Received: 7 February 2025 | Accepted: 24 August 2025

DOI: 10.1111/evj.70103

ORIGINAL ARTICLE

Equine Veterinary Journal WILEY

Reliability of three scoring systems for assessing quality of anaesthetic induction in horses

Marta Villalba-Díez | Leire Benavente-Sánchez | Rocío Bustamante | Isabel Santiago-Llorente | María Villalba-Orero

Department of Animal Medicine and Surgery, Faculty of Veterinary Medicine, Universidad Complutense de Madrid, Madrid, Spain

Correspondence
María Villalba-Orero, Department of Animal Medicine and Surgery (Faculty of Veterinary Medicine) and Complutense Veterinary Teaching Hospital, Universidad Complutense de Madrid, Madrid, Spain.
Email: mvorero@ucm.es

Funding information
Ministerio de Ciencia, Innovación y Universidades, Grant/Award Number: FPU22/03984; Public-private agreement UCM-Boehringer Ingelheim, Grant/Award Number: EM-0000000620

Abstract

Background: Several induction quality scoring systems (IQSS) have been described to evaluate drugs and risk factors of this anaesthetic period in horses, but no attempts to compare their reliability have been performed.

Objectives: To elucidate the reliability of three IQSS: the visual analogue scale (VAS), a simple descriptive scale (SDS), and a composite grading scale (CGS) proposed by the authors.

Study Design: Reliability study.

Methods: Eight randomly selected video-recorded anaesthetic inductions from horses that underwent general anaesthesia were evaluated twice by four blinded evaluators with experience in equine anaesthesia, with a 1-month interval between assessments using the three aforementioned IQSS. A total of 64 evaluations per scale were generated. To assess reliability, intra- and inter-rater intraclass correlation coefficient (ICC), and their 95% confidence intervals (CI) were calculated based on a mean rating ($k = 4$), absolute agreement, 2-way random-effects model.

Results: The inter-rater agreement was classified as moderate to good inter-rater reliability for all the scales, with the highest ICC found for the VAS (0.74 ± 0.11), followed by the CGS and the SDS (0.65 ± 0.22 and 0.63 ± 0.21, respectively). Intra-rater agreement results demonstrated very good reliability for both VAS and SDS (0.82 ± 0.08; 0.81 ± 0.18, respectively) and excellent reliability for the CGS (0.91 ± 0.08).

Main Limitations: The use of video-recordings instead of in situ evaluations, as the absence of audio may affect the assessment. Additionally, these findings are applicable only when free inductions are evaluated.

Conclusions: The VAS and the novel CGS are reliable IQSS in horses, as are the widely used SDS. As the SDS are inconsistent across the literature, the VAS would be advised if multiple evaluators assess induction quality for research purposes, whereas the CGS would be selected for studies involving a single observer. We suggest routine inclusion of the VAS in the evaluation of the anaesthetic induction in horses.

KEYWORDS
anaesthesia, horse, induction, reliability, scale

1 | INTRODUCTION

Induction of general anaesthesia in horses is a dangerous phase because of their size, weight, and temperament.[1,2] Controlled and smooth induction is important as the horse can injure itself and personnel, especially if the drugs used fail to provide a rapid and predictable sedation, hypnosis, and muscle relaxation.[3] Poor induction quality is associated with poor anaesthetic recoveries,[4,5] the highest anaesthetic mortality related period in horses.[6]

Several induction quality scoring systems (IQSS) have been described for equids to evaluate the efficacy of induction drugs and identify the main risk factors for injury during this anaesthetic period. The visual analogue scale (VAS) is a 10 cm scoring line with endpoints representing the worst and best possible quality. Although it is considered a simple and reliable scale for subjective events,[7] used in numerous clinical studies in horses,[8,9] it has not been utilised frequently to evaluate anaesthetic induction in equids.[5,7,10–14] Simple descriptive scales (SDS) are commonly used for assessing the anaesthetic induction in equids,[13–17] These scales evaluate ataxia, transition to lateral recumbency, the time until recumbency, muscle activity or rigidity, pedalling movements, and danger posed to the horse or handler during the entire period.[13,15,17,18] However, the descriptors and number of levels included in SDS vary widely, with five-level scales being the most frequently used.[14,16–18] In terms of reliability, inter-rater agreement has been reported to range from substantial to almost perfect for VAS when used as IQSS, and fair for SDS.[10,13] VAS have shown a greater sensitivity compared with SDS in both clinical and research settings.[7]

There is an extensive literature evaluating the quality of anaesthetic recovery in horses, aiming to improve this phase and reduce its morbidity and mortality,[15,19,20] and more complex scales, such as composite grading scales (CGS), have been developed.[21,22] CGS assign a score to individual features of the recovery period, which are summed to produce an overall score.[23] These user-friendly scales, not yet developed for assessing the quality of induction in horses, allow a detailed analysis of the different components of the recovery period. Although still subjective, they provide a more reliable methodology.[21] Despite the advantages of CGS, such scales have not been incorporated into routine evaluation of the anaesthetic induction phase. Various scoring systems have been validated for the recovery phase,[9,23,24] but the induction phase has not been studied.[25–28] Most studies on IQSS have involved only a single observer and, when multiple observers were included, only two studies assessed inter-rater agreement,[10,13] and none addressed intra-rater reliability or compared different IQSS for reliability.

