

COMPARISON OF PROPOFOL AND KETAMINE/DIAZEPAM ON THE QUALITY OF  
RECOVERY OF DOGS UNDERGOING ORTHOPEDIC SURGERY

by

JENNIFER ELIZABETH COPELAND

(Under the Direction of ERIK HOFMEISTER)

ABSTRACT

Recovery is an important part of the anesthetic episode, as monitoring is not as focused and complications can arise. This study investigated if there was a difference between the quality of recovery between induction with propofol or ketamine/diazepam. Dogs undergoing elective orthopedic procedures were induced with propofol ( $4 \text{ mg kg}^{-1}$ ) or a combination of ketamine/diazepam ( $5/0.25 \text{ mg kg}^{-1}$ ). Recoveries were video recorded and scored using a simple descriptive scale, numeric rating scale, and visual analog scale in real time. They were scored from the video by the live observer and three board-certified anesthesiologists. Two reviewers scored ketamine recoveries lower (better) than propofol ones. Averaging the scores given by all four video reviewers and only the anesthesiologists found lower scores for ketamine as well. Ketamine results in a statistically better recovery than propofol, but this difference may not be clinically significant.

INDEX WORDS: Propofol, Ketamine, Recovery, Anesthesia, Dogs, Video

COMPARISON OF PROPOFOL AND KETAMINE/DIAZEPAM ON THE QUALITY OF  
RECOVERY OF DOGS UNDERGOING ORTHOPEDIC SURGERY

by

JENNIFER ELIZABETH COPELAND

B.S., The University of Georgia, 2011

A.B., The University of Georgia, 2011

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment  
of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2014

© 2014

Jennifer Elizabeth Copeland

All Rights Reserved

COMPARISON OF PROPOFOL AND KETAMINE/DIAZEPAM ON THE QUALITY OF  
RECOVERY OF DOGS UNDERGOING ORTHOPEDIC SURGERY

by

JENNIFER ELIZABETH COPELAND

Major Professor: Erik Hofmeister

Committee: Benjamin Brainard  
Jane Quandt

Electronic Version Approved:

Maureen Grasso  
Dean of the Graduate School  
The University of Georgia  
May 2014

## DEDICATION

This thesis is dedicated to God, my parents, sister, grandparents, and other extended family. You have spent the last 25 years supporting, challenging, blessing, and praying for me. I will never know all of the sacrifices you have made for me, but I appreciate every one of them.

“Grace to you and peace from God our Father and the Lord Jesus Christ.”

## ACKNOWLEDGEMENTS

I first want to thank Dr. Jennifer Peterson, without whose suggestion, help, and prodding I would not have pursued this degree.

Dr. Erik Hofmeister was a better mentor than I could have picked for myself and I am indebted to him for the ways he challenged me. I hold in highest regard those people who can change the way I relate to the world on a daily basis and he did precisely that.

Drs. Brainard and Quandt helped make a most effective committee. I thank them for their guidance in this project and the effort they put into data collection and the editing of this paper.

## TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS .....	v
LIST OF TABLES .....	viii
LIST OF FIGURES .....	ix
CHAPTER	
1 INTRODUCTION .....	1
Validity .....	1
Recovery .....	5
2 MATERIALS AND METHODS.....	10
Validity .....	10
Recovery .....	11
3 RESULTS .....	16
Validity .....	16
Recovery .....	20
4 DISCUSSION.....	25
Validity .....	25
Recovery .....	27
REFERENCES .....	31
APPENDICES	
A RECOVERY SCORING SCALES.....	39

B PAIN SCORING SCALES.....	40
C KAPPA INTERPRETATION TABLE.....	41

## LIST OF TABLES

	Page
Table 1: Characteristics of dogs included in the study .....	22
Table 2: Surgery characteristics of dogs included in the study .....	22
Table 3: Average score given per group, for each rating scale, by each reviewer .....	23

## LIST OF FIGURES

	Page
Figure 1: Bland-Altman plot of difference vs average in the SDS between the live scoring session and recorded scoring session by the same reviewer.....	17
Figure 2: Bland-Altman plot of difference vs average in the SDS between the live scoring session and the anesthesiologists' sessions.....	17
Figure 3: Correlation between recovery scores on the NRS during the live session and each of the recorded sessions .....	18
Figure 4: Bland-Altman plot of difference vs average in the NRS between the live scoring session and recorded scoring session by the same reviewer.....	18
Figure 5: Bland-Altman plot of difference vs average in the NRS between the live scoring session and the anesthesiologists' sessions.....	19
Figure 6: Correlation between recovery scores on the VAS during the live session and each of the recorded sessions.....	20
Figure 7: Bland-Altman plot of difference vs average in the VAS between the live scoring session and recorded scoring session by the same reviewer.....	20
Figure 8: Bland-Altman plot of difference vs average in the VAS between the live scoring session and the anesthesiologists' sessions.....	20

## CHAPTER 1

### INTRODUCTION

#### Validity

A goal of veterinary medicine is to provide the best care possible for patients, but economic constraints also necessitate efficient use of employee hours. The time dogs spend recovering after general anesthesia is a high risk period. Potential complications include regurgitation, aspiration, hypothermia, vomiting, pain, hypersalivation, neurologic excitement, dysphoria, and respiratory and cardiovascular depression. (Bednarski et al. 2011, Tsai et al. 2007) Many clinics do not have the resources to have technicians and nurses spending extended periods of time sitting with a single patient to facilitate their recovery. In order to minimize this burden, studies have been undertaken to determine which anesthetic medications lead to the least eventful and complicated recoveries. (Aguiar et al. 2001, Andreoni and Hughes 2009, Barletta et al. 2011, Bufalari et al. 1998, Gupta et al. 2004, Hellebrekers et al. 1998, Jiménez et al. 2012, Karapinar et al. 2006, Kohrs and Durieux 1998, Leslie et al. 2003, Lozano et al. 2009, Mannarino et al. 2012, Matthews et al. 2004, Muhammed et al. 2009, Tsai et al. 2007, Ueyema et al. 2008) Recovery studies can be time intensive on the part of the researchers and blinding, which is important during subjective assessment, may be difficult to maintain for observers and raters. A potential way to bypass the time restrictions is through observing and rating recoveries via recorded video. Video recording allows researchers to review the recoveries during time available in their schedule and to review them more than once.

The purpose of this experiment is to determine the validity and reliability of using video recordings to evaluate anesthesia recovery in dogs. A measurement tool must be shown to be

valid for it to be of any benefit to those using it. Validity is a tool's ability to measure what it is supposed to measure. (Buckingham et al. 1921) It is typically used in circumstances when the output, unobservable features, cannot be directly or perfectly measured (e.g., IQ test or severity of symptoms). (Segen's Medical Dictionary) In 1954 the American Psychological Association added that validity "indicates to the test user the degree to which the test is capable of achieving certain aims." (American Psychological Association et al. 1954) Several aspects of a system have to be evaluated when determining its validity. The first characteristic is construct validity. Construct validity is the branch of validity used to create a new tool from a theory in the absence of a recognized gold standard and operational definitions. (Miller-Keane 2003) A subtype of construct validity, convergent validity, is also evaluated. This type of validity measures the overlap between tests that are intended to measure the same thing. Convergent validity for a novel assessment is important for ensuring multiple measurements that theoretically should be related are, in fact, related.

In addition to validity, reliability is important to know about a new test. Reliability describes the consistency of the measurements of a test (Buckingham et al. 1921). The results must be reproducible and consistent in order for a test to be useful. Reliability will be evaluated using test-retest reliability. (Holmes and Cockcroft 2008) This test has the same person evaluate the same event on separate occasions. The extent to which their two scores agree is then calculated to determine if the test is reliable.

It is possible for a test to be reliable but not valid. A test can produce the same results repeatedly, but if those results are not true, it is not valid. Inversely, if a test is not reliable, it cannot be valid. A test that is not producing consistent results cannot be producing the right results consistently. (Friis and Sellers 2004)

Validity and reliability can be measured using statistical tests such as correlation, kappa ( $\kappa$ ) tests, t-tests, and Bland-Altman plots. The first two are used to determine how well data points agree with each other, while the t-test evaluates differences between group means. (Moore and McCabe 2006) Bland-Altman plots also assess agreement, but between different methods instead of data points. (Bland and Altman 1986)

Correlation is a measure of association. If two events or points are related to each other, they will have high correlation values. Coefficient values range from -1 to +1 with values closer to  $\pm 1$  indicating a stronger relationship and 0 indicating no relationship. (Moore and McCabe 2006) Two common methods for testing correlation are Pearson's product-moment coefficient and Spearman's rank correlation coefficient. Pearson's test is the test normally thought of when correlation is mentioned. It tests for the presence and strength of a linear relationship between two variables by dividing the covariance of the variables by the product of their standard deviations. Spearman's coefficient tests whether two variables are related without requiring that they are linearly related. If both variables increase or decrease together, regardless of the shape of the distribution, they will have a high correlation coefficient. (Spearman 1904)

Cohen's kappa is a measure of agreement between two raters when investigating categorical items. It differs from correlation in that it takes into account agreement due to chance as if the raters had given answers randomly. If the raters agree no more often than is expected due to chance,  $\kappa$  will be equal to zero; perfect agreement results in a  $\kappa$  value of one. (Cohen 1960) The second measure of agreement is Fleiss' kappa. This test allows investigators to evaluate agreement between more than two raters when dealing with categorical or ordinal data. Like Cohen's  $\kappa$ , Fleiss'  $\kappa$  accounts for chance agreement between raters. (Fleiss 1971) This test

is useful for when measurement intervals are not necessarily equal, making it a generalization of Scott's pi. (Scott 1955)

The t-test is used to compare the means of groups. This can be done by comparing the mean of one group to a specified value (one-sample t-test) or by comparing two groups to each other (two-sample t-test). (Student 1908) If a two-sample t-test is used, the samples can be either independent (completely separate) or dependent, in which the same group is to be measured twice, or the groups can be matched for like characteristics before analysis. T-tests can also be either one- or two-tailed. A one-tailed test only evaluates change in a single direction whereas two-tailed considers any change in the analysis. To evaluate test-retest reliability, a two-sample, dependent, two-tailed t-test is used. Independent scores are generated on separate occasions (two-sample) by the same observer (dependent) and change in either direction is important (two-tailed).

