



**UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
FACULDADE DE MEDICINA**

Nuno Emanuel de Oliveira Figueiredo da Silva

**Validação da escala da Unesp-Botucatu para
avaliação da dor pós-operatória em ovinos**

**Tese apresentada à Faculdade de
Medicina, Universidade Estadual
Paulista “Júlio de Mesquita Filho”,
Câmpus de Botucatu, para obtenção
do título de Doutor em
Anestesiologia.**

Orientador: Prof. Titular Dr. Stelio Pacca Loureiro Luna

**Botucatu
2020**

Nuno Emanuel de Oliveira Figueiredo da Silva

Validação da escala da Unesp-Botucatu para
avaliação da dor pós-operatória em ovinos

Tese apresentada à Faculdade de
Medicina, Universidade Estadual
Paulista “Júlio de Mesquita Filho”,
Câmpus de Botucatu, para obtenção do
título de Doutor em Anestesiologia.

Orientador: Prof. Titular Dr. Stelio Pacca Loureiro Luna

Botucatu
2020

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO-CRB 8/7500

Silva, Nuno Emanuel Oliveira Figueiredo.

Validação da escala da Unesp-Botucatu para avaliação da dor pós-operatória em ovinos / Nuno Emanuel Oliveira Figueiredo Silva. - Botucatu, 2020

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Medicina de Botucatu

Orientador: Stelio Pacca Loureiro Luna
Capes: 50501011

1. Ovelha. 2. Bem-estar. 3. Analgesia. 4. Ovelha - Comportamento. 5. Dor - Medição. 6. Confiabilidade dos dados.

Palavras-chave: analgesia; bem-estar; comportamento; confiabilidade; ovelhas.

Nuno Emanuel de Oliveira Figueiredo da Silva

Validação da escala da Unesp-Botucatu para
avaliação da dor pós-operatória em ovinos

Tese apresentada à Faculdade de Medicina, Universidade
Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de Botucatu,
para obtenção do título de Doutor em 505.

Orientador: Prof. Titular Dr. Stelio Pacca Loureiro Luna

Comissão examinadora

Prof. Dr. Stelio Pacca Loureiro Luna
Faculdade de Medicina Veterinária e Zootecnia,
Universidade Estadual Paulista “Júlio de Mesquita Filho”
Câmpus de Botucatu, SP

Prof. Dr. José Paes de Oliveira Filho,
Faculdade de Medicina Veterinária e Zootecnia,
Universidade Estadual Paulista “Júlio de Mesquita Filho”
Câmpus de Botucatu, SP

Prof^a. Dra. Silvia Mitiko Nishida
Instituto de Biociências de Botucatu,
Universidade Estadual Paulista “Júlio de Mesquita Filho”

Prof^a. Dra. Renata Navarro Cassu
Faculdade de Ciências Agrárias, Universidade do Oeste Paulista,
Presidente Prudente, SP

Prof^a. Dra. Nadia Crosignani Outeda
Facultad de Veterinaria - UDeLaR – Montevideo, Uruguai

Data da defesa: Botucatu, 21 de fevereiro de 2020

DEDICATÓRIA

Aos meus Pais, pela vida, pela educação, sua formação, renúncia, paciência e testemunhos essenciais para Ser Humano e Ser Médico Veterinário.

Aos meus Avós pelos exemplos de humildade, perseverança e trabalho.

A Vivian Sauer por ser a Mulher tão especial que é, madura e desafiadora.

A Deus e a Jesus por me conceder mais esta oportunidade de evoluir.

AGRADECIMENTOS

Ao Amigo, Orientador e Mestre Stelio Luna pela sua confiança para este projeto. Por ser exemplo de liderança natural e carismática, pelos seus conselhos e críticas sempre construtivas, ajudando-me a focar no essencial.

Ao Amigo e Mestre Jean Joaquim por ser a minha referência como Médico Veterinário em termos clínicos e pela intenção de cura que revela em cada caso.

À Amiga e Professora Maria Luísa Cápuia pela sua confiança a nível profissional, familiar e pessoal.

Ao Renan pela parceria na pesquisa. À Paula, Nathália, Tabata e Alice pelo seu trabalho como anestesistas.

Ao Pedro Trindade pela inestimável colaboração nas análises estatísticas.

A Alice, Marilda e Maria Alice pela avaliação dos vídeos.

A Mayara, Renata, Mariana, Maira, Rubia, Alice e Pedro por fazerem parte de uma equipe unida e solidária.

Aos Professores Nadia Crosignani, Renata Cassu, Silvia Nishida e José Paes pelas considerações feitas na banca do exame de defesa, da mesma forma aos Profs Francisco Neto e Edson Siqueira na banca de exame de qualificação.

À Tatiane, Janete e Márcia da seção de Pós-graduação da FMB por todo amparo desde o primeiro dia de candidatura.

À Capes, pela concessão de 4 anos de bolsa de doutorado.

À FAPESP pelo Projeto Temático (2017/12815-0).

Ao Centro Espírita Caminho da Luz pelo amparo espiritual e oportunidade de trabalho íntimo.

À Dra. Gláucia Lima pelos apoios regulares.

EPIÍGRAFE

O médico verdadeiro não tem o direito de acabar a refeição, de escolher a hora, de inquirir se é longe ou perto. O que não atende por estar com visitas, por ter trabalhado muito e achar-se fatigado, ou por ser alta noite, mau o caminho e o tempo, ficar longe, ou no morro; o que sobretudo pede um carro a quem não tem como pagar a receita, ou diz a quem lhe chora à porta que procure outro – esse não é médico, é negociante de medicina, que trabalha para recolher capital e juros dos gastos da formatura.

Bezerra de Menezes

SILVA, N.E.O.F. **Validação da escala da Unesp-Botucatu para avaliação da dor pós-operatória em ovinos.** 77p. Tese (Doutorado) – Faculdade de Medicina, Universidade Estadual Paulista, Botucatu, 2020.

RESUMO

É importante desenvolver uma ferramenta com validação estatística robusta para diagnosticar a dor em ovinos e proceder a uma intervenção analgésica eficaz. Este estudo objetivou elaborar um etograma para avaliar o comportamento e validar uma escala para avaliar a dor aguda pós-operatória em ovinos. Realizou-se laparoscopia eletiva em 48 ovinos saudáveis, filmados em quatro momentos, um pré-operatório e três pós-operatórios antes e após a analgesia de resgate analgésico e 24 horas. Os vídeos foram aleatorizados e avaliados duas vezes com intervalo de um mês pelos quatro observadores. A análise estatística foi realizada no software R e as diferenças consideradas significativas quando $p < 0,05$. Baseada na associação múltipla adotou-se uma escala unidimensional. A confiabilidade intra e inter-observador variou de moderada a muito boa (coeficiente de correlação intraclassa $\geq 0,55$). A escala apresentou correlação de Spearman $> 0,80$ com as escalas numérica, descritiva simples e visual analógica e correlação de $0,59$ com a escala de expressão facial. Pelo teste de Friedman a escala foi responsiva, pelo aumento e diminuição dos escores de dor de todos os itens após a cirurgia e intervenção analgésica, respectivamente. Todos os itens da escala apresentaram correlação item-total de Spearman aceitável ($0,3 - 0,7$) e a sua consistência interna foi excelente (α de Cronbach = $0,84$). Todos os seus itens apresentaram especificidade $< 0,30$ e sensibilidade entre $0,67 - 0,85$, exceto para apetite. Segundo o índice de Youden o ponto de corte foi de ≥ 4 de 12 , com zona de incerteza diagnóstica de $2,5$ a $3,5$, que indica a acurácia da escala. A área sob a curva $> 0,95$ para todos os avaliadores demonstrou a excelente capacidade discriminatória do instrumento. Classificou-se os escores como dor intensidade leve (0 a 3); moderada (4 e 8) e severa (≥ 9). A escala facial teve boa confiabilidade intraobservador em três avaliadores e a interobservador variou de pobre a moderada; ela foi responsiva e teve muito boa consistência interna ($0,79$). O seu ponto de corte foi de ≥ 7 de 12 , com zona de incerteza diagnóstica de $2,5$ a $6,5$. Em conclusão, a escala composta da Unesp-Botucatu para avaliação da dor pós-operatória em ovinos é um instrumento válido, confiável, específico, sensível, com excelente consistência interna, acurácia e capacidade discriminatória.

Palavras-chave: analgesia, bem-estar, comportamento, confiabilidade, ovelhas.

SILVA, N.E.O.F. **Validation of the Unesp-Botucatu scale to assess postoperative pain in sheep.** 77p. Thesis (PhD) – School of Medicine, São Paulo State University, Botucatu, 2020.

ABSTRACT

The development of a tool with robust statistical validation is important to diagnose pain in sheep and carry out an effective analgesic intervention. The current study aimed to develop an ethogram to evaluate behavior and validate a scale to evaluate acute postoperative pain in sheep. Elective laparoscopy was performed in 48 healthy sheep, filmed at four moments, one preoperative and three postoperative before and after analgesic rescue analgesia and 24 hours after. The videos were randomized and assessed twice by the four evaluators, with an interval of one month between evaluations. The statistical analysis was performed using R software and differences were considered significant when $p < 0.05$. Based on the multiple association, a unidimensional scale was adopted. The intra- and inter-observer reliability ranged from moderate to very good (intraclass correlation coefficient ≥ 0.55). All items on the scale demonstrated an acceptable Spearman item-total correlation (0.3 - 0.7). The scale presented Spearman correlations > 0.80 with the numerical, simple descriptive, and visual analogue scales, and a correlation of 0.59 with the facial expression scale. According to the Friedman test, the scale was responsive, due to the increase and decrease in pain scores of all items after surgery and analgesic intervention, respectively. The internal consistency was excellent (Cronbach's $\alpha = 0.84$) and all items presented specificity < 0.30 and sensitivity between 0.67-0.85, except for appetite. According to the Youden index, the cut-off point was ≥ 4 out of 12, with a diagnostic uncertainty zone of 2.5 to 3.5, which indicates the accuracy of the scale. The area under the curve > 0.95 for all evaluators demonstrated excellent discriminatory capacity of the instrument. The scores of the scale were classified into three intensities of pain: mild (0 to 3); moderate (4 and 8) and severe (≥ 9). The facial scale had good intra-observer reliability in three evaluators and the interobserver ranged from poor to moderate; it was responsive and had very good internal consistency (0.79). Its cut-off point was ≥ 7 out of 12, with a diagnostic uncertainty range of 2.5 to 6.5. In conclusion, the Unesp-Botucatu composite scale to assess postoperative pain in sheep is a valid, reliable, specific, sensitive instrument, with excellent internal consistency, accuracy and discriminatory capacity.

Keywords: analgesia, welfare, behaviour, reliability, ovine.

LISTA DE FIGURAS

Figura 1 - Linha do tempo dos momentos para validação da escala Unesp-Botucatu de avaliação de dor pós-operatória em ovinos.....	23
Figura 2 - Fluxograma com as etapas de elaboração, refinamento e validação da escala Unesp-Botucatu de dor pós-operatória em ovinos. Testes estatísticos para o refinamento: 1) validação de conteúdo; 2) associação múltipla; 3) confiabilidade intraobservador e 4) interobservador; 5) validade de critério; 6) validade de construto; 7) consistência interna; 8) sensibilidade e especificidade. Para a validação final da escala utilizaram-se os 8 testes enumerados no refinamento (com exceção do 1) e o 9) que foi a determinação do ponto de resgate analgésico (vide tabela 1)	24
Figura 3 - Escalas unidimensionais utilizadas para avaliação da dor. 1 – EN, com escore 1 (sem dor) a 10 (pior dor possível); 2 - EDS com quatro níveis descritivos; 3 – EAV com escore 0mm (sem dor) à 100 mm (pior dor possível)	25
Figura 4 - Diagramas de caixa (box-plot) dos escores dos comportamentos que apresentaram diferença significativa relativamente ao valor basal no etograma de dor pós-operatória em ovinos..	33
Figura 5 - Distribuição da porcentagem da presença dos escores de cada item em cada momento da escala Unesp-Botucatu de dor pós-operatória em ovinos.	34
Figura 6 - Biplot da análise de componentes principais com os itens da escala Unesp-Botucatu de dor pós-operatória em ovinos, aplicada pelo observador padrão-ouro nos quatro momentos de avaliação	35
Figura 7 - Correlação entre os escores da escala Unesp-Botucatu de dor pós-operatória em ovinos e os das escalas: A) numérica; B) descritiva simples; C) analógica visual; D) de expressão facial (D) para todos os momentos.....	38
Figura 8 - Diagramas de caixa (box-plot) dos escores (mediana/amplitude) da escala Unesp-Botucatu de dor pós-operatória em ovinos (0-12), comparando os quatro momentos	39
Figura 9 - Curva ROC e área sob a curva (ASC) e curva ROC de dois gráficos com zona de incerteza diagnóstica para a escala Unesp-Botucatu de dor pós-operatória em ovinos	43
Figura 10 - Dendograma criado pela análise de agrupamento não hierárquica de acordo com a somatória total da Escala da Unesp-Botucatu, com formação de três grupos que foram nomeados (baixa, intermediária, alta) em função da intensidade da dor.....	45
Figura 11 - Diagramas de caixa (box-plot) para as três intensidades de dor (mild, moderate e severe) da escala Unesp-Botucatu de dor pós-operatória em ovinos.....	46
Figura 12 - Distribuição da porcentagem da presença dos escores de cada item em cada momento da escala de expressão facial em ovinos.....	47
Figura 13 - Análise de componentes principais com os itens da escala facial em ovinos aplicada pelo observador padrão-ouro nos quatro momentos de avaliação.....	48
Figura 14 - Curva ROC e área sob a curva (ASC) e curva ROC de dois gráficos com zona de incerteza diagnóstica para a escala de expressão facial em ovinos.....	55

LISTA DE TABELAS

Tabela 1 - Diretrizes do refinamento ^(R) e validação ^(V) estatística da escala da Unesp-Botucatu para avaliar a dor pós-operatória em ovinos e da escala de expressão facial ^(F)	27
Tabela 2 - Mediana e amplitude (mínimo – máximo) do percentual de duração de comportamentos de 48 ovinos antes e após a laparoscopia. As porcentagens de cada categoria comportamental foram calculadas com base no tempo total de cada momento de avaliação (20 mins).....	31
Tabela 3 - Escala final validada da Unesp-Botucatu para avaliação de dor aguda pós-operatória em ovinos	32
Tabela 4 - Associação múltipla dos itens da escala entre si após análise de componentes principais. Valores da carga fatorial, autovalor e variância de acordo com critério de Kaiser para todos os momentos da escala Unesp-Botucatu de dor pós-operatória em ovinos ...	34
Tabela 5 - Confiabilidade intraobservador em todos os momentos das duas fases de avaliação para: A) cada item da escala Unesp-Botucatu de dor pós-operatória em ovinos, as escalas numérica e descritiva simples (grau de confiabilidade kw); B) somatória da escala de dor e a escala analógica visual (consistência ICC)	36
Tabela 6 - Confiabilidade interobservador entre os quatro observadores para: A) cada item da escala de Unesp-Botucatu de dor pós-operatória em ovinos, as escalas numérica e descritiva simples (grau de confiabilidade kw); B) somatória da escala Unesp-Botucatu e escala analógica visual (consistência ICC)	36
Tabela 7 - Confiabilidade interobservador analisada pelo coeficiente de correlação intra-classe da matriz (consistência ICC) para o escore total da escala Unesp-Botucatu de dor pós-operatória em ovinos para a 2ª fase de avaliação dos quatro avaliadores entre si....	37
Tabela 8 - Correlação entre a escala Unesp-Botucatu de dor pós-operatória em ovinos e as escalas numérica, descritiva simples, analógica visual e a de expressão facial	38
Tabela 9 - Análise da validade de construto pela comparação dos escores entre os quatro momentos peri-operatórios para: A) cada item e total da escala Unesp-Botucatu em ovinos (0 -12); B) resgate analgésico (0-1), escalas numérica (1-10), descritiva simples (1-4) e analógica visual (0-100).....	39
Tabela 10 - Teste de correlação item-total da escala Unesp-Botucatu de dor pós-operatória em ovinos..	40
Tabela 11 - Consistência interna dos escores de cada item da escala Unesp-Botucatu de dor pós-operatória em ovinos em cada momento	40
Tabela 12 - Especificidade (em M1) e sensibilidade (em M2) de cada item da escala Unesp-Botucatu de dor pós-operatória em ovinos.....	41
Tabela 13 - Especificidade, sensibilidade e índice de Youden de cada observador para cada escore da escala Unesp-Botucatu de dor aguda pós-operatória em ovinos (A); Intervalos de confiança (IC) de 95% de 1001 repetições e de sensibilidade e especificidade > 0,90 aplicados para estimar a zona de incerteza diagnóstica do ponto de corte de cada avaliador, de acordo com o índice de Youden (B).....	42
Tabela 14 - Porcentagem de ovelhas para as quais foi indicada analgesia de resgate de acordo com a experiência clínica e de acordo com o índice de Youden da escala Unesp-Botucatu de dor pós-operatória em ovinos.....	44
Tabela 15 - Porcentagem de ovinos presentes na zona de incerteza diagnóstica de acordo com o índice de Youden escala Unesp-Botucatu de dor pós-operatória em ovinos	45
Tabela 16 - Mediana, mínimo e máximo de cada intensidade da dor da escala Unesp-Botucatu de dor pós-operatória em ovinos.....	45
Tabela 17 - Distribuição da frequência e porcentagem do número de animais em cada intensidade da dor.....	45

Tabela 18 - Associação múltipla dos itens da escala entre si após análise de componentes principais. Valores de carga fatorial, autovalor e variância de acordo com critério de Kaiser para todos os momentos da escala de expressão facial de ovinos.....	47
Tabela 19 - Confiabilidade intraobservador em todos os momentos das duas fases de avaliação para: A) cada item da escala de expressão facial em ovinos (grau de confiabilidade kw) e B) a somatória da escala facial (consistência ICC).....	49
Tabela 20 - Confiabilidade interobservador entre os quatro observadores para: A) cada item da escala de expressão facial em ovinos (grau de confiabilidade kw) e B) a somatória da escala facial em ovinos (consistência ICC)	49
Tabela 21 - Correlação entre a escala de expressão facial em ovinos e as escalas numérica, descritiva simples, analógica visual e a de dor aguda proposta.....	51
Tabela 22 - Análise da validade de construto pela comparação dos escores entre os quatro momentos na 2ª avaliação do observador padrão-ouro de cada item e do total da escala de expressão facial em ovinos (0-12).....	51
Tabela 23 - Teste de correlação item-total da escala de expressão facial.....	52
Tabela 24 - Consistência interna dos escores de cada item da escala de expressão facial em ovinos em cada momento.....	53
Tabela 25 - Especificidade (em M1) e da sensibilidade (em M2) de cada item da escala de expressão facial em ovinos.....	53
Tabela 26 - Especificidade, sensibilidade e índice de Youden de cada observador para cada escore da escala de expressão facial em ovinos (A); Intervalos de confiança (IC) de 95% de 1001 repetições e de sensibilidade e especificidade > 0,90 aplicados para estimar a zona de incerteza diagnóstica do ponto de corte de cada avaliador, de acordo com o índice de Youden (B).....	54
Tabela 27 - Frequência e porcentagem de animais presentes na zona de incerteza diagnóstica de todos os avaliadores da escala de expressão facial em ovinos.....	56

LISTA DE ABREVIATURAS E SIGLAS

ASC – área sob a curva

cm – centímetro

E – especificidade

EAV – escala analógica visual

EDS – escala descritiva simples

EN – escala numérica

et al – colaboradores

EF – escala facial

h – hora

IC – intervalo de confiança

ICC – coeficiente de correlação intraclass

IM – intramuscular

IV – intravenosa

IY – índice de Youden

Kg – quilograma

K_w – grau de confiabilidade kappa

MA – Momentos agrupados

mg – miligrama

min – minuto

ml – mililitro

mm – milímetro

N/A – não aplicável

°C – graus celcius

R – refinamento

RA – resgate analgésico

ROC – Curva Característica de Operação do Receptor

r_s - coeficiente de correlação de Spearman

S – sensibilidade

SC – subcutânea

V – validação

VO – via oral

LISTA DE SÍMBOLOS

α – alfa

= – igual

< – menor

$\leq$ – menor ou igual

> – maior

$\geq$ – maior ou igual

$\pm$ – mais ou menos

% – porcentagem

® – registrado

SUMÁRIO

1. INTRODUÇÃO	17
2. REVISÃO DE LITERATURA	18
3. HIPÓTESE E OBJETIVOS	21
4. MATERIAIS E MÉTODOS	22
4.1 Animais e delineamento experimental	22
4.2 Coleta de dados	22
4.3 Etograma	24
4.4 Escala de dor aguda	25
4.5 Análise estatística	26
5. RESULTADOS	30
5.1 Etograma	30
5.2 Escala de dor Unesp-Botucatu	30
5.3 Escala de expressão facial	46
6. DISCUSSÃO	57
6.1 Limitações	64
7. CONCLUSÃO	65
8. REFERÊNCIAS	66
APÊNDICES	73
Research article submitted to Plos One	78

1. INTRODUÇÃO

A incapacidade de expressar verbalmente a experiência dolorosa e a falta de instrumentos válidos e confiáveis para reconhecer e quantificar a dor em animais de produção compromete o seu estado de bem-estar e limita o uso de analgésicos nestas espécies (FLECKNELL, 2008; HUXLEY; WHAY, 2006; LIZARRAGA; CHAMBERS, 2012; LORENA et al., 2013). A convicção de que estes animais sentem menos dor do que os pequenos animais (RAEKALLIO et al., 2003), os motivos econômicos (DETECTION, 2005; HEWSON et al., 2007), a necessidade de períodos de carência para o consumo de leite e carne devido aos resíduos dos fármacos e a ausência de licenciamento de analgésicos são outros fatores que dificultam tratar a dor em animais de produção (LIZARRAGA; CHAMBERS, 2012).

Os ovinos são submetidos a diversos procedimentos que causam dor, muitas vezes sem a utilização devida de anestesia ou analgesia, como a castração, o corte da cauda, *mulesing* e marcação na orelha (FUTRO; MASŁOWSKA; DWYER, 2015). A dor em ovinos também ocorre em doenças como mastite, pododermatite, abscessos nas patas e miíases externas (DOLAN; FIELD; NOLAN, 2000; KALER et al., 2010; WEST, 2009), o que pode levar a dor crônica e neuropática (DOLAN et al., 2003; LEY; LIVINGSTON; WATERMAN, 1989; LEY; WATERMAN; LIVINGSTON, 1995).

A espécie ovina serve de modelo experimental pela semelhança de tamanho e peso com os humanos. Em 2014 na União Europeia utilizaram-se 60.209 ovinos em pesquisas, um aumento de 108%, em relação à 2011, apenas superado pelos suínos dentre os animais de produção. Os ovinos e suínos são mais usados como modelos do que cães e primatas não humanos para pesquisa e ensino, dada a limitação destes e questões éticas (GIGLIUTO et al., 2014; TAYLOR; REGO, 2016). Dentre os modelos experimentais que enquadram ovelhas estão os estudos pré-clínicos e translacionais em osteoporose (DIAS et al., 2018) e de regeneração óssea e osteointegração de implantes dentários (SARTORETTO et al., 2016).

Para tratar a dor necessita-se de instrumentos que a avaliem. Embora haja vários indicadores úteis para avaliar a nocicepção em situações experimentais, como a injeção de formalina no espaço interdigital (DOLAN et al., 2011), os filamentos de von Frey (WILKES et al., 2012), torniquetes (STUBSJØEN et al., 2010), estímulos elétricos (ONG et al., 1997) e estímulo pneumo-mecânico nos membros (MATHER et al., 2000), estes não são reprodutíveis e são difíceis de utilizar em situações clínicas. A actigrafia monitora os ovinos à distância, dispensando a presença do observador, porém, requer equipamento específico (MCLENNAN et al., 2015). Outras medidas fisiológicas como a concentração de cortisona no pêlo (STUBSJØEN et al., 2015a), a variabilidade da frequência cardíaca (STUBSJØEN et al., 2015b), a pressão arterial, a temperatura ocular e retal, a eletromiografia e a

eletroencefalografia são pouco exequíveis clinicamente e requerem contenção do animal (JONGMAN et al., 2000).

Quando se almeja avaliar a dor a expressão comportamental substitui a ausência de expressão verbal dos animais. O comportamento pode ser observado, não requer contenção do animal em situações de campo, o que não gera estresse, não requer equipamentos e não tem custo, portanto, é o método mais aplicável em situações clínicas ou mesmo experimentais (DE OLIVEIRA et al., 2014; LUNA ET AL., 2015; MOLONY; KENT; MCKENDRICK, 2002).

Contrariamente aos bovinos e caprinos, os ovinos quando sofrem dor sob contenção, tendem a permanecer mais em silêncio (MOLONY; KENT; MCKENDRICK, 2002) e apenas quando liberados expressam comportamentos de dor (MOLONY; KENT, 1997). Tais comportamentos são: redução da interação com o ambiente e com outros animais, anormalidades na marcha, claudicação, relaxamento dos membros pélvicos (que se movem lentamente e de forma alternada, quando os animais estão parados), girar da cabeça, hiporexia, vocalização anormal, movimento de lambes os lábios, lábios curvados, ranger de dentes, tremores, bater o pé no chão e abanar intenso da cauda (FUTRO; MASŁOWSKA; DWYER, 2015; GRANT, 2004; GUESGEN; BEAUSOLEIL; STEWART, 2013; MOLONY; KENT; MCKENDRICK, 2002).

2. REVISÃO DE LITERATURA

Toda escala é uma representação unidimensional e uma escala para a dor é uma experiência subjetiva e baseada na observação de comportamento que exige acurácia. Dois parâmetros são essenciais para se desenvolver e validar uma escala de dor: 1) validade, que indica se o instrumento mede efetivamente o atributo para o qual ele foi concebido (CRELLIN et al., 2007; MCDOWELL, 2006; MORTON et al., 2005) e 2) confiabilidade, que garante equivalência de resultados quando a medida é avaliada por um mesmo observador em ocasiões distintas, ou por diferentes observadores na mesma ocasião (BEYER; WELLS, 1989; CRELLIN et al., 2007; TAFFAREL et al., 2015). A escala deve ter também responsividade, em que os seus escores se alteram em função da variação do estímulo doloroso ou de analgesia (BRONDANI et al., 2012; BUSSIERES et al., 2008; VON BAEYER; SPAGRUD, 2007).

Escala de dor baseadas em dados comportamentais foram desenvolvidas em diferentes espécies: cães (FIRTH; HALDANE, 1999; HOLTON et al., 2001; MORTON et al., 2005), gatos (BRONDANI et al., 2013; BRONDANI; LUNA; PADOVANI, 2011a, 2011b), equinos (BUSSIERES et al., 2008; GRAUBNER et al., 2011; SUTTON et al., 2013a, 2013b; TAFFAREL et al., 2015), bovinos (DE OLIVEIRA et al., 2014) e suínos (LUNA et al., 2015). Em ovinos as escalas mais utilizadas para mensurar a dor pós-operatória ainda são as unidimensionais, como a numérica (EN), a descritiva simples (EDS) e a analógica visual (EAV). Entretanto estes instrumentos avaliam exclusivamente a intensidade da dor (VON BAEYER; SPAGRUD, 2007), enquanto que as escalas multidimensionais ou compostas, incluem qualidades sensoriais e emocionais (MURRELL et al., 2008).

Para desenvolver as escalas comportamentais, produz-se um etograma espécie-específico que quantifica a duração e/ou frequência dos comportamentos presentes antes e após um estímulo doloroso (BANKS, 1982). Embora estudos anteriores tenham reportado os descritores comportamentais de dor baseados em etograma após um estímulo nocivo (DWYER; LAWRENCE, 2000; FUTRO; MASŁOWSKA; DWYER, 2015; GRAHAM; KENT; MOLONY, 1997; GRANT, 2004; GUESGEN et al., 2016a; KENT; MOLONY; GRAHAM, 1998; KENT; MOLONY; ROBERTSON, 1995; LESTER et al., 1996; MCGLONE; STOBART, 1986; MELCHES et al., 2007; MELLEMA et al., 2006; MELLOR; MURRAY, 1989; MOLONY et al., 2012; MOLONY; KENT, 1997; MOLONY; KENT; ROBERTSON, 1993; SMALL et al., 2014; THORNTON; WATERMAN-PEARSON, 1999, 2002), de acordo com o nosso conhecimento, na literatura não há escalas comportamentais validadas para detectar a dor aguda em ovinos, seguindo análise estatística robusta (BELSHAW et al., 2015; MCDOWELL et al., 2006).

Até à presente pesquisa, os instrumentos desenvolvidos para avaliar a dor aguda na espécie ovina, se basearam em alterações comportamentais em cordeiros submetidos à orquiectomia e corte de cauda (MOLONY; KENT; MCKENDRICK, 2002) ou expressão facial

em ovinos com pododermatite e mastite (MCLENNAN et al., 2016). Nestes não se avaliaram as validades de critério, conteúdo e construto. Para se garantir uma mensuração fidedigna da dor é necessário desenvolver um instrumento com validação estatística robusta, como a realizada em gatos (BRONDANI et al., 2013), bovinos (DE OLIVEIRA et al., 2014), equinos (TAFFAREL et al., 2015) e suínos (LUNA et al., 2015), que utilizaram metodologia encoberta e aleatória, com evidências de validade, confiabilidade, sensibilidade, especificidade e ponto de intervenção analgésica definido (MEROLA; MILLS, 2016).