Before any assessment tool can be used for research or clinical applications, its reliability must be established.[29] The main objective of this study was to develop a CGS for evaluating the induction phase in horses and compare the reliability of three IQSS (VAS, SDS, and the novel CGS). We hypothesised that the VAS would demonstrate better inter- and intra-rater reliability compared with SDS and that the novel CGS would exhibit good inter- and intra-rater reliability.

2 | MATERIALS AND METHODS

2.1 | Study design

This study was conducted at the Complutense Veterinary Teaching Hospital. Eight video-recorded anaesthetic inductions, performed in a padded box, were randomly selected (www.randomizer.org) from a population of 235 client-owned horses that underwent general anaesthesia for elective and emergency procedures. The inclusion criteria for the study were: (i) recordings of unassisted anaesthetic inductions, (ii) visibility sufficient to observe drug administration, and (iii) the ability to observe the entire horse for at least 10 s after recumbency was achieved. The footage was edited to include part of the sedation prior to induction and a few seconds after the recumbency was reached, resulting in a video of <2.5 min to minimise evaluator fatigue. Any potential identification marks (e.g. date, time, induction box number) were removed. The video recordings were then distributed to the raters via a file-sharing cloud drive, with each evaluator receiving a unique order of video visualisation determined randomly to eliminate sequence bias (www.randomizer.org).

Data collected from the horses corresponding to the video recordings included age, weight, sex (male, female or castrated), breed, and American Society of Anesthesiologists (ASA) physical health status classification.

2.2 | Induction evaluation

General anaesthesia were evaluated twice by four blinded evaluators with experience in equine anaesthesia. The group consisted of one ECVAA resident, one equine anaesthetist with over 20 years of experience, and two predoctoral equine veterinarians specialising in anaesthesia. The raters were asked to familiarise themselves with the IQSS before performing the evaluation. To aid familiarisation, they were advised to apply the IQSS in their routine clinical anaesthetic cases prior to the video assessments, to become familiar with the key components of each scoring system. They were permitted to watch each video twice to fill in the evaluation form, but they were not instructed to use each IQSS in any specific order during the assessment of each video. All raters assessed the eight video recordings twice, with a 1-month interval between evaluations. All ratings were conducted independently.

Three IQSS were used (VAS, SDS and CGS). The VAS was a 10 cm line with the verbal descriptors 'worst induction imaginable' and 'perfect induction' at the 0 and 10 points, respectively. The criteria for using the VAS included transition to lateral recumbency, presence of limb movements or paddling, limb rigidity and danger to horse/handler.[17,18] The score was calculated as the distance (cm) from the left extremity to the assessor's mark. A VAS was also used to assess the quality of sedation prior to induction with the verbal descriptors 'worst sedation imaginable' and 'perfect sedation' at the 0 and 10 points, respectively; and using criteria such as the need for physical restraint after sedation, lowering of the head, degree of ataxia and spontaneous movements.[7] The SDS used was the five-level scale proposed by Mama et al.[18] where 1 represented vigorous struggling, paddling of the limbs and increased coordinated muscle activity during the transition to lateral recumbency, while 5 represented a smooth, timely collapse to lateral recumbency with good muscle relaxation. Finally, a novel 40-point CGS described in Table 1 was used. This scoring system is based on selecting the single descriptor that best matched the observed behaviour or condition in each section of the scale. Each descriptor was associated with a specific numerical value, and only one option can be selected per section. The assigned values are then summed to produce an overall score, which ranges from 3 (indicating the best quality induction) to 40 (indicating the poorest quality induction). The weighting of each descriptor was determined according to the criteria considered relevant in the most widely used IQSS for horses,[17,18] and adapted from previously described CGS for the assessment of recovery in equine anaesthesia.[21,22] In terms of sedation, the ideal level is considered to be moderate sedation. This is followed by deep sedation, which, although often effective, may exacerbate ataxia and affect induction negatively. The lowest level of sedation—light or barely perceptible—is scored highest, as insufficient sedation is a recognised cause of induction failure in horses.[2] For transition to recumbency, the optimal scenario involves a smooth and methodical descent followed by cases in which the horse takes a few steps before gently lying down. Poorer scores are given when the free induction becomes uncontrollable, although without posing a risk of injury to the handler or horse. The worst-case scenario involves a hazardous fall that could lead to injury for either party. Once the horse is in lateral recumbency, the ideal condition is a calm, immobile patient. A higher score is assigned when muscle stiffness is observed (in the limbs or neck), although no voluntary movement occurs. The next level includes mild, non-threatening movements (e.g., tongue, head or slight limb motion without paddling), which are indicative of a less optimal induction. The worst score in this section corresponds to evident paddling movements. Finally, an additional section allows for the inclusion of points if redosing with the induction agent was required. This evaluation is limited to the first few seconds post-induction, as the video recordings were edited to include only this timeframe. In this way, redosing is attributed to an inadequate induction or failure to achieve intubation, rather than to more subjective clinical preferences during the transfer of the patient to the surgical table.