The final tool included for assessment of reliability is the Bland-Altman plot. These graphs are used to compare measurements from different methods that are subject to error. (Hanneman 2008) A common application of this is assessing a new method against an established gold standard. The graphs show the difference in the two measurements of the same subject against the average of the two measurements. (Bland and Altman 1986)

The tools used for data collection are as important as the ways in which the data are analyzed. Two commonly used scales for assessing subjective criteria are the numeric rating scale (NRS) and visual analog scale (VAS). The NRS is a Likert scale that varies in size but often ranges from zero to ten. (Dawes 2008) It was originally developed by Huskisson (1974) and is commonly used in human medicine as a pain scale (McCaffery and Beebe 1993). The number that best represents the severity of the variable to be measured is selected from the list.

The VAS is a straight line, conventionally 10cm long, with verbal descriptors at the ends. Evaluators mark the place between the extremes that is most appropriate for what is being measured. (Downie et al. 1978)

A slightly more objective scale is the simple descriptive scale (SDS). This scale contains a set of, typically, four or five Likert numbers that also includes a verbal descriptor for each. The descriptors help to make it more objective than the previous scales, but they also decrease the sensitivity to minute changes. (Downie et al. 1978) Scorers choose the ranking based on the description provided rather than the placement of the number relative to the others.

### Recovery

Propofol (2,6-diisopropylphenol) is a short-acting hypnotic, belonging to the alkylphenol family, used in both humans and animals for the induction and maintenance of anesthesia. (Gupta et al. 2004; Langley and Heel 1988) It has a short duration of action due in part to its rapid distribution into body tissues. (Cockshott et al. 1992) Its lipophilic properties allow it to permeate the blood-brain barrier efficiently and have a rapid onset of action. (Langley and Heel 1988) It results in smooth anesthetic induction and anesthetic effects within 30 seconds of administration in dogs. (Aguilar et al. 2001) It has a short half-life and high clearance rate (Langley and Heel 1988), making it typically unsuitable for use as the sole anesthetic agent. It is prepared in a solution of propofol, glycerol, purified egg lecithin, and soybean oil for use. Ethylenediaminetetraacetic acid was later added to the mixture to deter microbial growth. (Thompson and Goodale 2000)

The pharmacokinetics of propofol contribute to making it a desirable anesthetic. In dogs given a 7 mg kg<sup>-1</sup> bolus of propofol, a model with three exponential phases was created. The

half-lives of the phases were 4.2, 31, and 303 minutes. The steady state volume of distribution was  $11.4 \text{ L kg}^{-1}$ . The total body clearance rate was  $76 \text{ mL min}^{-1} \text{ kg}^{-1}$ . (Cockshott et al. 1992) Cai et al. (2012) compared conventional propofol to a new microformulation and calculated distribution half-life to be 2.9 minutes, an elimination half-life of 56.6 minutes, the volume of distribution to be  $0.48 \text{ L kg}^{-1}$ , and the clearance rate to be  $22 \text{ mL min}^{-1} \text{ kg}^{-1}$  for the conventional preparation. A study done by Reid and Nolan (1993) investigated how premedication with acepromazine at  $0.05 \text{ mg kg}^{-1}$  IM affected the pharmacokinetics of propofol. In this study, propofol had a  $t_{1/2\alpha}$  of 0.78 minutes,  $t_{1/2\beta}$  of 70.4 minutes, volume of distribution at a steady state of  $3.6 \text{ L kg}^{-1}$ , and a clearance rate of  $39.6 \text{ mL min}^{-1} \text{ kg}^{-1}$ .

Propofol causes negative inotropic actions (Pagel and Wartier, 1993) producing the cardiovascular results of decreases in contractility (Nagashima et al. 2000; Puttick et al. 1992), systemic vascular resistance, cardiac output (Brüssel et al. 1989; Goodchild and Serrao 1989), and blood pressure. (Bufalari et al. 1998; Nagashima et al. 2000) Another effect of this action is good overall muscle relaxation. (Aguar et al. 2001) There is also evidence showing direct effects of propofol on arterial and venous vascular tone. (Pagel and Wartier, 1993) A hypothesis for the manner in which propofol causes dramatic hypotension is through inducing the resetting of baroreceptors to slow the heart rate despite the decreases in pressure in the arteries. (Brüssel et al. 1989)

A proposed method for the action of propofol is its interactions with gamma-aminobutyric acid (GABA) receptors. (Nolan and Reid 1993) By upregulating GABA transmission, propofol can produce both neuronal inhibition and muscle relaxation.

Propofol has been shown to lead to a quick recovery in dogs (Hall and Chambers 1987; Muhammed et al. 2009; Watkins 1987), but is occasionally accompanied by adverse side effects

and rough recoveries. In addition to the cardiovascular depressive effects, one of the most common complaints with propofol is pain on injection. (Tan and Onsiang, 1998) One study reported as many as 60% of human patients experiencing pain upon injection. (Jalota et al. 2011) It also has minimal analgesic properties that require it to be used in combination with other medications. Respiratory depressive effects are often also reported but normally remain within acceptable levels. (Aguiar et al. 2001; Ambros et al. 2008; Matthews et al. 2004)

Ketamine [2-(o-chlorophenyl)-2-(methylimino)cyclohexanone] is a sympathomimetic agent, (Wright 1982) derived from the dissociative drug phencyclidine, (Mair et al. 2009) that has been used for many years to induce anesthesia. It produces a cataleptoid state of anesthesia including reported poor muscle relaxation in small animals. (Haskins et al. 1985; Hellebrekers et al. 1998) The high lipid solubility is similar to propofol which allows it to be absorbed quickly by the body. (Wright 1982) As in propofol, this creates a rapid onset and short duration of anesthesia. (Haskins et al. 1985) Ketamine has good analgesic properties (Mair et al. 2009) and typically produces slight increases in ventilation, heart rate, and blood pressure. (Haskins et al. 1985; Wong and Jenkins 1974; Wright 1982)

The sympathomimetic properties of ketamine display themselves by inhibiting the reuptake of catecholamines. This inhibition can produce increases in cardiovascular values such as blood pressure, cardiac output, and heart rate. (Green 1999) One study reported that these increases are likely not a direct effect of the properties of the drug but are caused by other actions. (Wright 1982) Ketamine is an N-methyl D-aspartate (NMDA) receptor antagonist, which results in analgesia and possibly in some of its more severe side effects such as seizures. (Haskins et al. 1985)

Ketamine has many appealing anesthetic qualities since it causes no serious depression in any cardiovascular or respiratory parameters and has a rapid onset of action. However, it has been shown to have several unsatisfactory side effects as well. Excessive salivation and vomiting during recovery have been reported following ketamine administration. Poor muscle relaxation and spontaneous movement are also commonly encountered in small animals given ketamine. (Haskins et al. 1985; Hellebrekers et al. 1998)

Like propofol, ketamine has rapid pharmacokinetic properties. In dogs anesthetized with isoflurane and given a 3 mg/kg intravenous bolus of ketamine,  $t_{1/2\alpha}$  was 2.0 minutes and  $t_{1/2\beta}$  was 94.2 min. The volume of distribution was  $4.06 \text{ L kg}^{-1}$  and total body clearance rate was  $58.2 \text{ mL min}^{-1} \text{ kg}^{-1}$ . (Pypendop and Ilkiw 2005) These numbers vary from those found by Schweiger et al. (1991). In that study, a  $10 \text{ mg kg}^{-1}$  intravenous bolus was given to dogs anesthetized with enflurane. The  $t_{1/2\alpha}$  and  $t_{1/2\beta}$  times were increased relative to Pypendop et al. (2005), with values of 7.7 minutes and 121.9 minutes, respectively, while the volume of distribution ( $1.37 \text{ L kg}^{-1}$ ) and clearance rate ( $18.1 \text{ mL min}^{-1} \text{ kg}^{-1}$ ) were lower. (Schweiger 1991)