3. HIPÓTESE E OBJETIVOS

Diante da hipótese que a escala proposta neste estudo apresenta confiabilidade, responsividade e validades de conteúdo, construto e critério adequadas, o objetivo principal deste estudo foi desenvolver e validar uma escala comportamental para avaliar a dor aguda em ovinos, submetidos a cirurgia de tecidos moles (laparoscopia). Como objetivos secundários definiu-se um ponto de corte para a intervenção analgésica e aplicou-se a mesma análise estatística a uma escala de expressão facial já validada, para se correlacionar com a escala comportamental.

4. MATERIAIS E MÉTODOS

O projeto foi aprovado pela Comissão de Ética de Uso de Animais (CEUA) da Faculdade de Medicina Veterinária e Zootecnia, da UNESP, campus de Botucatu, sob o protocolo nº 0027/2017 e no qual foram utilizadas instalações desta instituição. O estudo seguiu as normas do ARRIVE (*Animal Research: Reporting of In Vivo Experiments*) (KILKENNY et al., 2010).

4.1 Animais e delineamento experimental

Utilizaram-se 48 ovinos (espécie *Ovis aries*, linhagem leiteira), da própria instituição, das raças Bergamácia, Lacaune e Dorper, cujas fêmeas tinham $3,5 \pm 1,8$ (1,5 – 6) anos de idade e peso de $58,5 \pm 17,3$ (34 - 75) kg. Como critérios de inclusão os ovinos foram considerados saudáveis pela avaliação clínica e laboratorial (hematócrito, proteína plasmática, glicose e lactato). Para se adaptarem ao ambiente, colocaram-se os ovinos nas baias, próximas ao cercado que habitavam 24 horas antes do início do estudo, período durante o qual estiveram em jejum alimentar de 12 horas de jejum hídrico. Em cada baia (3 x 2 x 1,10 m ou 2,20 x 2 x 1,20 m - comprimento x largura x altura) permaneceram de 4 a 5 ovinos ou 2 a 3 ovinos respectivamente.

Para o estudo piloto, filmaram-se e avaliaram-se as ovelhas antes e após a laparoscopia, para testar a posição das cameras e garantir a qualidade das filmagens. Para a fase experimental selecionaram-se os seguintes momentos para os registros comportamentais: M1 – uma hora antes da cirurgia; M2 – no momento previsto de maior dor, entre três e quatro horas após a cirurgia ; M3 – uma hora após a intervenção analgésica e M4 – 24 horas após a cirurgia.

Imediatamente antes da cirurgia administrou-se 30.000 UI/kg de penicilina benzatina (Pentabiótico®, Zoetis, São Paulo, SP, Brasil) por via intramuscular (IM). Após anestesia dissociativa com 0,5 mg/kg de diazepam (Compaz®, Cristália, Itapira, SP, Brasil) e 5 mg/kg de cetamina (Cetamin®, Syntec; Santana de Parnaíba, SP, Brasil) administrados por via intravenosa (IV), realizou-se anestesia epidural lombossacra com 0,1 ml/kg de lidocaína 1% sem vaso constritor (Xylestesin®, Cristália, Itapira, SP, Brasil) e infiltração anestésica com lidocaina 2% sem vasoconstritor (Xylestesin®, Cristália, Itapira, SP, Brasil) no local de incisão e posterior introdução de trocar. Quando os animais apresentaram alguma sensibilidade relacionada à cirurgia, caracterizada por movimentação de cabeça ou membros, ou aumento acima de 20% da frequência cardíaca em relação ao valor mensurado antes do início da cirurgia, complementou-se a anestesia dissociativa com 5 mg/kg de cetamina IV.

Em todos os animais um mesmo cirurgião experiente realizou as laparoscopias para aspiração folicular e reposição das células foliculares (DE ARAUJO et al., 2016; RODRIGUEZ

et al., 2015; TEIXEIRA et al., 2011b). Inseriram-se três trocâteres (5 mm) em três regiões retroumbilicais: a paramediana esquerda para introduzir a ótica do laparoscópio, a paramediana direita e a mediana retroumbilical para inserir as pinças atraumáticas de manipulação do útero, ovários e tubas uterinas e do sistema de aspiração. No local de inserção dos trocâteres suturou-se com fio de nylon 0, em pontos simples separados e aplicou-se spray de terramicina e hidrocortisona (Terracortril®, Pfizer, Itapevi, SP, Brasil) e repelente tópico. Para transferência intrafalopiana de oócitos incisou-se 1,5 cm de pele, tecido subcutâneo, musculatura abdominal e peritônio para introduzir uma pinça atraumática para apreensão da ponta do corno uterino. Realizou-se a intervenção analgésica após a avaliação de M2, com 0,5 mg/kg meloxicam 2% (Maxicam®, Ourofino, Cravinhos, SP, Brasil) e 0,2 mg/kg morfina (Dimorf®, Cristália, Itapira, SP, Brasil) IV em seringas separadas (Figura 1).

Figura 1 - Linha do tempo dos momentos para validação da escala Unesp-Botucatu de avaliação de dor pós-operatória em ovinos

Momentos	M1 1h antes da cirurgia	Anestesia (Diazepam + Cetamina IV Local: Lidocaína (epidural)	Cirurgia	M2 3-4h após cirurgia	Analgesia de resgate (Meloxicam + Morfina IV)	M3 1h após resgate analgésico	M4 24 h após cirurgia
Filmagens							

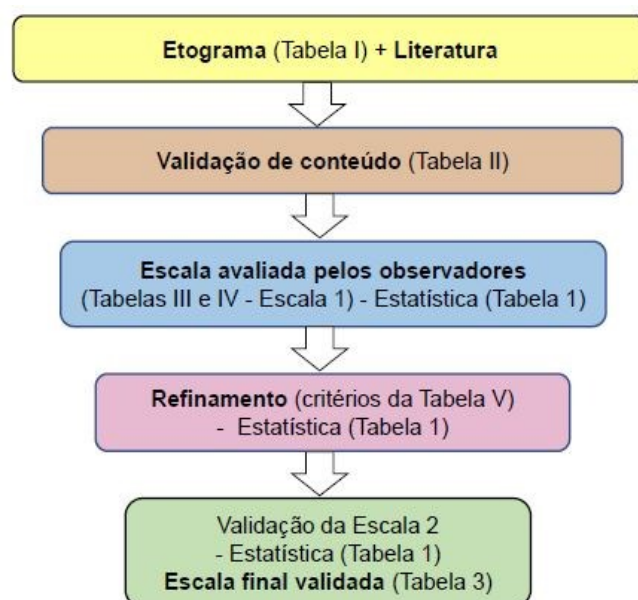
4.2 Coleta de dados

Foram operados de três a quatro animais por dia. Os procedimentos iniciaram-se às 9 horas da manhã e as avaliações comportamentais dos últimos animais finalizaram-se ao redor de 19 horas. O estudo foi realizado entre os meses de março e abril de 2017 e a temperatura e umidade medias do ambiente variaram entre 16 e 24 °C. e 68-92% respectivamente. O local tinha as seguintes características coordenadas geográficas: latitude - 22°51 'S; longitude - 48°26 'O; altitude - 818 m.observador presencial realizou as filmagens com câmera digital (Gopro Hero5 Black®) posicionada em um tripé. A câmera era ligada e o pesquisador presencial se retirava do local, de forma a minimizar a interferência humana no comportamento dos ovinos.

Para análise do comportamento relacionado à dor em ovinos, dividiu-se a pesquisa nas seguintes fases: 1) elaboração de um etograma (Tabela I) para caracterizar o comportamento dos animais antes e após o procedimento doloroso; 2) validação de conteúdo dos comportamentos normais e com dor com base em estudos prévios, estudo piloto e etograma (DWYER; LAWRENCE, 2000; FUTRO; MASŁOWSKA; DWYER, 2015; GRAHAM; KENT; MOLONY, 1997; GRANT, 2004; GUESGEN et al., 2016a; KENT; MOLONY; GRAHAM, 1998; KENT; MOLONY; ROBERTSON, 1995; LESTER et al., 1996; MCGLONE; STOBART, 1986; MELCHES et al., 2007; MELLEMA et al., 2006; MELLOR; MURRAY, 1989; MOLONY

et al., 2012; MOLONY; KENT, 1997; MOLONY; KENT; ROBERTSON, 1993; SMALL et al., 2014; THORNTON; WATERMAN-PEARSON, 1999, 2002) (Tabela II); 3) produção de uma escala de pré-refinamento (Tabela III e IV - escala 1), usada para avaliar os vídeos, que foram reavaliados após 30 dias, por quatro observadores encobertos quanto ao momento; 4) análise estatística da escala de pré-refinamento (Tabela III e IV) avaliada pelos observadores de acordo com os critérios da Tabela 1; 5) critério de refinamento aplicado a escala (Tabela IV) com base na análise estatística da Tabela 1; 6) validação da escala final (escala 2) após refinamento; 8) comparação da escala proposta com a escala facial e as três escalas unidimensionais (Figura 2).

Figura 2 - Fluxograma com as etapas de elaboração, refinamento e validação da escala Unesp-Botucatu de dor pós-operatória em ovinos.



Testes estatísticos para o refinamento: 1) validação de conteúdo; 2) associação múltipla; 3) confiabilidade intraobservador e 4) interobservador; 5) validade de critério; 6) validade de construto; 7) consistência interna; 8) sensibilidade e especificidade. Para a validação final da escala utilizaram-se os 8 testes enumerados no refinamento (com exceção do 1) e o 9 que foi a determinação do ponto de resgate analgésico (vide tabela 1).

4.3 Etograma

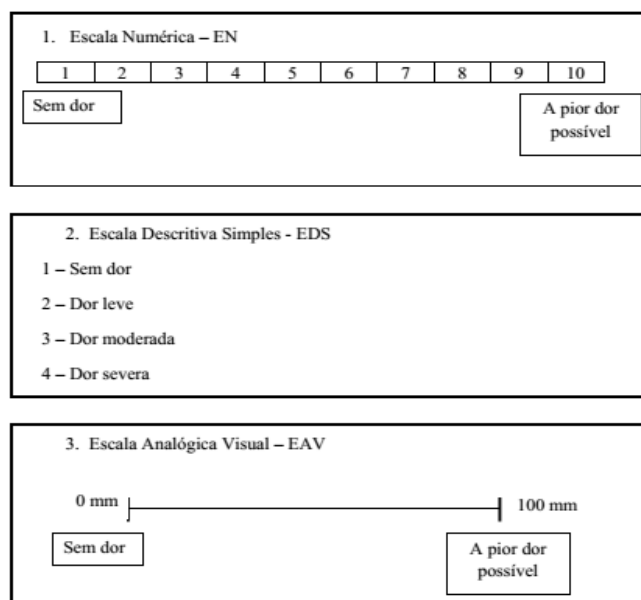
Para elaborar o etograma o pesquisador presencial avaliou por duas vezes vídeos de 20 minutos para cada um dos quatro momentos ($48 \text{ animais} \times 4 = 192 \text{ vídeos}$), num total de 64 horas de filmagens do estudo principal usando pelo método animal focal (MARTIN; BATESON, 1993). Estas avaliações objetivaram: 1) reconhecer as diferentes unidades de comportamentos com e sem dor e 2) medir a duração do tempo (em segundos) de cada comportamento e calcular a respectiva porcentagem de tempo em relação ao tempo total de filmagem de 20 minutos (Tabela I). A seguir os vídeos foram editados com inclusão dos

comportamentos predominantes para o período de cerca de três minutos em cada momento. Os vídeos editados foram avaliados por quatro observadores para o processo de validação da escala.

4.4 Escala de dor aguda

Os vídeos foram disponibilizados em plataforma virtual aos observadores, que avaliavam os quatro momentos de cada animal, de forma cega e aleatorizada quanto aos momentos. Ao final da observação de cada vídeo, os observadores, com base na sua experiência clínica, responderam se realizariam ou não a analgesia de resgate. Estes dados foram utilizados para determinar o ponto de corte relacionado à necessidade de intervenção analgésica. Na sequência determinou-se os escores de dor por meio de três escalas unidimensionais (EN, EDS e EAV). A EN variou de de “0” a “10”, onde “0” representa ausência de dor e “10” a pior dor possível ou imaginável; a EDS variou de 1 - sem dor, 2 – dor leve, 3 – dor moderada e 4 – dor intensa e a EAV se baseou numa linha reta com 100 mm de comprimento, em que “0” representa o animal sem dor e “100” a pior dor possível (Figura 3) (FIRTH; HALDANE, 1999; HOLTON et al., 2001; MORTON et al., 2005).

Figura 3 - Escalas unidimensionais utilizadas para avaliação da dor. 1 – EN, com escore 1 (sem dor) a 10 (pior dor possível); 2 - EDS com quatro níveis descritivos; 3 – EAV com escore 0mm (sem dor) à 100 mm (pior dor possível).



Como alguns subitens da Tabela III continham muito descritores agrupados (exemplo: “locomomoção”), eles foram subdivididos para avaliação individual da importância de cada um. O instrumento proposto para os quatro avaliadores baseou-se em uma escala composta por sete itens – cinco com três subitens, e dois com quatro subitens e alguns dos subitens com várias subdivisões, o que totalizou 33 variáveis comportamentais (Tabela IV).

No final, foi aplicada a escala facial de dor em ovinos (MCLENNAN et al., 2016), por meio do registro fotográfico capturado pelo observador presencial no final das filmagens de 20 minutos de cada momento (vide as fotografias das ovelhas correspondentes aos escores de cada um dos seis itens, com três subitens cada).

Para o refinamento e validação da escala de dor aguda e da escala facial, o padrão-ouro foi selecionado pelo que apresentou maior confiabilidade intra-observador e maior coerência ao indicar menos resgates analgésicos no momento sem dor (M1) e mais resgates no momento com dor (M2). Para o refinamento, todos os itens e subitens da Tabela IV foram avaliados de acordo com os critérios da Tabela 1 conforme consta na Tabela V do apêndice.

4.5 Análise estatística

As análises estatísticas foram realizadas no software R no ambiente de desenvolvimento integrado RStudio (RSTUDIOTEAM, 2016). Para todas as análises foi considerado α de 5%. Estimou-se o tamanho da amostra com poder do teste de 0,80 para teste de chi-quadrado (função “pwr.t.test” do pacote “pwr”). Para esta estimativa calculou-se o tamanho do efeito, aplicando a porcentagem de necessidade de resgate analgésico em M1 (sem dor; 9%) e em M2 (com dor; 91%; função “ES.w1” do pacote “htest”).

Para análise do etograma comparou-se o tempo gasto em segundos e as porcentagens de duração de cada categoria comportamental entre os quatro momentos pelo teste de Friedman (BRONDANI et al., 2013; DE OLIVEIRA et al., 2014). A tabela 1 apresenta as diretrizes do processo de refinamento e validação estatística da escala proposta para quantificar a dor aguda em ovinos e da escala de expressão facial, com a descrição dos tipos de análises efetuadas e os respectivos testes estatísticos.

A escala 1 (Tabela III) avaliada pelos observadores (Tabela IV) foi submetida ao refinamento conforme a Tabela 1. Para refinar a escala final validada seguiram-se os seguintes critérios de inclusão e exclusão de itens e subitens de acordo com 10 testes estatísticos a partir dos dados da segunda fase da avaliação do observador padrão-ouro (Tabela V): 1) etograma (maior escore do comportamento pelo teste de Friedman em M2 vs os outros momentos); 2) validação de conteúdo; 3) ao menos 15% de frequência de ocorrência de itens/subitens em M2; 4) associação múltipla (análise de componentes principais); 5) confiabilidade intraobservador e 6) interobservador; 7) validade de construto (testes de Friedman – maior escore do item/subitem no M2 vs os outros três momentos); 8) correlação item-total; 9) consistência interna; 10) especificidade e sensibilidade; Os subitens foram submetidos às análises 1, 2, 3, 4, 5 e 6. As variáveis que atingiram os critérios estipulados em mais de 50% destes testes estatísticos foram aceitas.

Para a escala facial analisaram-se seis descritores, por meio das análises 4 a 10 que se utilizaram para a validação da escala de dor aguda.

Tabela 1 - Diretrizes do refinamento ^(R) e validação ^(V) estatística da escala da Unesp-Botucatu para avaliar a dor pós-operatória em ovinos e da escala de expressão facial ^(F).

Adaptado da literatura (BRONDANI et al., 2013; DE OLIVEIRA et al., 2014; LUNA ET AL., 2015; NETEMEYER; BEARDEN; SHARMA, 2003; STREINER; NORMAN; CAIRNEY, 2015; TAFFAREL et al., 2015).

Tipo de análise	Descrição	Teste estatístico
Validação de conteúdo ^R	Um comitê composto por três especialistas analisou cada subitem dentro de cada item da escala em: relevante (+1), não sabe (0) ou irrelevante (-1).	Somou-se todos os valores de cada subitem (-1, 0 ou 1) e dividiu-se pelo número de observadores. Incluíram-se na escala os itens com escore total > 0,5.
Distribuição de escores ^{V F}	Construiu-se uma tabela com a distribuição da frequência da presença dos escores 0, 1 e 2 de cada item em cada momento e em nos MA pontuados na 2ª fase de avaliação do observador padrão-ouro.	Análise descritiva.
Associação múltipla ^{R V F}	Analisou-se a associação múltipla dos itens entre si em todos os momentos agrupados da escala comportamental e facial na 2ª fase de avaliação do observador padrão-ouro, por meio da análise de componentes principais, para definir o número de dimensões (ou domínios) determinados por diferentes variáveis que estabelecem a extensão da escala.	Análise de componentes principais (funções “princomp” e “get_pca_var” dos pacotes “stats” e “factoextra” respectivamente). De acordo com o critério de Kaiser (KAISER, 1958), selecionaram-se as dimensões representativas dos componentes com autovalor > 1 e variância > 20 e cada item da escala com valor de carga ≥ 0,50 ou ≤ - 0,50.
Confiabilidade intraobservador ^{R V F}	Repetibilidade - estimou-se o nível de concordância de cada observador consigo mesmo comparando as duas fases de avaliação, utilizando os escores de cada item, a soma total das escala comportamental, facial, numérica (EN), descritiva simples (EDS), analógica visual (EAV) e a necessidade de resgate analgésico. O observador com a maior somatória dos coeficientes de confiabilidade foi adotado como padrão-ouro.	Para os escores dos itens da escala comportamental e das EN e EDS, e necessidade de resgate analgésico usou-se o coeficiente de kappa ponderado (k_w) sendo que as discordâncias foram ponderadas de acordo com a sua distância ao quadrado da concordância perfeita. Estimou-se o intervalo de confiança (IC) de 95% do k_w (função “cohen.kappa” do pacote “psych”) Para a EAV foi usado o coeficiente de correlação intraclassa (ICC) tipo “agreement” e seu IC de 95% (função “icc” do pacote “irr”) (COHEN, 1968; LANDIS; KOCH, 1977; SCHUSTER, 2004) Para a somatória da escala comportamental foi usado o ICC tipo “consistency” e seu IC de 95%. Interpretação de k_w e ICC : muito boa 0,81 – 1,0; boa: 0,61 – 0,80; moderada: 0,41 – 0,60; razoável: 0,21 – 0,4; pobre < 0,2. Utilizou-se k_w e ICC > 0,50 como critério para refinar a escala (ALTMAN, 1991).
Confiabilidade interobservador ^{R V F}	1) Reprodutibilidade – estimou-se o nível de concordância entre o observador padrão-ouro e os outros três observadores na 2ª fase de avaliação, utilizando os escores de cada item; a soma total da escala comportamental, EN, EDS, EAV e a necessidade de resgate analgésico. 2) Matriz de concordância - gerou-se uma matriz para avaliar a concordância do escore total da escala comportamental na 2ª fase de avaliação de cada observador <i>versus</i> todos os outros observadores.	Para a somatória da escala comportamental foi usado o ICC tipo “consistency” e seu IC de 95%. Grau de confiabilidade ICC (consistência): muito boa: 0,81 - 1,0; boa: 0,61 - 0,80; moderada: 0,41 - 0,60; razoável: 0,21 - 0,4; pobre < 0,2 (ALTMAN, 1991). Aceitaram-se valores > 50.
Validade de critério ^{R V F}	Teste de validação concorrente (relação com método padrão-ouro): estimou-se a correlação da somatória da escala comportamental com as EN, EDS, EAV e escala facial de cada momento e também de todos momentos agrupados pontuados pelo observador padrão-ouro na 2ª fase de avaliação.	Coeficiente de correlação de postos de Spearman (r_s; função “rcorr” do pacote “Hmisc”). Interpretação do grau de correlação r_s (p<0,05): 0 – 0,35: correlação baixa; 0,35 – 0,7: correlação média; 0,7 – 1,0: correlação alta (BRONDANI et al., 2012).
Validade de construto ^{R V F}	Responsividade – comparou-se os comportamentos do etograma observados pelo pesquisador presencial nos quatro momentos de avaliação; comparou-se os escores de cada item das escala comportamental, facial, EN, EDS, EAV e a necessidade de resgate	Para variáveis dicotômicas (necessidade de resgate analgésico) aplicou-se a análise de regressão logística (função “glm” do pacote “stats”) usando o teste de Tukey (função “lsmeans” do pacote “lsmeans”) como teste post hoc. A normalidade de cada variável em cada momento foi

	analgésico ao longo do tempo (M1 vs M2 vs M3 vs M4) de acordo com a 2ª fase de avaliação do observador padrão-ouro. Interpretação: espera-se que as diferenças dos escores ocorram da seguinte forma: M2>M4≥M3>M1. Determinou-se a validade do construto pelo método do teste das três hipóteses: 1ª) se a escala realmente medir a dor, o escore após a cirurgia (M2) deve ser maior que o escore pré-operatório (M1 vs M2); 2ª) o escore deve diminuir após analgesia (M2 vs M3); 3ª) e ao longo do tempo (M2 vs M4).	avaliada por gráficos de caixas e histogramas (funções “boxplot” e “histogram” dos pacotes “graphics” e “lattice” respectivamente). A partir destes gráficos os itens da escala de dor, EN, EDS e EAV foram considerados não paramétricos. Desta forma utilizou-se o teste de Friedman (função “friedman.test” do pacote “stats”) em que o p-valor foi corrigido com o procedimento de Bonferroni (função “pairwiseSignTest” do pacote “rcompanion”) (29,46). Para a somatória da escala utilizou-se modelo linear misto (função “lmer” do pacote “lme4”), incluindo os momentos e as raças como efeitos fixos e o efeito aleatório de indivíduos, na qual o p-valor também foi corrigido pelo procedimento de Bonferroni. As diferenças entre os momentos foram identificadas por letras diferentes (função “cldList” do pacote “rcompanion”).
Correlação item-total ^{R V F}	Para análise da homogeneidade, relevância de cada item da escala e para investigar itens inflacionários, estimou-se a correlação de cada item com a somatória de todos os itens da escala, além da correlação entre cada item e a somatória de todos os outros itens (exceto este). Análise foi feita para cada momento e para todos os momentos agrupados (MA) na 2ª fase de avaliação do observador padrão-ouro.	Coefficiente de correlação de postos de Spearman (r_s; função “rcorr” do pacote “Hmisc”). Valores aceitáveis quando o grau de correlação r_s variou entre 0,3 - 0,7 (STREINER; NORMAN; CAIRNEY, 2015).
Consistência interna ^{R V}	Estimou-se a consistência (interrelação) dos escores de cada item da escala e o processo foi feito excluindo cada item. A análise foi feita para cada um dos 4 momentos e para todos eles agrupados (MA) na 2ª fase de avaliação do observador padrão-ouro.	Coefficiente alfa de Cronbach (α; função “cronbach” do pacote “psy”) (CRONBACH, 1951) Interpretação: 0,60-0,64, minimamente aceitável; 0,65-0,69 aceitável; 0,70-0,74 bom; 0,75-0,80 muito bom e > 0,80 excelente (JENSEN, 2003; STREINER, 2003; STREINER; NORMAN; CAIRNEY, 2015).
Especificidade ^{R V F}	Os escores das escalas comportamental e facial na 2ª fase de avaliação do observador padrão-ouro em M1 foram transformados em dicotômicos: (nível “0” - ausência de comportamento de expressão de dor para um dado item; níveis “1” e “2” - presença de comportamento de expressão de dor) e aplicados na respetiva equação.	$E_{M1} = \frac{VN}{VN + FP}$ 1 - E = especificidade. VN = verdadeiro negativo (escores que representavam comportamentos de ausência de dor – “0” – no momento em que os animais supostamente não apresentavam dor, por ser antes da cirurgia – M1). FP = falso positivo (escores que representavam comportamentos de expressão de dor – 1 ou 2 – no momento em que os animais supostamente não deveriam apresentar dor, por ser antes da cirurgia – M1). Interpretação: excelente 95 - 100%; boa 85 - 94.9%; moderada 70 - 84.9%; não é específica <70%. Só se incluíram itens ≥ 70% (40).
Sensibilidade ^{R V F}		$S_{M2} = \frac{VP}{VP + FN}$ S = sensibilidade. VP = verdadeiro positivo (escores que representavam comportamentos de expressão de dor – 1 ou 2 – no momento em que os animais deveriam apresentar dor, por ser após a cirurgia – M2). FN = falso negativo (escores que representavam comportamentos de ausência de dor – 0 – no momento em que os animais supostamente deveriam apresentar dor, por ser após a cirurgia – M2). Interpretação:

		excelente 95 - 100%; boa 85 - 94,9%; moderada 70 - 84,9% não é sensível < 70%. Só se incluíram itens $\geq 70\%$ (40)
Ponto de resgate analgésico VF	A necessidade de analgesia foi usada como valor verdadeiro e a soma da escala comportamental e facial como valor preditivo para construir uma curva de característica operacional do receptor (ROC). Em seguida, o ponto de corte para o resgate analgésico foi determinado com base no índice de Youden e na sua zona de incerteza diagnóstica, usando todos os momentos de avaliação da dor na escala comportamental pontuada por todos e todos os observadores na 2ª fase de avaliação. - Calculou-se a área sob a curva (ASC) que indica a capacidade discriminatória do teste. - Calculou-se a frequência e porcentagem dos animais pontuados na zona de incerteza diagnóstica do ponto de corte de todos os avaliadores para a escala, por meio de análise estatística descritiva. - Ponto de corte foi representado pelo máximo do índice de Youden usando todos os momentos da avaliação da dor nas escalas comportamental e facial pontuadas por todos os observadores na 2ª fase de avaliação.	$IY = S + E - 1$ IY = Índice de Youden; S = sensibilidade; E = especificidade. Análise da curva de característica de operação do receptor (ROC; função “roc” do pacote “pROC”) e da área sob a curva (ASC): representação gráfica da relação entre os “VP” (S) e os “FP” ($1-E$). IY é o ponto de maior sensibilidade e especificidade simultaneamente, determinado pela curva ROC. (DEYO; DIEHR; PATRICK, 1991). Interpretação: $ASC \geq 0,95$ - alta capacidade discriminatória da escala (STREINER; CAIRNEY, 2007). Determinou-se a zona de incerteza diagnóstica por dois métodos, calculando: 1º) o intervalo de confiança de 95% (IC) replicando 1001 vezes a curva ROC original pelo método bootstrap (funções “ci.coords” e “ci.auc” do pacote “pROC”) num; 2º) o intervalo entre os valores de sensibilidade e especificidade de 0,90. O valor mais baixo e mais alto desses dois métodos de todos os avaliadores foi considerado a zona de incerteza diagnóstica, que indica a acurácia diagnóstica (CANNESSON et al., 2011; CELEITA-RODRÍGUEZ et al., 2019; GALL et al., 2015; MALLAT et al., 2015).
Determinação da intensidade da dor ^V	Os escores foram classificados por intensidade: baixa, moderada ou severa , pontuados no momento de maior dor (M2) da escala comportamental por todos os observadores na 2ª fase de avaliação.	Empregou-se análise de agrupamento não hierárquica pelo método “máximo” calculando a distância entre a matriz pelo “centroide” (funções “dist” e “hclust” do pacote “stats”) (TRINDADE et al., 2020) seguido por teste de Kruskal-Wallis (função “kruskal” do pacote “agricolae”) para avaliar a diferença entre os grupos formados pela análise de agrupamento.

Escala comportamental – escala da Unesp-Botucatu para avaliar a dor pós-operatória em ovinos. Escala facial – escala de expressão facial. Escalas: numérica (EN), descritiva simples (EDS), analógica visual (EAV). As análises estatísticas foram realizadas no software R no ambiente de desenvolvimento integrado RStudio (RSTUDIOTEAM, 2016). Para todas as análises foi considerado α de 5%. M1 - pré-operatório; M2 – pós-operatório, antes do resgate analgésico; M3 - pós-operatório, após resgate analgésico; M4 – 24h após a cirurgia; MA - dados dos momentos agrupados (M1 + M2 + M3 + M4).

5. RESULTADOS

5.1 Etograma

Estimou-se o tamanho mínimo da amostra de 5 ovinos (67). Quanto ao etograma, a partir da descrição dos comportamentos (Tabela I – Apêndice), o tempo e porcentagens de cada comportamento observado (Tabela 2), comparando-se o momento de maior dor (M2) com o de suposta ausência de dor (M1) observaram-se as seguintes diferenças: em M2 diminuíram os tempos “comer”, “interação normal” e “cabeça acima da coluna” e aumentaram os tempos de “locomução alterada e ausente” e “dorso arqueado”. Após o resgate analgésico (M3), comparativamente ao M2, os tempos de “interação” e “locomução” normais e “apetite” aumentaram, e “cabeça abaixo da coluna” diminuiu. As categorias comportamentais referentes à provável presença de dor que apresentaram maior duração em M2 em relação ao M1 foram “locomução reduzida e anormal” e “arquear o dorso” (Figura 4).