TABLE 1 Proposed composite grading scale for assessing the quality of anaesthesia induction in horses.

I. Quality of sedation before induction (10)	
1	Marked sedation (moderate ataxia)
3	Profound sedation (severe ataxia)
7	Mild sedation (head lowered but auditive stimuli response)
10	No apparent sedation (raised head with auditive stimuli response)
II. Transition to lateral recumbency (10)	
1	Smooth, methodical
3	Few steps before recumbency, controlled
7	Uncontrolled, no danger to horse/handler
10	Uncontrolled, danger to horse/handler
III. Recumbency (10)	
1	Calm and quiet, without movements
3	Presence of rigidity or muscle fasciculations
7	Presence of mild movements, no danger to horse/handler
10	Presence of paddling, danger to horse/handler
IV. Need of an additional dose of induction drug (10)	
0	No
10	Yes
	Total score

3 | DATA ANALYSIS

The sample size was calculated using the Granmo calculator (https://www.datarus.eu/aplicaciones/granmo/), accepting an alpha risk of 0.05 and a beta risk of 0.2 in a two-sided test, to detect a difference ≥10% in the interclass correlation coefficient (ICC), using a standard deviation (SD) of 10%. The SD was chosen based on a previous study that compared two IQSS in dogs (a VAS and a SDS).[30] A sample size of eight video records, evaluated twice by the 4 raters, was determined, giving a total of 64 evaluations per scale.

Normality of the distribution was tested using the Shapiro–Wilk test for demographic data, and data were expressed as mean ± SD or median and percentiles (25th–75th), as appropriate. ICC estimates and their 95% confidence intervals (CI) were calculated based on a mean-rating ($k = 4$), absolute-agreement, 2-way random-effects model. For interpretation, ICC values were considered as follows: <0.4 as poor reliability, between 0.40 and 0.75 as moderate to good reliability, between 0.75 and 0.9 as very good reliability, and values >0.90 as excellent reliability.[31,32] All statistical analyses were conducted in IBM SPSS Statistics for Windows, Version 29.0.2.0 (20).

4 | RESULTS

4.1 | Descriptive analysis

All the randomly selected video-recorded inductions met the inclusion criteria. All variables followed a normal distribution, except for CGS

scores. The demographic data are presented in Table S1. The age and weight of the horses were 9 ± 6 years old and 466 ± 62 kg, respectively, and there were four stallions, three geldings and one mare. The most represented breed was Spanish Pure Breed, with four cases, followed by two crossbreeds, one hot-blooded, and one warm-blooded. Four cases (4) were considered as ASA grade I, two cases were ASA grade II and two cases were ASA grade III.

4.2 | Inter and intra-rater agreement

Inter-rater agreement is shown in Figure 1. VAS had moderate to good reliability (0.74 ± 0.11). The CGS and the SDS had moderate to good reliability (0.65 ± 0.22 and 0.63 ± 0.21, respectively). Individual ICCs obtained between the four raters and their 95% CI are shown in Table 2.

Intra-rater agreement results (Figure 2) demonstrated very good reliability for both VAS and SDS, with similar ICC values (0.82 ± 0.08; 0.81 ± 0.18, respectively); and excellent reliability for the CGS (0.91 ± 0.08). The individual ICCs and their 95% CI obtained from the two evaluations performed by the same rater are shown in Table 3.

5 | DISCUSSION

In this study, three different IQSS were evaluated to provide an overall view of their reliability and reproducibility and hence facilitate investigators' decisions when designing anaesthetic induction research in horses. Since inter-observer and intra-observer reliability reflect the reproducibility of a scoring system,[33] these results highlight that the VAS, although the simplest IQSS, is the most reproducible IQSS when multiple observers evaluate the anaesthetic induction in horses. This was supported by its higher inter-rater agreement. The novel CGS provides excellent repeatability when a single evaluator assesses this phase.