A common medication to pair with ketamine for induction of anesthesia is diazepam. Diazepam is a member of the benzodiazepene family. (Formanek et al. 1976) It can help reduce the cardiovascular changes that have been demonstrated with ketamine use. (Haskins et al. 1986) As with propofol and ketamine, diazepam is lipid soluble so it permeates the blood brain barrier quickly and easily. (Musulin et al. 2011) Diazepam also helps to reduce the muscle hypertonus caused by ketamine. (Haskins et al. 1986) However, it is not acceptable for use alone as Haskins et al. (1986) found that it caused unacceptable excitement in 36% of dogs and did not produce sedation or anesthesia in any. It also provides no analgesia. Diazepam does not cause many adverse cardiovascular or respiratory changes alone, but can potentiate the effects of drugs it is

paired with. (Haskins et al. 1986) Injections of diazepam can cause tenderness at the injection site (Siebke et al. 1976) and thrombophlebitis if it is injected into a smaller vein, due to the hyperosmolarity of the solution. (Langdon et al. 1973)

Since both of these anesthetics are commonly used for anesthesia induction, studies have been done involving these anesthetics, but not as a direct comparison of quality of recovery. Few studies have been done evaluating the recovery in dogs after ketamine anesthesia in any capacity. By comparing these common anesthetics, clinicians will be provided with the information to make informed decisions regarding what type of anesthetic is most suited to the individual patient.

The purpose of this study is to evaluate the speed and quality of recovery between patients anesthetized with propofol and ketamine/diazepam. The hypothesis is that propofol will result in a more rapid and smooth recovery.

## CHAPTER 2

### MATERIALS AND METHODS

#### Validity

From August to December 2013, 30 dogs presenting to the University of Georgia Veterinary Teaching Hospital anesthesia service were selected on the basis of a convenience sample for evaluation and video recording of recovery. Criteria for inclusion were dogs recovering in ICU following anesthesia. Dogs were excluded if they recovered in an oxygen cage or had other factors inhibiting observation such as aggression, resistant infection, or timing overlap with a concurrent investigation. All procedures followed standard hospital protocol and were approved by the Clinical Research Committee of the University of Georgia Veterinary Teaching Hospital.

Recovery monitoring began when dogs were extubated. Extubation occurred after dogs were able to swallow on their own. Dogs were video recorded and monitored for one hour. Scoring was done in real time at the end of the hour by an observer. Scoring was carried out using three different systems: an SDS, VAS, and NRS. The VAS scores were anchored at 0 (uneventful) to 10 cm (most eventful recovery) and measured in cm using a ruler. The SDS (Lozano 2009) had 5 (1 to 5) levels (Appendix A) and the NRS had 11 (0 to 10). The order in which the measurement systems were evaluated was randomized for each dog.

A library of videos was created of the recoveries in a random order on an external hard drive and distributed to three board-certified anesthesiologist raters and the original rater. The

raters independently reviewed the videos over a two week period. Recoveries were scored using the same three systems in a random order as the live scoring.

To evaluate construct validity, the correlation between scores given during the live session and those given during each of the recorded sessions were calculated. Bland-Altman plots were also constructed to measure agreement between the live and recorded scores given by the same reviewer and between the average of the veterinarians' scores and the live ones. Test-retest reliability utilized Cohen's kappa to assess agreement between the same observer's scores during the live and recorded sessions. The data was converted into a binary system of classifying recoveries as good or bad. Scales were considered to indicate a bad recovery if the designated score was 4 or more for the SDS, or 5 or more on the NRS and VAS. The overall recovery was considered bad if two of the three scales rated the recovery as bad when all five sessions were averaged together. Fleiss' kappa evaluated the agreement between the three anesthesiologists both with the categorical scores and the good/bad binary. Kappa values were interpreted using the guidelines set forth by Landis and Koch. (1977, Appendix C) The NRS and VAS were also compared to each other using linear regression. ANOVA was used to compare the relationship of the scores given on the SDS with those given on the NRS and VAS. Significance was set at  $\alpha < 0.05$ .

### Recovery

Before the study was started, a calibration study was done to determine the sample size needed to achieve the appropriate power. The calibration study focused on the three different recovery scoring systems that were to be used in the full study. These systems were a VAS (0-10 cm with 0 being uneventful and 10 being most eventful possible), NRS (0-10), and SDS (1-5,

Lozano 2009). After surgery, the quality of recovery of 30 dogs was evaluated on all three scales. The differences in the scales were evaluated to determine how large of a sample was needed to detect a  $\beta$  of 0.80 with an  $\alpha$  level of 0.95.

The calibration study determined 60 client-owned dogs of various breeds admitted to the Hospital for single-limb, elective orthopedic surgery needed to be enrolled in the study. Criteria for inclusion were as follows: American Society of Anesthesiologists physical status score of I or II and otherwise healthy on the basis of physical examination and minimal blood work (PCV, plasma total solids concentration, BUN concentration, and blood glucose concentration). Body condition score was assigned in accordance with a standardized system (Lund et al. 1999). Dogs were excluded if the attending anesthesiologist determined the dictated protocol would not be suitable for the patient based on previous anesthetic episodes or concomitant diseases or if they were to undergo a total hip replacement. The study protocol was approved by the Clinical Research Committee of the University of Georgia Veterinary Teaching Hospital and informed consent was obtained from the owners.

Prior to surgery, dogs were fasted for 12 hours and had free access to water during this time. They were randomly assigned using a random number table to receive either propofol (Group P) or ketamine/diazepam (Group K) as their induction agent. Dogs were pre-medicated with acepromazine ( $0.05 \text{ mg kg}^{-1}$ ), morphine ( $0.5 \text{ mg kg}^{-1}$ ), and glycopyrolate ( $0.01 \text{ mg kg}^{-1}$ ) IM. At least 30 minutes after premedication injection, an IV catheter was placed in a cephalic vein for induction of anesthesia. Anesthesia was induced with either propofol ( $4 \text{ mg kg}^{-1}$ ) or ketamine/diazepam ( $5/0.25 \text{ mg kg}^{-1}$ ), given IV, until endotracheal intubation was possible. Induction agents were given at 50% of the total volume over 10 seconds. After a 10s pause, the patient was then evaluated for the ability to intubate. If the patient could not be intubated, an

additional 10% of the total volume was given as a bolus and the patient re-evaluated after a 10s pause. This was repeated until the patient could be intubated or the full dose had been given. If the patient could not be intubated after receiving the full dose of drug, a further  $2 \text{ mg kg}^{-1}$  of propofol or  $5/0.25 \text{ mg kg}^{-1}$  ketamine/diazepam was given. Following endotracheal intubation, dogs were connected to a rebreathing anesthesia circuit. Anesthesia was maintained with isoflurane in 100% oxygen, given to achieve a surgical plane of anesthesia, as determined by the senior student and supervising technician. Indirect blood pressure monitoring was instituted using a Doppler device with an occlusive cuff and sphygmomanometer. An arterial line was placed in either the dorsal pedal or coccygeal artery to be used for direct arterial blood pressure measurement. If anesthesia preparation was delaying the beginning of surgery or multiple technicians and the attending anesthesiologist were unable to place the arterial catheter, direct blood pressure measurements were not collected. The surgical site was then clipped and prepared aseptically. An epidural block of duramorph ( $0.1 \text{ mg kg}^{-1}$ ) and bupivacaine ( $0.5 \text{ mg kg}^{-1}$ ) was administered for dogs undergoing hind limb procedures. Dogs having fore limb surgeries had brachial plexus blocks containing bupivacaine. The dogs were moved to the operating room and cardiovascular and respiratory parameters were measured every five minutes throughout surgery.

A sidestream capnograph was attached to an elbow piece on the endotracheal tube for end tidal carbon dioxide and isoflurane concentration while the patient was in the operating room. Continuous direct arterial blood pressure monitoring was performed using a pressure transducer attached to the same anesthetic monitor and systolic, diastolic, and mean blood pressures were recorded. The transducer was zero calibrated at the approximate level of the heart base. The pressure waveform was visually examined to ensure a lack of overdamping or underdamping. Pulse and respiratory rate were obtained either from the monitor or manually. Esophageal

temperature was measured using a probe connected to the same monitoring machine. Direct and indirect arterial blood pressure, pulse rate, respiratory rate, temperature, end-tidal carbon dioxide measurements, and end-tidal isoflurane concentration were recorded every 5 minutes.

After the completion of surgery, dogs were moved back into the prep area and then taken to radiology for post-operative radiographs. Thereafter, they were moved back to the prep area, the inhalant was discontinued, and the dogs were allowed to breathe 100% oxygen for 10 minutes. They were then disconnected from the anesthesia machine and taken to a cage in ICU. Extubation was done after dogs were able to swallow on their own. The time from the end of anesthesia to extubation and from extubation to head lift, sternal recumbency, and standing were all recorded. Administration of post-operative medication and rescue therapy, if necessary, were left to the discretion of the attending anesthesiologist. Success of the local block was determined by the technician and senior student based on end-tidal isoflurane concentration during surgery.