Já dentre as categorias comportamentais que sugerem o estado de normalidade e que apresentaram menor duração em M2 em relação ao M1 foram: “interação e locomução normal” e “cabeça acima da coluna”. O ato de “comer” aumentou em M3, em relação ao M2, pois os animais estavam sob jejum em M1 e desta forma não foi possível comparar M2 vs M1 para esta variável.

5.2 Escala de dor Unesp- Botucatu

De acordo com os critérios de inclusão/exclusão citados para o refinamento (Tabela V) excluíram-se os seguintes subitens: no item “locomução”, “pode andar para trás” e “andar em círculo”; no item “postura”, “escoiceia e bate com os membros no chão” e “estende um ou mais membros”; no item “miscelânea de comportamentos”, excluíram-se “tremores no corpo” e “em decúbito se arrasta sem se levantar.” Os dois subitens que se mantiveram da “miscelânea” substituíram os excluídos da “postura” e a miscelânea ficou assim com quatro subitens e passou se denominar “postura.”

A seguir procedeu-se à validação da versão final da escala da Unesp-Botucatu para avaliação de dor aguda pós-operatória em ovinos, que contém seis itens, cinco com três subitens – de escore “0” sem dor ou dor leve, “1” dor moderada e “2” dor intensa – e um item com quatro. (Tabela 3). O escore total foi baseado na soma de cada item, variando de zero (sem dor) a 12 (máxima dor),

Tabela 2 - Mediana e amplitude (mínimo – máximo) do percentual de duração de comportamentos de 48 ovinos antes e após a laparoscopia. As porcentagens de cada categoria comportamental foram calculadas com base no tempo total de cada momento de avaliação (20 mins).

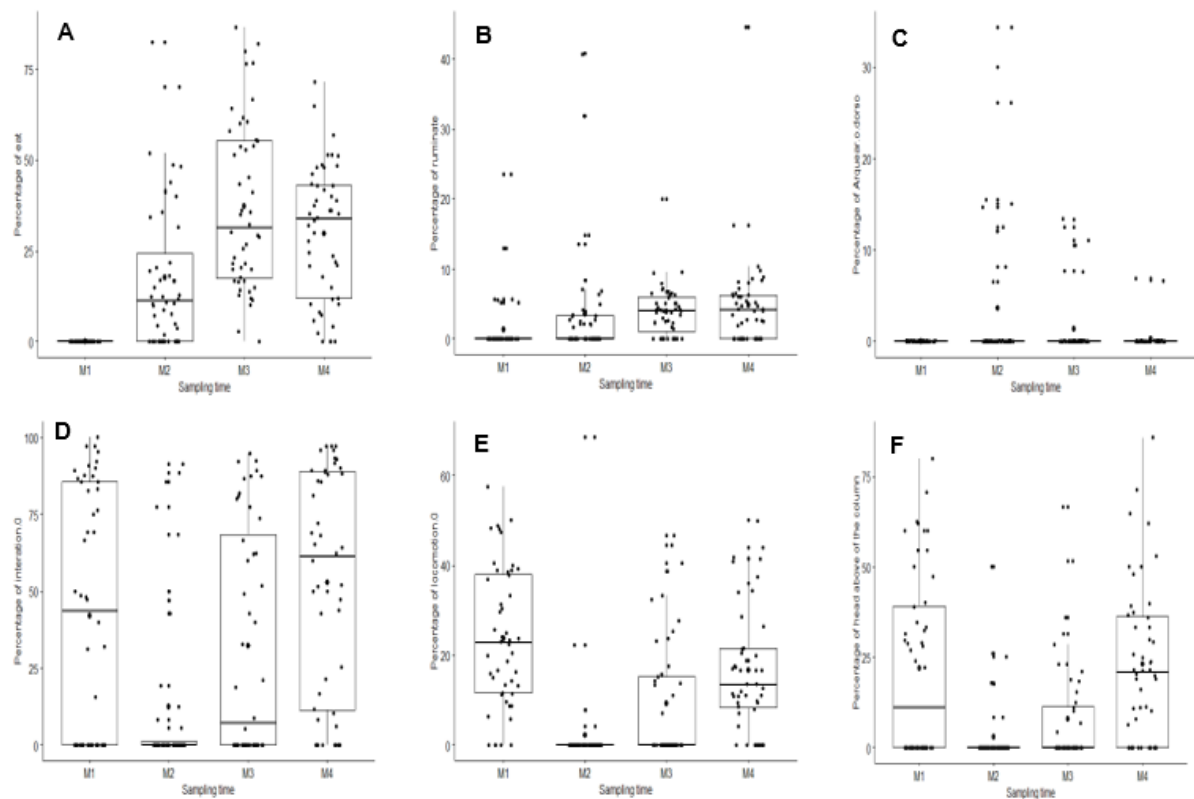
Momentos	M1		M2		M3		M4	
Comportamentos	Mediana	Amplitude	Mediana	Amplitude	Mediana	Amplitude	Mediana	Amplitude
Comer	0^c	0 – 0	11.23^b	0 – 82.26	31.25^a	0 – 86.36	33.71^{ab}	0 – 71.43
Ruminar	0^b	0 – 23.53	0^{ab}	0 – 40.68	4.04^a	0 – 20	4.13^a	0 – 44.44
Beber	0	0 – 6.76	0	0 – 25.93	0	0 – 15.52	0	0 – 2.17
Urinar	0	0 – 9.09	0	0 – 12	0	0 – 20.75	0	0 – 5.41
Defecar	0	0 – 0	0	0 – 18.52	0	0 – 0	0	0 – 0
Interação normal	43.68^a	0 – 100	0^b	0 – 91.3	7.15^a	0 – 94.74	61.25^a	0 – 97.3
Interação reduzida	0 ^a	0 – 96.77	8.09 ^a	0 – 94.29	0 ^a	0 – 100	0 ^b	0 – 97.3
Interação ausente	0	0 – 100	0	0 – 100	0	0 – 30	0	0 – 87.5
Locomoção normal	22.79^a	0 – 57.45	0^b	0 – 68.42	0^c	0 – 46.67	13.51^a	0 – 50
Locomoção reduzida/alterada	0 ^b	0 – 20.83	0 ^a	0 – 25	0 ^b	0 – 40	0 ^a	0 – 47.37
Locomoção anormal/ausente	0 ^b	0 – 70	2.13 ^a	0 – 100	0 ^{ab}	0 – 60	0 ^b	0 – 30
Cabeça acima da linha da coluna	11.11^a	0 – 80	0^b	0 – 50	0^b	0 – 66.67	20.89^a	0 – 85.71
Cabeça na linha da coluna	0	0 – 67.5	0	0 – 78.26	0	0 – 86.75	0	0 – 81.08
Cabeça abaixo da linha da coluna	9.82 ^a	0 – 69.77	12.7 ^a	0 – 81.08	6.46 ^b	0 – 60	5.48 ^b	0 – 25.00
Em pé parado	60.51 ^b	33.33 – 88.57	66.95 ^{ab}	0 – 93.75	73.87 ^a	0 – 100	74.26 ^{ab}	10.26 – 92
Em pé posição alterada	4.54 ^a	0 – 31.25	6.12 ^a	0 – 36.84	1.04 ^{ab}	0 – 60	0 ^b	0 – 8
Escoicear	0 ^b	0 – 0	0 ^{ab}	0 – 5	0 ^b	0 – 0	0 ^a	0 – 14.29
Estender a cabeça e pescoço e ou membro	0 ^{ab}	0 – 31.25	0 ^a	0 – 37.5	0 ^{ab}	0 – 60	0 ^b	0 – 10.53
Deitado	0	0 – 31.17	0	0 – 90	0	0 – 34.62	0	0 – 44.44
Deitado com alteração (cabeça rodada sobre si)	0	0 – 0	0	0 – 0	0	0 – 0	0	0 – 0
Deitado com a cabeça apoiada próxima ao solo	0	0 – 27.91	0	0 – 100	0	0 – 32	0	0 – 0
Olhar para ferida cirúrgica	0	0 – 0	0	0 – 0	0	0 – 0	0	0 – 0
Lamber a ferida cirúrgica	0	0 – 0	0	0 – 0	0	0 – 0	0	0 – 0
Movimento de cauda brusco e repetido	0	0 – 40	0	0 – 0	0	0 – 0	0	0 – 0
Mantér a cauda esticada	0	0 – 22.22	0	0 – 5.8	0	0 – 0	0	0 – 50
Arquear o dorso	0^b	0 – 0	0^a	0 – 34.29	0^{ab}	0 – 13.33	0^{ab}	0 – 6.9
Tremores no corpo	0	0 – 0	0	0 – 11.11	0	0 – 0	0	0 – 0

Letras diferentes expressam diferenças significativas entre momentos (valores em negrito expressam diferenças em M2 em relação ao M1 ou M3, especificamemnte para “comer”) sendo que a>b>c, de acordo com o teste de Friedman ($p < 0.05$) (BRONDANI et al., 2013; DE OLIVEIRA et al., 2014). M1: pré-operatório; M2: pós-operatório, antes da analgesia de resgate; M3: pós-operatório, após analgesia de resgate; M4: 24h pós-operatório;

Tabela 3 - Escala final validada da Unesp-Botucatu para avaliação de dor aguda pós-operatória em ovinos.

Item	Subitem (descritores)	Escore
Interação	Ativo, atento ao ambiente, interage e/ou acompanha os outros animais.	0
	Apático: pode permanecer próximo aos outros animais, mas interage pouco.	1
	Apático: isolado ou não interage com os outros animais, não se interessa no ambiente.	2
Atividade	Movimenta-se normalmente.	0
	Inquieto, movimenta-se acima do normal ou deita e levanta com frequência.	1
	Movimenta-se com menor frequência ou somente ao ser estimulado ou não se movimenta.	2
Locomoção	Move-se livremente, sem alteração de locomoção; quando parado, os membros pélvicos estão paralelos aos torácicos.	0
	Move-se com restrição e/ou com passos curtos e/ou com pausas e/ou claudica; quando parado os membros torácicos ou pélvicos podem estar mais abertos do que o normal.	1
	Dificuldade e/ou relutância em se levantar e/ou não se locomove e/ou anda de forma anormal e/ou claudica e/ou quando parado os membros podem estar mais abertos e para trás do que o normal; apoia-se contra uma superfície.	2
Posição da cabeça	Acima da linha da coluna.	0
	Na linha da coluna.	1
	Abaixo da linha da coluna.	2
Apetite	Normorexia e/ou ruminação presente	0
	Hiporexia	1
	Anorexia	2
Postura	a) Dorso arqueado. b) Estende a cabeça e pescoço. c) Se em decúbito ventral a cabeça se apoia no solo ou fica próxima ao solo na linha da coluna d) Movimenta a cauda brusca e repetidamente e/ou mantém a cauda esticada (exceto para defecar/urinar)	
	• Ausência destes comportamentos	0
	• Presença de um dos comportamentos relacionados	1
	• Presença de dois ou mais dos comportamentos relacionados	2

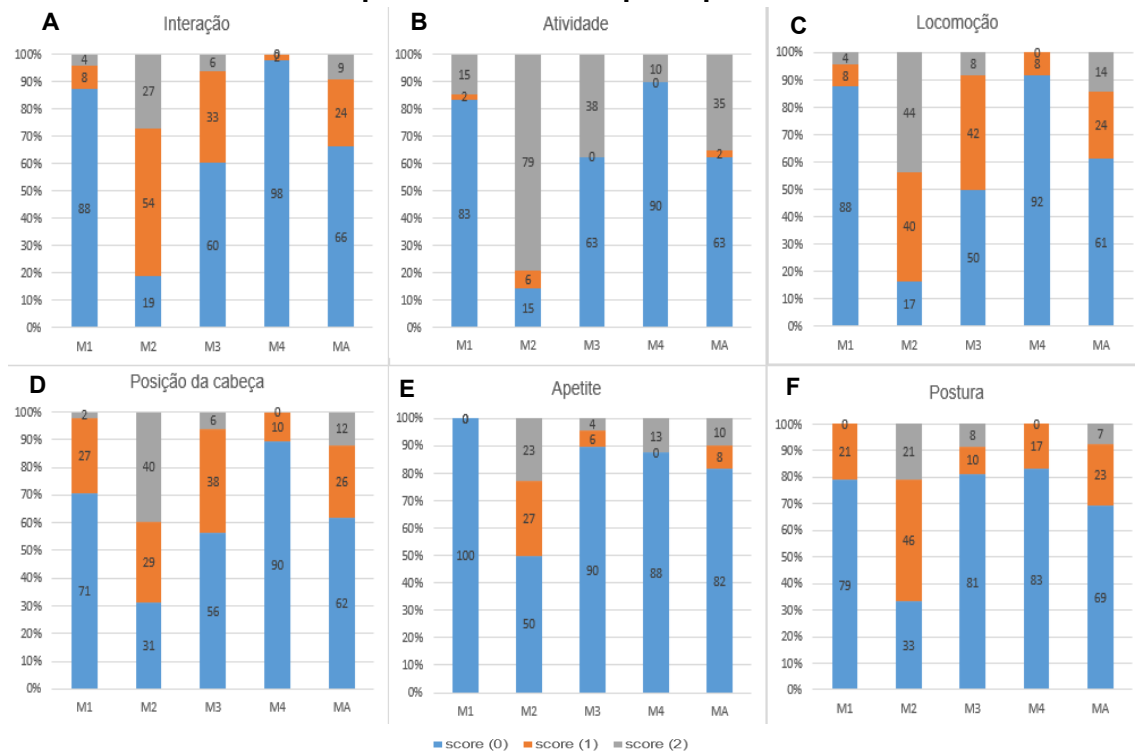
Figura 4 - Diagramas de caixa (box-plot) dos escores dos comportamentos que apresentaram diferença significativa relativamente ao valor basal no etograma de dor pós-operatória em ovinos.



A) comer; **B)** ruminar; **C)** arquear o dorso; **D)** interação normal; **E)** locomoção normal; **F)** cabeça acima da coluna. M1 - pré-operatório; M2 – pós-operatório, antes do resgate analgésico; M3 - pós-operatório, após resgate analgésico; M4 – 24h após a cirurgia (BRONDANI et al., 2013; DE OLIVEIRA et al., 2014).

A distribuição dos escores “0, 1 e 2” decorreu como esperado, de acordo com o grau de dor. O escore “0” predominou nos momentos M1, M3 e M4. Os escores “1” e “2” foram mais frequentes em M2 e se reduziram em M3 (Figura 5). Apenas os itens “atividade” e “apetite” de escore “1” foram pouco representativos quando se agrupou todos os momentos.

Figura 5 - Distribuição da porcentagem da presença dos escores de cada item em cada momento da escala Unesp-Botucatu de dor pós-operatória em ovinos.



A) interação; **B)** atividade; **C)** locomoção; **D)** posição da cabeça; **E)** apetite; **F)** postura. M1: pré-operatório; M2: pós-operatório, antes da analgesia de resgate; M3: pós-operatório, após analgesia de resgate; M4: 24h pós-operatório; MA - dados dos momentos agrupados (M1 + M2 + M3 + M4).

A associação múltipla dos itens da escala entre si em todos os momentos avaliada pela análise de componentes principais selecionou o componente principal 1, correspondente a uma dimensão representativa, motivo matemático pelo qual a escala foi considerada matematicamente unidimensional. (Tabela 4; Figura 6).

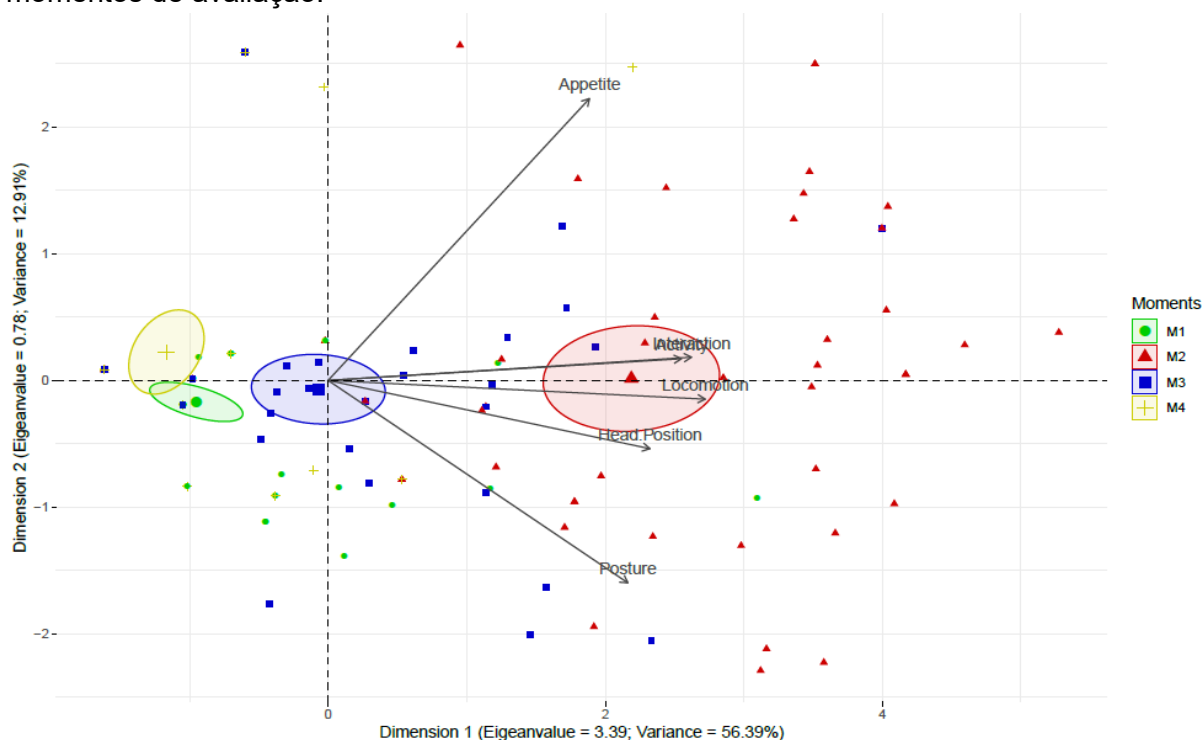
Tabela 4 - Associação múltipla dos itens da escala entre si após análise de componentes principais. Valores da carga fatorial, autovalor e variância de acordo com critério de Kaiser para todos os momentos da escala Unesp-Botucatu de dor pós-operatória em ovinos.

Itens	Dimensão 1	Dimensão 2
	Valor de carga	Valor de carga
Interação	0,82	-0,06
Atividade	0,80	-0,05
Locomoção	0,85	0,05
Posição da cabeça	0,73	0,17
Apetite	0,59	-0,70
Postura	0,68	0,50
Autovalor	3,39	0,78
Variância	56,39	12,91

Análise de componentes principais - A estrutura fatorial foi determinada considerando itens com carga fatorial $\geq 0,50$ ou $\leq -0,50$, com autovalor > 1 e variância $> 20\%$. A carga fatorial representa a correlação

entre os itens e os fatores. Os valores de carga em negrito indicam as variáveis que contribuem para cada fator e o respectivo autovalor e variância aceites (KAISER, 1958).

Figura 6 - Biplot da análise de componentes principais com os itens da escala Unesp-Botucatu de dor pós-operatória em ovinos, baseada no observador padrão-ouro nos quatro momentos de avaliação.



Elipses de confiança foram construídas de acordo com os momentos de avaliação da dor e os animais. Os dados foram obtidos a partir da análise do observador padrão-ouro nos quatro momentos de avaliação. As elipses foram construídas de acordo com os momentos de avaliação da dor (M1 – verde, M2 - vermelho, M3 – azul e M4 - amarelo). A elipse referente ao momento em que ovinos sofriam dor intensa (M2) foi posicionada na extremidade direita do gráfico, enquanto que no lado oposto estavam os momentos em que os ovinos provavelmente não sofriam dor (M1 e M4). O momento de dor moderada (M3) está posicionado ao meio. Todos os itens da escala são influenciados pelos momentos com dor (M2 e M3) uma vez que seus vetores estão posicionados na direção dessas elipses. M1 - pré-operatório; M2 – pós-operatório, antes do resgate analgésico; M3 - pós-operatório, após resgate analgésico; M4 – 24h após a cirurgia.

A confiabilidade intraobservador variou de razoável a boa para cada item da escala (com exceção do avaliador 3 que foi pobre a boa) e variou de boa a muito boa para o seu escore total em todos os momentos avaliados (Tabela 5). Selecionou-se como padrão-ouro o avaliador 2. A confiabilidade interobservador – entre o padrão-ouro e os outros três observadores – foi moderada (com exceção de uma que variou de pobre a moderada) para cada item da escala (Tabela 6). A concordância entre todos os observadores (correlação de matriz) na segunda fase da sua avaliação para o escore total da escala variou de moderada a boa (Tabela 7).

Tabela 5 - Confiabilidade intraobservador em todos os momentos das duas fases de avaliação para: A) cada item da escala Unesp-Botucatu de dor pós-operatória em ovinos, as escalas numérica e descritiva simples (grau de confiabilidade k_w); **B)** somatória da escala de dor e a escala analógica visual (consistência ICC).

A. Grau de confiabilidade k_w	Avaliador 1			Avaliador 2			Avaliador 3			Avaliador 4		
Itens da escala	k_w	Min	Max	k_w	Min	Max	k_w	Min	Max	k_w	Min	Max
Interação	0,64	0,55	0,74	0,66	0,56	0,77	0,52	0,41	0,64	0,65	0,56	0,74
Atividade	0,54	0,42	0,66	0,65	0,54	0,76	0,46	0,34	0,58	0,47	0,35	0,60
Locomoção	0,71	0,63	0,80	0,66	0,55	0,77	0,48	0,36	0,61	0,61	0,52	0,71
Posição da cabeça	0,67	0,56	0,77	0,71	0,62	0,80	0,48	0,35	0,61	0,59	0,48	0,69
Apetite	0,55	0,42	0,69	0,61	0,48	0,73	0,56	0,43	0,69	0,15	0,02	0,28
Postura	0,42	0,27	0,57	0,61	0,48	0,73	0,45	0,28	0,61	0,32	0,23	0,42
Resgate analgésico	0,67	0,56	0,77	0,75	0,65	0,85	0,53	0,41	0,65	0,53	0,40	0,65
Escala numérica	0,80	0,74	0,86	0,85	0,80	0,91	0,58	0,48	0,68	0,72	0,64	0,79
Escala descritiva simples	0,78	0,71	0,85	0,77	0,69	0,84	0,61	0,50	0,71	0,67	0,59	0,76

B. Coeficiente correlação intraclasse	Avaliador 1			Avaliador 2			Avaliador 3			Avaliador 4		
Escalas	ICC	Intervalo confiança		ICC	Intervalo confiança		ICC	Intervalo confiança		ICC	Intervalo confiança	
Escala Unesp-Botucatu	0,77	0,70	0,82	0,83	0,78	0,87	0,65	0,57	0,73	0,71	0,64	0,78
Escala analógica visual	0,80	0,74	0,85	0,81	0,76	0,86	0,56	0,44	0,64	0,71	0,63	0,77

Interpretação do grau de confiabilidade k_w ou ICC (consistência): muito boa: 0,81 - 1,0; boa: 0,61 - 0,80; moderada: 0,41 - 0,60; razoável: 0,21 - 0,4; pobre < 0,2 (ALTMAN, 1991). Negritos correspondem aos valores aceitáveis > 50.

Tabela 6. Confiabilidade interobservador entre os quatro observadores para: A) cada item da escala de Unesp-Botucatu de dor pós-operatória em ovinos, as escalas numérica e descritiva simples (grau de confiabilidade k_w); **B)** somatória da escala Unesp-Botucatu e escala analógica visual (consistência ICC).

A.	Avaliador 2 (padrão-ouro) vs								
Grau de confiabilidade k_w	Avaliador 1			Avaliador 3			Avaliador 4		
Itens da escala	k	Min	Max	k	Min	Max	k	Min	Max
Interação	0,52	0,42	0,62	0,60	0,49	0,70	0,45	0,35	0,55
Atividade	0,58	0,47	0,69	0,55	0,43	0,67	0,50	0,35	0,55

Locomoção	0,56	0,46	0,66	0,48	0,35	0,60	0,49	0,37	0,60
Posição da cabeça	0,54	0,42	0,65	0,48	0,36	0,61	0,43	0,31	0,54
Apetite	0,57	0,44	0,69	0,63	0,51	0,75	0,18	0,03	0,33
Postura	0,47	0,33	0,61	0,50	0,36	0,64	0,31	0,19	0,42
Resgate analgésico	0,53	0,41	0,64	0,63	0,52	0,74	0,46	0,35	0,57
Escala numérica	0,43	0,36	0,50	0,66	0,59	0,73	0,49	0,41	0,56
Escala descritiva simples	0,65	0,57	0,74	0,65	0,56	0,74	0,63	0,55	0,72

B	Avaliador 2 (padrão-ouro) vs							
	Avaliador 1			Avaliador 3			Avaliador 4	
Coeficiente de correlação intraclasse								
Escalas	ICC	Intervalo de confiança		ICC	Intervalo de confiança		ICC	Intervalo de confiança
Escala Unesp-Botucatu	0,72	0,65	0,78	0,76	0,70	0,82	0,63	0,53 - 0,70
Escala analógica visual	0,30	0,16	0,42	0,59	0,49	0,68	0,44	0,32 - 0,55

Interpretação do grau de confiabilidade k_w ou ICC (consistência): muito boa: 0,81 - 1,0; boa: 0,61 - 0,80; moderada: 0,41 - 0,60; razoável: 0,21 - 0,4; pobre < 0,2 (ALTMAN, 1991; BRONDANI et al., 2013). Negritos correspondem aos valores aceitáveis > 50.

Tabela 7. Confiabilidade interobservador analisada pelo coeficiente de correlação intra-classe da matriz (consistência ICC) para o escore total da escala Unesp-Botucatu de dor pós-operatória em ovinos para a 2ª fase de avaliação dos quatro avaliadores entre si.

	Avaliador 1		Avaliador 2		Avaliador 3		Avaliador 4	
Coeficiente de correlação intraclasse	ICC	IC	ICC	IC	ICC	IC	ICC	IC
Avaliador 1	-	-	0,72	0,65 - 0,78	0,71	0,64 - 0,78	0,74	0,67 - 0,80
Avaliador 2	0,72	0,65 - 0,78	-	-	0,76	0,70 - 0,82	0,63	0,53 - 0,70
Avaliador 3	0,71	0,64 - 0,78	0,76	0,70 - 0,82	-	-	0,63	0,54 - 0,71
Avaliador 4	0,74	0,67 - 0,80	0,63	0,53 - 0,70	0,63	0,54 - 0,71	-	-

Interpretação do grau de confiabilidade k_w ou ICC (consistência): muito boa: 0,81 - 1,0; boa: 0,61 - 0,80; moderada: 0,41 - 0,60; razoável: 0,21 - 0,4; pobre < 0,2 (ALTMAN, 1991; BRONDANI et al., 2013). Negritos correspondem aos valores aceitáveis > 50.

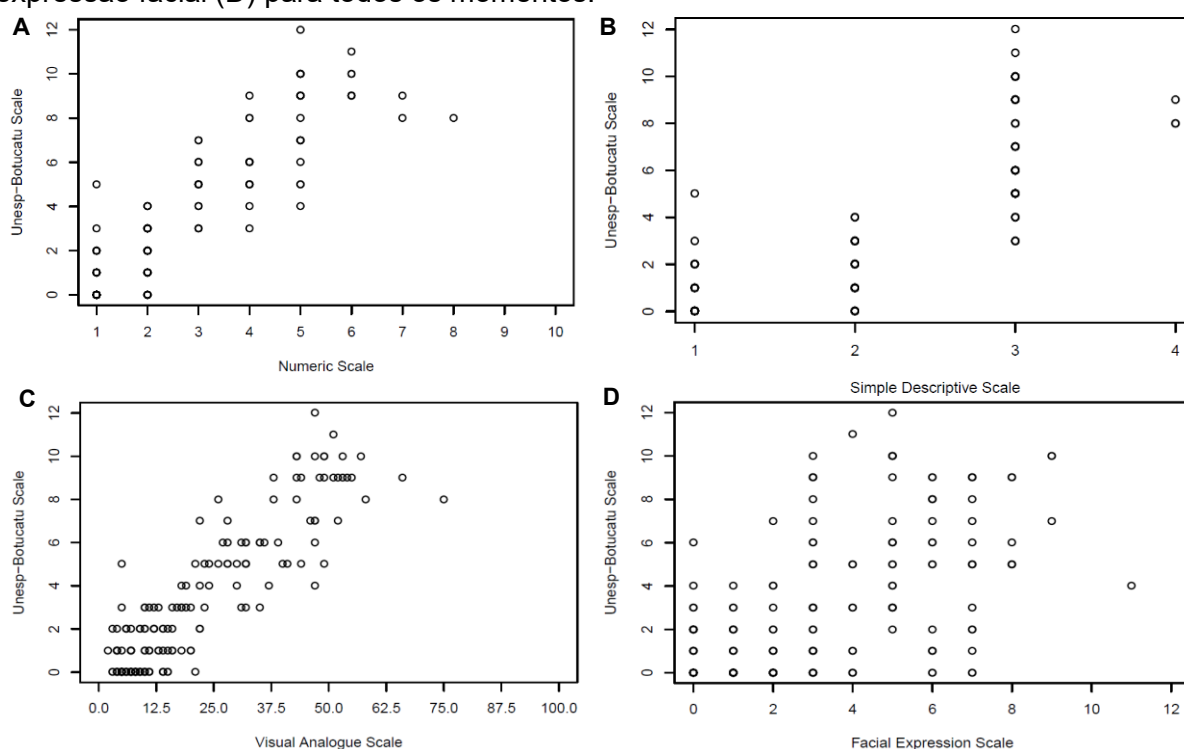
O teste de validação concorrente para analisar a validade de critério registrou alta correlação entre os escores da escala de dor e os da EN ($r = 0,87$), EDS ($r = 0,86$), EAV ($r = 0,81$) e média correlação com a escala facial ($r = 0,59$) para todos os momentos (MA) avaliados pelos quatro avaliadores (Tabela 8; Figura 7).