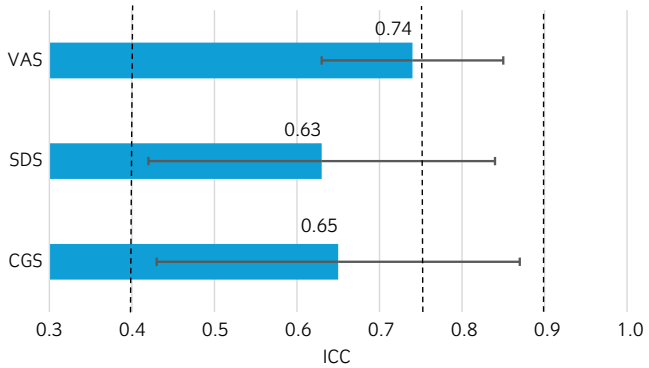

FIGURE 1 Average (mean ± SD) of the inter-rater ICC obtained for each anaesthetic induction quality scoring system assessed in a reliability study involving eight horses, evaluated twice by four raters. CGS, composite grading scale; SDS, simple descriptive scale; VAS, visual analogue scale. The dashed lines represent the cut-offs for ICC values interpretation as: poor reliability <0.4; moderate to good reliability = 0.4–0.75; very good reliability = 0.75–0.9; excellent reliability >0.9. ICC values were normally distributed using the Shapiro–Wilk test, so they are expressed as mean ± SD.

The VAS has been used as an IQSS in limited studies,[5,10–14] although it is considered a simple and reliable scale. It is more sensitive than present rating scales because it avoids defined levels that force observers to decide between a limited number of restricted categories.[7] Due to its simplicity, it is a widely used method for scoring subjective phenomena, and it has been successfully utilised in numerous clinical studies in horses, primarily related to pain scoring, but also for assessing recovery quality.[8,9,34] In this study, the highest agreement between observers was obtained using this scale, resulting in almost a very good reliability. Previous studies have reported inter-observer agreement between two raters of VAS as an IQSS in horses to range from substantial ($\kappa = 0.63$) to almost perfect ($r = 0.91$).[10,13] These findings, together with the current results, reinforce the importance of this scoring system, which has already been used successfully in other clinical contexts. A previous study using VAS for pain scoring reported a considerably variable inter-rater reliability, which was

TABLE 2 Individual inter-rater intraclass correlation coefficients (ICC) and 95% confidence interval calculation for each anaesthetic induction quality scoring system assessed in a reliability study involving eight horses, evaluated twice, by four raters.

		95% confidence interval	
	ICC	Lower bound	Upper bound
Simple Descriptive Scale			
R1-R2	**0.553**	−0.225	0.847
R1-R3	**0.906**	0.597	0.971
R1-R4	**0.689**	−0.226	0.912
R2-R3	**0.457**	−0.259	0.807
R2-R4	**0.373**	−0.145	0.770
R3-R4	**0.847**	0.171	0.957
Visual Analogue Scale			
R1-R2	**0.812**	0.479	0.934
R1-R3	**0.917**	0.760	0.971
R1-R4	**0.741**	0.086	0.917
R2-R3	**0.604**	−0.080	0.859
R2-R4	**0.684**	−0.233	0.915
R3-R4	**0.678**	0.121	0.885
Composite Grading Scale			
R1-R2	**0.701**	0.047	0.901
R1-R3	**0.967**	0.901	0.989
R1-R4	**0.663**	−0.239	0.904
R2-R3	**0.619**	−0.096	0.869
R2-R4	**0.261**	−0.208	0.663
R3-R4	**0.711**	−0.119	0.913

Note: R1, rater 1; R2, rater 2; R3, rater 3; R4, rater 4; ICC was calculated based on a mean-rating ($k = 4$), absolute-agreement, 2-way random-effects model. Bold indicates individual inter-rater intraclass correlation coefficients.

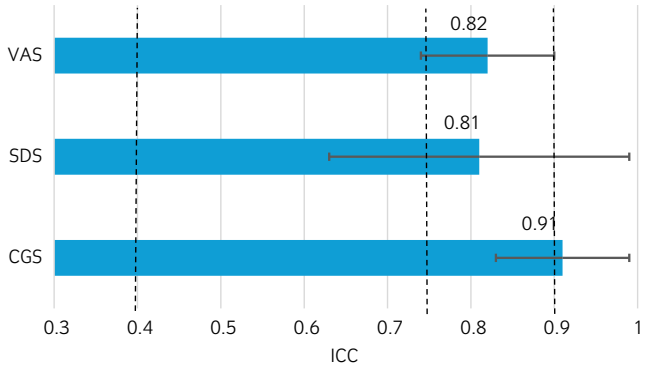

FIGURE 2 Average (mean ± SD) of the intra-rater ICC obtained for each anaesthetic induction quality scoring system assessed in a reliability study involving eight horses, evaluated twice by four raters. CGS, composite grading scale; SDS, simple descriptive scale; VAS, visual analogue scale. The dashed lines represent the cut-offs for ICC values interpretation as: poor reliability <0.4; moderate to good reliability = 0.4–0.75; very good reliability = 0.75–0.9; excellent reliability >0.9. ICC values were normally distributed using the Shapiro–Wilk test, so they are expressed as mean ± SD.