The sedation level of each dog was assessed at extubation and 1 hour using a modified composite scale in which a higher score indicates a greater level of sedation (Smith et al. 2001). Pain was scored at extubation and one hour using an interactive SDS where a low score indicates less pain (Moran 2013, Appendix B). At one hour, the recovery score was finalized, the observation components of the sedation score were conducted, followed by the interactive components of the sedation score, and finally the interactive pain score.

Recovery monitoring included the time between extubation and one hour after extubation. All monitoring and scoring was done in real time by a blinded observer and also video recorded. Scoring was carried out using three different systems: an SDS, VAS, and NRS. The VAS scores were anchored at 0 (uneventful) and 10 cm (most eventful recovery) and the

NRS had 11 (0 to 10) levels. The SDS (Lozano 2009) had 5 (1 to 5) levels (Appendix A). The order in which the measurement systems were evaluated was randomized for each dog.

A library of videos were created of the recoveries in a random order on an external hard drive and distributed to three board-certified anesthesiologist raters and the original rater. The raters independently reviewed the videos over a four week period. Recoveries were scored using the aforementioned three systems in a random order.

Normality was determined using the D'agostino-Pearson method. A 2-way unpaired t test was used for normally distributed data and a Mann-Whitney U test was used for non-normally distributed data to evaluate differences between the treatment groups. To test for a correlation between sedation and pain scores with quality of recovery, Pearson's or Spearman's correlations were used where appropriate. When classifying recoveries as simply "good" or "bad", the cutoffs were 4 for the SDS and 5 for the NRS and VAS with good being below the cutoff and bad being equal to or above. The overall recovery was considered bad if two of the three scales rated the recovery as bad when averaging all five sessions together. Significance was set at  $\alpha < 0.05$ .

## CHAPTER 3

### RESULTS

#### Validity

Of the 30 dogs studied, 18 were male (14 neutered, 4 intact) and 12 were female (10 spayed, 2 intact). They weighed mean  $\pm$  SD  $27.97 \pm 15.51$  kg with BCS scores of  $6.1 \pm 1.1$  out of nine and were  $77 \pm 48$  months of age. The breeds most represented were mixed breed (5), Labrador retriever (4), Corgi (2), German shepherd (2), and Rottweiler (2). There was also a single subject from each of the following breeds: American bulldog, Collie, French bulldog, Golden retriever, Siberian husky, Irish terrier, Maltese, Mastiff, Pit bull terrier, Pomeranian, Rat terrier, Shih-tzu, Visla and Weimaraner. The procedures included 13 orthopedic cases, 10 soft tissue surgeries, 4 diagnostic imaging sessions, and 3 other events. A complete list is as follows: tibial plateau leveling osteotomy (TPLO) (4), computed tomography (CT) (2), elbow arthroscopy (2), joint taps (2), mast cell tumor removal (2), magnetic resonance imaging (MRI) (2), abdominal exploratory and feeding tube placement (1), bone fragment removal (1), carpal fracture repair (1), cystotomy with liver biopsy and neuter (1), digit amputation (1), exploratory laparotomy (1), exploratory laparotomy for hemoabdomen (1), hydraulic sphincter implantation with spay (1), hygroma removal with neuter (1), laproscopic spay (1), parathyroidectomy with central line placement (1), shunt repair and neuter (1), stifle arthroscopy with neuter (1), total hip replacement (1), TPLO with contralateral implant removal (1), and wire implant removal (1).

Cohen's  $\kappa$  value when comparing the live scoring and recorded scoring done by the same observer was 0.79. Fleiss'  $\kappa$  among the three anesthesiologists for the SDS was 0.26. Evaluating

the live and recorded sessions by the same observer using Bland-Altman plots had a bias of  $0.133 \pm 0.507$  (Bias  $\pm$  SD, Figure 1). Assessing the three anesthesiologists against each other on the 5-point SDS showed a bias of a third of a point ( $0.33 \pm 0.717$ , Figure 2). Computing Fleiss' kappa with the three experts when categorizing recoveries simply as good and bad yielded a value of 0.71. The age of the patient was not associated with the score given on the SDS ( $p = 0.2412$ ).

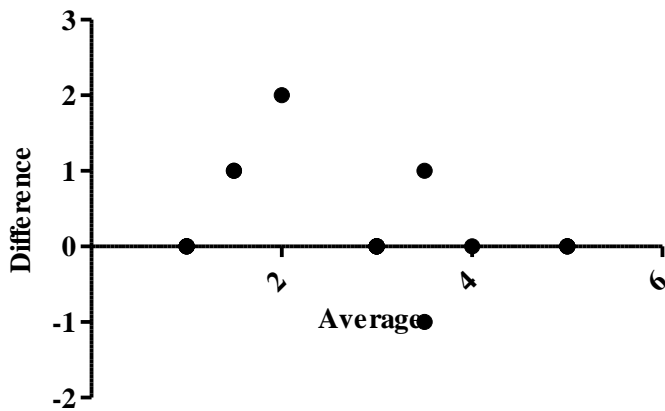


Figure 1. Bland-Altman plot of difference vs average in the SDS between the live scoring session and recorded scoring session by the same reviewer. Bias  $\pm$  SD:  $0.133 \pm 0.507$ .

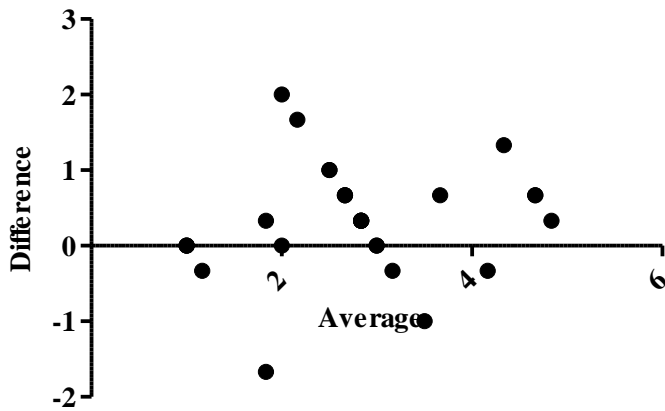


Figure 2. Bland-Altman plot of difference vs average in the SDS between the live scoring session and the anesthesiologists' sessions. Bias  $\pm$  SD:  $0.33 \pm 0.717$ .

Linear regression of the NRS scores showed a slope significantly different from zero ( $p < 0.0001$ ). The r-square value between the live scores and recorded scores by the same reviewer (M) was 0.74. R-square values comparing the live scores with the anesthesiologists' (B, Q, H) were 0.50, 0.57, and 0.60 (Figure 3). The NRS had lower  $\kappa$  values for both tests than the SDS. Fleiss'  $\kappa$  from the three veterinarians was 0.16. Cohen's kappa, looking at the same scorer live and recorded, was 0.29. Bland-Altman bias when calculating the live and recorded same observer was  $0.5 \pm 1.526$  (Figure 4) and for the three experts against the live scoring was  $1.178 \pm 1.641$  (Figure 5). Fleiss'  $\kappa$  for the agreement of veterinarians between good and bad recoveries was 0.54. The age of the patient was not associated with the score given on the NRS ( $p = 0.4012$ ).

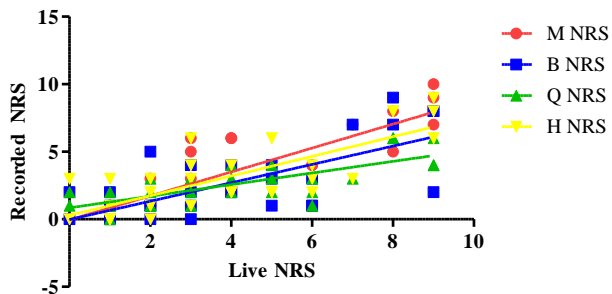


Figure 3. Correlation between recovery scores on the NRS during the live session and each of the recorded sessions. M:  $r^2 = 0.74$ ; B:  $r^2 = 0.50$ ; Q:  $r^2 = 0.57$ ; H:  $r^2 = 0.60$ .

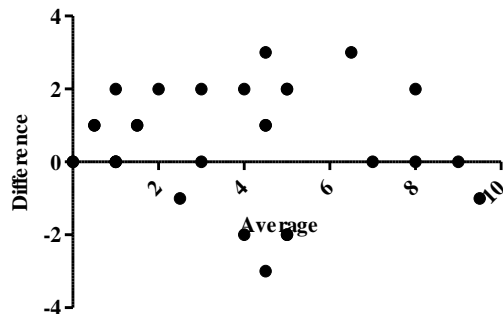


Figure 4. Bland-Altman plot of difference vs average in the NRS between the live scoring session and recorded scoring session by the same reviewer. Bias  $\pm$  SD:  $0.5 \pm 1.53$ .