Tabela 8 - Correlação entre a escala Unesp-Botucatu de dor pós-operatória em ovinos e as escalas numérica, descritiva simples, analógica visual e a de expressão facial.

Escalas	Escala Unesp-Botucatu				
Momentos	M1	M2	M3	M4	MA
Escala numérica	0,53	0,8	0,84	0,7	0,87
Escala descritiva simples	0,53	0,58	0,84	0,7	0,86
Escala analógica visual	0,38	0,76	0,81	0,41	0,81
Escala expressão facial	-0,03	0,32	0,53	0,02	0,59

Interpretação do coeficiente de correlação de Spearman ($p < 0,05$): 0 – 0,35: correlação baixa; 0,35 – 0,7: correlação média; 0,7 – 1,0: correlação alta (BRONDANI et al., 2013). Em negrito valores > 70 . M1: pré-operatório; M2: pós-operatório, antes da analgesia de resgate; M3: pós-operatório, após analgesia de resgate; M4: 24h pós-operatório; MA - dados dos momentos agrupados (M1 + M2 + M3 + M4).

Figura 7 - Correlação entre os escores da escala Unesp-Botucatu de dor pós-operatória em ovinos e os das escalas: A) numérica; B) descritiva simples; C) analógica visual; D) de expressão facial (D) para todos os momentos.



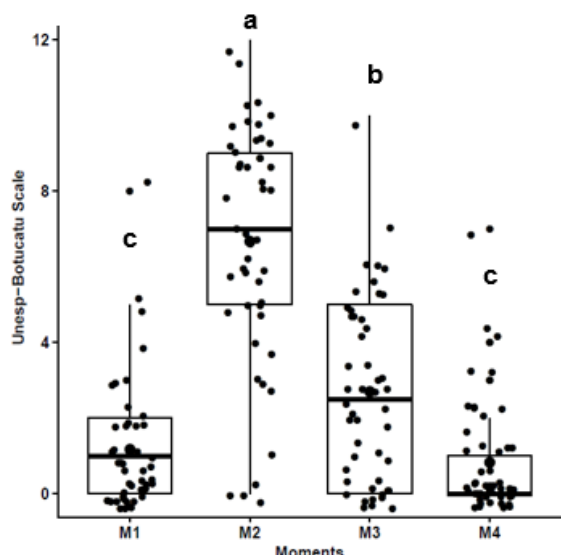
Os escores de todos os itens e da somatória da escala em M2 foram significativamente maiores do que em M1, M3 e M4, evidenciando a sua responsividade (validade de construto). As diferenças entre os escores totais da escala proposta, EN, EDS e EAV decorreram da seguinte forma: $M2 > M3 > M4 > M1$ (Tabela 9; Figura 8).

Tabela 9. Análise da validade de construto pela comparação dos escores entre os quatro momentos para: A) cada item e total da escala Unesp-Botucatu em ovinos (0 -12); B) resgate analgésico (0-1), escalas numérica (1-10), descritiva simples (1-4) e analógica visual (0-100).

Escalas/Momentos	M1		M2		M3		M4	
A. Itens	Mediana	Amplit.	Mediana	Amplit.	Mediana	Amplit.	Mediana	Amplit.
Interação	0 ^{bc}	0 - 2	1 ^a	0 - 2	0 ^b	0 - 2	0 ^c	0 - 1
Atividade	0 ^{bc}	0 - 2	2 ^a	0 - 2	0 ^b	0 - 2	0 ^c	0 - 2
Locomoção	0 ^c	0 - 2	1 ^a	0 - 2	0,5 ^b	0 - 2	0 ^c	0 - 1
Posição cabeça	0 ^{bc}	0 - 2	1 ^a	0 - 2	0 ^b	0 - 2	0 ^c	0 - 1
Apetite	0 ^{bc}	0 - 0	0,5 ^a	0 - 2	0 ^b	0 - 2	0 ^b	0 - 2
Postura	0 ^{bc}	0 - 1	1 ^a	0 - 2	0 ^b	0 - 2	0 ^b	0 - 1
Escala Unesp-Botucatu	2^c	0 - 8	7^a	0 - 12	2,5^b	0 - 10	0^c	0 - 7
B. RA	0^c	0-1	1^a	0-1	0^b	0-1	0^c	0-1
EN	1^c	1-4	5^a	1-8	2^b	1-5	1^c	1-5
EDS	1^c	1-3	3^a	1-3	2^b	1-3	1^c	1-3
EAV	8,5^{bc}	2-38	43^a	7-75	15^b	4-53	7^c	3-47

Letras diferentes expressam diferenças significativas entre momentos (valores em negrito) sendo que $a > b > c$, de acordo com o teste de Friedman ($p < 0.05$) (BRONDANI et al., 2013; DE OLIVEIRA et al., 2014). M1: pré-operatório; M2: pós-operatório, antes da analgesia de resgate; M3: pós-operatório, após analgesia de resgate; M4: 24h pós-operatório; MA - dados dos momentos agrupados (M1 + M2 + M3 + M4). RA: resgate analgésico; Escalas: numérica (EN), descritiva simples (EDS), analógica visual (EAV).

Figura 8 - Diagramas de caixa (box-plot) dos escores (mediana/amplitude) da escala Unesp-Botucatu de dor pós-operatória em ovinos (0-12), comparando os quatro momentos.



M1: pré-operatório; M2: pós-operatório, antes da analgesia de resgate; M3: pós-operatório, após analgesia de resgate; M4: 24h pós-operatório; MA - dados dos momentos agrupados (M1 + M2 + M3 + M4) (BRONDANI et al., 2013; DE OLIVEIRA et al., 2014).

As três diferentes raças de ovinos não tiveram efeito significativo na soma total da escala validada em nosso modelo estatístico.

A correlação item-total variou de 0,43 a 0,72 e desta forma todos os itens foram aceitos

(Tabela 10). O coeficiente α de Cronbach foi de 0,84 em MA, o que indica que o instrumento apresenta excelente consistência interna e reforça a possibilidade de utilização do escore total da escala para interpretar os resultados obtidos (Tabela 11).

Tabela 10 - Teste de correlação item-total da escala Unesp-Botucatu de dor pós-operatória em ovinos.

Itens/momentos	M1	M2	M3	M4	MA
Excluindo cada um dos itens					
Interação	0,30	0,52	0,58	0,29	0,72
Atividade	0,36	0,41	0,51	0,36	0,69
Locomoção	0,22	0,56	0,54	0,47	0,71
Posição da cabeça	0,17	0,53	0,31	0,11	0,54
Apetite	NA	0,47	0,26	0,05	0,43
Postura	0,46	0,42	0,22	0,22	0,53

Interpretação do coeficiente de correlação de postos de Spearman (r_s) - grau de correlação r_s : 0,3 - 0,7: valores aceitáveis em negrito (STREINER; NORMAN; CAIRNEY, 2015). M1: pré-operatório; M2: pós-operatório, antes da analgesia de resgate; M3: pós-operatório, após analgesia de resgate; M4: 24h pós-operatório; MA - dados dos momentos agrupados (M1 + M2 + M3 + M4). NA (não aplicável) = todos os escores foram "0."

Tabela 11 - Consistência interna dos escores de cada item da escala Unesp-Botucatu de dor pós-operatória em ovinos em cada momento.

Itens/momentos	M1	M2	M3	M4	MA
Todos os itens	0,6	0,78	0,64	0,52	0,84
Excluindo cada um dos itens					
Interação	0,5	0,73	0,56	0,46	0,79
Atividade	0,49	0,75	0,56	0,35	0,81
Locomoção	0,52	0,72	0,52	0,38	0,78
Posição da cabeça	0,62	0,74	0,62	0,50	0,82
Apetite	0,62	0,76	0,63	0,61	0,84
Postura	0,51	0,76	0,65	0,52	0,83

O coeficiente α de Cronbach foi calculado para o escore total e excluindo cada item da escala. Interpretação dos valores do coeficiente α : 0,60-0,64: minimamente aceitável; 0,65-0,69: aceitável; 0,70-0,74: boa; 0,75-0,80: muito boa; > 0,80: excelente (JENSEN, 2003; STREINER, 2003; STREINER; NORMAN; CAIRNEY, 2015). A negrito valores > 70. M1: pré-operatório; M2: pós-operatório, antes da analgesia de resgate; M3: pós-operatório, após analgesia de resgate; M4: 24h pós-operatório; MA - dados dos momentos agrupados (M1 + M2 + M3 + M4).

Todos os itens da escala apresentaram especificidade moderada a excelente. Os itens "interação", "atividade" e "locomoção" apresentaram sensibilidade moderada ou boa (Tabela 12).

Tabela 12 - Especificidade (em M1) e sensibilidade (em M2) de cada item da escala Unesp-Botucatu de dor pós-operatória em ovinos.

Itens	Especificidade (%)	Sensibilidade (%)
Interação	87	81
Atividade	83	85
Locomoção	87	83
Posição da cabeça	71	69
Apetite	100	50
Postura	79	67

Grau de E e S: excelente 95 - 100%; boa 85 - 94,9%; moderada 70 - 84,9% não é específico ou sensível < 70%. A negrito valores $\geq 70\%$ (BRONDANI et al., 2012).

Na análise da curva ROC para determinação do ponto de corte para diagnóstico da dor e recomendação de analgesia, o índice de Youden foi ≥ 3 para os avaliadores 1 e 3, de ≥ 4 para os avaliadores 2 (padrão-ouro) 4 e para todos avaliadores agrupados (escala de 0 – 12 pontos). A zona de incerteza diagnóstica foi de 2,5 a 3,5, sendo que a pontuação da escala para detectar ovinos verdadeiramente positivos (que realmente sofrem dor) foi $\geq 3,5$, enquanto que para detectar ovinos verdadeiramente negativos (sem dor) foi $\leq 2,5$. Os valores da área sob a curva foram altos (ASC > 0.95) para todos os avaliadores, indicando que o instrumento apresenta uma excelente capacidade discriminatória (Tabela 13; Figura 9).

A porcentagem de animais presentes na zona de incerteza diagnóstica foi baixa para todos os momentos de todos os avaliadores ($\leq 11\%$; 9 – 15%). Em M2 tal porcentagem de todos os avaliadores agrupados foi de 7% (0 a 13%), o que garante que se detectou com confiança que 93% dos ovinos apresentavam dor neste momento (Tabela 14). Em consonância com este dado, todos os avaliadores agrupados indicaram resgate analgésico em 93% dos animais em M2 (Tabela 15).

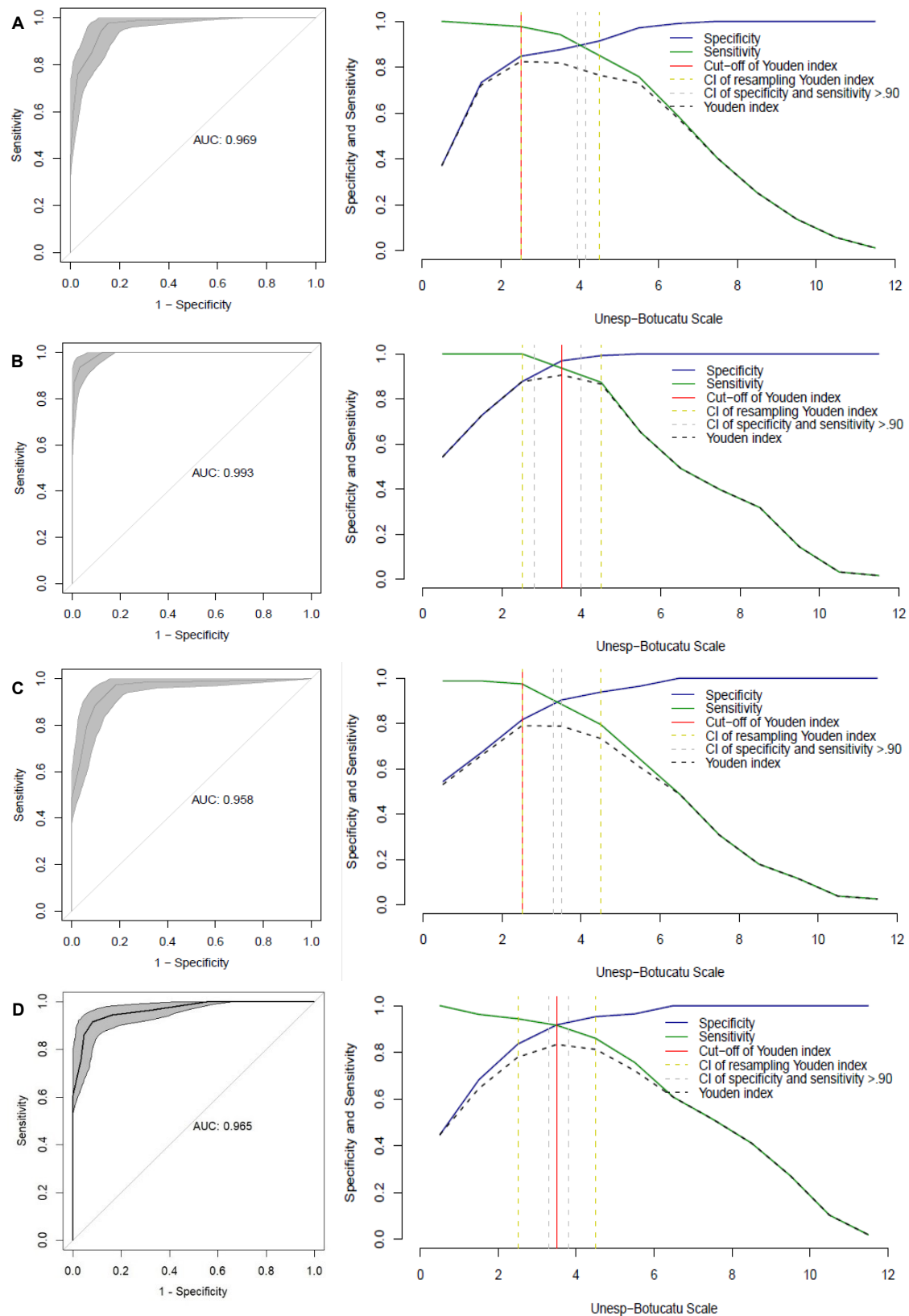
De acordo com a mediana, mínimo e máximo, as somatórias totais dividiram-se em três grupos (Figura 10) de intensidade de dor: leve (1; 0 - 3); moderada (7; 4 - 8) e intensa (10; 9 - 12) (Tabela 16; Figura 11). A partir desta classificação, em M2 observou-se que 10% das ovelhas tiveram intensidade baixa, 47% moderada e 43% intensa (Tabela 17).

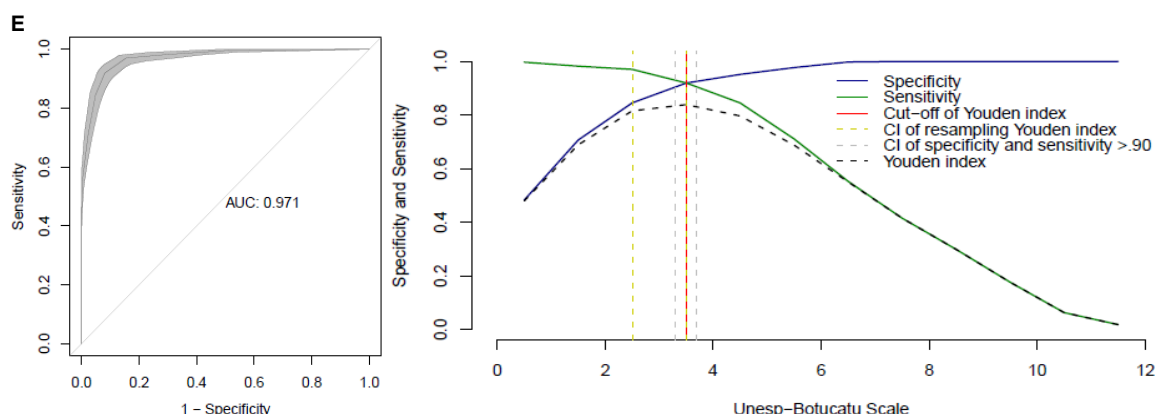
Tabela 13 - Especificidade, sensibilidade e índice de Youden de cada observador para cada escore da escala Unesp-Botucatu de dor aguda pós-operatória em ovinos (A); Intervalos de confiança (IC) de 95% de 1001 repetições e de sensibilidade e especificidade > 0,90 aplicados para estimar a zona de incerteza diagnóstica do ponto de corte de cada avaliador, de acordo com o índice de Youden (B).

A) Avaliadores	1			2			3			4			Todos		
Escore	E	S	IY	E	S	IY	E	S	IY	E	S	IY	E	S	IY
0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0
1	0,37	1	0,37	0,54	1	0,54	0,54	0,99	0,53	0,45	1	0,45	0,48	1	0,48
2	0,73	0,99	0,72	0,73	1	0,73	0,68	0,99	0,66	0,68	0,96	0,64	0,71	0,98	0,69
3	0,85	0,98	0,82	0,88	1	0,88	0,82	0,97	0,79	0,84	0,94	0,78	0,85	0,97	0,82
4	0,88	0,94	0,82	0,97	0,94	0,91	0,90	0,88	0,79	0,92	0,92	0,83	0,92	0,92	0,84
5	0,91	0,85	0,76	0,99	0,87	0,87	0,94	0,79	0,73	0,95	0,86	0,81	0,95	0,84	0,80
6	0,97	0,76	0,73	1	0,65	0,65	0,96	0,64	0,61	0,96	0,76	0,72	0,98	0,71	0,69
7	0,99	0,59	0,58	1	0,49	0,49	1	0,49	0,49	1	0,61	0,61	1	0,55	0,55
8	1	0,40	0,40	1	0,40	0,40	1	0,31	0,31	1	0,51	0,51	1	0,41	0,41
9	1	0,25	0,25	1	0,32	0,32	1	0,18	0,18	1	0,41	0,41	1	0,30	0,30
10	1	0,14	0,14	1	0,14	0,14	1	0,12	0,12	1	0,27	0,27	1	0,18	0,18
11	1	0,06	0,06	1	0,03	0,03	1	0,04	0,04	1	0,10	0,10	1	0,06	0,06
12	1	0,01	0,01	1	0,02	0,02	1	0,03	0,03	1	0,02	0,02	1	0,02	0,02
B)	Intervalo de confiança			Intervalo de confiança			Intervalo de confiança			Intervalo de confiança			Intervalo de confiança		
	Estimado	Mín.	Máx.	Estimado	Mín.	Máx.	Estimado	Mín.	Máx.	Estimado	Mín.	Máx.	Estimado	Mín.	Máx.
Replicação IC	2,5	2,5	4,5	3,5	2,5	4,5	3,5	2,5	4,5	3,5	2,5	4,5	3,5	2,5	3,5
IC > 0.90	-	3,95	4,15	-	2,8	4	-	3,3	3,5	-	3,3	3,8	-	3,3	3,7
ASC	0,969	0,95	0,99	0,993	0,99	1	0,958	0,93	0,98	0,965	0,94	0,99	0,971	0,96	0,98

E – Especificidade; **S** – Sensibilidade; **IY** – Índice de Youden. Valores a negrito: na linha do ponto de corte de cada avaliador (A) e dos valores estimados (B).

Figura 9 - Curva ROC e área sob a curva (ASC) [à esquerda] e curva ROC de dois gráficos com zona de incerteza diagnóstica para a escala Unesp-Botucatu de dor pós-operatória em ovinos [à direita].





A – Avaliador 1; **B** - Avaliador 2; **C** – Avaliador 3; **D** - Avaliador 4; **E** – Todos os avaliadores. Curva ROC (característica de operação do receptor) com intervalo de confiança de 95% (IC) (DEYO; DIEHR; PATRICK, 1991) calculada a partir de 1001 replicações e área sob a curva (ASC) [à esquerda]. Interpretação da ASC $\geq 0,95$ - alta capacidade discriminatória (STREINER; CAIRNEY, 2007). Curva ROC de dois gráficos, IC de 1001 replicações e sensibilidade e especificidade $> 0,90$ aplicados para estimar a zona de incerteza diagnóstica do ponto de corte de cada avaliador, de acordo com o índice de Youden [à direita] para a Escala Unesp-Botucatu de ovinos (CANNESSON et al., 2011; CELEITA-RODRÍGUEZ et al., 2019; GALL et al., 2015; MALLAT et al., 2015).

Considerando a validade de critério preditiva, fundamentado no índice de Youden, entre 83 e 98% das ovelhas (90% considerando todos os avaliadores agrupados) receberiam analgesia de resgate no momento de dor mais intensa (M2) (Tabela 14). A analgesia desnecessária seria indicada em 8 a 23% das ovelhas no M1, com base na pontuação atribuída por cada avaliador e 15% considerando todos os avaliadores agrupados. Esse resultado demonstrou por um lado que a escala foi sensível na distinção da dor, e por outro foi específica ao distinguir de ovelhas que não tinham dor. Não houve diferença entre a indicação de analgesia de resgate de acordo com a experiência clínica e o índice de Youden da escala de dor ($p = 0,74$ - teste exato de Fisher).

Tabela 14 - Porcentagem de ovelhas para as quais foi indicada analgesia de resgate de acordo com a experiência clínica e de acordo com o índice de Youden da escala Unesp-Botucatu de dor pós-operatória em ovinos.

Avaliador	1		2 (ouro)		3		4		Todos	
RA	Exp	IY	Exp	IY	Exp	IY	Exp	IY	Exp	IY
M1	13	13	8	8	23	15	19	25	16	15
M2	90	90	83	83	83	90	98	98	89	90

Cálculo realizado com base em 4 avaliadores x 48 ovinos = 192 ovinos. RA – Indicação de analgesia de resgate de acordo com a experiência clínica pontuada no final de cada análise de vídeo (Exp) e de acordo com o índice de Youden da USAPS (pontuação ≥ 4). Ouro - observador padrão ouro; M1: pré-operatório; M2: pós-operatório, antes da analgesia de resgate (valor a negrito em destaque para análise). O índice de Youden ≥ 4 é representativo do ponto de corte para indicação de analgesia de resgate (consulte a Tabela 13 para obter resultados do índice de Youden).

A porcentagem de animais presentes na zona de incerteza diagnóstica era baixa em todos os momentos de todos os avaliadores (11%; 9 - 15%). No M2, esse percentual para

todos os avaliadores agrupados foi de 7% (0 a 13%), o que garante que 93% das ovelhas com dor foram detectadas com confiança neste momento (Tabela 15).

Tabela 15. Porcentagem de ovinos presentes na zona de incerteza diagnóstica de acordo com o índice de Youden da escala Unesp-Botucatu de dor pós-operatória em ovinos.

Avaliador	1	2 (ouro)	3	4	Total
M1	6	8	23	13	13
M2	13	10	4	0	7
M3	8	21	17	13	15
M4	10	8	15	15	12
MA	9	12	15	10	11

Cálculo realizado com base em 4 avaliadores x 48 ovinos = 192 ovinos. M1: pré-operatório; M2: pós-operatório, antes da analgesia de resgate (valor a negrito em destaque para análise; M3: pós-operatório, após analgesia de resgate; M4: 24h pós-operatório; MA - dados dos momentos agrupados (M1 + M2 + M3 + M4). Os escores da zona de incerteza diagnóstica variaram de 2,5 a 3,5; portanto, $\leq 2,5$ indica dor verdadeiramente negativa (ovelha sem dor) e $\geq 3,5$ indica dor verdadeiramente positiva (ovelha que sofre de dor).

Tabela 16 - Mediana, mínimo e máximo de cada intensidade da dor da escala Unesp-Botucatu de dor pós-operatória em ovinos.

Escala Unesp-Botucatu	Mediana	Mínimo	Máximo
Leve	1	0	3
Moderada	7	4	8
Intensa	10	9	12

Figura 10 - Dendrograma criado pela análise de agrupamento não hierárquica de acordo com a somatória total da Escala da Unesp-Botucatu, com formação de três grupos que foram nomeados (leve, moderada, severa) em função da intensidade da dor.



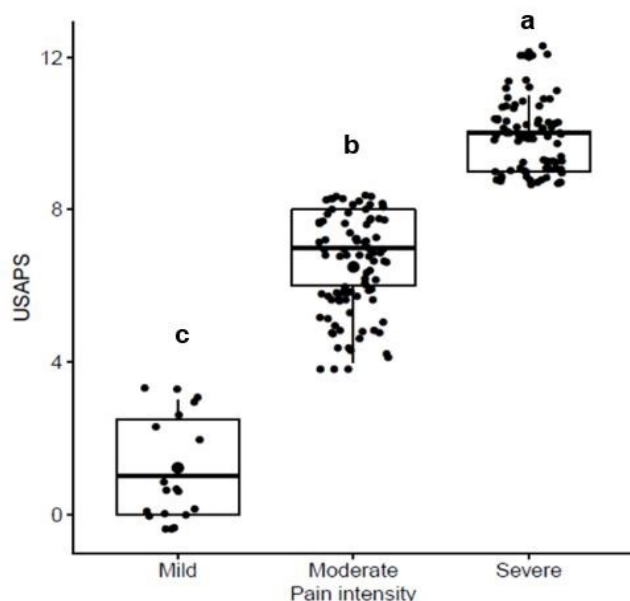
Tabela 17. Frequência e porcentagem de animais em cada intensidade da dor da escala Unesp-Botucatu de dor pós-operatória em ovinos.

Intensidade da dor	Severa		Moderada		Baixa		Total	
Avaliador	Nº	%	Nº	%	Nº	%	Nº	%
1	18	38	25	52	5	10	48	100
2	19	40	21	44	8	16	48	100

3	12	25	31	70	5	10	48	100
4	33	69	14	29	1	2	48	100
Total	82	43	91	47	19	10	192	100

Nº frequência; % porcentagem. Agrupamento com base nos escores da escala da Unesp-Botucatu pontuados no momento de maior dor (M2) da escala comportamental por todos os observadores na 2ª fase de avaliação.

Figura 11 - Diagramas de caixa (box-plot) para as três intensidades de dor (*mild*, *moderate* e *severe*) da escala Unesp-Botucatu de dor pós-operatória em ovinos.

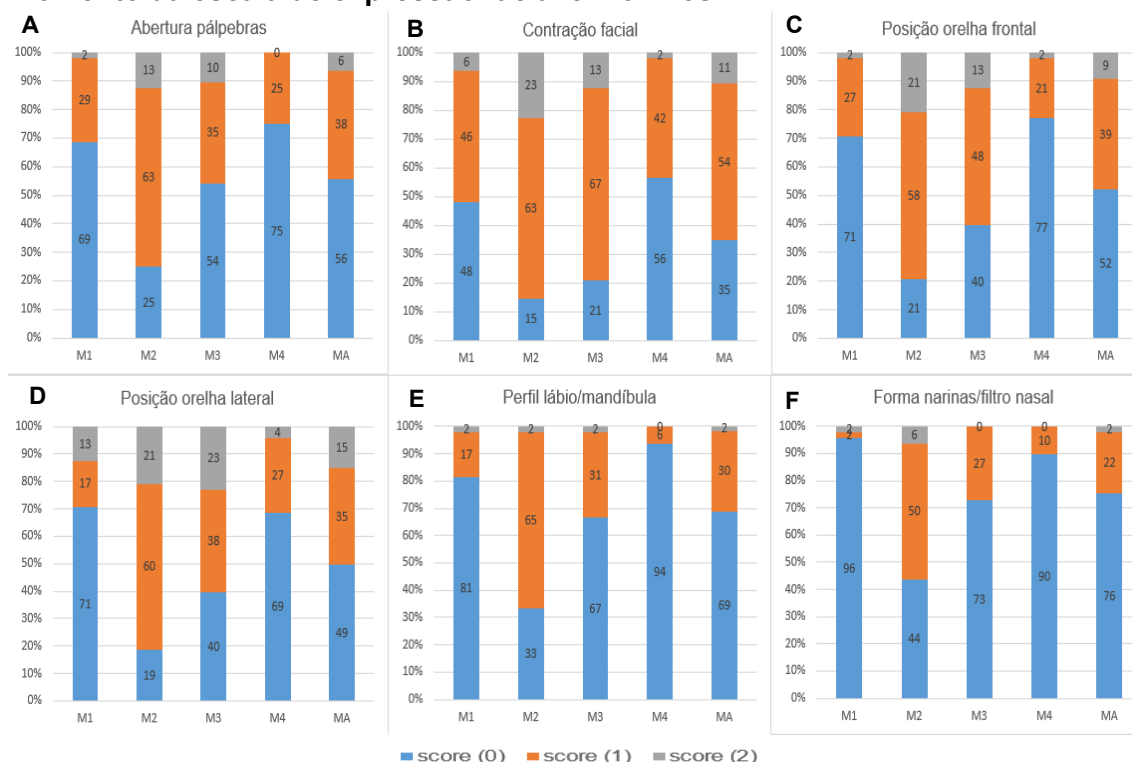


De acordo com a mediana, mínimo e máximo obtidos pela análise de agrupamento não hierárquica seguida por teste de Kruskal-Wallis, os escores dividiram-se em 3 grupos: mild - intensidade leve (1; 0 - 3); moderada - dor moderada (7; 4 - 8) e high - dor de intensidade alta (10; 9 - 12). Letras diferentes indicam diferença estatística pelo teste de Kruskal-Wallis sendo que $a > b > c$.