attributed to the influence of observation duration.[35] This factor is unlikely to influence its use as an IQSS because the anaesthetic induction phase has a short duration. Furthermore, implementing this scale for anaesthetic induction assessment would facilitate standardised measures in teaching hospitals, as experience does not appear to affect VAS scores.[9,36]

On the other hand, most studies that use an IQSS use SDS types,[13,14,16,37] but these are highly variable, with definitions, the number of levels, and considered criteria differing significantly. This inconsistency impedes comparisons between studies. Results of the present study showed the lowest inter-and intra-rater agreement of SDS compared with the other scales. A fair inter-rater agreement for SDS has been previously described for both induction ($\kappa = 0.38$) and recovery quality evaluation ($\kappa = 0.19$),[13,23] likely associated with the lack of consensus when assigning brief and enclosed definitions to describe such a complex situation. Similarly, SDS has been identified as the least sensitive scale for detecting discrete changes in pain in horses when compared with VAS.[38] In addition, the training required to interpret SDS makes them less reliable when inexperienced observers perform the evaluation.[39] Therefore, SDS are limited in assessing intricate clinical events. The SDS proposed by Mama et al.[18] was chosen for this study as it is the most commonly used scale for assessing equine anaesthetic inductions.[15,40–42]

This is the first report describing a CGS for anaesthetic induction evaluation in horses. The development of this scale was based on previous scales used for assessing recovery of anaesthesia in horses[21,22] and on the criteria of the most utilised IQSS.[17,18] The level of sedation was also taken into account because an inadequate sedation is considered to be a cause of induction failure, as horses that do not respond appropriately to sedative drugs are likely to exhibit poor response to injectable anaesthetics.[2] Good inter-rater and an excellent intra-rater reliability was found, making it a feasible and reproducible IQSS for use in horses. When compared with VAS, the inter-rater agreement observed for CGS was lower, probably due to its complexity, as ease of use is inversely related to the reliability of a scale.[24] Nevertheless, the excellent intra-rater reliability described here suggests that CGS is likely the most suitable IQSS for situations where a single anaesthesiologist assesses induction quality.

As expected, intra-observer reliability was higher than inter-observer reliability for all the scoring systems evaluated, a finding consistent with previous reports.[39] Interestingly, in this study, the VAS showed a very good intra-rater reliability, in contrast to the excellent reliability determined for the CGS. This might be linked to the continuous nature of the VAS measurement, which increases the likelihood of variation in a single rater's assessments and allows small differences to be recorded at the expense of reduced intra-rater reproducibility.[36] In contrast, SDS and CGS are categorical scales, which may limit scoring variability and dispersion.

One of the raters (rater 2), despite being an experienced clinician, had notably lower agreement indices both in comparison with the other raters and between his/her own repeated evaluations. While this might be partially explained by individual interpretation of the scoring systems, it may also reflect a potential learning or adaptation effect throughout the evaluation process. This highlights how prior experience does not always guarantee consistency when using new or subjective scoring systems and reinforces the

TABLE 3 Intraclass correlation coefficients (ICC) and 95% confidence interval calculation for each anaesthetic induction quality scoring system assessed in a reliability study involving eight horses, evaluated twice, by four raters.

		95% confidence interval	
	ICC	Lower bound	Upper bound
Simple Descriptive Scale			
R1	**0.966**	0.847	0.993
R2	**0.563**	−1.019	0.911
R3	**0.920**	0.549	0.984
R4	**0.809**	0.159	0.961
Visual Analogue Scale			
R1	**0.918**	0.621	0.983
R2	**0.820**	0.007	0.965
R3	**0.729**	−0.283	0.948
R4	**0.814**	0.008	0.963
Composite Grading Scale			
R1	**0.985**	0.929	0.997
R2	**0.849**	0.211	0.970
R3	**0.967**	0.833	0.993
R4	**0.839**	0.114	0.968

Note: R1, rater 1; R2, rater 2; R3, rater 3; R4, rater 4; ICC was calculated based on a mean-rating ($k = 4$), absolute-agreement, 2-way random-effects model. Bold indicates individual inter-rater intraclass correlation coefficients.

importance of specific training and familiarisation when implementing such tools. A formal training session prior to data collection might have reduced this bias source.