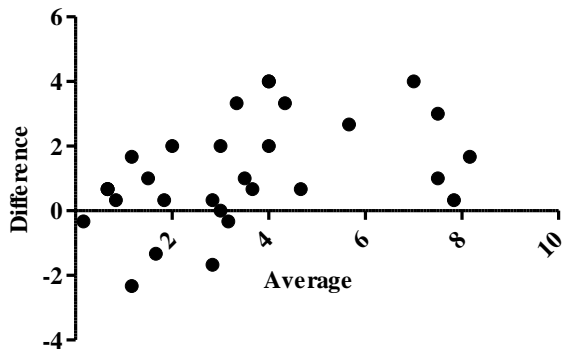


Figure 5. Bland-Altman plot of difference vs average in the NRS between the live scoring session and the anesthesiologists' sessions. Bias  $\pm$  SD:  $1.178 \pm 1.64$ .

The last scale, the VAS, is continuous data, so Fleiss'  $\kappa$  was not applied to the anesthesiologists' scores. The value of Cohen's  $\kappa$  was 0.71 between the live and recorded scores from the same reviewer. Correlation between the recorded and live scoring had slopes significantly different from zero for the VAS as well ( $p < 0.0001$ ). The same reviewer had an r-square value of 0.82 between live and recorded sessions and the anesthesiologists were 0.50, 0.61, and 0.63 (Figure 6). The Bland-Altman plots for the two data arrangements (live vs. recorded same observer and live vs. recorded veterinarians) had bias of one third of a point for the same observer ( $0.313 \pm 1.169$ , Figure 7) and  $0.840 \pm 1.472$  when comparing the veterinarians to the live scores (Figure 8). Agreement among the anesthesiologists' scorings as good versus bad as assessed by Fleiss'  $\kappa$  was 0.6. The age of the patient was not associated with the score given on the VAS ( $p = 0.3642$ ).

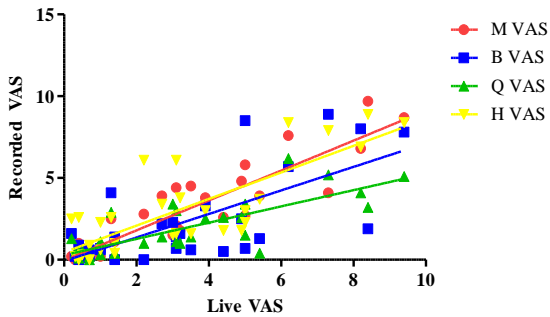


Figure 6. Correlation between recovery scores on the VAS during the live session and each of the recorded sessions. M:  $r^2 = 0.82$ ; B:  $r^2 = 0.50$ ; Q:  $r^2 = 0.61$ ; H:  $r^2 = 0.63$ .

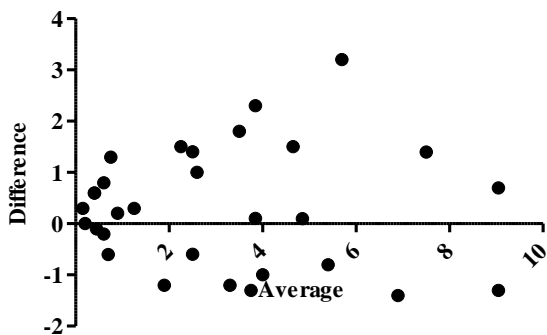


Figure 7. Bland-Altman plot of difference vs average in the VAS between the live scoring session and recorded scoring session by the same reviewer. Bias  $\pm$  SD:  $0.313 \pm 1.17$ .

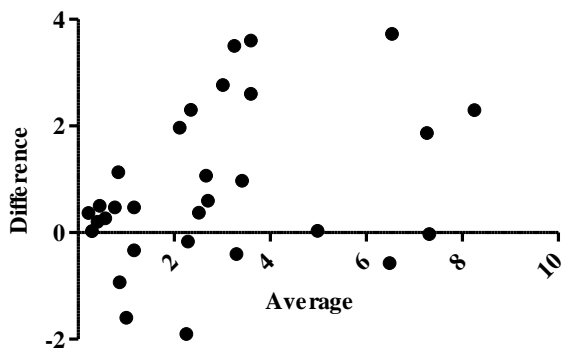


Figure 8. Bland-Altman plot of difference vs average in the VAS between the live scoring session and the anesthesiologists' sessions. Bias  $\pm$  SD:  $0.84 \pm 1.47$ .

### Recovery

The dogs enrolled in the study were 50% male, weighed  $28.7 \pm 12.5$  kg, aged  $63.67 \pm 32.9$  months, and had a BCS of  $6.4 \pm 1.2$ . Breeds included Labrador retriever (16), Mixed breed (9), Golden retriever (5), American pit bull terrier (4), Australian shepherd (3), Boston terrier (2),

German shepherd (2), and Yorkshire terrier (2). American bulldog, American Staffordshire terrier, Beagle, Boxer, Boykin spaniel, Cane Corso, Cavalier King Charles spaniel, Cocker spaniel, Collie, English bulldog, Jack Russell terrier, Maltese, Pointer, Pug, Rottweiler, Visla, and Weimaraner were each represented by a single subject. Surgeries included TPLO (39), correction of a medially luxating patella (6), fracture repair (4), lateral suture (3), femoral head ostectomy (2), tibial tubercle transposition (1), medial patella luxation repair combined with a lateral suture (1), medial collateral ligament repair (1), surgical implant repair (1), partial carpal arthrodesis (1), and TPLO with a medial patella luxation correction (1). (Table 1) Breed and surgery differences were evaluated by dividing the dogs into weight class (above or below 18 kg) and as TPLO or non-TPLO, respectively. Differences between groups were insignificant for all characteristics (weight,  $p = 0.5332$ ; age,  $p = 0.2466$ ; BCS,  $p = 0.6697$ ; breed,  $p = 0.7656$ ; procedure,  $p = 0.1760$ ) except gender ( $p = 0.0389$ ). There were also no differences between treatment groups for minutes spent under anesthesia (Group P:  $261.1 \pm 51.83$ ; Group K:  $260.4 \pm 57.84$ ;  $p = 0.6467$ ), time to extubation (Group P:  $21.5 \pm 16.26$ ; Group K:  $21.33 \pm 14.97$ ;  $p = 0.3471$ ), time to head lift (Group P:  $6.53 \pm 16.04$ ; Group K:  $11.43 \pm 18.34$ ;  $p = 0.7256$ ), time to sternal recumbency (Group P:  $33.03 \pm 22.73$ ; Group K:  $40.03 \pm 22.81$ ;  $p = 0.2461$ ), or time to voluntary motion (Group P:  $52.27 \pm 17.34$ ; Group K:  $56.6 \pm 12.04$ ;  $p = 0.2838$ ). Sedation and pain scores at extubation and the scores an hour later were also not significantly different between groups ( $p = 0.7891$ , 1.00, 0.0689, and 0.2439, respectively). Anesthesia characteristics are summarized in Table 2.

First, the scores given during each of the five scoring sessions were compared independently of the other sessions (Table 3). Differences in the live scores (SDS:  $p = 0.422$ ;

Table 1. Characteristics of dogs included in the study. BCS = body condition score. \* =  $p < 0.05$

	Sex*		Surgery		Breed		Weight	BCS (/9)	Age (mo)
	Female	Male	TPLO	Other	>18kg	<18kg	(kg)		
<b>Group P</b>	19	11	22	8	22	8	27.69±11.57	6.47±1.196	68.67±29.98
<b>Group K</b>	11	19	17	13	23	7	29.74±13.75	6.33±1.213	58.67±35.94

Table 2. Surgery characteristics of dogs included in the study.

	Length of Anesth. (min) (p = 0.6467)	Extubation (min) (p = 0.3471)	Head Lift (min) (p = 0.7256)	Sternal Recumbency (min) (p = 0.2641)	Motion (min) (p = 0.2838)	Sedation Extubation (/14) (p = 0.7981)	Pain Extubation (/5) (p = 1.00)	Sedation Hour (/14) (p = 0.0689)	Pain Hour (/5) (p = 0.2439)
<b>Group P</b>	261.1 ± 51.83	21.5 ± 16.26	6.533 ± 16.04	33.03 ± 22.73	52.27 ± 17.34	11.93 ± 2.42	1.1 ± 0.3051	8.6 ± 3.297	1.367 ± 0.6687
<b>Group K</b>	260.4 ± 57.84	21.33 ± 14.97	11.43 ± 18.34	40.03 ± 22.81	56.6 ± 12.04	12.23 ± 1.995	1.1 ± 0.3051	10.1 ± 3.377	1.2 ± 0.4842

NRS:  $p = 0.982$ ; VAS:  $p = 0.6842$ ), recorded scores with the same reviewer (SDS:  $p = 0.2836$ ; NRS:  $p = 0.3259$ ; VAS:  $p = 0.2131$ ), and those given by one of the other reviewers (SDS:  $p = 0.0703$ ; NRS:  $p = 0.0867$ ; VAS:  $p = 0.1257$ ) were all insignificant for each scale. The other two video reviewers found a difference between groups on all three scales to be significant ( $p < 0.033$  for each).