5.3 Escala de expressão facial

Os escores dos seis itens distribuíram-se da seguinte forma: o “0” predominou em todos os itens de M1 e M4; o “1” foi o mais frequente em todos os itens de M2; de M2 para M3 o escore “0” aumentou em todos os itens; em M3 houve três itens em que o escore “1” foi mais frequente que o “0” (Figura 12).

Figura 12 - Distribuição da porcentagem da presença dos escores de cada item em cada momento da escala de expressão facial em ovinos.



M1: pré-operatório; M2: pós-operatório, antes da analgesia de resgate (valor a negrito em destaque para análise; M3: pós-operatório, após analgesia de resgate; M4: 24h pós-operatório; MA - dados dos momentos agrupados (M1 + M2 + M3 + M4).

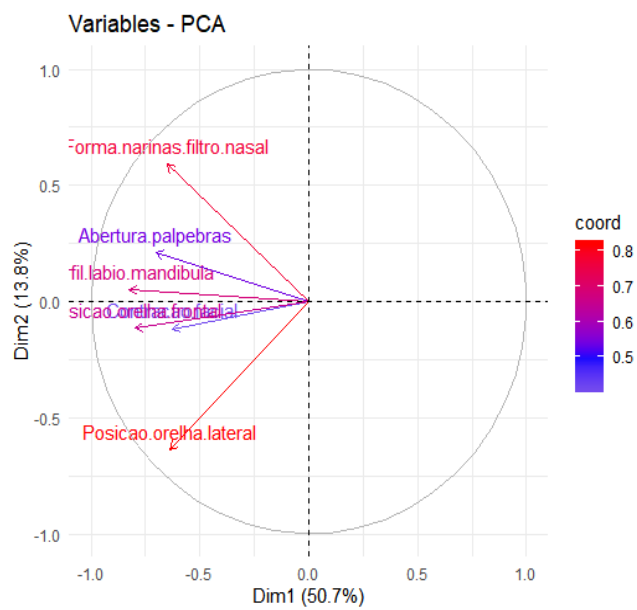
A análise de componentes principais gerou um fator com valor de 3,05. Para esta dimensão todos os itens apresentaram valores de carga aceitáveis. Definiu-se a escala como unidimensional (Tabela 18; Figura 13).

Tabela 18. Associação múltipla dos itens da escala entre si após análise de componentes principais. Valores de carga fatorial, autovalor e variância de acordo com critério de Kaiser para todos os momentos da escala de expressão facial de ovinos.

Itens	Dimensão 1	Dimensão 2
	Valor de carga	Valor de carga
Abertura das pálpebras	-0,70	0,21
Contração facial	-0,63	-0,12
Posição da orelha frontal	-0,80	-0,12
Posição da orelha lateral	-0,64	-0,64
Perfil do lábio e mandíbula	-0,82	0,05
Forma das narinas e filtro nasal	-0,65	0,59
Autovalor	3,05	0,83
Variância	50,67	13,85

Análise de componentes principais - A estrutura fatorial foi determinada considerando itens com carga fatorial $\geq 0,50$ ou $\leq -0,50$, com autovalor > 1 e variância $> 20\%$. A carga fatorial representa a correlação entre os itens e os fatores. Os valores de carga em negrito indicam as variáveis que contribuem para cada fator e o respectivo autovalor e variância aceites (KAISER, 1958).

Figura 13 - Análise de componentes principais com os itens da escala facial em ovinos aplicada pelo observador padrão-ouro nos quatro momentos de avaliação.



PCA – Análise de components principais

Tabela 19 - Confiabilidade intraobservador em todos os momentos das duas fases de avaliação para: A) cada item da escala de expressão facial em ovinos (grau de confiabilidade k_w) e B) a somatória da escala facial (consistência ICC).

A Grau de confiabilidade k_w	Avaliador 1			Avaliador 2			Avaliador 3			Avaliador 4		
Itens da escala	k_w	min.	máx.	k_w	min.	máx.	k_w	min.	máx.	k_w	min.	máx.
Abertura das pálpebras	0,49	0,37	0,61	0,47	0,34	0,60	0,40	0,27	0,54	0,47	0,36	0,59
Contração facial	0,67	0,58	0,76	0,53	0,41	0,64	0,44	0,34	0,53	0,59	0,44	0,66
Posição da orelha frontal	0,72	0,62	0,83	0,53	0,42	0,65	0,54	0,44	0,64	0,55	0,44	0,66
Posição da orelha lateral	0,72	0,63	0,81	0,61	0,49	0,72	0,50	0,41	0,60	0,65	0,54	0,76
Perfil do lábio e mandíbula	0,25	0,13	0,37	0,47	0,34	0,59	0,33	0,21	0,45	0,47	0,38	0,57
Forma das narinas e filtro nasal	0,42	0,30	0,53	0,40	0,27	0,53	0,44	0,33	0,55	0,34	0,20	0,47

B	Avaliador 1			Avaliador 2			Avaliador 3			Avaliador 4		
Coeficiente correlação intraclass	ICC	Intervalo de confiança		ICC	Intervalo de confiança		ICC	Intervalo de confiança		ICC	Intervalo de confiança	
Escala de expressão facial	0,63	0,54	0,71	0,74	0,67	0,80	0,54	0,44	0,64	0,72	0,64	0,78

Grado de confiabilidade k_w ou ICC (consistência): muito boa: 0,81 - 1,0; boa: 0,61 - 0,80; moderada: 0,41 - 0,60; razoável: 0,21 - 0,4; pobre < 0,2. (ALTMAN, 1991). A negrito os valores aceitáveis > 50.

Tabela 20 - Confiabilidade interobservador entre os quatro observadores para: A) cada item da escala de expressão facial em ovinos (grau de confiabilidade k_w) e B) a somatória da escala facial em ovinos (consistência ICC).

A.	Avaliador 2 (padrão-ouro) vs								
Grau de confiabilidade k_w	Avaliador 1			Avaliador 3			Avaliador 4		
Itens da escala	k_w	min.	máx.	k_w	min.	máx.	k_w	min.	máx.
Abertura das pálpebras	0,51	0,41	0,62	0,51	0,39	0,62	0,38	0,26	0,51
Contração facial	0,40	0,32	0,49	0,43	0,32	0,54	0,46	0,35	0,56
Posição da orelha frontal	0,21	0,14	0,28	0,53	0,43	0,63	0,38	0,28	0,48
Posição da orelha lateral	0,57	0,47	0,67	0,42	0,31	0,53	0,58	0,46	0,70
Perfil do lábio e mandíbula	0,07	0,02	0,13	0,16	0,09	0,24	0,26	0,14	0,38
Forma das narinas e filtro nasal	- 0,01	-0,11	0,08	0,12	0,02	0,23	0,11	-0,01	0,23

B	Avaliador 2 (padrão-ouro) vs								
	Avaliador 1			Avaliador 3			Avaliador 4		
Coeficiente correlação intraclass	ICC	Intervalo de confiança		ICC	Intervalo de confiança		ICC	Intervalo de confiança	
Escala de expressão facial	0,03	-0,11	0,17	0,39	0,26	0,50	0,60	0,51	0,69

Grau de confiabilidade k_w ou ICC (consistência): muito boa: 0,81 - 1,0; boa: 0,61 - 0,80; moderada: 0,41 - 0,60; razoável: 0,21 - 0,4; pobre < 0,2 (ALTMAN, 1991).
A negrito os valores aceitáveis > 50.

A confiabilidade intraobservador foi boa para três avaliadores e moderada para um (Tabela 19) e a interobservador variou de pobre a moderada (Tabela 20).

A correlação entre os escores da escala facial e das escalas EN ($r=0,65$), EDS ($r=0,64$), EAV ($r=0,67$) e escala comportamental ($r=0,59$) foi média para todos os momentos avaliados pelo observador padrão-ouro (Tabela 21).

Tabela 21 - Correlação entre a escala de expressão facial em ovinos e as escalas numérica, descritiva simples, analógica visual e a de dor aguda proposta.

Escalas	Escala de expressão facial				
Momentos	M1	M2	M3	M4	MA
Escala numérica	0,27	0,41	0,55	0,23	0,65
Escala descritiva simples	0,27	0,55	0,54	0,23	0,64
Escala analógica visual	0,33	0,47	0,57	0,34	0,67
Escala Unesp-Botucatu	-0,03	0,32	0,53	0,02	0,59

$p = 0,000$. Interpretação do coeficiente de correlação de Spearman: 0 – 0,35: correlação baixa; 0,35 – 0,7: correlação moderada; 0,7 – 1,0: correlação alta (BRONDANI et al., 2013). A negrito valores > 70 . M1: pré-operatório; M2: pós-operatório, antes da analgesia de resgate; M3: pós-operatório, após analgesia de resgate; M4: 24h pós-operatório; MA - dados dos momentos agrupados (M1 + M2 + M3 + M4).

Os escores de todos os itens (com exceção de um) e do total da escala em M2 foram significativamente maiores do que em M1 e M4 evidenciando a responsividade ou validade de construto da escala. De M2 para M3 o escore total da escala e de dois itens diminuíram significativamente (tabela 22).

Tabela 22 - Análise da validade de construto pela comparação dos escores entre os quatro momentos para cada item e do total da escala de expressão facial em ovinos (0-12).

Momentos	M1		M2		M3		M4	
Itens da escala	Mediana	Amplit.	Mediana	Amplit.	Mediana	Amplit.	Mediana	Amplit.
Abertura pálpebras	0 ^{bc}	0-2	1 ^a	0-2	0 ^a	0-2	0 ^b	0-1
Contração facial	1 ^{ab}	0-2	1 ^c	0-2	1 ^{ac}	0-2	0 ^b	0-2
Posição orelha frontal	0 ^b	0-2	1 ^a	0-2	1 ^a	0-2	0 ^b	0-2
Posição orelha lateral	0 ^b	0-2	1 ^a	0-2	1 ^a	0-2	0 ^b	0-2
Perfil lábio mandíbula	0 ^{bc}	0-2	1 ^a	0-2	0 ^c	0-2	0 ^b	0-1
Forma narinas filtro nasal	0 ^b	0-2	1 ^a	0-2	0 ^c	0-1	0 ^{bc}	0-1
Escala expressão facial	1 ^c	0-7	5 ^a	0-9	3,5 ^b	0-11	1 ^c	0-7

Letras diferentes expressam diferenças significativas (valores a negrito) entre momentos pelo Teste de Friedman ($p<0,05$) sendo que $a>b>c$ (BRONDANI et al., 2013; DE OLIVEIRA et al., 2014). M1: pré-operatório; M2: pós-operatório, antes da analgesia de resgate; M3: pós-operatório, após analgesia de resgate; M4: 24h pós-operatório; MA - dados dos momentos agrupados (M1 + M2 + M3 + M4).

A correlação item-total variou de 0,46 a 0,69, desta forma todos os itens foram aceitos (Tabela 23). O coeficiente α de Cronbach da escala foi de 0,79 em MA, o que indica que o instrumento apresenta muito boa consistência interna (Tabela 24).

Tabela 23. Teste de correlação item-total da escala de expressão facial.

Itens/momentos	M0	M1	M2	M3	MA
Excluindo cada um dos itens					
Abertura das pálpebras	0,27	0,52	0,43	0,23	0,51
Contração facial	0,33	0,27	0,45	0,08	0,46
Posição da orelha frontal	0,45	0,43	0,77	0,45	0,69
Posição da orelha lateral	0,36	0,26	0,49	0,37	0,52
Perfil do lábio e mandíbula	0,59	0,67	0,61	0,13	0,68
Forma das narinas e filtro nasal	0,19	0,4	0,37	0,15	0,47

Interpretação do coeficiente de correlação de postos de Spearman (r_s) - grau de correlação r_s : 0,3 - 0,7: valores aceitáveis a negrito (STREINER; NORMAN; CAIRNEY, 2015). M1: pré-operatório; M2: pós-operatório, antes da analgesia de resgate; M3: pós-operatório, após analgesia de resgate; M4: 24h pós-operatório; MA - dados dos momentos agrupados (M1 + M2 + M3 + M4).

Tabela 24 - Consistência interna dos escores de cada item da escala de expressão facial em ovinos em cada momento.

Itens/momentos	M0	M1	M2	M3	MA
Todos os itens	0,65	0,67	0,77	0,53	0,79
Excluindo cada um dos itens					
Abertura das pálpebras	0,63	0,59	0,75	0,48	0,76
Contração facial	0,60	0,67	0,74	0,58	0,78
Posição da orelha frontal	0,60	0,63	0,66	0,32	0,73
Posição da orelha lateral	0,62	0,69	0,76	0,44	0,78
Perfil do lábio e mandíbula	0,52	0,56	0,70	0,51	0,73
Forma das narinas e filtro nasal	0,67	0,63	0,76	0,51	0,78

O coeficiente α de Cronbach foi calculado para o escore total e excluindo cada item da escala. Interpretação dos valores do coeficiente α : 0,60-0,64: minimamente aceitável; 0,65-0,69: aceitável; 0,70-0,74: boa; 0,75-0,80: muito boa; > 0,80: excelente (JENSEN, 2003; STREINER, 2003; STREINER; NORMAN; CAIRNEY, 2015). A negrito valores > 70. M1: pré-operatório; M2: pós-operatório, antes da analgesia de resgate; M3: pós-operatório, após analgesia de resgate; M4: 24h pós-operatório; MA - dados dos momentos agrupados (M1 + M2 + M3 + M4).

Os itens “posição da orelha frontal” e “posição da orelha lateral” apresentaram simultaneamente especificidade e sensibilidade. “Abertura das pálpebras” e “contração facial” apresentaram apenas sensibilidade e “perfil do lábio e mandíbula” e “forma das narinas e filtro nasal” apresentaram apenas especificidade (Tabela 22).

Tabela 25 - Especificidade (em M1) e da sensibilidade (em M2) de cada item da escala de expressão facial em ovinos.

Itens	Especificidade (%)	Sensibilidade (%)
Abertura das pálpebras	69	75
Contração facial	48	85
Posição da orelha frontal	71	79
Posição da orelha lateral	71	81
Perfil do lábio e mandíbula	81	67
Forma das narinas e filtro nasal	99	56

Grau de E e S: excelente 95 - 100%; boa 85 - 94,9%; moderada 70 - 84,9% não é específico ou sensível < 70%. A negrito valores $\geq 70\%$ (BRONDANI et al., 2012).

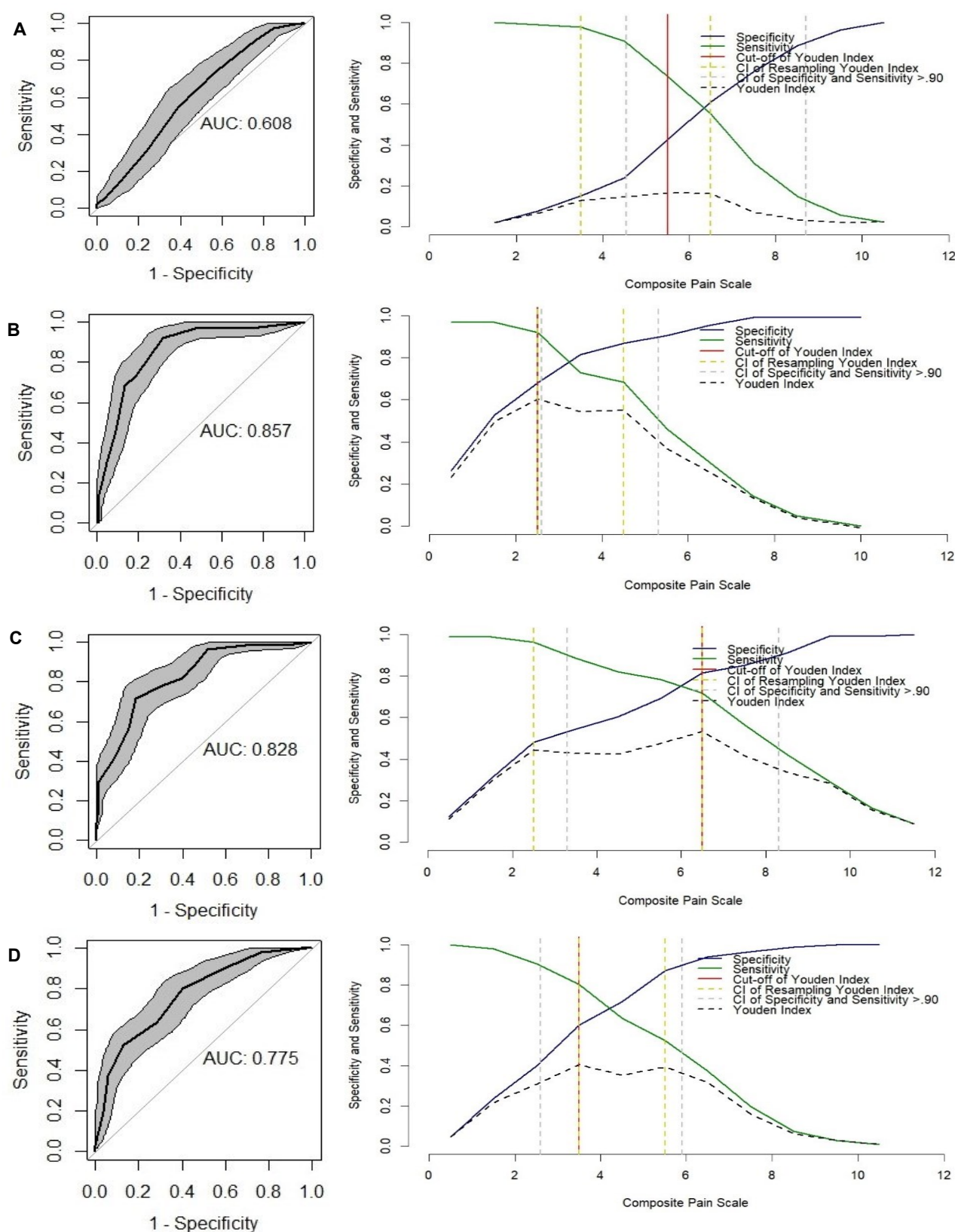
O ponto de corte para diagnóstico da dor e de recomendação para analgesia determinada pelo índice de Youden foi de ≥ 6 , ≥ 3 (padrão-ouro), ≥ 7 e ≥ 4 para os quatro avaliadores respectivamente (escala de 0 – 12 pontos). A zona de incerteza diagnóstica foi de 2,5 a 6,5, sendo que a pontuação da escala para detectar ovinos verdadeiramente positivos (que realmente sofrem dor) foi $\geq 6,5$, enquanto que para detectar ovinos verdadeiramente negativos (sem dor) foi $\leq 2,5$. Desta forma o ponto de resgate analgésico ficou definido como ≥ 7 . Observando a área sob a curva (ASC $> 0,95$), apenas um avaliador apresentou valores próximos de uma habilidade discriminatória excelente (Tabela 26; Figura 14).

Tabela 26 - Especificidade, sensibilidade e índice de Youden de cada observador para cada escore da escala de expressão facial em ovinos (A); Intervalos de confiança (IC) de 95% de 1001 repetições e de sensibilidade e especificidade > 0,90 aplicados para estimar a zona de incerteza diagnóstica do ponto de corte de cada avaliador, de acordo com o índice de Youden (B).

A	Avaliador 1			Avaliador 2			Avaliador 3			Avaliador 4		
Escore	E	S	IY	E	S	IY	E	S	IY	E	S	IY
0	0	1	0	0	1	0	0	1	0	0	1	0
1	0	1	0	0,26	0,97	0,23	0,12	0,99	0,11	0,05	1	0,05
2	0,02	1	0,02	0,53	0,97	0,50	0,31	0,99	0,29	0,24	0,98	0,22
3	0,08	0,99	0,06	0,68	0,92	0,60	0,48	0,96	0,44	0,40	0,91	0,31
4	0,15	0,98	0,13	0,81	0,73	0,54	0,54	0,88	0,43	0,60	0,80	0,40
5	0,24	0,91	0,15	0,87	0,68	0,55	0,61	0,82	0,43	0,72	0,64	0,35
6	0,43	0,74	0,16	0,91	0,46	0,37	0,69	0,78	0,48	0,87	0,52	0,39
7	0,61	0,55	0,16	0,95	0,30	0,26	0,82	0,72	0,53	0,94	0,37	0,32
8	0,76	0,31	0,07	0,99	0,14	0,14	0,85	0,56	0,41	0,96	0,20	0,16
9	0,89	0,15	0,04	0,99	0,05	0,04	0,91	0,42	0,34	0,99	0,07	0,06
10	0,96	0,06	0,02	0,99	0	-0,01	0,99	0,29	0,29	1	0,03	0,03
11	1	0,02	0,02	0,99	0	-0,01	0,99	0,17	0,16	1	0,01	0,01
12	1	0	0	1	0	0	1	0,09	0	1	0	0
B	Intervalo de confiança			Intervalo de confiança			Intervalo de confiança			Intervalo de confiança		
	Estimado	Mínimo	Máximo	Estimado	Mínimo	Máximo	Estimado	Mínimo	Máximo	Estimado	Mínimo	Máximo
Replicação IC	5,5	3,5	6,5	2,5	2,5	4,5	6,5	2,5	6,5	3,5	3,5	5,5
IC > 0.90	-	4,55	8,65	-	2,6	5,3	-	3,3	8,3	-	2,6	5,9
ASC	0,608	0,530	0,687	0,857	0,801	0,912	0,829	0,771	0,886	0,776	0,711	0,840
Ponte de corte	5,5	-	-	2,5	-	-	6,5	-	-	3,5	-	-

E – Especificidade; S – Sensibilidade; IY – Índice de Youden. Valores a negrito: na linha do ponto de corte de cada avaliador (A) e dos valores estimados (B).

Figura 14 - Curva ROC e área sob a curva (ASC) [à esquerda] e curva ROC de dois gráficos com zona de incerteza diagnóstica para a escala de expressão facial em ovinos [à direita].



A – Avaliador 1; **B** - Avaliador 2; **C** – Avaliador 3; **D** - Avaliador 4. Curva ROC (característica de operação do receptor) com intervalo de confiança de 95% (IC) (DEYO; DIEHR; PATRICK, 1991) calculada a partir de 1001 replicações e área sob a curva (ASC) [à esquerda]. Interpretação da ASC $\geq 0,95$ - alta acurácia do ponto de resgate (STREINER; CAIRNEY, 2007). Interpretação da ASC $\geq 0,95$ - alta acurácia do ponto de resgate (STREINER; CAIRNEY, 2007). Curva ROC de dois gráficos, IC de 1001 replicações e sensibilidade e

especificidade > 0,90 aplicados para estimar a zona de incerteza diagnóstica do ponto de corte de cada avaliador, de acordo com o índice de Youden [à direita] para a escala de expressão facial em ovinos (CANNESSON et al., 2011; GALL et al., 2015; MALLAT et al., 2015).

A porcentagem de animais que estiveram na zona de incerteza diagnóstica foi alta para todos os momentos de todos os avaliadores, sobretudo em M2, resultado que, ao contrário da escala de dor comportamental, não garante que se detecte se os ovinos apresentem ou não dor com rigor, já que 85% se encontraram na zona de incerteza (Tabela 27).

Tabela 27 - Frequência e porcentagem de animais presentes na zona de incerteza diagnóstica de todos os avaliadores da escala de expressão facial em ovinos.

Momentos	M1	M2	M3	M4	MA
Nº de animais	113/192	163/192	143/192	108/192	527/768
Porcentagem	59	85	74	56	69

Cálculo realizado com base em 4 avaliadores x 48 ovinos = 192 ovinos. M1: pré-operatório; M2: pós-operatório antes da analgesia de resgate (valor a negrito em destaque para análise); M3: pós-operatório, após analgesia de resgate; M4: 24h pós-operatório; MA - dados dos momentos agrupados (M1 + M2 + M3 + M4).

6. DISCUSSÃO

A criação de ferramentas espécie-específicas válidas para avaliar a dor é pré-requisito para reconhecer o fenômeno e determinar a necessidade e eficácia do tratamento antálgico. Neste prisma pode-se considerar a escala comportamental de dor aqui proposta, como um instrumento confiável, válido e com ponto de intervenção analgésica definido, que pode ser utilizado para avaliar a dor pós-operatória em ovinos. Este instrumento apresenta aplicabilidade clínica, pois pode garantir que os ovinos se beneficiem de analgesia quando necessário, e aplicabilidade experimental, por exemplo para executar estudos que comparem a eficácia analgésica de fármacos, bem como estudos translacionais (FLECKNELL, 2008).

O processo de validação de um instrumento que avalie a dor se embasa na investigação de comportamentos e, quando possível de dados fisiológicos específicos da espécie presentes durante situações de dor, seguido de cotejamento de tais alterações com o estado de normalidade (FLECKNELL, 2008). Tal metodologia foi usada neste estudo no qual se construiu um etograma que abrangeu o período pré-operatório, em que os animais estavam supostamente desprovidos de dor, seguido do pós-operatório em que os animais provavelmente apresentavam dor intensa, seguida de analgesia de resgate para redução da dor e avaliação após 24 horas. Desta forma o delineamento experimental apresentou o intuito de se testar o instrumento em intensidades diferentes de dor. O etograma, juntamente com os comportamentos de expressão de dor em ovinos descritos na literatura, serviram de base para a construção da escala e a seguir, após a análise de conteúdo pelo comitê de pesquisadores experientes na área, definiu-se o primeiro instrumento incluindo-se os comportamentos relevantes e excluindo-se os irrelevantes, para tornar o instrumento o mais simples e representativo possível.

As filmagens por videocâmeras agregam valor aos dados pois permitem arquivar o material para futuras pesquisas e minimizam a influência do observador na avaliação, o que evita possíveis alterações comportamentais que o animal possa apresentar inerentes à presença do observador. Por outro lado ressalta-se que diante de uma situação clínica, muitas vezes não é possível se fazer uma avaliação remota, e a presença do observador poderia interferir no comportamento dos animais (LEACH et al., 2009; LESTER; FANSELOW, 1985; PINHO RH et al., 2019; SORGE et al., 2014). Desta forma ainda se necessita validar a escala proposta em situações clínicas e com a presença do observador, para garantir que estes resultados sejam reproduzíveis.

De acordo com o etograma, ovinos com dor se alimentaram menos e consequentemente ruminaram menos em relação aos momentos após o resgate analgésico, o que justificou a introdução do apetite como um dos critérios a serem avaliados na escala. Como, no momento pré-operatório, os animais estavam sob jejum e não havia alimento disponível, obviamente este comportamento não estava presente. Estudos anteriores também demonstraram a diminuição do apetite em ovinos submetidos à castração e laparoscopia (24,57,88), entretanto, ovinos com dor moderada, podem apresentar apetite e ruminação (66, 89). A melhora do apetite após a terapia

antálgica fortalece a necessidade de analgesia pós-operatória em ovinos (TEIXEIRA et al., 2011a), tal como sucedeu em bovinos (DE OLIVEIRA et al., 2014).

Como reportado pela literatura (ADAMS, 2017; MANTECA et al., 2017; MOLONY; KENT, 1997; MOLONY; KENT; ROBERTSON, 1993; PAULL et al., 2008) neste estudo também observou-se, nos ovinos com dor, redução da interação com o ambiente e da atividade, cabeça baixa e arqueamento dorsal, o que confirma que a laparoscopia promove alterações comportamentais indicativas de dor aguda (TEIXEIRA et al., 2011b). Alguns destes comportamentos só retornaram ao normal após 24 horas da cirurgia.

Os achados do etograma se assemelharam aos da literatura. Cordeiros submetidos ao procedimento de *mulesing* ficaram curvados e andaram menos no pós-operatório (PAULL et al., 2008) e após a orquiectomia se observaram diferentes comportamentos e indicadores fisiológicos de dor pós-operatória (DINNISS et al., 1999; MELLOR et al., 2002). A orquiectomia e o corte de cauda desencadeiam comportamentos relacionados à diversas intensidades de dor. Na dor intensa ocorre movimentos de membros, cauda e cabeça e posturas com extensão total dos membros pélvicos; na dor moderada observa-se vocalização, posições parado, sentado e deitado com extensão parcial dos membros pélvicos ou com tremor; na dor leve ou sem dor as posturas podem ser normais (MOLONY; KENT; MCKENDRICK, 2002).

A vocalização poderia ser um possível indicador de dor como reportado em outros ruminantes, como os bovinos (CURRAH; HENDRICK; STOOKEY, 2009; DE OLIVEIRA et al., 2014; MILLMAN, 2013; WATTS; STOOKEY, 2000) e caprinos (MILLMAN, 2013), entretanto, em ovinos este comportamento está mais relacionado ao isolamento social, contenção e no transoperatório. No etograma deste estudo não se incluiu a vocalização pois ela não foi observada, exceto nos momentos de oferta de alimento. Assim a vocalização não é, na espécie ovina, um indicador de dor pós-operatória, sobretudo em animais adultos e mesmo após estímulos dolorosos intensos (FERGUSON; LEE; FISHER, 2017; MOLONY; KENT; MCKENDRICK, 2002).

Um diferencial deste estudo que apresenta ônus e bônus, é que a maioria dos estudos que avaliam a dor aguda em ovinos foi em cordeiros (ADAMS, 2017; FUTRO; MASŁOWSKA; DWYER, 2015; GRANT, 2004; GUESGEN; BEAUSOLEIL; STEWART, 2013; MANTECA et al., 2017; MOLONY; KENT, 1997; MOLONY; KENT; MCKENDRICK, 2002; MOLONY; KENT; ROBERTSON, 1993; PAULL et al., 2008), o que pode limitar a extrapolação dos resultados aos animais adultos. Alguns comportamentos mais específicos de cordeiros citados na literatura, como “saltar como coelho” não ocorreram, pois são mais frequentes em animais jovens, até cerca de 5 meses de idade, do que em animais adultos. Um comportamento comum em bovinos (DE OLIVEIRA et al., 2014), e citado em cordeiros (27,47,93), que é o de olhar o flanco e lamber a região dolorida, não foi observado nas ovelhas adultas desta pesquisa.

