of 25 studies were included and reported in this systematic review. The characteristics of the populations and pain scoring instruments are reported in Tables 4 and 5, respectively.

A total of 15 unique instruments scoring acute pain were identified from an investigation of 25 studies. Short or modified instruments were evaluated separately and independently. In cats, 6 OMIs were included: the Feline Grimace Scale (FGS)[15–21] with 7 studies, the UNESP-Botucatu Multidimensional Pain Scale (MCPS)[22–25] with 4 studies and its short version, UNESP-Botucatu Feline Pain Scale (UFEPS-SF)[24,26] with 2 studies, the revised Glasgow Composite Measure Pain Scale—Feline (rCMPS-F)[27] with 1 study, the embedded facial image version of CMPS-Feline (CMPS-F)[19,28] with 2 studies, and the Colorado State University Feline Acute Pain Scale (CSU-FAPS)[19,29] with 2 studies (Table 5). In dogs, 6 instruments were investigated: the Glasgow Composite Measure Pain Scale—Canine (CMPS)[30,31] with 2 studies, its short form (CMPS-SF) with 3 studies,[32–34] its modified form with 1 study,[35] the University of Melbourne Pain Scale (UMPS)[36] with 1 study, and the 4A-VET[37] and its modified form[38] with 1 study each. Three additional OMIs were evaluated: the Visual Analog Scale (VAS) with 3 studies,[25,34,39] the Numerical Rating Scale (NRS)[34,40] with 2 studies, and the Simple Descriptive Scale (SDS)[34] with 1 study (Table 5).

Measurement properties

The summary of the number of studies, methodological quality, the summarized quality of results for each measurement property, and evidence of an OMI are presented in Table 6. The pain scales developed for assessing acute pain in cats had more comprehensively reported measurement properties and were supported by higher quality evidence than those in dogs. The MCPS, UFEPS-SF, and FGS[15,22–24,26] presented high-quality evidence and sufficient results across all 7 measurement properties, supported by a robust number of studies compared to other instruments. In contrast, instruments scoring acute pain in dogs generally lacked development studies, thorough reporting and showed weaker evidence and lower quality of findings across their 7 measurement properties than the instruments applied to cats.

There was a variability in the available evidence across the different instruments and their measurement properties (Table 6). Reliability, construct validity, and criterion validity were the most frequently reported and evaluated in 11, 10, and 8 instruments, respectively. Appropriate information on the scale development of the CMPS-F, CSU-FAPS, 4A-VET, modified 4A-VET, and modified CMPS was unavailable.

Discussion

This systematic review identified and evaluated 15 OMIs assessing acute pain in cats and dogs based on 25 eligible studies. The included OMIs encompassed multidimensional (eg, affective-emotional, cognitive, and sensory aspects of pain) and unidimensional (eg, pain intensity) tools, with pain scoring instruments for cats generally supported by a greater volume and quality

Table 1 Criteria for the assessment of methodological quality, adapted from COnsensus-based standards for the selection of health status measurement INstruments guidelines.

Instrument component	Categories	Rating criteria
Scale development	1a. General design requirements	1. Is a clear description provided of the construct to be measured? 2. Is the origin of the construct clear: was a theory, conceptual framework or disease model used or clear rationale provided to define the construct to be measured? 3. Is a clear description provided of the target population and context for which the scale was developed?
	1b. Scale development	4. Was the scale development study performed in a sample representing the target population? 1. Was an appropriate method used to identify relevant items for a new scale? 2. Was a skilled observer or group of observers (experts in the field) used to define the items?
Measurement properties	2a. Internal consistency	1. Was the internal consistency calculated and reported? 2. Were there any other important flaws?
	2b. Reliability	1. Was INTER-rater reliability reported? Was the number of raters appropriate for INTER-rater reliability testing? a. Was the statistical method for calculating INTER-rater reliability appropriate? 2. Was INTRA-rater reliability reported? Was the time interval appropriate for INTRA-rater reliability testing? a. Were the test conditions similar for the measurements? eg, type of administration, environment, instructions b. Was the statistical method for calculating INTRA-rater reliability appropriate? 3. Were there any other important flaws?
	2c. Measurement error	1. Were sensitivity, specificity and/or accuracy determined? 2. Were there any other important flaws?
	2d. Criterion validity	1. Was criterion validity reported? 2. Is it clear what the gold standard or other method measure(s)? 3. Were the measurement properties of the gold standard or other validated method adequate? 4. Was the statistical method appropriate for the hypotheses to be tested? 5. Were there any other important flaws?
	2e. Hypothesis testing for construct validity	1. Was construct validity reported? 2. Was an adequate description provided of important characteristics of the subgroups? 3. Was the statistical method appropriate for the hypotheses to be tested? 4. Were there any other important flaws?
	2 f. Responsiveness	1. Was responsiveness reported? 2. Was an adequate description provided of the intervention given? 3. Was the statistical method appropriate for the hypotheses to be tested? 4. Were there any other important flaws?

of evidence than those for dogs. A subset of these instruments (MCPS, UFEPS-SF and FGS) presented robust measurement properties.[15,22–24,26] Our findings highlight an apparent disparity in the quality of psychometric evaluation among existing pain scales, with critical properties such as content validity and responsiveness underreported in many cases. These results underscore the ongoing need for rigorous development and validation of pain assessment tools and provide a foundation for targeted improvements in clinical and research practice.

Content validity can be determined using relevance, comprehensiveness, and comprehensibility.[9] It measures the extent of the appropriateness of the content reflecting the construct to be measured (ie, acute pain). Content validity of the VAS, NRS, and SDS was not rated, as they are not species-specific and report numerical values assigned from the assessor's subjective point of view.[6,41] The MCPS, UFEPS-SF, rCMPS-F, FGS, UMPS, CMPS, and CMPS-SF (n = 7) reported scale development studies, while CMPS-F, CSU-FAPS, 4A-VET, modified 4A-VET, and modified CMPS (n = 5) lacked similar reporting. The 4A-VET and its modified form mentioned a scale development study, but the full manuscript could not be found on online databases.[37,38] The development studies for the CMPS-F and the modified CMPS have not been reported,[28,35] likely because the 2 instruments were created based on existing scales (ie, rCMPS-F and CMPS, respectively). However, the COSMIN manual indicates that sub-scales should be assessed separately.[9] The lack of development studies is a critical limitation.[9] If a scale does not appropriately measure what it claims, other measurement properties might be regarded as less reliable.