Limitations of this study include the absence of audio in the video recordings, as sound can help to determine the force with which animals collide with the induction box walls.[9] The use of video recordings was essential for assessing intra-observer agreement, but inter-observer results might differ when anaesthetic inductions are assessed in situ, as the lack of adequate visualisation and the development of fatigue after assessing several video recordings could affect the rater's evaluation. The small sample size was calculated based on ICC results from previous studies in dogs, as there was no available data on the ICC of IQSS in horses. Additionally, these findings are applicable only when free inductions are assessed. On the other hand, no assessment of the consistency of the novel CGS was performed prior to this study, as the aim of this study was merely to introduce a potential new method of assessment based on similar recovery CGS. The precision and interpretability of the inter-rater ICC value for the CGS can be affected by its deviation from normality and the presence of potential outliers. The observers involved had varying levels of experience, which may have influenced the reliability of the scores as rater experience could be a relevant factor in future assessments. Finally, this was a single-centre study, which may limit the generalisability of the results to other clinical settings.

In conclusion, in research studied on induction quality, the VAS is recommended for use by multiple evaluators, whereas the CGS is appropriate for studies involving a single observer. Based on these results, for routine clinical use, we suggest the VAS should be used to evaluate the anaesthetic induction in horses.

FUNDING INFORMATION

Ministry of Science, Innovation and Universities; Training of University Professors Grant (FPU22/03984). Public-private agreement UCM-Boehringer Ingelheim_EM-0000000620.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

Leire Benavente-Sánchez: Methodology; writing – review and editing; visualization. **Rocío Bustamante:** Writing – review and editing; methodology; visualization. **Isabel Santiago-Llorente:** Conceptualization; investigation; funding acquisition; writing – review and editing; methodology; visualization; resources. **María Villalba-Orero:** Conceptualization; investigation; funding acquisition; methodology; formal analysis; writing – review and editing; supervision; resources. **Marta Villalba-Díez:** Writing – original draft; formal analysis.

DATA INTEGRITY STATEMENT

Marta Villalba Díez had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

ETHICAL ANIMAL RESEARCH

The director of the centre signed an authorisation for research use of Complutense Veterinary Teaching Hospital clinical data.

INFORMED CONSENT

Explicit owner consent for inclusion of animals in this study was not obtained. Owners or their agents were made aware that case information may be used for research in general.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/evj.70103.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Figshare (doi: 10.6084/m9.figshare.28380278).

ORCID

Marta Villalba-Díez https://orcid.org/0000-0001-6088-0837

REFERENCES

1. Brock N, Hildebrand SV. A comparison of xylazine-diazepam-ketamine and xylazine-guaifenesin-ketamine in equine anesthesia. Vet Surg. 1990;19:468–74. https://doi.org/10.1111/j.1532-950x.1990.tb01237.x
2. Hubbell JAE, Muir WW, Gorenberg E, Hopster K. A review of equine anesthetic induction: are all equine anesthetic inductions 'crash' inductions? J Equine Vet Sci. 2024;139:105130. https://doi.org/10.1016/j.jevs.2024.105130
3. Muir WW, Hubbell JAE. Equine anesthesia: monitoring and emergency therapy. 2nd ed. St. Louis: Saunders Elsevier; 2009.
4. Santiago-Llorente I. Variables anestésicas como predictores de mortalidad en caballos operados de cólico. 2016. https://docta.ucm.es/entities/publication/9ffda340-c5df-48c3-948f-02112878a9d1
5. Casoni D, Spadavecchia C, Wampfler B, Thormann W, Levionnois OL. Clinical and pharmacokinetic evaluation of S-ketamine for intravenous general anaesthesia in horses undergoing field castration. Acta Vet Scand. 2015;57:21. https://doi.org/10.1186/s13028-015-0112-4
6. Gozalo-Marcilla M, Bettschart-Wolfensberger R, Johnston M, Taylor PM, Redondo JI. Data collection for the fourth multicentre confidential enquiry of perioperative fatalities (CEPEF4) study: new technology and preliminary results. Animals. 2021;11:https://doi.org/10.3390/ani11092549
7. Corletto F, Raisis BAA, Brearley JC. Comparison of morphine and butorphanol as pre-anaesthetic agents in combination with romifidine for field castration in ponies. Vet Anaesth Analg. 2005;32:16–22. https://doi.org/10.1111/j.1467-2995.2004.00184.x
8. Alonso B, Carregaro A, Cuypers C, Michielsen A, Gasthuys F, Schauvliege S. Effects of detomidine or romifidine during maintenance and recovery from isoflurane anaesthesia in horses. Vet Anaesth Analg. 2022;49:624–33. https://doi.org/10.1016/j.vaa.2022.07.004
9. Farmer E, Chase-Topping M, Lawson H, Clutton RE. Factors affecting the perception of recovery quality in horses after anaesthesia. Equine Vet J. 2014;46:328–32. https://doi.org/10.1111/evj.12133
10. Ferreira TH, Brosnan RJ, Shilo-Benjamini Y, Moore SB, Hollingsworth SR. Effects of ketamine, propofol, or thiopental administration on intraocular pressure and qualities of induction of and recovery from anesthesia in horses. Am J Vet Res. 2013;74:1070–7. https://doi.org/10.2460/ajvr.74.8.1070