No que concerne à validação da escala, submeteu-se o instrumento a uma metodologia encoberta e aleatória, por meio do mesmo delineamento experimental já abordado em gatos (BRONDANI et al., 2013), bovinos (DE OLIVEIRA et al., 2014) equinos (TAFFAREL et al., 2015) e

suínos (LUNA et al., 2015), e que apresenta reconhecida robustez científica (MEROLA; MILLS, 2016).

A validade e a confiabilidade são os atributos essenciais para um instrumento identificar e quantificar a dor dos animais. Observadores familiarizados com o comportamento da espécie procederam à validação de conteúdo da escala, por meio da avaliação da representatividade de cada item. Esta análise mede o quanto o instrumento reflete o fenômeno de interesse, neste caso, a dor (BRONDANI et al., 2013; BRONDANI; LUNA; PADOVANI, 2011b; BUSSIERES et al., 2008; ISON et al., 2016; MURRELL et al., 2008). Os avaliadores foram selecionados pela sua experiência, o que melhora a confiabilidade, isto é, a precisão da ferramenta baseada na repetibilidade e reprodutibilidade (ROBERTSON; KENT; MOLONY, 1994).

A escala de dor aguda inicialmente proposta continha 33 variáveis incluindo os itens, subitens e suas subdivisões (tabela IV), que foram excluídos de acordo com os critérios dos testes estatísticos citados (Tabela 1). Este processo de refinamento da escala composta teve várias etapas e identificou as 12 variáveis mais relevantes e adequadas para medir a dor em ovinos. O refinamento da escala de dor por meio da análise dos seus itens foi essencial para a melhoria da qualidade da escala final validada, em versão mais simples e objetiva para uso (BOATENG; NEILANDS; FRONGILLO, 2018).

A análise de distribuição de escores ao proporcionar um panorama da ocorrência de cada escore em cada momento, indica a importância de cada escore. Os resultados foram dentro do esperado, já que o escore "0" relacionado à ausência de dor, predominou antes e após 24h da cirurgia, os escores "1" e "2" ocorreram mais no pós operatório e após o resgate analgésico e o escore "2" foi mais evidente sobretudo após a cirurgia sugerindo uma intensidade maior de dor. Os resultados de cada item acompanharam de forma geral os resultados da somatória da escala. Apenas o escore "1" do item atividade foi pouco evidente, o que demonstra que é raro ovinos se movimentarem acima do normal ou deitar e se levantar com frequência, como ocorre em outras espécies (BRONDANI et al., 2013; DE OLIVEIRA et al., 2014; LUNA ET AL., 2015; TAFFAREL et al., 2015).

A análise de componentes principais relaciona as variáveis da ferramenta de forma agrupada (GRACEY, 1992) e calcula o número de dimensões determinados por diferentes variáveis (BRONDANI et al., 2013) que estabelecem a extensão ou dimensionalidade da escala (JENSEN, 2003). Estas variáveis são relacionadas para que os itens que definem partes específicas do construto sejam agrupados por meio da associação múltipla (DEVON et al., 2007). De acordo com o critério de Kaiser selecionou-se um componente principal e considerou-se a escala como unidimensional (KAISER, 1958). As escalas desenvolvidas em bovinos (DE OLIVEIRA et al., 2014) e suínos (LUNA et al., 2015) são unidimensionais e a de felinos é multidimensional pois gerou três dimensões (BRONDANI et al., 2013; BRONDANI; LUNA; PADOVANI, 2011b). Um instrumento é multidimensional quando além da intensidade da dor, inclui características qualitativas e temporais, como dimensões sensoriais, emocionais e cognitivas, que apresentam alta correlação na

experiência da dor (GRACELY, 1992; NETEMEYER; BEARDEN; SHARMA, 2003).

Em uma validação de escala de dor aguda de cordeiros, a análise de componentes principais gerou dois componentes principais (MOLONY; KENT; MCKENDRICK, 2002). As escalas unidimensionais não são tão satisfatórias como as que apresentam mais de uma dimensão, pois avaliam apenas a intensidade da dor. No entanto, dada a simplicidade e praticidade são facilmente aplicáveis. Neste estudo é prematuro concluir sobre a dimensão da escala de dor proposta, uma vez que apenas se avaliou um tipo de dor, em tecidos moles, por meio de apenas um modelo estatístico. A escala proposta para ovinos contempla diversas vertentes biológicas da dor, tais como a fisiológica (apetite), sensorial (postura, atividade), emocional (interação com outros animais e atenção no ambiente) e temporal (resposta à analgesia). Estudos futuros abordando diferentes modelos de dor, como a ortopédica e com outras abordagens estatísticas poderão confirmar o número de dimensões desta escala.

Todo instrumento novo deve ser cotejado com outra ferramenta já consagrada e validada, considerada padrão-ouro (MEROLA; MILLS, 2016). Como este não é o caso em ovinos, pois não há uma escala validada com estatística robusta, para avaliar a dor pós-operatória, adotou-se a metodologia baseada na concordância entre os escores de dor atribuídos pelos avaliadores e um avaliador padrão-ouro, seguindo-se o mesmo critério aplicado em gatos (BRONDANI et al., 2013), equinos (TAFFAREL et al., 2015), bovinos (DE OLIVEIRA et al., 2014) e suínos (LUNA et al., 2015).

A confiabilidade intra e interobservador para cada item e para o escore total da escala de ovinos foi similar à de bovinos (DE OLIVEIRA et al., 2014) e suínos (LUNA et al., 2015). Tais resultados foram inferiores aos de gatos (BRONDANI et al., 2013) e superiores aos de equinos (TAFFAREL et al., 2015). Quando se compara com outros instrumentos desenvolvidos na espécie ovina, a escala aqui proposta apresentou uma confiabilidade similar à escala reportada em cordeiros submetidos à dor aguda (73% intraobservador e 79% interobservador) (MOLONY; KENT; MCKENDRICK, 2002). Relativamente a uma escala de locomoção de ovelhas (cujo escore variou de 0 a 6) que teve confiabilidades intra (91%) e interobservadores (93%) muito boas (KALER; WASSINK; GREEN, 2009), o instrumento proposto apresentou resultados inferiores, no qual o item “locomoção” teve confiabilidade intraobservador boa e a interobservador apenas moderada. Já outro estudo em que 10 veterinários e 10 criadores de ovelhas pontuaram uma escala de locomoção, obteve valores muito bons e bons de confiabilidade intraobservador e interobservadores, no entanto, a confiabilidade para os escores individuais de locomoção variou de razoável a moderada (ANGELL et al., 2015).

A validade indica que o instrumento mede com precisão o que foi avaliado, como demonstrou a escala proposta. Há vários tipos de validade. A validade de critério avalia a eficácia da medição de uma escala e um outro teste para analisar este atributo é o de validação concorrente que compara o instrumento às escalas já existentes e validadas (HOLTON et al., 2001). A validade de critério pode ser preditiva, que avalia o critério após o teste ou concorrente, que avalia o instrumento e o critério simultaneamente, que foi o caso desse estudo (PASQUALI, 2009). Como mencionado

anteriormente não há um instrumento validado para avaliar a dor pós-operatória em ovinos que possa servir de comparação, e tal como em bovinos (DE OLIVEIRA et al., 2014), equinos (TAFFAREL et al., 2015) e suínos (LUNA et al., 2015), comparou-se o instrumento proposto com a EN, EDS, EAV, quando se observou uma alta correlação. Na espécie ovina a correlação de uma escala de claudicação com a EN e EAV foi boa (WELSH; GETTINBY; NOLAN, 1993). Ao se comparar a escala proposta com a escala facial em ovinos, que embora não totalmente validada, foi considerada para comparação, a correlação foi moderada.

A validade de construto ou responsividade reflete a capacidade de resposta da escala e examina se o instrumento detecta diferenças previsíveis entre grupos, ou momentos como nesse estudo (CRELLIN et al., 2007). O método testa a hipótese de que o tempo e a intervenção cirúrgica e analgésica devem alterar os escores de dor (BRONDANI et al., 2013) e tem sido usado para validar escalas em medicina veterinária (BRONDANI et al., 2012, 2013; DE OLIVEIRA et al., 2014; LUNA et al., 2015; TAFFAREL et al., 2015). As diferenças observadas entre os escores de dor entre os momentos, e especialmente no momento esperado de maior dor comparado aos demais, confirmam que a escala proposta é responsiva tanto para identificar graus intensos de dor, como graus de intensidade moderada, ocorrido após o resgate analgésico ou mesmo leve, ocorrido 24 horas após a cirurgia. Nos bovinos as diferenças entre escores ($M2 > M4 > M3 > M1$) (DE OLIVEIRA et al., 2014) demonstraram uma redução do efeito analgésico após 24 horas da cirurgia, devido à presença de dor. Em ovinos a dor em M4 diminuiu após M3 ($M2 > M3 > M4 > M1$), mas o protocolo cirúrgico, anestésico e analgésico foi diferente nas duas espécies de ruminantes,. O aumento e a diminuição dos escores de dor após a cirurgia e analgesia de resgate, respectivamente, também ocorreu em gatos (BRONDANI et al., 2013), equinos (TAFFAREL et al., 2015) e suínos (LUNA et al., 2015). Como se realizou este estudo em três raças diferentes e não houve efeito das raças na pontuação dos escores, aparentemente pode-se utilizar o instrumento em diferentes raças de ovinos.

Todos os itens da escala proposta tiveram uma correlação item-total aceitável, tal como em suínos (LUNA ET AL., 2015), o que demonstra as suas relevâncias individuais e assegura a homogeneidade da ferramenta. A consistência interna da escala proposta foi excelente, com resultados muito similares a gatos - 0.86 (BRONDANI et al., 2013), bovinos - 0.87 (DE OLIVEIRA et al., 2014) e suínos - 0.82 (LUNA et al., 2015), o que assegura que os escores dos itens que compõem a escala podem ser somados e garantem que o escore total seja representativo da intensidade da dor (BRONDANI et al., 2013). A semelhança dos valores quando se exclui cada item demonstra que eles apresentam tendência e importância semelhantes (JENSEN, 2003).

A escala foi específica para todos os itens, entretanto, foi sensível apenas para três. Destes três itens, os valores da posição da cabeça e da postura apresentaram-se muito próximos dos considerados como de boa sensibilidade (0.70). Apenas o apetite não apresentou sensibilidade, desta forma ovinos mesmo com dor podem se alimentar, como descrito em outras espécies (DE OLIVEIRA et al., 2014; LUNA ET AL., 2015). Para garantir a identificação de animais verdadeiros

positivos, ou seja aqueles que sofrem dor, o critério de sensibilidade é mais relevante do que a especificidade, pois se a escala for sensível, pode-se diagnosticar a dor, para poder tratá-la. Quando pouco específica, pode-se diagnosticar uma dor inexistente, e tratar o animal sem necessidade. Desta forma do ponto de vista do bem-estar do animal, é melhor tratar um animal que não tenha dor do que não fazê-lo quando tem.

A análise da curva ROC (STREINER; CAIRNEY, 2007) estimou o ponto de corte para a intervenção analgésica em ovinos tal como se fez nas escalas de dor em gatos (BRONDANI et al., 2013), bovinos (DE OLIVEIRA et al., 2014) e suínos (LUNA et al., 2015). A determinação de escores indicativos da necessidade do uso de analgésicos auxilia na decisão clínica do profissional, para ratificar a eficácia do tratamento analgésico (BRONDANI et al., 2013) e previne o sofrimento desnecessário em animais. De acordo com a zona de incerteza diagnóstica de todos os avaliadores, assegura-se que ovinos com escore de 4,5 de um total de 12 pontos (≥ 4) estão realmente com dor, enquanto os que têm escore $\leq 2,5$ não apresentam dor.

A baixa porcentagem de animais dentro da zona de incerteza diagnóstica garante uma boa confiabilidade na tomada de decisão sobre a indicação de analgesia de resgate em animais que realmente apresentam dor e, que portanto deveriam receber analgesia. Desta forma, a escala proposta apresenta uma excelente acurácia diagnóstica. Embora a definição do escore referente ao ponto de intervenção analgésica seja uma boa ferramenta, ressalta-se mesmo que os escores sejam < 4 , pode ser necessária analgesia adicional de acordo com a avaliação clínica em alguns casos, a critério do observador.

O ponto de corte foi $> 4/10$ em bovinos (DE OLIVEIRA et al., 2014), $> 4/18$ em suínos (LUNA et al., 2015) e para a subescala “expressão de dor” em gatos foi $> 2/12$ (BRONDANI et al., 2013). Em estudo empírico recente pontuou-se a dor em cirurgia cardíaca de ovinos; para escores de 0-2/25 não houve intervenção, de 3-9/25 realizaram resgate analgésico e $\geq 10/25$ realizaram analgesia multimodal (IZER et al., 2018). Em outro estudo se submeteram 18 ovelhas a laparoscopia e avaliou-se a dor por 24 horas com base em três itens: diminuição do apetite, dificuldade de locomoção e arqueamento do dorso, porém a escala de dor variando de 0 a 6 foi pouco sensível, com 90% dos animais com escore “0” e 10% com “1” (TEIXEIRA et al., 2011a).

Em consonância à baixa porcentagem de animais dentro da zona de incerteza diagnóstica, as altas áreas sob a curva observadas neste estudo (> 0.95) indicam que a escala possui alta capacidade discriminatória, isto é, ela classifica corretamente os indivíduos com ou sem dor (STREINER; CAIRNEY, 2007), resultados que se assemelharam aos bovinos (DE OLIVEIRA et al., 2014), suínos (LUNA et al., 2015) e na subescala “expressão de dor” em gatos (BRONDANI et al., 2013), usando a mesma metodologia.

Dentro do nosso conhecimento, até o momento as escalas que avaliam a dor em várias espécies de animais não classificam a sua intensidade com base em seus escores, a não ser de forma empírica (REID et al., 2007; BRONDANI et al., 2013; DE OLIVEIRA et al., 2014; LUNA et al., 2015)). Apesar dessas escalas terem um ponto de corte para o reconhecimento da dor, isso não é

quantificável para todos os escores e sua intensidade. Na escala deste estudo, a dor de intensidade leve correspondeu aos escores de [0 - 3] e a dor moderada aos escores [4 - 8]. A zona de incerteza diagnóstica foi de [2,5 – 4,5], no limite máximo da dor leve e no limite mínimo da dor moderada. Tal achado corrobora o fato de a dor de intensidade leve ser mais difícil de diagnosticar. Os escores ≥ 9 correspondem à dor de intensidade severa, que prevaleceram em animais de laboratório submetidos à cirurgia ortopédica, de grande invasividade (KOHN et al., 2007). A aplicação clínica da nossa escala a cirurgias ortopédicas de ovinos poderá investigar se nesta espécie ocorre o mesmo.

Isto demonstra a importância de aplicar escalas validadas ao se quantificar a dor, já que esta é uma experiência individual (FILLINGIM, 2017), que não deve ser generalizada de acordo com o grau de invasividade cirúrgica. A possibilidade de se avaliar e quantificar a dor individualmente, garante o tratamento adequado à sua intensidade, com melhor seleção de fármacos e doses (KOHN et al., 2007).

A expressão facial para avaliar a dor é uma ferramenta prática e usada em animais humanos ou não. A escala facial para detectar a dor em ovelhas adultas foi considerada validada, confiável e precisa (MCLENNAN et al., 2016). Apresentou correlação positiva dos escores totais de dor com pododermatite e mastite, e o respectivo tratamento reduziu o seu escore total. A intenção inicial de se aplicar esta escala neste estudo foi para reforçar a validação da escala comportamental.

A análise de componentes principais gerou um fator e considerou-se a escala unidimensional. A confiabilidade intraobservador para todos os momentos foi boa para três avaliadores, já a interobservador foi moderada em apenas um caso, resultados estes inferiores à muito boa (0,86) do estudo original (MCLENNAN et al., 2016) e à moderada/boa de dois experimentos da mesma pesquisa (GUESGEN et al., 2016b).

A correlação da escala facial com os escores das outras quatro escalas foi moderada. A escala facial original foi validada com base em fotos, mesmo método aqui adotado, mas a filmagem da expressão facial (MCLENNAN et al., 2016) por meio de extração de imagens dos vídeos tem-se revelado mais fidedigna para avaliar a dor (FINKA et al., 2019; FUTRO; MASŁOWSKA; DWYER, 2015; GLEERUP et al., 2015; GRANT, 2004; GUESGEN; BEAUSOLEIL; STEWART, 2013; MOLONY; KENT; MCKENDRICK, 2002).

A escala facial foi responsiva e apresentou excelente consistência interna. Todos os itens da escala foram aceitos na correlação item-total. Os itens “posição da orelha frontal e lateral” foram os únicos que apresentaram tanto especificidade, como sensibilidade. Essas alterações na posição da orelha são relevantes, como demonstrado após o corte de cauda de cordeiros, em que as orelhas para trás se associaram à dor, ao contrário das orelhas para a frente ou para o lado (GUESGEN et al., 2016a). Um outro estudo relacionou a posição das orelhas às respectivas emoções: horizontal – neutro; trás – medo; cima – raiva; assimétricas – surpresa (BOISSY et al., 2011).

O ponto de corte foi ≥ 6 , considerando todos os avaliadores, enquanto no estudo que validou a escala facial esse ponto foi $> 1,5$ - tanto para pododermatite como para mastite - (MCLENNAN et

al., 2016), porque apenas se considerou o avaliador mais experiente. A acurácia de dois avaliadores foi similar ao estudo original ($\approx 0,80$) (MCLENNAN et al., 2016). Em estudo recente estimou-se o nível de dor de ovelhas pela deteção de unidade de ação facial com uma acurácia de 67% (MCLENNAN; MAHMOUD, 2019).

6.1 Limitações

O estudo apresentou algumas limitações. A principal é que a escala comportamental foi validada apenas para um tipo específico de cirurgia de tecidos moles (laparoscopia) e em fêmeas. São necessários mais estudos para testar esta ferramenta em diferentes procedimentos, como cirurgia ortopédica e em circunstâncias clínicas, para afiançar a sua versatilidade. Para garantir que o instrumento seja válido em condições de campo, também é necessária a validação clínica com observadores menos experientes. Como a maioria dos estudos que avaliaram a dor aguda em ovinos foi em cordeiros, isto pode limitar o cotejamento dos dados, o que requer que o instrumento seja testado em cordeiros.

Outra limitação é que a análise de vídeo não se equipara necessariamente a presencial. A observação por vídeo tem como desvantagem carecer de alguns detalhes observados em tempo real, e como vantagem pode-se rever mais vezes. De acordo com a experiência anterior dos autores, a análise de vídeo fornece material para desenvolver o etograma e é um passo importante para validar uma escala, como relatado em gatos (BRONDANI et al., 2013), equinos (TAFFAREL et al., 2015), bovinos (DE OLIVEIRA et al., 2014) e suínos (LUNA et al., 2015), sobretudo para avaliar a confiabilidade intraobservador. De acordo com estudo de nosso grupo em gatos, os resultados da análise de vídeo também foram reproduzíveis em um cenário clínico (BRONDANI et al., 2013; BRONDANI; LUNA; PADOVANI, 2011a, 2011b).

Algumas limitações se referem especificamente aos vídeos. Embora o estudo tenha sido encoberto, alguns vídeos podem sugerir sobre qual momento se tratava: no momento basal, as ovelhas estavam sob jejum, sem ração disponível, daí a dificuldade dos observadores interpretarem se os animais não se alimentavam dada à falta de alimento ou se realmente apresentavam anorexia; ao redor de 21% dos vídeos em M3 foi filmagem noturna com luz artificial, o que poderia sugerir que correspondessem à M3; variações no ciclo circadiano poderiam alterar alguns comportamentos como a atividade, desta forma a redução na atividade observada à noite pode não ter relação com dor ou desconforto, mas sim com a redução natural da atividade no período noturno (WYSE et al., 2018); dada à pequena diferença no tamanho das baias, a densidade de animais variava um pouco, o que poderia influenciar os comportamentos de interação e locomoção; a lâ escura de alguns animais também pode ter dificultado avaliar alguns itens dos vídeos/fotos, tornando a análise menos precisa, sobretudo da escala facial.

Para melhorar a confiabilidade dos dados, os autores sugerem que os observadores passem por um período de treinamento, já que a instrução e o treinamento melhoraram o reconhecimento da dor ao menos em animais de laboratório (ROUGHAN; FLECKNELL, 2003).

7. CONCLUSÃO

Conclui-se que, após o refinamento da escala originalmente proposta, a escala composta da Unesp-Botucatu para avaliação da dor aguda pós-operatória em ovinos é um instrumento válido, confiável, responsivo, específico e sensível, com excelente consistência interna e capacidade discriminatória. A indicação do ponto de corte bem para a analgesia de resgate e a classificação da intensidade da dor da escala embasa a identificação da dor e o adequado tratamento analgésico pós-operatório em ovelhas. Para avaliar a aplicabilidade clínica da escala e garantir a sua versatilidade, recomenda-se avaliá-la em outros procedimentos cirúrgicos e em cordeiros.

8. REFERÊNCIAS

- ADAMS, J. Assessment and management of pain in small ruminants. **Livestock**, v. 22, p. 324–328, 2 nov. 2017.
- ALTMAN, D. Some common problems in medical research. In: HALL, C. & (Ed.). . **Practical statistics for medical research**. London: Chapman and Hall, 1991. p. 404–408.
- ANGELL, J. W. et al. A practical tool for locomotion scoring in sheep: Reliability when used by veterinary surgeons and sheep farmers. **Veterinary Record**, v. 176, n. 20, p. 521, 2015.
- ANIL, S. S. et al. Validation of a method for assessment of an acute pain in lambs. **Applied Animal Behaviour Science**, v. 76, n. 3, p. 215–238, 2002.
- BANKS, E. M. Behavioral research to answer questions about animal welfare. **Journal of animal science**, v. 54, n. 2, p. 434–446, fev. 1982.
- BELSHAW, Z. et al. Quality of life assessment in domestic dogs: An evidence-based rapid review. **Veterinary Journal**, v. 206, n. 2, p. 203–212, 2015.
- BEYER, J. E.; WELLS, N. The assessment of pain in children. **Pediatric clinics of North America**, v. 36, n. 4, p. 837–854, ago. 1989.
- BOATENG, G. O.; NEILANDS, T. B.; FRONGILLO, E. A. Best Practices for Developing and Validating Scales for Health , Social , and Behavioral Research : A Primer. v. 6, n. June, p. 1–18, 2018.
- BOISSY, A. et al. Cognitive sciences to relate ear postures to emotions in sheep. **Animal Welfare**, v. 20, n. 1, p. 47- 56 (10), 2011.
- BRONDANI, J. T. et al. Validade e responsabilidade de uma escala multidimensional para avaliação de dor pós-operatória em gatos. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 64, n. 6, p. 1529–1538, 2012.
- BRONDANI, J. T. et al. Validation of the English version of the UNESP-Botucatu multidimensional composite pain scale for assessing postoperative pain in cats. **BMC Veterinary Research**, v. 9, n. 1, p. 1, 2013.
- BRONDANI, J. T.; LUNA, S. P. L.; PADOVANI, C. R. Refinement and initial validation of a multidimensional composite scale for use in assessing acute postoperative pain in cats. **American journal of veterinary research**, v. 72, n. 2, p. 174–183, fev. 2011a.
- BRONDANI, J. T.; LUNA, S.; PADOVANI, C. R. Assessing Acute Postoperative Pain in Cats. **American Journal of Veterinary Research**, v. 72, n. 2, p. 174–183, 2011b.
- BUSSIERES, G. et al. Development of a composite orthopaedic pain scale in horses. **Research in veterinary science**, v. 85, n. 2, p. 294–306, out. 2008.
- CANNESSON, M. et al. Assessing the diagnostic accuracy of pulse pressure variations for the prediction of fluid responsiveness: a “gray zone” approach. **Anesthesiology**, v. 115, n. 2, p. 231–241, ago. 2011.
- CELEITA-RODRÍGUEZ, N. et al. Comparison of the diagnostic accuracy of dynamic and static preload indexes to predict fluid responsiveness in mechanically ventilated, isoflurane anesthetized dogs. **Veterinary Anaesthesia and Analgesia**, v. 46, n. 3, p. 276–288, 1 maio 2019.
- COHEN, J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. **Psychological bulletin**, v. 70, n. 4, p. 213–220, out. 1968.
- CRELLIN, D. et al. Analysis of the validation of existing behavioral pain and distress scales for use in the procedural setting. **Paediatric anaesthesia**, v. 17, n. 8, p. 720–733, ago. 2007.
- CRONBACH, L. J. Coefficient alpha and the internal structure of tests. **Psychometrika**, v. 16, p. 297–334, 1951.

- CURRAH, J. M.; HENDRICK, S. H.; STOOKEY, J. M. The behavioral assessment and alleviation of pain associated with castration in beef calves treated with flunixin meglumine and caudal lidocaine epidural anesthesia with epinephrine. **The Canadian veterinary journal = La revue veterinaire canadienne**, v. 50, n. 4, p. 375–382, abr. 2009.
- DE ARAUJO, E. A. B. et al. Aspiração folicular videolaparoscópica comparativa em ovelhas Dorper e Santa Inês. **Ciencia Animal Brasileira**, v. 17, n. 1, p. 98–104, 2016.
- DE OLIVEIRA, F. A. et al. Validation of the UNESP-Botucatu unidimensional composite pain scale for assessing postoperative pain in cattle. **BMC Veterinary Research**, v. 10, n. 1, p. 1–14, 2014.
- DETECTION, P. Journal of Applied Animal Welfare Pain Detection and Amelioration in Animals on the Farm : Issues and Options. v. 8, n. 776099595, p. 37–41, 2005.
- DEVON, H. A. et al. A psychometric toolbox for testing validity and reliability. **Journal of nursing scholarship : an official publication of Sigma Theta Tau International Honor Society of Nursing**, v. 39, n. 2, p. 155–164, 2007.
- DEYO, R. A.; DIEHR, P.; PATRICK, D. L. Reproducibility and responsiveness of health status measures statistics and strategies for evaluation. **Controlled Clinical Trials**, v. 12, n. 4, p. S142–S158, 1 ago. 1991.
- DIAS, I. R. et al. Preclinical and Translational Studies in Small Ruminants (Sheep and Goat) as Models for Osteoporosis Research. **Current osteoporosis reports**, v. 16, n. 2, p. 182–197, abr. 2018.
- DINNISS, A. S. et al. The behaviour pattern of lambs after castration using a rubber ring and/or castrating clamp with or without local anaesthetic. **New Zealand veterinary journal**, v. 47, n. 6, p. 198–203, dez. 1999.
- DOLAN, S. et al. Up-regulation of metabotropic glutamate receptor subtypes 3 and 5 in spinal cord in a clinical model of persistent inflammation and hyperalgesia. **Pain**, v. 106, n. 3, p. 501–512, dez. 2003.
- DOLAN, S. et al. Activation of metabotropic glutamate receptor 7 in spinal cord inhibits pain and hyperalgesia in a novel formalin model in sheep. **Behavioural Pharmacology**, v. 22, n. 5–6, p. 582–588, set. 2011.
- DOLAN, S.; FIELD, L. C.; NOLAN, A. M. The role of nitric oxide and prostaglandin signaling pathways in spinal nociceptive processing in chronic inflammation. **Pain**, v. 86, n. 3, p. 311–320, jun. 2000.
- DWYER, C. M.; LAWRENCE, A. B. Maternal behaviour in domestic sheep (ovis aries): Constancy and change with maternal experience. **Behaviour**, v. 137, n. 10, p. 1391–1413, 2000.
- EC, J. et al. EEG changes in 4-week-old lambs in response to castration, tail docking and mulesing. **Australian Veterinary Journal**, v. 78, n. 5, p. 339–343, 2000.
- FERGUSON, D. M.; LEE, C.; FISHER, A. **Advances in Sheep Welfare**. [s.l: s.n.].
- FILLINGIM, R. B. Individual differences in pain: Understanding the mosaic that makes pain personal. **Pain**, 2017.
- FINKA, L. R. et al. Geometric morphometrics for the study of facial expressions in non- human animals , using the domestic cat as an exemplar. **Scientific Reports**, n. June, p. 1–12, 2019.
- FIRTH, A. M.; HALDANE, S. L. Development of a scale to evaluate postoperative pain in dogs. **Journal of the American Veterinary Medical Association**, v. 214, n. 5, p. 651–659, mar. 1999.
- FLECKNELL, P. Analgesia from a veterinary perspective. **British journal of anaesthesia**, v. 101, n. 1, p. 121–124, jul. 2008.
- FUTRO, A.; MASŁOWSKA, K.; DWYER, C. M. Ewes direct most maternal attention towards lambs that show the greatest pain-related behavioural responses. **PLoS ONE**, v. 10, n. 7, p. 1–15, 2015.
- GALL, O. et al. Postoperative pain assessment in children: a pilot study of the usefulness of the analgesia nociception index. **British journal of anaesthesia**, v. 115, n. 6, p. 890–895, dez. 2015.

GIGLIUTO, C. et al. Pain assessment in animal models: do we need further studies? **Journal of pain research**, v. 7, p. 227–236, 2014.

GLEERUP, K. B. et al. **An equine pain face**. Veterinary anaesthesia and analgesia. **Anais...**United States: jan. 2015

GRACEY, R. H. **Evaluation of multi-dimensional pain scales**.PainUnited States, mar. 1992.