In order to evaluate differences between the treatment groups while looking at more than one scoring session, the scores given by each reviewer on the same scale to a patient were averaged. When evaluating the scores given during all five sessions, the NRS and VAS show no significant differences between groups P and K ( $p = 0.114$  and  $0.0746$ , respectively). However, the average SDS scores for recoveries in group K are lower than those for group P ( $p = 0.0406$ ). The average scores from the set of four video reviews showed significant differences on all three scales with group K having better recoveries (SDS:  $p = 0.0214$ ; NRS:  $p = 0.0428$ ; VAS:  $p = 0.0476$ ). Evaluating the scores given by the three anesthesiologists also had significantly better

Table 3. Average score given per group, for each rating scale, by each reviewer.

		<b>Group P</b>	<b>Group K</b>
<b>Live 1</b>	SDS	2.867±1.137	2.633±1.098
	NRS	3.433±2.176	3.5±2.013
	VAS	3.012±2.095	2.697±1.882
<b>Recorded 1</b>	SDS	2.767±1.104	2.467±1.042
	NRS	3.433±2.239	2.9±1.918
	VAS	2.833±2.132	2.217±1.628
<b>2</b>	SDS	2.6±1.133	1.933±0.9803
	NRS	2.967±2.282	1.622±1.52
	VAS	2.303±2.361	1.093±1.252
<b>3</b>	SDS	2.667±0.8422	2.233±0.9714
	NRS	2.0±1.539	2.3±1.664
	VAS	2.4±1.579	1.88±1.828
<b>4</b>	SDS	2.933±1.015	2.233±1.223
	NRS	3.233±2.542	1.9±2.155
	VAS	3.333±2.562	2.033±2.297

recoveries for group K (SDS:  $p = 0.0097$ ; NRS:  $p = 0.0227$ ; VAS:  $p = 0.0320$ ). The scores given by the same reviewer during the live session and later from the recording did not show differences between the treatment groups (SDS:  $p = 0.3118$ ; NRS:  $p = 0.6438$ ; VAS:  $p = 0.4507$ ).

Sedation and pain scores at an hour post-extubation were correlated with recovery scores from all three scales. Sedation scores had  $r$  square values of  $-0.4913$  ( $p < 0.0001$ ),  $-0.4982$  ( $p < 0.0001$ ), and  $-0.4799$  ( $p = 0.0001$ ) for the SDS, NRS, and VAS scales, respectively. Pain score  $r$  square values were  $0.3219$  ( $p = 0.0121$ ) for SDS,  $0.3637$  ( $p = 0.0043$ ) for NRS, and  $0.3550$  ( $p = 0.0054$ ) for VAS. The pain score at extubation was also correlated with the SDS scale ( $r^2 = 0.2848$ ,  $p = 0.0274$ ), but not with NRS ( $p = 0.0693$ ) and VAS ( $p = 0.0913$ ).

Possible effects of hydromorphone on timing of recovery, specifically to extubation were evaluated. The time elapsed from the last dose of intraoperative hydromorphone to the end of anesthesia differed significantly between the groups ( $p = 0.0153$ ) with group K and group P receiving the last dose  $115 \pm 18.5$  and  $55.88 \pm 13.4$  minutes before the end of anesthesia, respectively. There was no difference in number of dogs receiving an opioid intraoperatively or after the end of anesthesia in the two treatment groups ( $p = 0.6023$  and  $0.4321$ , respectively). Comparing the mean time to extubation of dogs in both groups, both receiving hydromorphone and not, revealed no difference between the groups ( $p = 0.5407$ ). Hydromorphone was also not related to an overall good recovery as five of the eight bad recoveries received it.

Linear regression shows length of time to extubation was not impacted by the patient's age ( $p = 0.4786$ ) but quality of recovery was (SDS:  $p = 0.0108$ ; NRS:  $p = 0.0062$ ; VAS:  $p = 0.0083$ ). Length of hypotension (mean arterial pressure  $< 60$  mmHg) in patients did not vary significantly ( $p = 0.96$ ) and the number of patients receiving treatment for hypotension was not significant ( $p = 0.3006$ ). The possible effect of the success or failure of local block administration was investigated, but since half of the overall bad recoveries had successful local blocks, no statistical tests were run. The number of dogs in each group receiving rescue sedation during the recovery period did not differ significantly ( $p = 0.1287$ ). Of the eight bad recoveries, two were in group K and six were from group P but comparing the groups described no significant difference ( $p = 0.1287$ ).

## CHAPTER 4

### DISCUSSION

#### Validity

The SDS had higher agreement than the NRS and VAS since it had fewer scoring options. The SDS was also less subjective and had specific benchmarks including vocalization, trembling, paddling, and vomiting. The two lowest scores were reserved for animals who did not, at any time, exhibit any of these characteristics. If an animal vocalized at any point during the hour observation, it was automatically given a 3 even if the rest of the recovery was smooth. The level of agreement decreased when looking at only the veterinarians' scores. This is expected since the number of reviewers decreased by 40% and agreement was only considered if it was perfect, not simply close.

The increased agreement between SDS scores relative to the NRS and VAS is expected given that there were half as many point options to choose from on the scale. The considerable agreement (0.7) between the scores given on the SDS by the same reviewer during the live session and the recorded reviewing session is indicative of the robustness of the scale. Fleiss'  $\kappa$  among the anesthesiologists was relatively low at 0.26. Even though there are only five scoring options, there were fewer raters (3) than categories. This naturally leads to a lower  $\kappa$  because the mathematical probability of agreeing is lower under these conditions. Fleiss'  $\kappa$  showed quite good agreement between the anesthesiologists' scores when bifurcated into good and bad on the SDS. This is to be expected since each of the raters are experts in the field and should be able to accurately determine if a dog is having a good recovery or a bad recovery. The Bland-Altman

plots also showed substantial robustness when comparing the same reviewer. Calculations showed that the recoveries were typically scored 0.13 points higher during the live session than the recorded session. The three expert veterinarians, scoring from videos, also agreed with the live scorer quite well, only scoring recoveries a third of a point better in general. This agreement is quite good between the different assessment methods especially considering the relative inexperience of the rater doing the live scoring.

Since the NRS had 11 scoring options, linear regression was done to determine the correlation between different rating sessions. The two sessions by the same reviewer (live and recorded) correlated well with each other. Cohen's  $\kappa$  between these two sessions was not as impressive (0.29). This lower value for the NRS from the SDS is not surprising since there were twice as many ranking options. Of the 30 NRS pairs, 43% of them differed by more than 9% (one point) and only three differed by more than 18% (two points), but those were not counted as agreements in the  $\kappa$  calculations since it was not exact agreement. The anesthesiologists' scores were fairly correlated with the live observer's with r-square values around 0.55, but the value for Fleiss'  $\kappa$  was not high. There were six of the 30 subjects in which the anesthesiologists rated the recoveries at three consecutive values. While the  $\kappa$  statistic would consider this to be zero agreement, the range of scores is <20% of the scale, which is not broad. As with each of the other statistics, Fleiss'  $\kappa$  of the binary (good vs bad) recovery quality was lower for the NRS than the SDS. However, it was still in an acceptable range of agreement. Both Bland-Altman plots for the NRS showed minimal bias. The first observer scored recorded recoveries half a point better and the veterinarians a little more than a point better compared with the live scores.

The value of Cohen's  $\kappa$  for the two sessions of the same reviewer was even higher for the VAS than that of the SDS. Linear regression also showed significant correlation between the

scores given in those two sessions, with the VAS having a higher  $r^2$  than the NRS. This is interesting especially since the scores were continuous instead of categorical and the range of scores was increased from the SDS. The anesthesiologists' scores also correlated better with the live scores on the VAS than on the NRS. Better correlation in spite of continuous data is likely because the VAS scores were skewed heavily towards lower scores and did not have many high values. VAS scores have been found to be consistently lower than NRS scores in several studies (Bashir et al. 2013; Chanques et al. 2010; Reich et al. 2012) The Bland-Altman plots for the live session against the recorded ones showed surprisingly little bias again, with less than one centimeter of bias being present in all scoring sessions. When comparing the veterinarians' ratings as good or bad, the value of Fleiss'  $\kappa$  showed moderate agreement for the VAS.

In theory, the scoring done live should be the most accurate because the observer was fully aware of all that was happening around the dog, and had a better context from which to understand the reactions. Since the video sessions are correlated well with the live scores, especially those sessions done by the same reviewer, and the Bland-Altman plots show bias of approximately 10% or less, video recording is a valid way to judge recoveries in dogs. The agreement calculated by Cohen's  $\kappa$  values between the live and recorded sessions of the same reviewer and the substantial agreement found between the anesthesiologists using Fleiss'  $\kappa$  on the binary data assures the reliability of this reviewing tool.

### Recovery

The sedation and pain scores assigned to a patient at the end of observation were correlated with the quality of recovery. As in Jiménez et al. (2012), dogs with higher sedation scores (more sedate) had better recoveries. Animals with a lower pain score (less painful) also

had better recovery scores. Sedation scores varied between patients, being distributed along the length of the scale, and showed approximately a 50% correlation with each of the recovery scales. Pain scores were skewed towards low pain scores so the r-square value was not as high. Age was positively correlated with recovery scores and was not related to recovery time

Post-operative administration of hydromorphone was not related to the time it took dogs to recover in this study. This contrasts with papers which have shown hydromorphone administration to lead to a prolonged recovery. A study done in humans found that it took longer for patients to not experience unpleasant effects when administered hydromorphone than when given other medications. (Walker and Zacny 1999) Grubb et al. (2013) claimed that time of recovery was extended in all cats in their study regardless of administration of hydromorphone.