Internal consistency refers to the correlation among the scale items, generally determined by Cronbach's α.[9,42] Other methods are possible, such as item-total correlation, where each item of the scale can be correlated when 1 item is omitted, and the items can be adjusted based on correlations.[42] The COSMIN risk of bias checklist indicates Cronbach's α as the criterion for quality of findings. The MCPS, UFEPS-SF, FGS, 4A-VET, and CMPS (n = 5) reported internal consistency.[15,22,23,26,30,37] Among the 5 instruments, all showed high quality of evidence with sufficient (+) results except the CMPS,[30] which had insufficient (−) results with a Cronbach's α value $<.70$. This indicates that items of the instrument were

Table 2 Criteria for the quality of findings and summary, adapted from COnsensus-based standards for the selection of health status measurement INstruments guidelines.

Instrument component	Categories	Rating criteria
Scale development	1a. General design requirements	(+): The model/stimulus are relevant AND all items refer to relevant aspects of the construct to be measured AND are relevant for the target population AND context of use. (?): Not all information for (+) reported OR potential biases identified (−): Criteria for (+) not met AND substantial bias identified.
Measurement properties	2a. Internal consistency	(+): At least low-quality evidence for sufficient unidimensionality AND Cronbach's $\alpha \geq .70$ (?): Criteria for "at least low evidence for sufficient unidimensionality" not met OR Evidence for insufficient unidimensionality OR Not enough information Reported (−): At least low-quality evidence for sufficient unidimensionality AND Cronbach's $\alpha < .70$.
	2b. Reliability	(+): ICC or (weighted) κ or Pearson/Spearman correlation $\geq$0.70 (?): Not enough information reported (−): ICC or (weighted) κ or Pearson/Spearman correlation $<$ 0.70.
	2c. Measurement error	(+) Accuracy $>$ 80% (?) Not defined OR $>$ 60 and $<$ 80% (−) Accuracy $<$ 60%.
	2d. Criterion validity	(+): Correlation with gold standard $\geq$0.70 OR AUC $\geq$0.70 (?): Not enough information reported (−): Correlation with gold standard $<$ 0.70 OR AUC $<$ 0.70.
	2e. Hypothesis testing for construct validity	(+): $\geq$75% of the results are in accordance with the predefined hypothesis (?): No relevant results were found (−): $\geq$75% of the results deviates from the predefined hypothesis.
	2f. Responsiveness	(+): $\geq$75% of the results are in accordance with the predefined hypothesis OR AUC $\geq$0.70 (?): No relevant results were found (−): $\geq$75% of the results deviate from predefined hypotheses OR AUC $<$ 0.70.

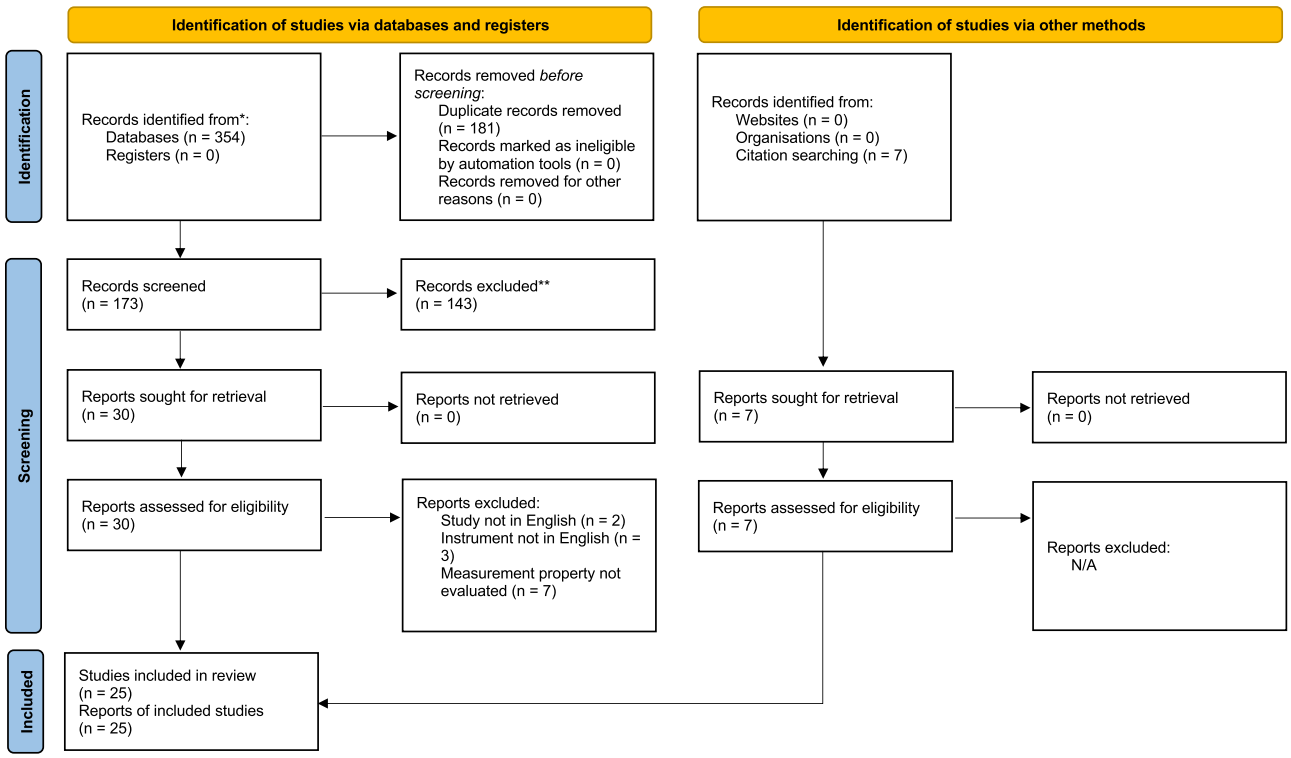

Figure 1 Preferred reporting items for systematic reviews and meta-analyses- COnsensus-based standards for the selection of health status measurement INstruments flow diagram for new systematic reviews which included searches of databases, registers and other sources. Source: Page MJ, et al. BMJ 2021;372:n71. doi: 10.1136/bmj.n71. This work is licensed under CC BY 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

Table 3 Definitions of quality levels from COnsensus-based standards for the selection of health status measurement INstruments guidelines.

Quality level	Definition
High	We are very confident that the true measurement property lies close to that of the estimate* of the measurement property
Moderate	We are moderately confident in the measurement property estimate: the true measurement property is likely to be close to the estimate of the measurement property, but there is a possibility that it is substantially different
Low	Our confidence in the measurement property estimate is limited: the true measurement property may be substantially different from the estimate of the measurement property
Very low	We have very little confidence in the measurement property estimate: the true measurement property is likely to be substantially different from the estimate of the measurement property

*Estimate of the measurement property refers to the pooled or summarized result of the measurement property of an Outcome Measurement Instrument. These definitions were adapted from the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach.

not homogeneous. Internal consistency is usually reported with the development of the OMI as it evaluates whether the items are appropriate for the measured construct (ie, acute pain). This might partially explain the limited number of instruments to report internal consistency, as development or content validity has not been reported for some scales. Additionally, internal consistency is irrelevant for unidimensional scales with single numerical or descriptive values (ie, VAS, NRS, and SDS).