GRAHAM, M. J.; KENT, J. E.; MOLONY, V. Effects of four analgesic treatments on the behavioural and cortisol responses of 3-week-old lambs to tail docking. **Veterinary journal (London, England : 1997)**, v. 153, n. 1, p. 87–97, jan. 1997.

GRANT, C. Behavioural responses of lambs to common painful husbandry procedures. **Applied Animal Behaviour Science**, v. 87, n. 3, p. 255–273, 2004.

GRAUBNER, C. et al. Clinical application and reliability of a post abdominal surgery pain assessment scale (PASPAS) in horses. **Veterinary journal (London, England : 1997)**, v. 188, n. 2, p. 178–183, maio 2011.

GUESGEN, M. J. et al. Lambs show changes in ear posture when experiencing pain. **Animal Welfare**, v. 25, n. 2, p. 171–177, 2016a.

GUESGEN, M. J. et al. Coding and quantification of a facial expression for pain in lambs. **Behavioural Processes**, v. 132, p. 49–56, nov. 2016b.

GUESGEN, M. J.; BEAUSOLEIL, N. J.; STEWART, M. Effects of early human handling on the pain sensitivity of young lambs. **Veterinary Anaesthesia and Analgesia**, v. 40, n. 1, p. 55–62, 2013.

HEWSON, C. J. et al. Canadian veterinarians' use of analgesics in cattle, pigs, and horses in 2004 and 2005. **The Canadian veterinary journal = La revue veterinaire canadienne**, v. 48, n. 2, p. 155–164, fev. 2007.

HOLTON, L. et al. Development of a behaviour-based scale to measure acute pain in dogs. **The Veterinary record**, v. 148, n. 17, p. 525–531, abr. 2001.

HUXLEY, J. N.; WHAY, H. R. Current attitudes of cattle practitioners to pain and the use of analgesics in cattle. **The Veterinary record**, v. 159, n. 20, p. 662–668, nov. 2006.

ISON, S. H. et al. **A Review of Pain Assessment in Pigs** **Frontiers in Veterinary Science** , 2016. Disponível em: <<https://www.frontiersin.org/article/10.3389/fvets.2016.00108>>

IZER, J. M. et al. Development of a Pain Scoring System for Use in Sheep Surgically Implanted with Ventricular Assist Devices. **Journal of Investigative Surgery**, v. 1939, p. 1–10, 2018.

JENSEN, M. P. Questionnaire validation: a brief guide for readers of the research literature. **The Clinical journal of pain**, v. 19, n. 6, p. 345–352, 2003.

KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, v. 23, p. 187–200, 1958.

KALER, J. et al. Factors associated with changes of state of foot conformation and lameness in a flock of sheep. **Preventive veterinary medicine**, v. 97, n. 3–4, p. 237–244, dez. 2010.

KALER, J.; WASSINK, G. J.; GREEN, L. E. The inter- and intra-observer reliability of a locomotion scoring scale for sheep. **Veterinary Journal**, v. 180, n. 2, p. 189–194, 2009.

KENT, J. E.; MOLONY, V.; GRAHAM, M. J. Comparison of methods for the reduction of acute pain produced by rubber ring castration or tail docking of week-old lambs. **Veterinary Journal**, v. 155, n. 1, p. 39–51, jan. 1998.

KENT, J. E.; MOLONY, V.; ROBERTSON, I. S. Comparison of the Burdizzo and rubber ring methods for castrating and tail docking lambs. **The Veterinary record**, v. 136, n. 8, p. 192–196, fev. 1995.

- KILKENNY, C. et al. Animal research: reporting in vivo experiments: the ARRIVE guidelines. **British journal of pharmacology**, v. 160, n. 7, p. 1577–1579, ago. 2010.
- KOHN, D. F. et al. Guidelines for the assessment and management of pain in rodents and rabbits. **Journal of the American Association for Laboratory Animal Science**, v. 46, n. 2, p. 97–108, 2007.
- LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, n. 1, p. 159–174, mar. 1977.
- LEACH, M. C. et al. Research in Veterinary Science Behavioural effects of ovariohysterectomy and oral administration of meloxicam in laboratory housed rabbits. **Research in Veterinary Science**, v. 87, n. 2, p. 336–347, 2009.
- LESTER, L. S.; FANSELOW, M. S. Exposure to a Cat Produces Opioid Analgesia in Rats. v. 99, p. 756–759, 1985.
- LESTER, S. J. et al. Behavioural and cortisol responses of lambs to castration and tailing using different methods. **New Zealand veterinary journal**, v. 44, n. 2, p. 45–54, abr. 1996.
- LEY, S. J.; LIVINGSTON, A.; WATERMAN, A. E. The effect of chronic clinical pain on thermal and mechanical thresholds in sheep. **Pain**, v. 39, n. 3, p. 353–357, 1989.
- LEY, S. J.; WATERMAN, A. E.; LIVINGSTON, A. A field study of the effect of lameness on mechanical nociceptive thresholds in sheep. **The Veterinary record**, v. 137, n. 4, p. 85–87, jul. 1995.
- LIZARRAGA, I.; CHAMBERS, J. P. Use of analgesic drugs for pain management in sheep. **New Zealand Veterinary Journal**, v. 60, n. 2, p. 87–94, 2012.
- LORENA, S. E. R. S. et al. Attitude of Brazilian veterinarians in the recognition and treatment of pain in horses and cattle. **Veterinary Anaesthesia and Analgesia**, v. 40, n. 4, p. 410–418, 2013.
- LUNA, S; ARAUJO, A; NETO, P; OLIVEIRA, F; BRONDANI, J; TELLES, F; AZEREDO, L. **Validation of the UNESP-Botucatu unidimensional composite pain scale for assessing postoperative pain in pigs.** 12th World Congress of Veterinary Anesthesiology. **Anais...Kyoto**: 2015
- MALLAT, J. et al. Decrease in pulse pressure and stroke volume variations after mini-fluid challenge accurately predicts fluid responsivenessdagger. **British journal of anaesthesia**, v. 115, n. 3, p. 449–456, set. 2015.
- MANTECA, X. et al. Évaluation De La Douleur Chez Les Ovins. n. 1, p. 1–2, 2017.
- MARTIN, P.; BATESON, P. P. G. **Measuring behaviour: An introductory guide, 2nd ed.** New York, NY, US: Cambridge University Press, 1993.
- MATHER, L. E. et al. Lack of secondary hyperalgesia and central sensitization in an acute sheep model. **Regional Anesthesia and Pain Medicine**, v. 25, n. 2, p. 174–180, 2000.
- MCDOWELL, I. The Theoretical and Technical Foundations of Health Measurement. In: **Measuring Health**. 3. ed. New York: Oxford University Press, 2006.
- MCDOWELL, I. et al. The development and assessment of temperament tests for adult companion dogs. **Veterinary journal (London, England : 1997)**, v. 1, n. 3, p. 94–108, 1 dez. 2006.
- MCGLONE, J. J.; STOBART, R. H. A quantitative ethogram of behavior of yearling ewes during two hours post-parturition. **Applied Animal Behaviour Science**, v. 16, n. 2, p. 157–164, 1986.
- MCLENNAN, K. M. et al. Technical note: Validation of an automatic recording system to assess behavioural activity level in sheep (*Ovis aries*). **Small Ruminant Research**, v. 127, p. 92–96, 2015.
- MCLENNAN, K. M. et al. Development of a facial expression scale using footrot and mastitis as models of pain in sheep. **Applied Animal Behaviour Science**, v. 176, p. 19–26, 2016.

- MCLENNAN, K.; MAHMOUD, M. Development of an Automated Pain Facial Expression Detection System for Sheep (*Ovis Aries*). **Animals**, v. 9, n. 4, p. 196, abr. 2019.
- MELCHES, S. et al. Castration of lambs: A welfare comparison of different castration techniques in lambs over 10 weeks of age. **Veterinary Journal**, v. 173, n. 3, p. 554–563, 2007.
- MELLEMA, S. C. et al. Influence of local anaesthesia on pain and distress induced by two bloodless castration methods in young lambs. **Veterinary Journal**, v. 172, n. 2, p. 274–283, 2006.
- MELLOR, D. J. et al. A comparison of catecholamine and cortisol responses of young lambs and calves to painful husbandry procedures. **Australian veterinary journal**, v. 80, n. 4, p. 228–233, abr. 2002.
- MELLOR, D. J.; MURRAY, L. Effects of tail docking and castration on behaviour and plasma cortisol concentrations in young lambs. **Research in veterinary science**, v. 46, n. 3, p. 387–391, maio 1989.
- MEROLA, I.; MILLS, D. S. Systematic review of the behavioural assessment of pain in cats. **Journal of Feline Medicine and Surgery**, v. 18, n. 2, p. 60–76, 2016.
- MILLMAN, S. T. Behavioral responses of cattle to pain and implications for diagnosis, management, and animal welfare. **The Veterinary clinics of North America. Food animal practice**, v. 29, n. 1, p. 47–58, mar. 2013.
- MOLONY, V. et al. Pain in lambs castrated at 2days using novel smaller and tighter rubber rings without and with local anaesthetic. **Veterinary Journal**, v. 193, n. 1, p. 81–86, 2012.
- MOLONY, V.; KENT, J. E. Assessment of acute pain in farm animals using behavioral and physiological measurements. **Journal of animal science**, v. 75, n. 1, p. 266–272, jan. 1997.
- MOLONY, V.; KENT, J. E.; MCKENDRICK, I. J. Validation of a method for assessment of an acute pain in lambs. **Applied Animal Behaviour Science**, v. 76, n. 3, p. 215–238, 2002.
- MOLONY, V.; KENT, J. E.; ROBERTSON, I. S. Behavioural responses of lambs of three ages in the first three hours after three methods of castration and tail docking. **Research in veterinary science**, v. 55, n. 2, p. 236–245, set. 1993.
- MORTON, C. M. et al. Application of a scaling model to establish and validate an interval level pain scale for assessment of acute pain in dogs. **American journal of veterinary research**, v. 66, n. 12, p. 2154–2166, dez. 2005.
- MURRELL, J. C. et al. Application of a modified form of the Glasgow pain scale in a veterinary teaching centre in the Netherlands. **The Veterinary record**, v. 162, n. 13, p. 403–408, mar. 2008.
- NETEMEYER, R.; BEARDEN, W.; SHARMA, S. Scaling Procedures. Issues and Applications. **Sage**, 2003.
- ONG, R. M. et al. Behavioural and EEG changes in sheep in response to painful acute electrical stimuli. **Australian Veterinary Journal**, v. 75, n. 3, p. 189–193, 1997.
- PASQUALI, L. Psychometrics. **Revista da Escola de Enfermagem da USP**, v. 43, n. spe SE-Original Article, 1 dez. 2009.
- PAULL, D. R. et al. Effects of meloxicam or tolfenamic acid administration on the pain and stress responses of Merino lambs to mulesing. **Australian veterinary journal**, v. 86, n. 8, p. 303–311, ago. 2008.
- PINHO RH, LEACH MC, ROCHA FDL, ARMANI DJC, FERREIRA-SOUZA G, DISSENHA A, CALCIOLARI K, MINTO BW, LUNA, S. **Influence of the presence or absence of an observer on pain assessment in rabbits submitted to orthopaedic surgery**. AVA Autumn meeting. **Anais...Ghent**: 2019
- RAEKALLIO, M. et al. Pain alleviation in animals: attitudes and practices of Finnish veterinarians. **Veterinary journal (London, England : 1997)**, v. 165, n. 2, p. 131–135, mar. 2003.
- REID, J. et al. Development of the short-form Glasgow Composite Measure Pain Scale (CMPS-SF) and derivation of an analgesic intervention score. **Animal Welfare**, v. 16, p. 97–104, maio 2007.

- ROBERTSON, I. S.; KENT, J. E.; MOLONY, V. Effect of different methods of castration on behaviour and plasma cortisol in calves of three ages. **Research in veterinary science**, v. 56, n. 1, p. 8–17, jan. 1994.
- RODRIGUEZ, M. et al. Aspiração Folicular Por Laparoscopia Em Ovinos. **Revista Investigação**, v. 14, n. 2, p. 55–60, 2015.
- ROUGHAN, J. V; FLECKNELL, P. A. Evaluation of a short duration behaviour-based post-operative pain scoring system in rats. **European journal of pain (London, England)**, v. 7, n. 5, p. 397–406, 2003.
- RSTUDIOTEAM. **RStudio Team**, 2016. Disponível em: <<http://www.rstudio.com/>>
- SARTORETTO, S. C. et al. SHEEP AS AN EXPERIMENTAL MODEL FOR BIOMATERIAL IMPLANT EVALUATION. **Acta ortopédica brasileira**, v. 24, n. 5, p. 262–266, 2016.
- SCHUSTER, C. A Note on the Interpretation of Weighted Kappa and its Relations to Other Rater Agreement Statistics for Metric Scales. **Educational and Psychological Measurement**, v. 64, n. 2, p. 243–253, 1 abr. 2004.
- SMALL, A. H. et al. Efficacy of a buccal meloxicam formulation for pain relief in Merino lambs undergoing knife castration and tail docking in a randomised field trial. **Australian Veterinary Journal**, v. 92, n. 10, p. 381–388, 2014.
- SORGE, R. E. et al. Olfactory exposure to males , including men , causes stress and related analgesia in rodents. v. 11, n. 6, 2014.
- STREINER, D. L. Starting at the beginning: an introduction to coefficient alpha and internal consistency. **Journal of personality assessment**, v. 80, n. 1, p. 99–103, fev. 2003.
- STREINER, D. L.; CAIRNEY, J. What's under the ROC? An introduction to receiver operating characteristics curves. **Canadian journal of psychiatry. Revue canadienne de psychiatrie**, v. 52, n. 2, p. 121–128, fev. 2007.
- STREINER, D.; NORMAN, G.; CAIRNEY, J. **Health Measurement Scales: A Practical Guide To Their Development and Use**. [s.l: s.n.]. v. 117
- STUBSJØEN, S. M. et al. Applying fractal analysis to heart rate time series of sheep experiencing pain. **Physiology and Behavior**, v. 101, n. 1, p. 74–80, 2010.
- STUBSJØEN, S. M. et al. Assessment of chronic stress in sheep (part I): The use of cortisol and cortisone in hair as non-invasive biological markers. **Small Ruminant Research**, v. 132, n. part I, p. 25–31, 2015a.
- STUBSJØEN, S. M. et al. Assessment of chronic stress in sheep (part II): Exploring heart rate variability as a non-invasive measure to evaluate cardiac regulation. **Small Ruminant Research**, v. 133, p. 30–35, 2015b.
- SUTTON, G. A. et al. A behaviour-based pain scale for horses with acute colic: scale construction. **Veterinary journal (London, England : 1997)**, v. 196, n. 3, p. 394–401, jun. 2013a.
- SUTTON, G. A. et al. Validation of two behaviour-based pain scales for horses with acute colic. **Veterinary journal (London, England : 1997)**, v. 197, n. 3, p. 646–650, set. 2013b.
- TAFFAREL, M. O. et al. Refinement and partial validation of the UNESP-Botucatu multidimensional composite pain scale for assessing postoperative pain in horses. **BMC Veterinary Research**, v. 11, n. 1, 2015.
- TAYLOR, K.; REGO, L. EU statistics on animal experiments for 2014. **ALTEX - Alternatives to animal experimentation**, v. 33, n. 4 SE-Letters, 1 nov. 2016.
- TEIXEIRA, P. P. M. et al. Laparoscopic ovum collection in sheep: gross and microscopic evaluation of the ovary and influence on oocyte production. **Animal reproduction science**, v. 127, n. 3–4, p. 169–175, set. 2011a.
- TEIXEIRA, P. P. M. M. et al. Ovariectomy by laparotomy, a video-assisted approach or a complete

laparoscopic technique in Santa Ines sheep. **Small Ruminant Research**, v. 99, n. 2, p. 199–202, 2011b.

THORNTON, P. D.; WATERMAN-PEARSON, A. E. Quantification of the pain and distress responses to castration in young lambs. **Research in veterinary science**, v. 66, n. 2, p. 107–118, abr. 1999.

THORNTON, P. D.; WATERMAN-PEARSON, A. E. Behavioural responses to castration in lambs. **Animal Welfare**, v. 11, n. 2, p. 203–212, 2002.

TRINDADE, P. H. E. et al. Effect of work on body language of ranch horses in Brazil. **PLOS ONE**, v. 15, n. 1, p. e0228130, jan. 2020

VON BAEYER, C. L.; SPAGRUD, L. J. Systematic review of observational (behavioral) measures of pain for children and adolescents aged 3 to 18 years. **Pain**, v. 127, n. 1–2, p. 140–150, jan. 2007.

WATTS; STOOKEY. Vocal behaviour in cattle: the animal's commentary on its biological processes and welfare. **Applied animal behaviour science**, v. 67, n. 1–2, p. 15–33, mar. 2000.

WELSH, E. M.; GETTINBY, G.; NOLAN, A. M. Comparison of a visual analogue scale and a numerical rating scale for assessment of lameness, using sheep as a model. **American journal of veterinary research**, v. 54, n. 6, p. 976–983, jun. 1993.

WEST, D. M. B. **The Sheep Book Health Disease and Production**. [s.l.] New Zealand Veterinary Association Vetlearn., 2009.

WILKES, D. et al. A large animal neuropathic pain model in sheep: A strategy for improving the predictability of preclinical models for therapeutic development. **Journal of Pain Research**, v. 5, p. 415–424, 2012.

WYSE, C. A. et al. Circadian rhythms of melatonin and behaviour in juvenile sheep in field conditions: Effects of photoperiod, environment and weaning. **Physiology & behavior**, v. 194, p. 362–370, out. 2018.

APÊNDICES

Tabela I - Etograma com a descrição da duração dos comportamentos analisados nos momentos peri-operatórios em 48 ovinos submetidos a laparoscopia.

Comportamento	Descrição
Manutenção	
Comer	Ato de comer ração
Ruminar	Ato de ruminar
Beber	Ato de beber água
Urinar	Ato de urinar
Defecar	Ato de defecar
Comportamentos típicos	
Interação social	Ativo, atento ao ambiente, interage e/ou acompanha os outros animais
Locomoção	Move-se livremente, sem alteração de locomoção; Quando parado, os membros pélvicos estão paralelos aos torácicos.
Postura em estação normal	Posição quadrupedal parada
Cabeça acima da linha da coluna	Posição da cabeça levantada
Comportamentos indicativos de dor moderada	
Interação reduzida	Apático: pode permanecer próximo aos outros animais, mas interage pouco.
Locomoção reduzida/alterada	Move-se com restrição e/ou com passos curtos e/ou com pausas e/ou manca com apoio do membro; quando parado os membros pélvicos podem estar mais abertos e para trás do que o normal.
Postura em estação alterada	Posição quadrupedal com dorso arqueado e membros pélvicos rígidos e/ou estendidos caudalmente
Cabeça na linha da coluna	Posição da cabeça para a frente
Escoicear e bater com os membros no chão	Escoiceando em estação
Estende a cabeça e pescoço e/ou membro(s)	Esticar a cabeça e pescoço e membros torácicos o pélvicos
Deitado	Deitado sozinho e quieto no canto da baia
Movimento de cauda brusco e repetido	Enérgico balançar de cauda
Mantém a cauda esticada	Cauda esticada (sem estar a urinar ou a defecar)
Arquear o dorso	Arqueamento do dorso
Comportamentos indicativos de dor intensa	
Interação ausente	Apático: está isolado ou não acompanha os outros animais, não se interessa no ambiente.
Locomoção ausente/anormal	Reluta em se levantar ou se levanta com dificuldade; não se locomove ou apresenta andar instável ou rígido e/ou manca com pouco apoio ou sem apoio do membro; pode andar para trás ou andar em círculo ou apoiar-se contra uma superfície; quando parado os membros pélvicos podem estar mais abertos e para trás do que o normal.
Cabeça abaixo da linha da coluna	Cabeça baixa
Deitado com alteração (cabeça rodada sobre si)	Deitado com cabeça rodando para trás
Deitado com a cabeça apoiada/próxima ao solo	Deitado com aparente desconforto, quieto, sozinho no canto da baia
Tremores no corpo	Dorso e abdômen (sobretudo) tremendo

Adaptado da literatura (DWYER; LAWRENCE, 2000; FUTRO; MASŁOWSKA; DWYER, 2015; GRAHAM; KENT; MOLONY, 1997; GRANT, 2004; GUESGEN et al., 2016a; KENT; MOLONY; GRAHAM, 1998; KENT; MOLONY; ROBERTSON, 1995; LESTER et al., 1996; MCGLONE; STOBART, 1986; MELCHES et al., 2007; MELLEMA et al., 2006; MELLOR; MURRAY, 1989; MOLONY et al., 2012; MOLONY; KENT, 1997; MOLONY; KENT; ROBERTSON, 1993; SMALL et al., 2014; THORNTON; WATERMAN-PEARSON, 1999, 2002).

Tabela II - Validação de conteúdo da escala proposta para avaliação da dor aguda pós-operatória em ovinos por 3 avaliadores. De acordo com a experiência de cada avaliador, para cada item da escala tem relevância ou não para avaliar a dor em ovinos foi dada a pontuação com -1= item irrelevante; 0= não sabe ou 1= item relevante.

Variável	Critério	Avaliador 1	Avaliador 2	Avaliador 3	Escore
Interação com o ambiente	Ativo, atento ao ambiente, interage e/ou acompanha os outros animais	1	1	1	1
	Apático: pode permanecer próximo aos outros animais, mas interage pouco	1	1	1	1
	Apático: está isolado ou não acompanha os outros animais, não se interessa no ambiente	1	1	1	1
Ingestão de Alimentos	Normorexia e/ou ruminação presente	1	1	1	1
	Hiporexia	1	0	1	0,66
	Anorexia	1	1	1	1
Atividade	Movimenta-se normalmente	1	1	1	1
	Inquieto, movimenta-se acima do normal; ou deita-se e levanta-se com frequência	1	1	1	1
	Movimenta-se com menor frequência ou somente ao ser estimulado ou não se movimenta	1	1	1	1
Locomoção	Move-se livremente, sem alteração de locomoção; quando parado, os membros pélvicos estão paralelos aos torácicos	1	1	1	1
	Move-se com restrição e/ou com passos curtos e/ou com pausas de mais de 10 segundos e/ou manca com apoio do membro; quando parado os membros pélvicos podem estar mais abertos e para trás do que o normal.	1	1	1	1
	Reluta em se levantar ou se levanta com dificuldade; não se locomove ou apresenta andar instável ou rígido e/ou manca com pouco ou sem apoio do membro; pode andar para trás ou sobre os joelhos; pode saltar como coelho, ou andar em círculo ou apoiar-se contra uma superfície ou cair; quando parado os membros pélvicos podem estar mais abertos e para trás do que o normal.	1	0	1	0,66
Posição da cabeça	Acima da linha da coluna	1	0	1	0,66
	Na linha da coluna	1	0	1	0,66
	Abaixo da linha da coluna	1	0	1	0,66
Atenção ao local da dor	Não olha	1	1	1	1
	Vira a cabeça e olha	1	1	1	1
	Lambe ou tenta lambear ou evita contato da área com superfícies ou outros animais	1	1	1	1
Expressão facial	Orelha direcionada para trás e mais aberta	-1	1	0	0,33
	Olhos semi-fechados	1	1	1	1
	Dobra o(s) lábio(s), incluindo reflexo de flehmen	0	1	1	0,66
Postura em posição quadrupedal	Escoiceia e bate com os membros no chão	-1	1	1	0,33
	Estende a cabeça e pescoço e/ou um ou mais membros	1	0	1	0,66
	Tremores no corpo (sem considerar as orelhas)	-1	1	0	0
Posição em decúbito	Em pé ou em decúbito, relaxado	1	1	1	1
	Estende o pescoço e o corpo para frente	1	1	0	0,66
	Em decúbito ventral a cabeça se apoia no solo ou fica próxima ao solo na linha da coluna	1	1	1	1
Miscelânea de comportamentos	Movimenta a cauda brusca e repetidamente, exceto quando mama ou mantém a cauda esticada, exceto quando defeca ou urina	1	0	1	0,66
	Dorso arqueado	1	1	1	1
	Roda sobre si mesmo parcial ou totalmente, sem se levantar	1	0	0	0

Tabela III - Proposta para escala Unesp-Botucatu para avaliação da dor aguda pós-operatória em ovinos após a validação de conteúdo

Interação com o ambiente	(0) Ativo, atento ao ambiente, interage e/ou acompanha os outros animais. (1) Apático: pode permanecer próximo aos outros animais, mas interage pouco. (2) Apático: está isolado ou não acompanha os outros animais, não se interessa no ambiente.
Atividade	(0) Movimenta-se normalmente. (1) Inquieto, movimenta-se acima do normal; ou deita e levanta com frequência. (2) Movimenta-se com menor frequência ou somente ao ser estimulado ou não se movimenta.
Locomoção	(0) Move-se livremente, sem alteração de locomoção; quando parado, os membros pélvicos estão paralelos aos torácicos. (1) Move-se com restrição e/ou com passos curtos e/ou com pausas e/ou manca com apoio do membro; quando parado os membros pélvicos podem estar mais abertos e para trás do que o normal. (2) Reluta em se levantar ou se levanta com dificuldade; não se locomove ou apresenta andar instável ou rígido e/ou manca com pouco ou sem apoio do membro; pode andar para trás ou sobre os joelhos; pode saltar como coelho, ou andar em círculo ou apoiar-se contra uma superfície ou cair; quando parado os membros pélvicos podem estar mais abertos e para trás do que o normal.
Posição da cabeça	(0) Acima da linha da coluna (1) Na linha da coluna (2) Abaixo da linha da coluna
Postura	a) Escoiceia e bate com os membros no chão b) Em decúbito estende a cabeça e pescoço e/ou um ou mais membros. c) Em decúbito ventral a cabeça se apoia no solo ou fica próxima ao solo na linha da coluna. (0) Ausência destes comportamentos (1) Presença de um dos comportamentos acima (a) e/ou (b) e/ou (c) e/ou (d) (2) Presença de dois ou mais comportamentos acima (a) e/ou (b) e/ou (c)
Ingestão de alimentos	(0) Normorexia e/ou ruminção presente. (1) Hiporexia. (2) Anorexia.
Miscelânea de comportamentos	a) Movimenta a cauda brusca e repetidamente, exceto quando mama ou mantém a cauda esticada, exceto quando defeca ou urina b) Dorso arqueado c) Tremores no corpo (sem considerar as orelhas) d) Roda sobre si mesmo parcial ou totalmente, sem se levantar. (0) Ausência destes comportamentos (1) Presença de um dos comportamentos acima (a) e/ou (b) e/ou (c) e/ou (d) (2) Presença de dois ou mais comportamentos (a) e/ou (b) e/ou (c) e/ou (d)

Tabela IV - Escala proposta para avaliação da dor aguda pós-operatória em ovinos preenchida pelos observadores com 33 variáveis (itens, subitens e subdivisões) para refinamento e redução dos descritores.

Item principal	Sub-itens (com sub-divisões em alguns sub-itens)	Vídeo 1	Vídeo 2	Vídeo 3	Vídeo 4
Interação	A. Ativo, atento ao ambiente, interage e/ou acompanha os outros animais.				
	B. Apático: pode permanecer próximo aos outros animais, mas interage pouco.				
	C. Apático: está isolado ou não acompanha os outros animais, não se interessa no ambiente.				
Atividade	A. Movimenta-se normalmente.				
	B. Inquieto, movimenta-se acima do normal ou deita e levanta com frequência.				
	C. Movimenta-se com menor frequência ou somente ao ser estimulado ou não se movimenta.				
Locomoção	A. Move-se livremente, sem alteração de locomoção; quando parado, os membros pélvicos estão paralelos aos torácicos.				
	B. Move-se com restrição e/ou com passos curtos e/ou com pausas e/ou claudica; quando parado os membros podem estar mais abertos do que o normal.				
	C. 1.1 Reluta em se levantar e/ou				
	1.2 se levanta com dificuldade.				
	2.1 não se locomove e/ou				
	2.2 apresenta andar anormal e/ou				
	2.3 e/ou claudica				
	3.1 pode andar para trás e/ou				
	3.2 andar em círculo e/ou				
	3.3 apoiar-se contra uma superfície.				
Posição da cabeça	A. Acima da linha da coluna.				
	B. Na linha da coluna.				
	C. Abaixo da linha da coluna.				
Postura	1. Escoiceia ou bate com um ou mais membros no chão.				
	2.1. Estende a cabeça e pescoço e/ou				
	2.2. Estende um ou mais membros.				
	3. Se em decúbito ventral a cabeça se apoia no solo ou fica próxima ao solo na linha da coluna.				
Apetite	A1. Normorexia				
	A2. Ruminação presente.				
	B. Hiporexia.				
	C. Anorexia.				
Miscelânea de comportamentos	1.1. Movimenta a cauda brusca e repetidamente				
	1.2. Mantém a cauda esticada, exceto quando defeca ou urina.				
	2. Dorso arqueado.				
	3. Tremores no corpo (sem considerar as orelhas).				
	4. Em decúbito se arrasta, sem se levantar.				