The lack of difference in the number of hypotensive episodes between groups is interesting since that is one of the typically defining characteristics of propofol. (Bufalari et al. 1998; Nagashima et al. 2000) This could be due to the dose of acepromazine ( $0.05 \text{ mg kg}^{-1}$ ) given during the premed. This dose has produced hypotension in dogs undergoing anesthesia induced with propofol and maintained with isoflurane. (Sinclair and Dyson 2012) Since this dose was standard for all subjects, its hypotensive effects might have masked those that were caused by propofol. Blood pressure was also not measured in the immediate peri-induction period, missing the time when the most severe propofol-related hypotension is normally seen. (Brüssel et al. 1989; Bufalari et al. 1996; Nolan and Reid 1993)

The various combinations of data yielded a wide range of results in this study. Thirteen of the 27 (48%) tests found significant difference between the treatment groups and three of the 14 insignificant tests neared significance ( $p = 0.0703, 0.0746, \text{ and } 0.0867$ ). The least experienced scorer did not have statistical significance between the treatment groups for any test during live

scoring, scoring from the recorded video, or when averaging the two sessions together. This could be due to the lack of recoveries experienced, relative to the other scorers, prior to the beginning of the study and the accompanying learning curve. Analyzing all five scoring sessions together showed a difference on only one of the tests (SDS). This was the only data set that split scales between significant and non-significant. The trend of the SDS having the lowest p-value in each group held true for 7/9 calculations. The SDS had descriptors associated with each score, making it the most objective of the scales used. The structure introduced by the descriptors may have helped to homogenize the scores, making it easier to delineate between the treatment groups.

Calculations combining the scores for the four sets of video scoring showed significantly better recovery scores for the dogs in group K than those in group P. The mean of SDS scores was over a half-point (10%) lower in group K. NRS and VAS values were almost a point lower for the ketamine treatment. Considering just the three veterinarians' scores showed significant decreases by over half a point (SDS), over a full point (NRS), and a point (VAS) in group K.

The better recoveries in group K could be due to the analgesic properties of ketamine that propofol does not have. This analgesia may be demonstrated by the longer time elapsed for dogs in group K from the last dose of hydromorphone to the end of anesthesia. However, the number of doses of hydromorphone given intraoperatively was identical between groups. Even though there was no difference in pain scores between the groups, pain might not have been perceptible to an observer due to the sedation level of the patient.

The results of this study should be considered in light of some limitations of this study design. The SDS was scored based on the worst moments during the recovery. An animal vocalizing at any point automatically scored a rank of at least at a three even if the remainder of

the recovery was smooth. Another of the descriptors on the SDS was “trembling” which was occasionally challenging to distinguish from shivering caused by hypothermia, and may have led to inflated scores. The VAS was anchored at “most eventful recovery” on the top end. This could have led to minor elevations in scores as some recoveries were eventful but not necessarily bad. The nature of the videos and noise of the ICU also made it difficult to determine if the audible vocalizations were coming from the dog under observation or not. This could typically be differentiated by looking at respiration patterns, but this was not always to differentiate. The video reviewers did not always have the benefit of knowledge of what sedative or analgesic medications had been given to patients. This should not have had much effect on scores directly, but it could have, especially when a patient was observed on the video to receive an injection that may or may not have been additional sedative medication. Finally, only eight of the 60 recoveries reviewed were considered bad overall (being rated as bad on 2/3 scales by the average of all five sessions’ scores). A difference was still found in this study, but an even larger sample size may be necessary to have a normally distributed range of scores, especially for full validation of the scoring systems across all scenarios.

In this study, anesthetic induction with a combination of ketamine and diazepam produced a better quality recovery than induction with propofol. The difference, however, was approximately a 10% decrease in recovery score on each scale. While this difference was statistically significant, the clinical significance may be negligible.

## REFERENCES

- Aguiar AJA, Luna SPL, Oliva VNLS et al. (2001) Continuous infusion of propofol in dogs premedicated with methotrimeprazine. *Vet Anesth Analg* 28, 220-224.
- Ambros B, Duke-Novakovski, T, Pasloske KS. (2008) Comparison of the anesthetic efficacy and cardiopulmonary effects of continuous rate infusions of alfaxalone-2-hydroxypropyl- $\beta$ -cyclodextrin and propofol in dogs. *Am J Vet Res* 69, 1391-1398.
- American Psychological Association, American Educational Research Association, & National Council on Measurements Used in Education. (1954) Technical recommendations for psychological tests and diagnostic techniques. *Psychol Bull* 51, 1-38.
- Andreoni V, Hughes JML. (2009) Propofol and fentanyl infusions in dogs of various breeds undergoing surgery. *Vet Anesth Analg* 36, 523-531.
- Barletta M, Austin BR, Ko JC, et al. (2011) Evaluation of dexmedetomidine and ketamine in combination with opioids as injectable anesthesia for castration in dogs. *J Am Vet Med Assoc* 238, 1159-1167.
- Bashir MS, Khade A, Borkar P, et al. (2013) A comparative study between different pain rating scales in patients of osteoarthritis. *Indian J Physiol Pharmacol* 57, 205-208.
- Bednarski R, Grimm K, Harvey R, et al. (2011) AAHA Anesthesia Guidelines for Dogs and Cats. *J Am Anim Hosp Assoc* 47, 377-385.
- Bland JM, Altman DG. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 8, 307-310.

- Brüssel T, Theissen JL, Vigfusson G et al. (1989) Hemodynamic and cardiodynamic effects of propofol and etomidate: negative inotropic properties of propofol. *Anesth Analg* 69, 35-40.
- Buckingham BR, McCall WA, Otis AS, et al. (1921) Report of the Standardization Committee. *J Educ Res* 4, 78–80.
- Bufalari A, Miller SM, Giannoni C, et al. (1998) The Use of Propofol as an Induction Agent for Halothane and Isoflurane Anesthesia in Dogs. *J Am Anim Hosp Assoc* 34, 85-91.
- Cai WH, Deng WD, Yang HH, et al. (2102) A propofol microemulsion with low free propofol in the aqueous phase: Formulation, physicochemical characterization, stability and pharmacokinetics. *Int J Pharm* 436, 536-544.
- Chanques G, Viel E, Constantin J, et al. (2010) The measurement of pain in intensive care unit: Comparison of 5 self-report intensity scales. *Pain* 151, 711–721.
- Cockshott ID, Douglas EJ, Plummer GF, Simons PJ. (1992) The pharmacokinetics of propofol in laboratory animals. *Xenobiotica* 22, 369-375.
- Cohen J. (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20, 37-46.
- Dawes, J. (2008) Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *Int J Market Res* 50, 61-77.
- Downie WW, Leatham PA, Rhind VM, et al. (1978) Studies with pain rating scales. *Annals of the Rheumatic Diseases* 37, 378-381.
- Fleiss JL. (1971) Measuring nominal scale agreement among many raters. *Psychol Bull* 76 378–382.
- Formanek A, Korbuly D, Moore R, et al. (1976) The effect of valium on myocardial blood flow. *Radiology* 121, 541-544.

- Friis RH, Sellers TA. (2004). *Epidemiology for Public Health Practice*. Boston, MA: Jones and Bartlett Publishers.
- Goodchild CS, Serrao JM. (1989) Cardiovascular effects of propofol in the anaesthetized dog. *Br J Anaesth* 63, 87-92.
- Green S. Dissociative agents. (1999) In: Krauss B (ed). *Pediatric Procedural Sedation and Analgesia*. Lippincott Williams & Wilkins, Philadelphia; 47-58.
- Grubb TL, Greene SA, Perez TE. (2013) Cardiovascular and respiratory effects, and quality of anesthesia produced by alfaxalone administered intramuscularly to cats sedated with dexmedetomidine and hydromorphone. *J Feline Med Surg* 15, 858-865.
- Gupta A, Stierer T, Zuckerman R et al. (2004) Comparison of Recovery Profile After Ambulatory Anesthesia with Propofol, Isoflurane, Sevoflurane and Desflurane: A Systematic Review. *Anesth Analg* 98, 632–641.
- Hall LW and Chambers JP. (1987) A clinical trial of propofol infusion anaesthesia in dogs. *J Small Anim Pract* 28, 623-637.
- Haskins SC, Farver TB, Patz JD. (1985) Ketamine in Dogs. *Am J Vet Res* 46, 1855-1860.
- Haskins SC, Farver TB, Patz JD. (1986) Cardiovascular changes in dogs given diazepam and diazepam-ketamine. *Am J Vet Res* 47, 795-798.
- Hellebrekers LJ, van Herpen H, Hird JFR et al. (1998) Clinical efficacy and safety of propofol or ketamine anaesthesia in dogs premedicated with medetomidine. *Vet Rec* 142, 631-634.
- Holmes M, Cockcroft P. (2008). *Handbook of Veterinary Clinical Research*. Ames, IA: Blackwell Publishing.
- Huskisson EC. (1974) Measurement of pain. *Lancet*, 2, 1127-1131.