Measurement error is the precision of a score.[9] In this systematic review, if the study reported sensitivity, specificity, accuracy, or area under the curve, they were considered measurement error. The COSMIN guidelines state that any of these parameters above 80% or 90% is considered sufficient. Three instruments (MCPS, UFEPS-SF, and FGS) reported measurement error and were rated as high-quality evidence with sufficient results.[23–25] Measurement error is an area that OMIs require more attention, as it has been vastly underreported. For example, it is vital to determine the ability of a pain scoring instrument to discriminate true positive and true negative scores during pain assessment.

Reliability refers to whether the scores remain consistent across different raters (ie, inter-rater reliability) or within 1 rater over time (ie, intra-rater reliability).[9] The COSMIN checklist does not distinguish intra- or inter-rater reliability. Therefore, both were investigated separately in this systematic review. Reliability can be assessed by evaluating the interclass correlation coefficient, κ value or correlation (ie, Pearson or Spearman).[42] Values $\geq$0.70 are considered sufficient (+). Reliability was the most reported measurement property and was included in 11 out of 15 instruments. The MCPS, UFEPS-SF, FGS, CSU-FAPS, CMPS-SF, VAS, and NRS (n = 7) were considered high-quality evidence.[15,17–19,23–26,29] Inconsistent (+/−) quality of results was recorded for the CSU-FAPS, CMPS-SF, and VAS (n = 3).[19,25,29,33,34] For these 3 OMIs, there were only 2 studies investigating reliability; 1 showed sufficient results and the other showed insufficient results.[19,25,29,32–34,39] According to the COSMIN guidelines, an option exists to consider differences in study quality. However, the authors considered it inappropriate to summarize very good and adequate results based only on 2 studies. Hence, the summarized result was rated as inconsistent. The CMPS-F and modified 4A-VET (n = 2) showed low quality of evidence.[19,38] The 4A-VET and SDS (n = 2) were rated very low with insufficient summarized findings.[37,39] The 4 OMIs listed each had only 1 study assessing reliability. In this study, the quality of evidence is downgraded due to the severe risk of bias when only 1 study is available, which is rated as doubtful or inadequate.[9]

Criterion validity is assessed by analyzing the correlation of the proposed instrument with a gold-standard OMI.[9] The MCPS, UFEPS-SF, rCMPS-F, FGS, CSU-FAPS, VAS, and NRS (n = 7) reported criterion validity.[15,23,24,26,27,29,34] All displayed high quality of evidence except rCMPS-F and CSU-FAPS (n = 2), which were rated low.[27,29] The latter was attributed to poor methodological quality. Furthermore, the lack of a universally accepted gold standard instrument for evaluating acute pain in dogs and cats presents inherent challenges in establishing criterion validity. This limitation is particularly evident with older OMIs, such as the UMPS or CMPS.[30,36] As a result, these instruments might be unable to report criterion validity according to contemporary methodological standards. Furthermore, OMIs have been developed and investigated using differing comparators with varying criteria, as there was no existing set standard to reference. This heterogeneity in validation approaches poses a challenge when attempting to rate the evidence of instruments over time. For instance, some instruments were compared to NRS or SDS,[27,31,34,35] which are unidimensional scales that are more prone to subjectivity by the assessor.[44] Other instruments assessed the correlation between psychometric instruments within the same species.[15,24,26,29] Criterion validity is better assessed using robustly validated scoring instruments assessing the same construct (ie, acute pain) within the same population (ie, cats or dogs).

Responsiveness is the ability of the instrument to detect changes over time, especially after an intervention.[9] In this systematic review, responsiveness was considered a significant decrease in pain scores after the administration of analgesia. Responsiveness was the second least reported measurement property with studies including the MCPS, UFEPS-SF, rCMPS-F, and FGS.[15,16,18,23,24,26,27] Most of these instruments recorded a high quality of evidence with sufficient results, describing the provided analgesic intervention and confirming their hypothesis using statistical tests such as paired t-tests or the Wilcoxon signed rank test to compare pain scores. The summarized findings were also sufficient, indicating that $\geq$75% of the results were in accordance with the predefined hypothesis of the study. The rCMPS-F displayed low evidence with sufficient results due to the methodology of 1 study being rated as doubtful.[27] Several pain scoring instruments lack the reporting of responsiveness and should be the focus of future research; it is fundamental to know if the instrument is responsive after appropriate analgesia so treatment efficacy can be evaluated.

Construct validity assesses the extent to which the instrument measures its intended outcome by comparing different known

Table 4 Characteristics and demographics of the population involved in studies reporting measurement properties of acute pain scoring instruments in cats and dogs.