11. Hubbell JAE, Hinchcliff KW, Schmall M, Muir WW, Robertson JT, Sams RA. Anesthetic, cardiorespiratory, and metabolic effects of four intravenous anesthetic regimens induced in horses immediately after maximal exercise. Am J Vet Res. 2000;61:1545–52. https://doi.org/10.2460/ajvr.2000.61.1545
12. Larenza MP, Ringer SK, Kutter APN, Conrot A, Theurillat R, Kummer M, et al. Evaluation of anesthesia recovery quality after low-dose racemic or S-ketamine infusions during anesthesia with isoflurane in horses. Am J Vet Res. 2009;70:710–8. https://doi.org/10.2460/ajvr.70.6.710
13. Allison A, Robinson R, Jolliffe C, Taylor PM. Evaluation of the use of midazolam as a co-induction agent with ketamine for anaesthesia in sedated ponies undergoing field castration. Equine Vet J. 2018;50:321–6. https://doi.org/10.1111/evj.12759
14. Balko JA, Fogle C, Stuska SJ, Fogle JE, Posner LP. Retrospective and prospective assessment of butorphanol, azaperone and medetomidine (BAM™) for immobilisation of feral horses (Equus ferus caballus). Equine Vet J. 2022;54:549–55. https://doi.org/10.1111/evj.13490
15. Hector RC, Rezende ML, Mama KR, Hess AM. Recovery quality following a single post-anaesthetic dose of dexmedetomidine or romifidine in sevoflurane anaesthetised horses. Equine Vet J. 2020;52:685–91. https://doi.org/10.1111/evj.13238
16. Smith MC, Bass L, Damone J, Mama K, Rao S. Comparison of xylazine and detomidine in combination with midazolam/ketamine for field castration in quarter horses. Equine Vet J. 2020;52:516–21. https://doi.org/10.1111/evj.13232
17. Klöppel H, Leece EA. Comparison of ketamine and alfaxalone for induction and maintenance of anaesthesia in ponies undergoing castration. Vet Anaesth Analg. 2011;38:37–43. https://doi.org/10.1111/j.1467-2995.2010.00584.x
18. Mama KR, Steffey EP, Pascoe PJ. Evaluation of propofol as a general anesthetic for horses. Vet Surg. 1995;24:188–94. https://doi.org/10.1111/j.1532-950x.1995.tb01317.x
19. Kälin I, Henze IS, Ringer SK, Torgerson PR, Bettschart-Wolfensberger R. Comparison of recovery quality following medetomidine versus xylazine balanced isoflurane anaesthesia in horses: a retrospective analysis. Animals. 2021;11:2440. https://doi.org/10.3390/ani11082440
20. Gozalo-Marcilla M, Steblaj B, Schauvliege S, Duchateau L, Gasthuys F. Comparison of the influence of two different constant-rate infusions (dexmedetomidine versus morphine) on anaesthetic requirements, cardiopulmonary function and recovery quality in isoflurane anaesthetized horses. Res Vet Sci. 2013;95:1186–94. https://doi.org/10.1016/j.rvsc.2013.09.014
21. Clark-Price SC, Posner LP, Gleed RD. Recovery of horses from general anesthesia in a darkened or illuminated recovery stall. Vet Anaesth Analg. 2008;35:473–9. https://doi.org/10.1111/j.1467-2995.2008.00414.x
22. Donaldson LL, Dunlop GS, Holland MS, Burton BA. The recovery of horses from inhalant anesthesia: a comparison of halothane and isoflurane. Vet Surg. 2000;29:92–101. https://doi.org/10.1111/j.1532-950x.2000.00092.x
23. Clark-Price SC, Lascola KM, Carter JE, Da Cunha AF, Donaldson LL, Doherty TJ, et al. Assessment of agreement among diplomates of the American College of Veterinary Anesthesia and Analgesia for scoring the recovery of horses from anesthesia by use of subjective grading scales and development of a system for evaluation of the recovery of horses from anesthesia by use of accelerometry. Am J Vet Res. 2017;78:668676. https://doi.org/10.2460/ajvr.78.6.668
24. Suthers JM, Christley RM, Clutton RE. Quantitative and qualitative comparison of three scoring systems for assessing recovery quality after general anaesthesia in horses. Vet Anaesth Analg. 2011;38:352–62. https://doi.org/10.1111/j.1467-2995.2011.00629.x
25. Posner LP, Kasten JI, Kata C. Propofol with ketamine following sedation with xylazine for routine induction of general anaesthesia in horses. Vet Rec. 2013;173:550. https://doi.org/10.1136/vr.101906