- Jalota L, Kalira V, George E et al. (2011) Prevention of pain on injection of propofol: Systematic review and meta-analysis. *BMJ* 342, d1110.
- Jiménez CP, Mathis A, Mora SS et al. (2012) Evaluation of the quality of the recovery after administration of propofol or alfaxalone for induction of anaesthesia in dogs anaesthetized for magnetic resonance imaging. *Vet Anesth Analg* 39, 151-159.
- Karapinar B, Yilmaz D, Demirag K et al. (2006) Sedation with intravenous ketamine and midazolam for painful procedures in children. *Pediatr Int* 48, 146-151.
- Kohrs R, Durieux ME. (1998) Ketamine: Teaching an Old Drug New Tricks. *Anesth Analg* 87, 1186-1193.
- Landis JR and Koch GG. (1977) The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.
- Langdon DE, Harlan JR, Bailey RL. (1973) Thrombophlebitis with diazepam used intravenously. *J Amer Med Assoc* 223, 184-185.
- Langley MS, Heel RC. (1988) Propofol. A review of its pharmacodynamic and pharmacokinetic properties and use as an intravenous anaesthetic. *Drugs* 35, 334-372.
- Leslie K, Troedel S, Irwin K, et al. (2003) Quality of Recovery from Anesthesia in Neurosurgical Patients. *Anesthesiology* 99, 1158-1165.
- Lozano AJ, Brodbelt DC, Borer KE, et al. (2009) A comparison of the duration and quality of recovery from isoflurane, sevoflurane and desflurane anaesthesia in dogs undergoing magnetic resonance imaging. *Vet Anaesth Analg* 36, 220–229.
- Lund EM, Armstrong PJ, Kirk CA et al. (1999) Health status and population characteristics of dogs and cats examined at private veterinary practices in the United States. *J Am Vet Med Assoc* 214, 1336-1341.

- Mair AR, Pawson P, Courcier E et al. (2009) A comparison of the effects of two different doses of ketamine used for co-induction of anaesthesia with a target-controlled infusion of propofol in dogs. *Vet Anaesth Analg* 36, 532-538.
- Mannarino R, Luna SPL, Monteiro ER, et al. (2012) Minimum infusion rate and hemodynamic effects of propofol, propofol-lidocaine and propofol-lidocaine-ketamine in dogs. *Vet Anesth Analg* 39, 160-173.
- Matthews NS, Brown RM, Barling KS et al. (2004) Repetitive Propofol Administration in Dogs and Cats. *J Am Anim Hosp Assoc* 40, 255-260.
- McCaffery M, and Beebe A. (1993) *Pain: Clinical Manual for Nursing Practice*. Baltimore: V.V. Mosby Company.
- Miller-Keane Encyclopedia and Dictionary of Medicine, Nursing, and Allied Health, Seventh Edition. © 2003 by Saunders, an imprint of Elsevier, Inc.
- Moore DS, McCabe GP. (2006) *Introduction to the Practice of Statistics*. New York, NY: W. H. Freeman and Company.
- Moran CE, Hofmeister EH. (2013) Prevalence of pain in a university veterinary intensive care unit. *J Vet Emerg Crit Care (San Antonio)* 23, 29-36.
- Muhammad N, Zafar MA, Muhammad G et al. (2009) Comparative anaesthetic efficacy of propofol, thiopental sodium and combination of propofol with ketamine hydrochloride in dogs. *Pakistan Vet. J* 29, 11-15.
- Musulini SE, Mariani CL, Papich MG. (2011) Diazepam pharmacokinetics after nasal drop and atomized nasal administration in dogs. *J Vet Pharmacol Ther* 34, 17-24.
- Nagashima Y, Furukawa Y, Chiba S. (2000) Propofol decreases contractility of isolated blood-perfused left ventricular muscle in the dog. *J Anesth* 14, 45-47.

- Nolan A, Reid J (1993) Pharmacokinetics of propofol administered by infusion in dogs undergoing surgery. *Br J Anaesth* 70, 546–551.
- Pagel PS, Warltier DC. (1993) Negative inotropic effects of propofol as evaluated by the regional preload recruitable stroke work relationship in chronically instrumented dogs. *Anesthesiology* 78 100-108.
- Puttick RM, Diedericks J, Sear JW et al. (1992) Effect of graded infusion rates of propofol on regional and global left ventricular function in the dog. *Br J Anaesth* 69, 375-381.
- Pypendop BH, Ilkiw JE. (2005) Pharmacokinetics of ketamine and its metabolite, norketamine, after intravenous administration of a bolus of ketamine to isoflurane-anesthetized dogs. *Am J Vet Res* 66, 2034-2038.
- Reich A, Heisig M, Phan NQ, et al. (2012) Visual analogue scale: evaluation of the instrument for the assessment of pruritus. *Acta Derm Venereol* 92, 497-501.
- Reid J, Nolan AM. (1993) Pharmacokinetics of propofol in dogs premedicated with acepromazine and maintained with halothane and nitrous oxide. *J Vet Pharmacol Ther* 16, 501-505.
- Schwieger IM, Szlam F, Hug CC Jr. (1991) The pharmacokinetics and pharmacodynamics of ketamine in dogs anesthetized with enflurane. *J Pharmacokinet Biopharm* 19, 145–156.
- Scott, W. (1955). Reliability of content analysis: The case of nominal scale coding. *Public Opin Quart* 19, 321–325.
- Segen's Medical Dictionary. (2012). Farlex, Inc.
- Siebke H, Ellersten BB, Lind B. (1976) Reactions to intravenous injections of diazepam. *Br J Anaesth* 48, 1187-1189.
- Sinclair MD, Dyson DH. (2012) The impact of acepromazine on the efficacy of crystalloid,

- dextran or ephedrine treatment in hypotensive dogs under isoflurane anesthesia. *Vet Anaesth Analg* 39, 563-573.
- Smith LJ, Ya JK, Bjorling DE et al. (2001) Effects of hydromorphone or oxymorphone, with or without acepromazine, on preanesthetic sedation, physiologic values, and histamine release in dogs. *J Amer Vet Med Assoc* 218, 1101-1105.
- Spearman C. (1904) The proof and measurement of association between two things. *Amer J Psychol* 15, 72-101.
- Student. (1908) The probable error of a mean. *Biometrika* 6, 1-25.
- Tan CH, Onsiong MK. (1998) Pain on injection of propofol. *Anaesthesia* 53, 468-476.
- Thompson KA, Goodale DB. The recent development of propofol (DIPRIVAN). (2000) *Intensive Care Med* 26 S400-404.
- Tsai TC, Wang LY, Yeh LS. (2007) Clinical Comparison of Recovery from Total Intravenous Anesthesia with Propofol and Inhalation Anesthesia with Isoflurane in Dogs. *J Vet Med Sci* 69, 1179-1182.
- Ueyema Y, Waselau AC, Wiese AJ, et al. (2008) Anesthetic and cardiopulmonary effects of intramuscular morphine, medetomidine, ketamine injection in dogs. *Vet Anesth Analg* 35, 480-487.
- Walker DJ, Zacny JP. (1999) Subjective, psychomotor, and physiological effects of cumulative doses of opioid mu agonists in healthy volunteers. *J Pharmacol Exp Ther* 289, 1454-1464.
- Watkins, SB, Hall LW, Clarke KW. Propofol as an intravenous anaesthetic agent in dogs. (1987) *Vet Rec* 120, 326-329.

Wong DH. Jenkins LC. (1974) An experimental study of the mechanism of action of ketamine on the central nervous system. *Can Anaesth Soc J* 21, 57-67.

Wright M. (1982) Pharmacologic effects of ketamine and its use in veterinary medicine. *J Am Vet Med Assoc* 180, 1462-1471.

APPENDIX A  
RECOVERY SCORING SCALES

VAS

Uneventful \_\_\_\_\_ Most Eventful Possible

Numerical

0    1    2    3    4    5    6    7    8    9    10

SDS

- 1    Very smooth, no excitement, vocalization, trembling or vomiting. No convulsions.
- 2    Quite smooth, a little excitement. No paddling, vocalization, trembling or vomiting. No convulsions.
- 3    Moderately smooth with excitement. Some paddling, vocalization, trembling or vomiting observed. No convulsions.
- 4    Not smooth and with excitement. Paddling, vocalization, trembling or vomiting observed. No convulsions.
- 5    Extreme excitement observed with aggression, vocalization, violent movements or convulsions observed. Rescue sedation or anticonvulsant therapy needed.

APPENDIX B  
PAIN SCORING SCALE

0. No Pain
1. Mild Pain
2. Moderate Pain
3. Severe Pain
4. Worst Possible Pain

## APPENDIX C

### KAPPA INTERPRETATION TABLE

<b><math>\kappa</math></b>	<b>Interpretation</b>
< 0	Poor agreement
0.01 – 0.20	Slight agreement
0.21 – 0.40	Fair agreement
0.41 – 0.60	Moderate agreement
0.61 – 0.80	Substantial agreement
0.81 – 1.00	Almost perfect agreement