Species, instrument	Reference	Number of animals (and/or groups)	Age	Sex (%Female; F)	Breed	Source and duration of pain
Cat, FGS	15	Development: 51 cats Validation: 55 cats	6.3 ± 3.6 years	Controls: F (75%) Painful F (54%)	Various	Naturally occurring acute pain, video-recorded and assessed. Painful cats were assessed an hour after analgesic treatment.
Cat, FGS	16	65 cats	1.37 ± 0.9 years	F (100%)	Various	Pain from ovariohysterectomy, evaluated at baseline, 15 min after sedation, and at 0.5, 1, 2, 3, 4, 6, 8, 12, and 24 h after extubation using the FGS in real-time (FGS-RT). Cats were video recorded simultaneously at baseline, 15 min after sedation, and at 2, 6, 12, and 24 h after extubation for subsequent image assessment.
Cat, FGS	17	24 cats	6 ± 3.3 years	F (54.2%)	Various	Pain from oral treatment (dental extractions). Images of cat faces were captured from video recordings at 6 h postoperatively (day 1), day 6, and before and after rescue analgesia.
Cat, FGS	18	36 cats	Mean: 15 ± 4.7 weeks	F (100%)	Various	Pain from ovariohysterectomy, assessed in real time before surgery and 1, 2, 4, 6, 8, 12, and 24 h postoperative (UFEPS-SF) and videotaped (FGS).
Cat, FGS; CMPS-F; CSU-FAPS	19	20 cats	Information unavailable	Information unavailable	Information unavailable	Routine neutering (spay or castration), assessed pre- and post-operatively.
Cat, FGS	20	10 (images of) cats	Information unavailable	Information unavailable	Information unavailable	Differing degrees of naturally occurring pain secondary to medical conditions or surgery.
Cat, FGS	21	100 (images of) cats	Information unavailable	Information unavailable	Information unavailable	Differing degrees of naturally occurring pain secondary to medical conditions or surgery.
Cat, MCPS	22	40 cats under ovariohysterectomy	Information unavailable	F (100%)	Information unavailable	Pain from ovariohysterectomy was assessed an hour after, then at regular intervals for 7 days after analgesia.
Cat, MCPS	23	Phase 1: 30 cats Phase 2: 28 cats	Phase 1: 14.1 ± 5.2 months Phase 2: 8.5 ± 4.2 months; 11.4 ± 5.5 months	F (100%)	Phase 1: Mixed Phase 2: Domestic Short Hair	Pain from ovariohysterectomy was observed at 4 time points (preoperative, 30 min to 1, 4, and 24 h postoperatively).
Cat, MCPS; UFEPS-SF	24	40 cats	3.8 ± 4.3 years	F (34%)	Mixed	Pain management for clinical (n = 20) or surgical (n = 20) care. Evaluated before premedication and at every hour from 1 to 6 (or 6.5 in case of rescue analgesia) after extubation.
Cat, MCPS; VAS	25	45 cats	Information unavailable	F (100%)	Mixed	Pain from ovariohysterectomy was assessed prior to anesthesia and 0.5, 1, 2, 3, 4, 6, and 8 h postoperatively.
Cat, UFEPS-SF	26	30 cats	Information unavailable	F (100%)	Information unavailable	Pain from ovariohysterectomy was assessed preoperatively, 30-60 min after surgery, and 4 h after intervention analgesia.
Cat, rCMPS-F	27	Study 1: 25 cats Study 2: 116 cats	Study 1: Range: 8 weeks-19 years; Mean: 5 years 8 month Study 2: Range: 9 weeks-22 years	Study 1: F (68%) Study 2: F (44%)	Mixed	Study 1: Admitted for surgery, traumatic or medical conditions requiring analgesic treatment. Study 2: Postoperative care or admitted for any acutely painful trauma or medical condition in multiple locations.

(continued)

Table 4 Continued

Species, instrument	Reference	Number of animals (and/or groups)	Age	Sex (%Female; F)	Breed	Source and duration of pain
Cat, CMPS-F	28	119 cats	Range: 1-240 months; Median: 63 months	F (45.4%)	Information unavailable	Undergoing postoperative care or having been admitted for surgery, acute painful trauma, or medical condition.
Cat, CSU-FAPS	29	68 cats 21 cats for inter rater, 47 for validity	Range: 0.3-4 years; Median: 1 year	F (100%)	Mixed	Pain from ovariohysterectomy. Inter-rater reliability: Pain assessed 6 h postoperatively. Convergent validity: Pain assessed in the morning and evening of days 1 and 2, then the morning of day 3.
Dog, UMPS	36	48 dogs	Range: 5 months-6.5 years; Mean: 19 months	Information unavailable	21 purebred, 12 mixed	Dogs undergoing general anesthesia only or also undergoing ovariohysterectomy. Scored at 0, 1, 2, 3, 6, 8, 12, and 18 h after extubation.
Dog, CMPS	30	No animals investigated	-	-	-	-
Dog, CMPS	31	80 dogs	Range: 5 months-15 years	F (48.8%)	Variety	Pain after orthopedic surgery, soft tissue surgery, and medical conditions. Assessed the following day after surgery (19-29 h after surgery).
Dog, CMPS-SF	32	122 dogs	No restrictions	No restrictions	No restrictions	Undergone surgery with no restriction of type of surgical procedure.
Dog, CMPS-SF	33	45 dogs	Range: 3-195 months; Median: 75 months	F (51.1%)	16 breeds involved	Undergone different surgical procedures or had underlying medical conditions expected to be painful.
Dog, VAS; NRS; CMPS-SF	34	31 dogs	Information unavailable	Information unavailable	Information unavailable	After recovery from anesthesia, videotaped over a 2-month period.
Dog, Modified CMPS	35	60 dogs	Range: 3 months-13 years; Mean: 6.2 years	F (51.7%)	Variety	Pain from soft tissue or orthopedic surgery.
Dog, 4A-Vet	37	25 dogs	Information unavailable	Information unavailable	Information unavailable	Postoperative pain managed with placebo, preemptive and multimodal analgesia, and preemptive analgesia with oral tramadol, assessed within a 48-hour frame.
Dog, 4A-VET modified	38	16 dogs	Range: 2-5 years	F (50%)	16 beagles	Three methods of bone marrow aspiration: iliac under sedation, sternum with and without sedation.
Dog, NRS, SDS, VAS	39	50 dogs	Information unavailable	Information unavailable	Mixed	Underwent various surgical procedures, assessed 1 h after surgery (group 1), 21-27 h after surgery (group 2), and both (group 3).

groups.[9] This study focused on the significant differences in pain scores between painful and pain-free groups (ie, pain scores should be higher in painful than in pain-free individuals). Construct validity was the second most reported measurement property, including 10 OMIs (MCPS, UFEPS-SF, CMPS-F, FGS, 4A-VET and its modified form, UMPS, CMPS, CMPS-SF and its modified form).[15,16,18,22–24,26,28,32,35–38] Results were mixed regarding the quality of evidence, but they generally scored "high" or "moderate," and most results were deemed sufficient. The CMPS-F, 4A-VET, and UMPS were rated "low" regarding quality of evidence. rCMPS-F reported construct validity,[27] but it was regarded as testing responsiveness rather than construct validity, since subgroups were not compared. Overall, the reporting of measurement properties was incomplete across many instruments. This could be partially attributed to the advent of subscales or modified versions of original instruments rated independently.[9] As a result, the number of studies and their measurement properties might be dispersed across related, but distinct instruments. This dispersion might have led to an apparent lack of comprehensive reporting for specific instruments, as evidence that might otherwise be considered collectively was instead fragmented across subscales. For example, a single study used

Table 5 Characteristics of the acute pain scoring instruments in cats and dogs investigated in this study.