26. Keates HL, van Eps AW, Pearson MR. Alfaxalone compared with ketamine for induction of anaesthesia in horses following xylazine and guaifenesin. Vet Anaesth Analg. 2012;39:591–8. https://doi.org/10.1111/j.1467-2995.2012.00756.x
27. Jarrett MA, Bailey KM, Messenger KM, Prange T, Gaines B, Posner LP. Recovery of horses from general anesthesia after induction with propofol and ketamine versus midazolam and ketamine. J Am Vet Med Assoc. 2018;253:101–7. https://doi.org/10.2460/javma.253.1.101
28. Brosnan RJ, Steffey EP, Escobar A, Palazoglu M, Fiehn O. Anesthetic induction with guaifenesin and propofol in adult horses. Am J Vet Res. 2011;72:1569–75. https://doi.org/10.2460/ajvr.72.12.1569
29. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15:155–63. https://doi.org/10.1016/j.jcm.2016.02.012
30. Wolfe KL, Hofmeister EH, Clark-Price SC, Reed R, Quandt J. Development of the Auburn Induction Scale for evaluating induction quality in dogs. Vet Anaesth Analg. 2022;49:608–14. https://doi.org/10.1016/j.vaa.2022.08.010
31. Fleiss JL. The design and analysis of clinical experiments. Wiley; 1999. https://onlinelibrary.wiley.com/doi/book/10.1002/9781118032923
32. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. 3rd ed. Upper Saddle River, NJ: Pearson/Prentice Hall; 2015.
33. van Loon JPAM, Van Dierendonck MC. Objective pain assessment in horses (2014–2018). Vet J. 2018;242:1–7. https://doi.org/10.1016/j.tvjl.2018.10.001
34. van Loon JPAM, Van Dierendonck MC. Monitoring acute equine visceral pain with the equine Utrecht University scale for composite pain assessment (EQUUS-COMPASS) and the equine Utrecht University scale for facial assessment of pain (EQUUS-FAP): a scale-construction study. Vet J. 2015;206:356–64. https://doi.org/10.1016/j.tvjl.2015.08.023
35. de Grauw JC, van Loon JPAM. Systematic pain assessment in horses. Vet J. 2016;209:14–22. https://doi.org/10.1016/j.tvjl.2015.07.030
36. Vettorato E, Chase-Topping ME, Clutton RE. A comparison of four systems for scoring recovery quality after general anaesthesia in horses. Equine Vet J. 2010;42:400–6. https://doi.org/10.1111/j.2042-3306.2010.00093.x
37. Mama KR, Steffey EP, Pascoe PJ. Evaluation of propofol for general anesthesia in premedicated horses. Am J Vet Res. 1996;57:512–6.
38. Ashley FH, Waterman-Pearson AE, Whay HR. Behavioural assessment of pain in horses and donkeys: application to clinical practice and future studies. Equine Vet J. 2010;37:565–75. https://doi.org/10.2746/042516405775314826
39. Viñuela-Fernández I, Jones E, Chase-Topping ME, Price J. Comparison of subjective scoring systems used to evaluate equine laminitis. Vet J. 2011;188:171–7. https://doi.org/10.1016/j.tvjl.2010.05.011
40. de Grauw J, van Loon T. Clinical effects of two doses of butorphanol with detomidine for intravenous premedication of healthy warmblood horses. Vet Anaesth Analg. 2020;47:681–5. https://doi.org/10.1016/j.vaa.2019.11.010
41. Tokushige H, Araki M, Kusano K, Arima D, Ito H, Yamazaki Y, et al. A retrospective comparison of induction with thiopental/guaifenesin and propofol/ketamine in thoroughbred racehorses anesthetized with sevoflurane and medetomidine during arthroscopic surgery. J Equine Sci. 2019;30:25–31. https://doi.org/10.1294/jes.30.25

42. Aoki M, Wakuno A, Kushiro A, Mae N, Kakizaki M, Nagata S, et al. Evaluation of total intravenous anesthesia with propofol-guaifenesin-medetomidine and alfaxalone-guaifenesin-medetomidine in Thoroughbred horses undergoing castration. J Vet Med Sci. 2017; 79:2011–8. https://doi.org/10.1292/jvms.16-0658

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Villalba-Díez M, Benavente-Sánchez L, Bustamante R, Santiago-Llorente I, Villalba-Orero M. Reliability of three scoring systems for assessing quality of anaesthetic induction in horses. Equine Vet J. 2025. https://doi.org/10.1111/evj.70103