Species, instrument Refs	Source/model of pain	Number of items	Method of scoring	Response options	Range of scores (total)	Analgesic intervention
Feline, FGS[15–21]	Several (naturally occurring)	5—ear position, orbital tightening, muzzle tension, whiskers change, and head position	Screenshots from videos	0—1—2	Sum/10 = 0-1.0 (1.0)	YES; > 0.39/1.0 *
Feline, MCPS[22–25]	Ovariohysterectomy	4—psychomotor, protection of wound, physiologic, and vocal expression	Videotaping	0-1-2-3	Sum: 30	YES; > 7/30
Feline, UFEPS-SF[24,26]	Ovariohysterectomy	4—posture, comfort, reaction to palpation, and miscellaneous behaviors	Videotaping and real time	0-1-2-3	Sum: 12	YES; ≥4/12
Feline, rCMPS-F[27]	Several (naturally occurring)	4—observation, approach and interaction, palpation, general impression		0-1-2-3-(4)	Sum: 16	YES; ≥4/16
Feline, CMPS-F[19,28]	Several (naturally occurring)	4—observation, approach and interaction, palpation, and general impression	Photographs	0-1-2-3-(4)	Sum: 20	YES; ≥5/20
Feline, CSU-FAPS[19,29]	Ovariohysterectomy	3—psychological & behavioral, response to palpation, and body tension	Real time	0-1-2-3-4	Sum: 4	NO
Canine, UMPS[36]	Postoperative pain (Ovariohysterectomy)	6—physiologic data, response to palpation, activity, mental status, posture, and vocalization	Video recording	0-1-2-3	Sum: 27	YES; Score not indicated
Canine, CMPS[30,31]	Medical and surgical pain	7—posture, comfort, vocalization, attention to wound, demeanor, mobility, and response to touch	Real time	Ticking checkboxes	Sum: 29 checkboxes	YES; 10-14
Canine, CMPS-SF[32–34]	Postoperative pain	6—vocalization, attention to wound, mobility, response to touch, demeanor, and posture/activity	Real time	0-1-2-3-4-(5)	Sum: 24; 20 (if mobility cannot be assessed)	YES; ≥6/24; ≥5/20 (mobility not assessed)
Canine, Modified CMPS[35]	Pain from orthopedic or soft tissue surgery	4—looking at dog, putting lead on dog and out of kennel, and palpation of wound	Real time	Ticking checkboxes	Information not available	NO
Canine, 4A-VET[37]	Several, from surgery	6—global subjective appreciation, general attitude, interactive behavior, gait evaluation, reaction to the surgical wound palpation, and intensity of the reaction	Real time	0-1-2-3	Sum: 18	YES; 11 > 18
Canine, 4A-VET modified[38]	Bone marrow aspiration	4—Global subjective pain score, general attitude, reaction to handling of the affected area, and intensity of the reaction	Real time	0-1-2-3	Sum: 12	NO
Feline/Canine, VAS[25,34,39]	Several	1	Variable	100 mm	Sum: 100 mm	NO
Feline/Canine, NRS[34,40]	Several	1	Variable	0-10	Sum: 10	NO
Feline/Canine, SDS[34]	Several	1	Variable	Variable	Variable	NO

References are reported in superscript. Abbreviations: FGS = Feline Grimace Scale; MCPS = UNESP-Botucatu Multidimensional Pain Scale; UFEPS-SF = UNESP-Botucatu Feline Pain Scale- Short Form; rCMPS-F = Revised Glasgow Composite Measure Pain Scale—Feline; CMPS-F = Glasgow Composite Measure Pain Scale—Feline; CSU-FAPS = Colorado State University Feline Acute Pain Scale; CMPS = Glasgow Composite Measure Pain Scale—Canine; CMPS-SF = Glasgow Composite Measure Pain Scale—Canine, Short Form; UMPS = University of Melbourne Pain Scale; 4A-VET = 4A-Vet Pain Scale; VAS = Visual Analog Scale; NRS = Numerical Rating Scale; SDS = Simple Descriptive Scale.

the modified CMPS and reported only 1 measurement property.[35] Additionally, pain scoring instruments supported by only 1 study often failed to report a comprehensive range of measurement properties. This poses a problem as the quality assessment is based on limited data, which might not fully represent the instrument's performance. As a result, some instruments such as the CMPS-F, 4A-VET and its modified form showed low evidence quality and insufficient findings.[28,37,38]

This systematic review has limitations that could alter the interpretation and generalizability of the findings. Our

Table 6 Summary of the consensus ratings given for each acute pain scoring instrument regarding their measurement properties including the number of studies, methodological quality, quality of results and overall quality of evidence.

Species, instrument Refs	Measurement Property	Total number of studies	Methodological quality: number of studies	Summarized quality of results	Overall quality of evidence
Cat, FGS[15–21]	General design requirements and development	1	A:1	+	High
	Internal consistency	1	V:1	+	High
	Reliability	6	V:5 D:1	+	High
	Measurement error	1	V:1	+	High
	Criterion validity	1	V:1	+	High
	Construct validity	3	V:3	+	High
	Responsiveness	3	V:3	+	High
Cat, MCPS[22–25]	General design requirements and development	2	V:2	+	High
	Internal consistency	2	V:2	+	High
	Reliability	3	V:2	+	High
			D:1	?	?
	Measurement error	3	V:2	+	High
			I: 2	?	?
	Criterion validity	2	V:1 A:1	+	High
	Construct validity	3	V:1 A:1 D:1	+	High
	Responsiveness	2	V:2	+	High
Cat, UFEPS-SF[24,26]	General design requirements and development	1	V:1	+	High
	Internal consistency	1	V:1	+	High
	Reliability	2	V:2	+	High
	Measurement error	2	V:2	+	High
	Criterion validity	2	V:2	+	High
	Construct validity	2	V:1 A:2	+	High
	Responsiveness	2	V:2	+	High
Cat, rCMPS-F[27]	General design requirements and development	1	V:1	+	High
	Internal consistency	0	No studies available	?	?
	Reliability	0	No studies available	?	?
	Measurement error	0	No studies available	?	?
	Criterion validity	1	D:1	+	Low
	Construct validity	0	I:1	?	?
	Responsiveness	1	D:1	+	Low
Cat, CMPS-F[19,28]	General design requirements and development	0	No studies available	?	?
	Internal consistency	0	No studies available	?	?
	Reliability	1	D:1	+	Low
	Measurement error	0	No studies available	?	?
	Criterion validity	0	No studies available	?	?
	Construct validity	1	D:1	+	Low
	Responsiveness	0	No studies available	?	?
Cat, CSU-FAPS[19,29]	General design requirements and development	0	No studies available	?	?
	Internal consistency	0	No studies available	?	?
	Reliability	2	V:1 D:1	+/−	High
	Measurement error	0	No studies available	?	?

(continued)

Table 6 Continued.

Species, instrument [Refs]	Measurement Property	Total number of studies	Methodological quality: number of studies	Summarized quality of results	Overall quality of evidence
	Criterion validity	1	D:1	-	Low
	Construct validity	0	No studies available	?	?
	Responsiveness	0	No studies available	?	?
Dog, UMPS[36]	General design requirements and development	1	A:1	+	Moderate
	Internal consistency	0	No studies available	?	?
	Reliability	0	No studies available	?	?
	Measurement error	0	D:1	?	?
	Criterion validity	0	No studies available	?	?
	Construct validity	1	D:1	+	Low
	Responsiveness	0	No studies available	?	?
Dog, CMPS[30,31]	General design requirements and development	1	I:1	+	Low
	Internal consistency	1	V:1	−	High
	Reliability	0	No studies available	?	?
	Measurement error	0	No studies available	?	?
	Criterion validity	0	No studies available	?	?
	Construct validity	1	V:1	+	High
	Responsiveness	0	No studies available	?	?
Dog, CMPS-SF[32–34]	General design requirements and development	1	A:1	+	Moderate
	Internal consistency	0	No studies available	?	?
	Reliability	2	V:2	+/-	High
	Measurement error	2	A:2	?	?
	Criterion validity	1	V:1	?	?
	Construct validity	1	V:1	+	High
	Responsiveness	0	No studies available	?	?
Dog, modified CMPS[35]	General design requirements and development	0	No studies available	?	?
	Internal consistency	0	No studies available	?	?
	Reliability	0	No studies available	?	?
	Measurement error	0	No studies available	?	?
	Criterion validity	0	No studies available	?	?
	Construct validity	1	A:1	+	Moderate
	Responsiveness	0	No studies available	?	?
Dog, 4A-VET[37]	General design requirements and development	0	No studies available	?	?
	Internal consistency	1	V:1	+	High
	Reliability	1	I:1	?	Very Low
	Measurement error	0	No studies available	?	?
	Criterion validity	0	D:1	?	?
	Construct validity	1	A:1	−	Moderate
	Responsiveness	0	No studies available	?	?
Dog, modified 4A-VET[38]	General design requirements and development	0	No studies available	?	?
	Internal consistency	0	No studies available	?	?
	Reliability	1	D:1	−	Low
	Measurement error	0	No studies available	?	?
	Criterion validity	0	D:1	?	?
	Construct validity	1	D:1	−	Low
	Responsiveness	0	No studies available	?	?
Cat/Dog, VAS[25,34,39]	General design requirements and development	0	No studies available	?	?

(continued)

Table 6 Continued.

Species, instrument [Refs]	Measurement Property	Total number of studies	Methodological quality: number of studies	Summarized quality of results	Overall quality of evidence
	Internal consistency	0	No studies available	?	?
	Reliability	2	V:1 D:1	+/−	High
	Measurement error	0	No studies available	?	?
	Criterion validity	1	V:1	−	High
	Construct validity	0	No studies available	?	?
	Responsiveness	0	No studies available	?	?
Cat/Dog, NRS[34,40]	General design requirements and development	0	No studies available	?	?
	Internal consistency	0	No studies available	?	?
	Reliability	1	V:1	+	High
	Measurement error	1	A:1	?	?
	Criterion validity	1	V:1	+	High
	Construct validity	0	No studies available	?	?
	Responsiveness	0	No studies available	?	?
Cat and Dog, SDS[34]	General design requirements and development	0	No studies available	?	?
	Internal consistency	0	No studies available	?	?
	Reliability	1	I:1	−	Very low
	Measurement error	0	No studies available	?	?
	Criterion validity	0	No studies available	?	?
	Construct validity	0	No studies available	?	?
	Responsiveness	0	No studies available	?	?

Refs: References, written in superscript. Methodological quality: "V" (very good), "A" (adequate), "D" (doubtful) and "I" (inadequate). Quality of results: "+" (sufficient), "−" (insufficient), "+/−" (inconsistent), or "?" (indeterminate). Quality of Evidence: High, Moderate, Low, Very Low or Unknown. Note: Data retrieved from the articles included in this systematic review and reported herein are subject to bias or error attributable to any misinterpretation or unclear reporting of the results. Abbreviations: FGS = Feline Grimace Scale; MCPS = UNESP-Botucatu Multidimensional Pain Scale; UFEPS-SF = UNESP-Botucatu Feline Pain Scale- Short Form; rCMPS-F = Revised Glasgow Composite Measure Pain Scale—Feline; CMPS-F = Glasgow Composite Measure Pain Scale—Feline; CSU-FAPS = Colorado State University Feline Acute Pain Scale; CMPS = Glasgow Composite Measure Pain Scale—Canine; CMPS-SF = Glasgow Composite Measure Pain Scale—Canine, Short Form; UMPS = University of Melbourne Pain Scale; 4A-VET = 4A-Vet Pain Scale; VAS = Visual Analog Scale; NRS = Numerical Rating Scale; SDS = Simple Descriptive Scale.

assessments might be biased due to the subjective nature of data and methodology evaluation. An attempt to reduce bias was made by using 2 investigators. Additionally, the lack of a third reviewer with expertise in the field is also a limitation that could have contributed to a biased interpretation when reaching consensus. Moreover, an important conceptual point is that an OMI is never absolutely validated as evidence accumulates over time as studies continue to test and support the instrument's measurement properties. Some information in the original manuscripts might have been missing or inadequately reported, which limits the ability to fairly and consistently assess the quality of evidence. Consequently, assessments might have been less reliable, particularly for instruments lacking sufficient validation studies. Instruments assessing pain in cats have been developed more recently compared to that in dogs. This temporal difference might reflect an improvement in the understanding of health measurement scale development and validation. Earlier instruments were often created through face validity. In contrast, more recent cat instruments, such as the FGS and the UFEPS-SF, were developed using robust psychometric principles and rigorous validation processes. The evolution of methodological standards should be considered when interpreting the findings of this review. The COSMIN guidelines are rigorous, but its criteria are not without limitations. COSMIN's latest iteration reflects continuing refinement of measurement properties and OMIs assessments, but it is important to acknowledge that the field of psychometrics is dynamic, and methodological frameworks and standards likely evolve over time. Inclusion criteria were restricted to manuscripts and instruments developed in English, which might reduce the generalizability of the findings to non-English speaking contexts. The decision to focus on English-language sources was made for practical and methodological reasons, as translation resources and linguistic expertise would have been required. Furthermore, focusing on instruments developed in English helped with consistency when evaluating measurement properties, as translation and cultural adaptation can introduce variability.[9] Although this exclusion might omit some relevant studies, this systematic review synthesizes the available English-language literature. Consequently, cross-cultural validity, which is a measurement property assessing the extent to which the translated or culturally adapted version of the OMI appropriately reflects the original version,[9] was not evaluated. Future studies might assess the cross-cultural validity, as different language versions (ie, Spanish, Italian) of the OMIs are currently available

for evaluating acute pain in cats and dogs.[45,46] This systematic review did not address interpretability and feasibility, which COSMIN identifies as essential for the practical application of OMIs, even though they are not technically classified as measurement properties. Interpretability pertains to how clearly scores from a pain assessment tool can be understood within a practical or clinical framework.[9] Instruments scoring acute pain should focus on offering guidance on interpreting both individual scores and changes over time. Feasibility considers factors like usability, required training, and time needed for application, and this should also be regarded as part of development studies.[9] Furthermore, establishing specific thresholds for analgesic interventions through receiver operating characteristic curve analysis between painful and pain-free individuals could further enhance clinical utility by identifying cut-offs that signal the need for the administration of analgesia.[42]

Conclusion

Using the COSMIN guidelines, the measurement properties of 15 instruments scoring acute pain in cats and dogs were evaluated. The MCPS, UFEPS-SF and FGS demonstrated high-quality evidence and findings in cats. Overall, the present evidence for measurement properties of instruments assessing pain in cats is superior to in dogs. The lack of a universally accepted gold standard, inconsistencies in methodological quality and incomplete reporting of measurement properties limited the comparability and reliability of many instruments. Our findings provide valuable insights into the current landscape of OMIs assessing acute pain in cats and dogs and emphasize the need for rigorous, standardized validation processes for practical clinical use.

Author contributions

Jungyoon Lee (Conceptualization, Data curation, Formal analysis, Investigation, Validation, Writing—original draft). Paulo V Steagall (Conceptualization, Data curation, Formal analysis, Investigation, Validation, Review original draft).

Conflicts of interest

Authors declare no conflicts of interest.

Funding

This study was awarded with the Morris Animal Foundation Veterinary Student Scholarship (recipient Jungyoon Lee; 25FEANVS-0014).

Off-label antimicrobial declaration

Authors declare no off-label use of antimicrobials.

Institutional animal care and use committee or other approval declaration

Authors declare no institutional animal care and use committee or other approval was needed.

Human ethics approval declaration

Authors declare human ethics approval was not needed.

References

1. Monteiro BP, Lascelles BDX, Murrell J, Robertson S, Steagall PVM, Wright B. 2022 WSAVA guidelines for the recognition, assessment and treatment of pain. *J Small Anim Pract*. 2023;64:177-254. https://doi.org/10.1111/jsap.13566
2. Mellor DJ, Beausoleil NJ, Littlewood KE, et al. The 2020 five domains model: including human-animal interactions in assessments of animal welfare. *Animals (Basel)*. 2020;10:10. https://doi.org/10.3390/ani10101870
3. Steagall PV, Monteiro BP. Acute pain in cats: recent advances in clinical assessment. *J Feline Med Surg*. 2019;21:25-34. https://doi.org/10.1177/1098612X18808103
4. Gruen ME, Lascelles BDX, Colleran E, et al. 2022 AAHA pain management guidelines for dogs and cats. *J Am Anim Hosp Assoc*. 2022;58:55-76. https://doi.org/10.5326/JAAHA-MS-7292
5. Treede R-D, Rief W, Barke A, et al. Chronic pain as a symptom or a disease: the IASP classification of chronic pain for the international classification of diseases (ICD-11). *PAIN*. 2019;160:19-27. https://doi.org/10.1097/j.pain.0000000000001384
6. Tomacheuski RM, Monteiro BP, Evangelista MC, Luna SPL, Steagall PV. Measurement properties of pain scoring instruments in farm animals: a systematic review using the COSMIN checklist. *PLoS One*. 2023;18:e0280830. https://doi.org/10.1371/journal.pone.0280830
7. Evangelista MC, Monteiro BP, Steagall PV. Measurement properties of grimace scales for pain assessment in nonhuman mammals: a systematic review. *PAIN*. 2022;163:e697-e714. https://doi.org/10.1097/j.pain.0000000000002474
8. Steagall PV, Robertson S, Simon B, Warne LN, Shilo-Benjamini Y, Taylor S. 2022 ISFM consensus guidelines on the management of acute pain in cats. *J Feline Med Surg*. 2022;24:4-30. https://doi.org/10.1177/1098612X211066268
9. Mokkink LB, Elsman EBM, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures version 2.0. *Qual Life Res*. 2024;33:2929-2939. https://doi.org/10.1007/s11136-024-03761-6
10. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27:1147-1157. https://doi.org/10.1007/s11136-018-1798-3
11. Elsman EBM, Mokkink LB, Terwee CB, et al. Guideline for reporting systematic reviews of outcome measurement instruments (OMIs): PRISMA-COSMIN for OMIs 2024. *Qual Life Res*. 2024;33:2029-2046. https://doi.org/10.1007/s11136-024-03634-y
12. Mokkink LB, Boers M, van der Vleuten CPM, et al. COSMIN risk of bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. *BMC Med Res Methodol*. 2020;20:293. https://doi.org/10.1186/s12874-020-01179-5
13. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic

reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2012;21:651-657. https://doi.org/10.1007/s11136-011-9960-1
14. Brozek JL, Canelo-Aybar C, Akl EA, et al. GRADE guidelines 30: the GRADE approach to assessing the certainty of modeled evidence-an overview in the context of health decision-making. *J Clin Epidemiol*. 2021;129:138-150. https://doi.org/10.1016/j.jclinepi.2020.09.018
15. Evangelista MC, Watanabe R, Leung VSY, et al. Facial expressions of pain in cats: the development and validation of a Feline Grimace Scale. *Sci Rep*. 2019;9:19128. https://doi.org/10.1038/s41598-019-55693-8
16. Evangelista MC, Benito J, Monteiro BP, et al. Clinical applicability of the Feline Grimace Scale: real-time versus image scoring and the influence of sedation and surgery. *PeerJ*. 2020;8:e8967. https://doi.org/10.7717/peerj.8967
17. Watanabe R, Doodnaught GM, Evangelista MC, Monteiro BP, Ruel HLM, Steagall PV. Inter-Rater reliability of the Feline Grimace Scale in cats undergoing dental extractions. *Frontiers in Veterinary Science*. 2020;7:302. https://doi.org/10.3389/fvets.2020.00302
18. Cheng AJ, Malo A, Garbin M, Monteiro BP, Steagall PV. Construct validity, responsiveness and reliability of the Feline Grimace Scale in kittens. *J Feline Med Surg*. 2023;25:1098612x231211765. https://doi.org/10.1177/1098612X231211765
19. Adami C, Filipas M, John C, Skews K, Dobson E. Inter-observer reliability of three feline pain scales used in clinical practice. *J Feline Med Surg*. 2023;25:1098612X231194423. https://doi.org/10.1177/1098612X231194423
20. Monteiro BP, Lee NH, Steagall PV. Can cat caregivers reliably assess acute pain in cats using the Feline Grimace Scale? A large bilingual global survey. *J Feline Med Surg*. 2023;25:1098612x221145499. https://doi.org/10.1177/1098612X221145499
21. Evangelista MC, Steagall PV. Agreement and reliability of the Feline Grimace Scale among cat owners, veterinarians, veterinary students and nurses. *Sci Rep*. 2021;11:5262. https://doi.org/10.1038/s41598-021-84696-7
22. Brondani JT, Luna SP, Padovani CR. Refinement and initial validation of a multidimensional composite scale for use in assessing acute postoperative pain in cats. *Am J Vet Res*. 2011;72:174-183. https://doi.org/10.2460/ajvr.72.2.174
23. Brondani JT, Mama KR, Luna SPL, et al. Validation of the English version of the UNESP-Botucatu multidimensional composite pain scale for assessing postoperative pain in cats. *BMC Vet Res*. 2013;9:143. https://doi.org/10.1186/1746-6148-9-143
24. Belli M, de Oliveira AR, de Lima MT, Trindade PHE, Steagall PV, Luna SPL. Clinical validation of the short and long UNESP-Botucatu scales for feline pain assessment. *PeerJ*. 2021;9:e11225. https://doi.org/10.7717/peerj.11225
25. Benito J, Monteiro BP, Beauchamp G, Lascelles BDX, Steagall PV. Evaluation of interobserver agreement for postoperative pain and sedation assessment in cats. *J Am Vet Med Assoc*. 2017;251:544-551. https://doi.org/10.2460/javma.251.5.544
26. Luna SPL, Trindade PHE, Monteiro BP, et al. Multilingual validation of the short form of the Unesp-Botucatu Feline Pain Scale (UFEPS-SF). *PeerJ*. 2022;10:e13134. https://doi.org/10.7717/peerj.13134
27. Calvo G, Holden E, Reid J, et al. Development of a behaviour-based measurement tool with defined intervention level for assessing acute pain in cats. *J Small Anim Pract*. 2014;55:622-629. https://doi.org/10.1111/jsap.12280
28. Reid J, Scott EM, Calvo G, Nolan AM. Definitive Glasgow acute pain scale for cats: validation and intervention level. *Vet Rec*. 2017;180:449-449. https://doi.org/10.1136/vr.104208
29. Shipley H, Guedes A, Graham L, Goudie-DeAngelis E, Wendt-Hornickle E. Preliminary appraisal of the reliability and validity of the Colorado State University Feline Acute Pain Scale. *J Feline Med Surg*. 2019;21:335-339. https://doi.org/10.1177/1098612X18777506
30. Holton L, Reid J, Scott EM, et al. Development of a behaviour-based scale to measure acute pain in dogs. *Vet Rec*. 2001;148:525-531. https://doi.org/10.1136/vr.148.17.525
31. Morton CM, Reid J, Scott EM, Holton LL, Nolan AM. Application of a scaling model to establish and validate an interval level pain scale for assessment of acute pain in dogs. *Am J Vet Res*. 2005;66:2154-2166. https://doi.org/10.2460/ajvr.2005.66.2154
32. Reid J, Nolan AM, Hughes JML, Lascelles D, Pawson P, Scott EM. Development of the short-form Glasgow Composite Measure Pain Scale (CMPS-SF) and derivation of an analgesic intervention score. *Anim Welfare*. 2007;16:97-104. https://doi.org/10.1017/S096272860003178X
33. Marco-Martorell M, Duffy N, Martinez M, Maddox T, Robson K. Agreement of pain assessment using the short form of the Canine Glasgow Composite Measure Pain Scale between veterinary students, veterinary nurses, veterinary surgeons, and ECVAA-diplomates. *Animals*. 2024;14:2310. https://doi.org/10.3390/ani14162310
34. Hofmeister EH, Barletta M, Shepard M, Brainard BM, Trim CM, Quandt J. Agreement among anesthesiologists regarding postoperative pain assessment in dogs. *Vet Anesth Analg*. 2018;45:695-702. https://doi.org/10.1016/j.vaa.2018.04.001
35. Murrell JC, Psatha EP, Scott EM, Reid J, Hellebrekers LJ. Application of a modified form of the Glasgow pain scale in a veterinary teaching centre in the Netherlands. *Vet Rec*. 2008;162:403-408. https://doi.org/10.1136/vr.162.13.403
36. Firth AM, Haldane SL. Development of a scale to evaluate postoperative pain in dogs. *J Am Vet Med Assoc*. 1999;214:651-659. https://doi.org/10.2460/javma.1999.214.05.651
37. Rialland P, Authier S, Guillot M, et al. Validation of orthopedic postoperative pain assessment methods for dogs: a prospective, blinded, randomized, placebo-controlled study. *PLoS One*. 2012;7:e49480. https://doi.org/10.1371/journal.pone.0049480
38. Guillot M, Rialland P, Nadeau M, del Castillo JRE, Gauvin D, Troncy E. Pain induced by a minor medical procedure (bone marrow aspiration) in dogs: comparison of pain scales in a pilot study. *J Vet Intern Med*. 2011;25:1050-1056. https://doi.org/10.1111/j.1939-1676.2011.00786.x
39. Holton LL, Scott EM, Nolan AM, et al. Comparison of three methods used for assessment of pain in dogs. *J Am Anim Hosp Assoc*. 1998;212:61-66. https://doi.org/10.2460/javma.1998.212.01.61
40. Holton LL, Scott EM, Nolan AM, Reid J, Welsh E. Relationship between physiological factors and clinical pain in dogs scored using a numerical rating scale. *J Small Anim Pract*. 1998;39:469-474. https://doi.org/10.1111/j.1748-5827.1998.tb03681.x

41. Price DD, Bush FM, Long S, Harkins SW. A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. *PAIN*. 1994;56:217-226. https://doi.org/10.1016/0304-3959(94)90097-3
42. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A Practical Guide to Their Development and Use*. 6th ed. Oxford University Press; 2024. https://doi.org/10.1093/med/9780192869487.001.0001
43. Suzuki T. Does the combination use of two pain assessment tools have a synergistic effect? *J Intensive Care*. 2017;5:1. https://doi.org/10.1186/s40560-016-0195-7
44. Conzemius MG, Hill CM, Sammarco JL, Perkowski SZ. Correlation between subjective and objective measures used to determine severity of postoperative pain in dogs. *J Am Vet Med Assoc*. 1997;210:1619-1622. https://doi.org/10.2460/javma.1997.210.11.1619
45. Brondani J, Luna S, Crosignani N, et al. Validation of the Spanish version of the UNESP-Botucatu multidimensional composite pain scale to assess postoperative pain in cats. *Arch Med Vet*. 2014;46:477-486. https://doi.org/10.4067/S0301-732X2014000300020
46. Della Rocca G, Catanzaro A, Conti MB, et al. Validation of the Italian version of the UNESP-Botucatu multidimensional composite pain scale for the assessment of postoperative pain in cats. *Vet Ital*. 2018;54:49-61. https://doi.org/10.12834/VetIt.567.2704.22

Journal of Veterinary Internal Medicine, 2026, 1–15
https://doi.org/10.1093/jvimsj/aalaf062
Systematic Review