Tabela V. Processo de refinamento para inclusão e exclusão de itens e subitens na escala Unesp - Botucatu para avaliação de dor aguda pós-operatória a ser validada, de acordo com a 2ª fase de avaliação do observador padrão-ouro

Itens / Testes	Soma	Eto. M2	CV	%M2 < 15	ACP	Intra	Inter	F (M1xM2)	F (M2xM3)	F (M2xM4)	CIT	CI	Sp/S
Interação	10		1		1	1	1	1	1	1	1	1	1
Atividade	9		1		1		1	1	1	1	1	1	1
Locomoção	10		1		1	1	1	1	1	1	1	1	1
Locomoção 1	3			1		1	1						
Locomoção 2	3			1		1	1						
Locomoção 3	3			1		1	1						
Locomoção 4	3			1		1	1						
Posição da cabeça	7		1		1	1	1				1	1	1
Postura	2					1					1		
Postura A	2			1			1						
Posture B1	3	1	1	1									
Posture B2	2		1	1									
Posture C	4		1	1		1	1						
Apetite	9		1		1	1	1	1	1	1	1	1	
Miscelânea de comportamentos	4						1	1			1	1	
Miscelânea A	3		1	1			1						
Miscelânea B	3	1	1	1									
Miscelânea C	0												
Miscelânea D	2					1	1						

Testes estatísticos: **Eto** – Etograma em M2; **% M2 < 15** – Porcentagem de ao menos 15% de ocorrência de scores em M2; **VC** – validação de conteúdo; **VC** – validação de conteúdo; **ACP** – Análise de componentes principais. **Intra** – Confiabilidade intraobservador; **Inter** – Confiabilidade interobservador; **F (M)** – teste de Friedman entre momentos; **CIT** – correlação item-total; **CI** – Consistência interna; **E** – Especificidade; **S** – Sensibilidade; **Soma** – somatória de 10 testes estatísticos (teste de Friedman foi dividido em 3).

Locomoção: 1 - Dificuldade e/ou relutância em se levantar; 2 - não se locomove e/ou anda de forma anormal e/ou claudica e/ou quando parado os membros podem estar mais abertos e para trás do que o normal; 3 - apoia-se contra uma superfície; 4 - quando parado os membros podem estar mais abertos e para trás do que o normal. **Postura**: A - Escoiceia ou bate com um ou mais membros no chão; B1 - estende a cabeça e pescoço; B2 - e/ou um ou mais membros; C - Se em decúbito ventral a cabeça se apoia no solo ou fica próxima ao solo na linha da coluna. **Miscelânea de comportamentos**: A - Movimenta a cauda brusca e repetidamente e/ou mantém a cauda esticada; B - Dorso arqueado; C - Tremores no corpo (sem considerar as orelhas); D - Em decúbito se arrasta, sem se levantar.

Momentos - M1: pré-operatório; M2: pós-operatório antes da analgesia de resgate; M3: pós-operatório, após analgesia de resgate; M4: 24h pós-operatório.

nº 1 - foi incluído de acordo com os critérios de cada teste. Itens e subitens incluídos na escala final em negrito. Os principais itens foram submetidos a 12 testes (incluídos no mínimo 7 inclusivé); os subitens foram submetidos a 5 testes (incluídos no mínimo 3 inclusivé).

Research article submitted to Plos One

Validation of the Unesp-Botucatu sheep acute composite pain scale (USAPS)

Nuno Emanuel Oliveira Figueiredo Silva^{1¶}, Pedro Henrique Esteves Trindade^{1¶}, Alice Rodrigues Oliveira^{1&}, Marilda Onghero Taffarel^{3&}, Maria Alice Pires Moreira^{4&}, Renan Denadai^{2&}, Paula Barreto Rocha^{1&}, Stelio Pacca Loureiro Luna^{1¶*}

¹ Department of Anesthesiology, School of Medicine, São Paulo State University (Unesp), Botucatu, São Paulo, Brazil

² Department of Veterinary Surgery and Animal Reproduction, School of Veterinary Medicine and Animal Science, São Paulo State University (Unesp), Botucatu, São Paulo, Brazil.

³ Department of Veterinary Medicine, State University of Maringá, Umuarama, Paraná, Brazil.

⁴ Goiano Federal Institute, Urutaí Campus, Department of Veterinary Medicine, Urutaí, GO, Brazil.

¶These authors contributed equally to this work.

&These authors also contributed equally to this work.

* Corresponding and Senior author

Email: stelio.pacca@unesp.br

Abstract

A scale with robust statistical validation is essential to diagnose pain and guarantee effective analgesia. This blind, randomized, prospective and opportunist study aimed to develop an ethogram to evaluate behavior and validate a scale to assess acute ovine postoperative pain. Elective laparoscopy was performed in 48 healthy sheep, filmed at one preoperative and three postoperative moments, before and after analgesic rescue and 24 hours after. The videos were randomized and assessed twice by four evaluators, with a one-month interval between evaluations. Statistical analysis was performed using R software and differences were considered significant when $p < 0.05$. The intra- and inter-observer reliability ranged from moderate to very good (intraclass correlation coefficient ≥ 0.55). The scale presented Spearman correlations > 0.80 with the numerical, simple descriptive, and visual analogue scales, and a correlation of 0.59 with the facial expression scale. According to the Friedman test, the scale was responsive, due to the increase and decrease in pain scores of all items after surgery and analgesic intervention, respectively. Based on the multiple association, a unidimensional scale was adopted. All items on the scale demonstrated an acceptable Spearman item-total correlation (0.3-0.7). The internal consistency was excellent (Cronbach's $\alpha = 0.84$) and all items presented specificity > 0.70 and sensitivity between 0.67-0.85, except for appetite. According to the Youden index, the cut-off point was ≥ 4 out of 12, with a diagnostic uncertainty zone of 2.5 to 3.5, which indicates the accuracy of the scale. The area under the curve > 0.95 for all evaluators demonstrated excellent discriminatory capacity of the instrument. Total scores of the scale were classified as mild (0-3), moderate (4-8) and severe (≥ 9) pain. In

conclusion, the Unesp-Botucatu sheep acute composite pain scale (USAPS) is valid, reliable, specific, sensitive, with excellent internal consistency, accuracy, discriminatory capacity, and a defined cut-off point.

Introduction

The lack of valid and reliable instruments to recognize and quantify pain in farm animals compromises their welfare state and limits the use of analgesics in these species [1–4]. The conviction that these animals feel less pain compared to small animals [5], economic reasons [6,7], the need for withholding periods for the consumption of milk and meat due to residues of the drugs, and the absence of licensed analgesics are other factors that make it difficult to treat pain in farm animals [3].

Sheep are subjected to several painful procedures, often without appropriate use of anesthesia or analgesia, such as castration, tail cutting, mulesing, and ear marking [8]. Pain in sheep also occurs from diseases such as mastitis, pododermatitis, abscesses in the feet, and external myiasis [9–11], which can lead to chronic and neuropathic pain [12–14].

The sheep species is suitable as an experimental model due to its similarity in size and weight to humans. In 2014, 60,209 sheep were used in research in the European Union, an increase of 108% compared to 2011, only surpassed by pigs among farm animals. Sheep and pigs are more commonly used than dogs and non-human primates as non-rodent models for research and teaching, as their limitations to using these other species as models, including ethical issues [15,16]. Among the experimental models that include sheep are preclinical and translational studies on osteoporosis [17] and bone regeneration and osteointegration of dental implants [18].

To treat pain, instruments are needed to assess it. Although there are several useful indicators to assess nociception in experimental situations, such as the injection of formalin into the interdigital space [19], von Frey filaments [20], tourniquets [21], electrical stimuli [22], and pneumo-mechanical stimulus in limbs [23], these are not reproducible and are difficult to use in clinical situations. Actigraphy can be used to monitor sheep from a distance, which excludes the presence of the observer, however, this method requires specific equipment [24]. Other physiological measures, such as hair cortisone concentration [25], heart rate variability [26], blood pressure, ocular and rectal temperature, electromyography, and electroencephalography, are not clinically feasible and require the animal to be restrained [27].

The behavioral expression of pain replaces the absence of verbal expression of the animals. Behavior is easy to observe, does not require restraint of the animal in field situations, and, therefore, does not generate stress, does not require equipment, and has no cost, thus representing the most applicable method in both clinical or experimental situations [28–30].

In contrast to cattle and goats, sheep, when suffering pain under restraint, tend to remain more silent [28] and only express pain behaviors when released [31]. These behaviors are: reduced interaction with the environment and with other animals, gait abnormalities, lameness, relaxation of the pelvic limbs (which move slowly and alternately when the animals are stationary), turning of the head, hyporexia, abnormal vocalization, lip-licking movement, curved lips, gnashing of teeth, tremors, stamping feet on the ground, and strong tail wagging [8,28,32,33].

Two analyzes are essential to develop and validate a pain scale: validity, which indicates whether the instrument effectively measures the attribute for which it was designed [34–36], and reliability, which guarantees equivalence of results when the measure is evaluated by the same observer on different occasions, or by different observers on the same occasion [36–38]. The scale must also be responsive, meaning its scores change according to a painful stimulus or analgesia [39–41]. Behavior-based pain scales have been developed in dogs [34,42,43],[44] cats [41,45–47], horses [38,40,48–50], cattle [29], and pigs [30]. The most commonly used scales to measure postoperative pain in sheep are still unidimensional, such as the numerical (NS), simple descriptive (SDS), and visual analogue scales (VAS). However, these instruments exclusively evaluate the intensity of pain [39], whereas multidimensional or composite scales include sensory, motor and emotional qualities [45,47].

To develop the scales, a species-specific ethogram is produced to quantifies the duration and/or frequency of the behaviors present before and after a painful stimulus [51]. Although previous studies have reported the behavioral descriptors of pain based on an ethogram after a nociceptive stimulus [8, 28,31,32,52–66], according to our knowledge, there are no validated behavioral scales in the literature to detect acute pain in sheep, following solid statistical analysis [67,68]. The instruments developed to evaluate acute pain in sheep were based on behavioral changes in lambs submitted to orchiectomy and tail cutting [28] or facial expression in sheep with pododermatitis and mastitis [69]. Another facial scale (sheep grimace) was published after the beginning of our research [70]. None of these studies evaluated the criteria, content, and construct validity. To guarantee the reliable measurement of pain, it is necessary to develop an instrument with in-depth statistical validation, such as that performed in in cats [47], cattle [29] horses [38], and pigs [30], which used a blind and random methodology, with evidence of validity, reliability, sensitivity, specificity, and a defined analgesic intervention point [71].

Given the hypothesis that the scale proposed in the current study presents reliability, and content, construct and criterion validities, the main objective of this study was to validate a behavioral scale to assess acute pain in sheep undergoing soft tissue surgery (laparoscopy), constructed from the literature and an ethogram, followed by refinement and subsequent validation, with definition of the cut-off point for analgesic intervention.

Material and methods

This was a blind, randomized, prospective and opportunist study. The study was approved by the Ethics Committee on Animal Use from the School of Veterinary Medicine and Animal Science, São Paulo State University (Unesp), Botucatu, São Paulo, Brazil, under protocol 0027/2017 and followed the recommendations of ARRIVE [72]. In total, 48 sheep (*Ovis aries* species, dairy line) were used, from the institution itself, of the breeds Bergamácia, Lacaune, and Dorper, 3.5 ± 1.8 (1.5 - 6) years old and weighing 58.5 ± 17.3 (34 - 92) kg. As inclusion criteria, the sheep were required to be considered healthy through clinical and laboratory evaluation (hematocrit, plasma protein, glucose and lactate). The sheep were placed in stalls, close to the pen they lived in and where they were used to stay like a shelter when atmospheric conditions were extrem, 24 hours before the start of the study, during which they fasted for feed, and for 12 hours they fasted for water. In each stall (3 x 2 x 1.10 m or 2.20 x 2 x 1.20 m - length x width x height) 6 to 8 sheep or 2 to 4 sheep were housed, respectively.

For the pilot study, sheep were filmed and evaluated before and after laparoscopy, to test the position of the cameras and other adjustments, in order to guarantee the quality of filming. For the main study, the following moments were selected: M1 - one hour before surgery; M2 - at the predicted time of greatest pain, between three and four hours after the end of surgery; M3 - one hour after analgesic intervention; and M4 - 24 hours after surgery.

Immediately before surgery, 30,000 IU/kg of benzathine penicillin (Pentabiótico®, Zoetis, São Paulo, SP, Brazil) was administered intramuscularly (IM). After dissociative anesthesia with 0.5 mg/kg of diazepam (Compaz®, Cristália, Itapira, SP, Brazil) and 5 mg/kg of ketamine (Cetamin®, Syntec; Santana de Parnaíba, SP, Brazil) administered intravenously (IV), lumbosacral epidural anesthesia was performed with 0.1 ml/kg of 1% lidocaine without vasoconstrictor (Xylestesin®, Cristália, Itapira, SP, Brazil) and anesthetic infiltration with 2% lidocaine without vasoconstrictor (Xylestesin®, Cristália, Itapira, SP, Brazil) at the incision site and subsequent introduction of a trocar. When the animals demonstrated any sympathetic response related to the surgery, characterized by an increase of more than 20% in heart rate in relation to the value observed before the beginning of the surgery, dissociative anesthesia was supplemented with 5 mg/kg of ketamine IV.

In all animals, the same experienced surgeon performed a laparoscopy for follicular aspiration and replacement of follicular cells [73–75], by inserting three trocars (5mm) in three retro-umbilical regions. The postoperative analgesic intervention was performed after the M2 evaluation, with 0.5 mg/kg meloxicam 2% (Maxicam®, Ourofino, Cravinhos, SP, Brazil) and 0.2 mg/kg morphine (Dimorf®, Cristália, Itapira, SP, Brazil) IV in separate syringes (Fig 1).

Moments Filming	M1 1h before surgery	Anesthesia (Diazepam + Ketamine IV) Local anesthesia: Lidocaine (epidural)	Surgery	M2 3-4h after surgery	Rescue Analgesia (Meloxicam + Morphine IV)	M3 1h after rescue analgesia	M4 24 h after surgery
---------------------------	--------------------------------	---	----------------	---------------------------------	--	--	---------------------------------

Fig 1. Timeline of moments for validation of the Unesp-Botucatu sheep acute composite pain scale (USAPS).

Data collection

Three to four animals underwent surgery per day. The procedures started at 9 am and the evaluations of the last animals ended around 7 pm. The study was carried out in the months of April and May 2017 and the mean temperature and humidity of the environment varied between 16 - 24°C and 68 - 92 %, respectively. The local had the following geographic coordinates: latitude - 22°51' S; longitude - 48°26' O; altitude - 818 m. The presential observer made the recordings using a digital camera (GoPro Hero5 Black®) positioned on a tripod. The camera was turned on and the presential researcher left the place, and stayed at least 10 m in order to minimize human interference in the behavior of the sheep.

To analyze the pain-related behavior in sheep, the research was divided into the following phases: 1) elaboration of an ethogram to characterize the behavior of the animals before and after the painful procedure (S1 Table); 2) content validation of the normal and pain behaviors based on previous studies, the pilot study, and the ethogram [8,28,31,32,52–66] (S2 Table); 3) production of a pre-refinement scale (S3 Table - scale 1), used to evaluate the videos, which were reevaluated after 30 days, by four observers who were blinded regarding the moments; 4) statistical analysis of the pre-refinement scale (S3 Table) evaluated by the observers according to the criteria in Table 1; 5) refinement criteria applied the scale (S4 Table), based on the statistical analysis of Table 1; 6) validation of the final scale (scale 2) after refinement (Table 3); 7) comparison of the final scale with the facial scale and the three unidimensional scales (Fig 2).

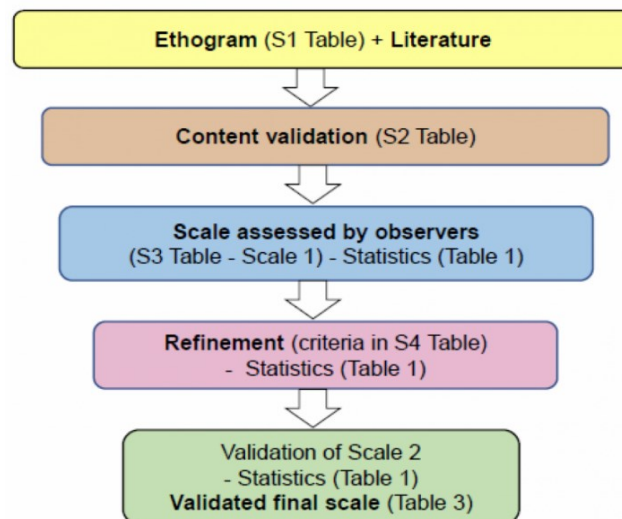


Fig 2. Flowchart with the stages of elaboration, refinement, and validation of the USAPS. Statistical tests (table 1) used in refinement and final validation of the scale: 1) content validation (only in refinement); 2) multiple association; 3) intra-observer reliability; 4) inter-observer reliability; 5) criterion validity; 6) construct validity; 7) internal consistency; 8) sensitivity and specificity; 9) determination of the analgesic rescue point (only in refinement).

Ethogram

To elaborate the ethogram, the presential researcher evaluated, twice, 20-minute videos of each of the four moments (48 animals x 4 = 192 videos), for a total of 64 hours of filming, through recordings continuously evaluated using the focal animal method [76]. These assessments aimed to: 1) recognize the different behaviors and 2) measure the length of time (in seconds) of each behavior presented in S1 Table and its respective percentage of time in relation to the total recording time of 20 minutes. Next the videos were edited with the inclusion of the predominant behaviors for the period of about three minutes at each moment. The edited videos were evaluated by four observers for the scale validation process.

Pain scale

The videos were made available on a virtual platform to the observers, who evaluated the four moments of each animal, in a blind and randomized manner regarding the moments. At the end of the observation of each video, the observers, based on their clinical experience, answered whether or not they would administer rescue analgesia (0 = no and 1 = yes to do rescue analgesia). These data were used to determine the cut-off point related to the need for analgesic intervention. Next, pain scores were determined using three unidimensional scales (NS, SDS, and VAS). The NS ranges from "0" to "10", where "0" represents no pain and "10" the worst possible or imaginable pain; the SDS ranges from 1 - no pain, 2 - mild pain, 3 - moderate pain, and 4 - severe pain; and the VAS is based on a straight line 100 mm long, where "0" represents the animal without pain and "100" the worst possible pain [34,42,43].

As some sub-items in S3 Table used to evaluate the videos, contained a lot of grouped

descriptors (example: “locomotion”), they were subdivided for individual assessment of the importance of each one. The scale assessed by the four evaluators was based on a scale composed of seven items - five with three sub-items and two with four sub-items. In addition, some of the sub-items had several sub-divisions, which totaled 33 behavioral variables. Finally, the facial pain scale in sheep [69], was applied, through the photographic record captured by the presential observer at the end of the 20-minute recording of each moment.

For the refinement and validation of the acute pain scale and facial scale, the gold standard observer was selected through that which presented greater intra-observer reliability and greater consistency when indicating less rescue analgesics at the moment without pain (M1) and more rescue analgesics at the moment with pain (M2). For refinement, all items and sub-items of the pre-refinement scale (S3 Table) were evaluated according to the statistical analysis of Table 1, and refinement criteria described in S4 Table.

Statistical analysis

Statistical analyzes were performed using R software in the RStudio integrated development environment [77]. For all analyzes, an α of 5% was considered. The sample size was estimated in 13 with a 0.90 test power, $\alpha = 0.0001$; for the chi-square test (<http://biomath.info/power/chsq.htm>). For this estimate, the effect size was calculated, applying the percentage of need for rescue analgesia at M1 (without pain; 8%) and M2 (in pain; 92%) according to the data of the 2nd assessment of the gold standard evaluator (Table 13).

For the analysis of the ethogram, the time spent in seconds and the percentages of duration of each behavioral category between the four moments were compared using the Friedman test [29,47]. Table 1 presents the guidelines for the refinement and statistical validation process of the proposed scale to quantify acute pain in, with a description of the types of analysis performed and the respective statistical tests.

Pre-refinement scale (S3 Table) was submitted to refinement according to Table 1. To refine the validated final scale, the subsequent inclusion and exclusion criteria of items and sub-items were followed according to 10 statistical tests, based on data from the second phase of the gold standard observer assessment (S4 Table): 1) ethogram (highest score of behavior by the Friedman test at M2 vs the other moments); 2) content validation; 3) at least 15% frequency of occurrence of items/sub-items at M2; 4) multiple association (principal component analysis); 5) intra-observer reliability; 6) inter-observer reliability; 7) construct validity (Friedman's tests - highest score of item/sub-item at M2 vs the other three moments); 8) item-total correlation; 9) internal consistency; 10) specificity and sensitivity. The items and sub-items were submitted to these analyses according to the criteria of Table 1. The variables that met the criteria stipulated in more than 50% of these statistical tests were accepted. Data supporting the results are available in supplementary material.

Table 1. Guidelines for statistical refinement (^R) and validation (^V) of the Unesp-Botucatu sheep acute pain scale (USAPS).

Type of analysis	Description	Statistical test
Content validation^R	The following steps were performed: 1) a list of pain-related behaviors reported in the literature and 2) behaviors observed in the ethogram were scored by 3) a committee composed of three experienced veterinarians used to assess pain in ruminants who analyzed each sub-item within each item of the scale into relevant (+1), do not know (0), or irrelevant (-1)	All the values of each sub-item (-1, 0, or 1) were added and the total was divided by the number of observers. Items with a total score > 0.5 were included in the scale [78].
Intra-observer reliability^{RV}	Repeatability - the level of agreement of each observer with themselves was estimated by comparing the two phases of assessment, using the scores of each item, the total sum of the USAPS, numerical (NS), simple descriptive (SDS), and visual analogue scales (VAS) as well as the need for analgesic rescue.	For the scores of the items of the USAPS and the NS and SDS, and the need for analgesic rescue, the weighted kappa coefficient (k_w) was used; the disagreements were weighted according to their distance to the square of perfect agreement. The 95% confidence interval (CI) k_w ("cohen.kappa" function of the "psych" package) was estimated. For the VAS, the intraclass correlation coefficient (ICC) type "agreement" was used and its 95% CI ("icc" function of the "irr" package) [79–81]. For the sum of the behavioral scale , the consistency type ICC and its 95% CI were used. Interpretation of k_w and ICC: very good 0.81 - 1.0; good: 0.61 - 0.80; moderate: 0.41 - 0.60; reasonable: 0.21 - 0.4; poor < 0.2. The k_w and ICC > 0.50 was used as a criterion to refine the scale [47,82,84].
Inter-observer reliability^{RV}	1) Reproducibility – the level of agreement between the gold standard observer and the other three observers in the 2nd evaluation phase was estimated, using the scores for each item; the total sum of the behavioral scale, NS, SDS, VAS, and the need for rescue analgesic. 2) Agreement matrix - a matrix was generated to assess the agreement of the total score of the USAPS in the 2nd evaluation phase of each observer versus all other observers.	
Criterion validity^{RV}	1) Concurrent criterion validity (relationship with gold standard method): the correlation of the sum of the behavioral scale was estimated with the NS, SDS, VAS, and facial expression scale of all grouped moments scored by the gold standard observer in the 2nd evaluation phase. 2) Concurrent criterion validity: agreement between the gold standard and the other evaluators (reproducibility) 3) Predictive criterion validity was assessed by the number of sheep that should receive rescue analgesia according to the Youden Index (described below) in the moment of greatest pain (M2).	Spearman rank correlation coefficient (rs; "rcorr" function of the "Hmisc" package). Interpretation of the degree of correlation rs (p<0.05): 0 - 0.35: low correlation; 0.35 - 0.7: moderate correlation; 0.7 - 1.0: high correlation [47]. Please see above the item Inter-observer reliability: 1) Reproducibility. Descriptive analysis.
Construct validity^{RV}	Responsiveness – the behavior from the ethogram observed by the presential researcher at the four moments of evaluation was compared; the scores of each item and the total score of the behavioral scale, NS, SDS, VAS, and the need for rescue analgesic over time (M1 vs M2 vs M3 vs M4) were compared according to the 2nd phase of the gold standard observer assessment. Interpretation: differences in scores are expected to occur as follows: M2>M4≥M3>M1. Construct validity was determined using the three hypothesis test method: 1) if the scale really measures pain, the score after surgery (M2) should be higher than the preoperative score (M1 vs M2); 2) the score should decrease after analgesia (M2 vs M3); 3) and over time (M2 vs M4).	For dichotomous variables (need for rescue analgesic) logistic regression analysis ("glm" function of the "stats" package) was applied using the Tukey test ("lsmeans" function of the "lsmeans" package) as a post hoc test. The normality of each variable at each moment was assessed by graphs of boxes and histograms ("boxplot" and "histogram" functions of the "graphics" and "lattice" packages, respectively). From these graphs, the pain scale items, NS, SDS, and VAS were considered nonparametric. Thus, we used the Friedman test (function "friedman.test" of the package "stats") in which the p-value was corrected with the Bonferroni procedure (function "pairwiseSignTest" of the package "rcompanion") [29,47]. For the total score of the scale , a mixed linear model ("lmer" function of the "lme4" package) was used, including the moments (M1, M2, M3 and M4) and breeds as fixed effects and the random effect of individuals, in which the p-value was also corrected by the Bonferroni procedure. The differences between the

		moments were identified by different letters ("cldList" function of the "rcompanion" package).
Multiple association^{RV}	The multiple association of the items with each other was analyzed at all moments grouped on the behavioral scale in the 2nd phase of the gold standard observer assessment, through the analysis of main components, to define the number of dimensions-determined by different variables that establish the scale extension.	Principal component analysis ("princomp" and "get_pca_var" functions from the "stats" and "factoextra" packages respectively). According to the Kaiser criterion [83], representative dimensions of the components were selected with eigenvalue > 1 and variance > 20 and each item on the scale with a load value ≥ 0.50 or ≤ - 0.50.
Item-total correlation^{RV}	Correlation coefficient of item score with total score. To analyze homogeneity, the relevance of each item on the scale, and to investigate inflationary items, the correlation of each item with the sum of all other items on the scale, by excluding the evaluated item, was estimated). Analysis was performed for all moments grouped (MG) in the 2nd phase of the gold standard observer assessment.	Spearman rank correlation coefficient (r_s ; "rcorr" function of the "Hmisc" package). Interpretation of correlation r_s : suitable values 0.3 - 0.7 [84]. Items were accepted if >0,3.
Internal consistency^{RV}	The consistency (interrelation) of the scores for each item on the scale was estimated and the process was performed again to exclude each item. The analysis was performed for each of the 4 moments and for all of them grouped (MG) in the 2nd phase of the gold standard observer assessment.	Cronbach's alpha coefficient (α ; "cronbach" function of the "psy" package) [84] Interpretation: 0.60-0.64, minimally acceptable; 0.65-0.69 acceptable; 0.70-0.74 good; 0.75-0.80 very good; and > 0.80 excellent [84–86].
Specificity^{RV}	The scores of the behavioral scale in the 2nd phase of assessment of the gold standard observer at M1 were transformed into dichotomous (level "0" - absence of pain expression behavior for a given item; levels "1" and "2" - presence of pain expression behavior) and applied to the respective equation.	$Sp_{M1} = \frac{TN}{TN + FP}$ Sp = specificity. TN = true negative (scores that represented painless behaviors - "0" - at the time when the animals were expected to have no pain, since it was before surgery - M1). FP = false positive (scores that represented pain expression behaviors - 1 or 2 - at the time when the animals were expected to have no pain, since it was before surgery - M1). Interpretation: excellent 95 - 100%; good 85 - 94.9%; moderate 70 - 84.9%; not sensitive <70%. Only items ≥ 70% were included pain [84].
Sensitivity^{RV}	The scores of the behavioral scale in the 2nd phase of assessment of the gold standard observer at M2 were transformed into dichotomous (as described for specificity) and applied to the respective equation.	$S_{M2} = \frac{TP}{TP + FN}$ S = sensitivity. TP = true positive (scores that represented pain expression behaviors - 1 or 2 - at the time the animals were expected to have pain, since it was after surgery - M2). FN = false negative (scores representing painless behaviors - 0 - at the time the animals were expected to have pain, since it was after surgery - M2). Interpretation: excellent 95 - 100%; good 85 - 94.9%; moderate 70 - 84.9%; not sensitive <70%. Only items ≥ 70% were included [40,84].
Distribution of scores^V	A table was constructed with the distribution of the frequency of the presence of the scores 0, 1, and 2 of each item at each moment and in the MG scored in the 2nd phase of the gold standard observer assessment.	Descriptive analysis.
Rescue analgesic point^V	The need for analgesia according to the clinical experience after the analysis of the videos was used as the true value and the sum of the behavior scale as a predictive value to build a receiver operating characteristic curve (ROC) . Then, the cut-off point for analgesic rescue was determined based on the Youden index and its diagnostic	$YI = (S + Sp) - 1$ YI = Youden Index ; S = sensitivity; Sp = specificity. Analysis of the receiver operating characteristic curve (ROC; "roc" function of the "pROC" package) and the area under the curve (AUC): graphical representation of the relationship between the "TP" (S) and the "FP" (1-Sp). YI is the point of

	uncertainty zone using all moments of pain assessment on the behaviour scale scored by each and all observers in the 2nd evaluation phase. Cut-off point was represented by the Youden index using all moments of pain assessment on the behaviour scale, NS, SDS and VAS scored by all observers in the 2nd evaluation phase. The area under the curve (AUC) was calculated, which indicates the discriminatory capacity of the test. The frequency and percentage of animals scored in the diagnostic uncertainty zone of the cut-off point of all evaluators for the behaviour scale, NS, SDS, VAS were calculated using descriptive statistical analysis.	greatest sensitivity and specificity simultaneously, determined by the ROC curve [87]. Interpretation: AUC $\geq$ 0.95 - high discriminatory capacity of the scale [88]. The diagnostic uncertainty zone was determined by two methods, calculating: 1st) the 95% confidence interval (CI) replicating the original ROC curve 1,001 times by the bootstrap method ("ci.coords" and "ci.auc" functions of "pROC" package); 2) the interval between the sensitivity and specificity values of 0.90. The highest interval of one these two methods of all the evaluators were considered the diagnostic uncertainty zone, which indicates the diagnostic accuracy [89–92].
Determination of pain intensity ^v	Scores were classified by intensity: low, intermediate or high, in 2nd evaluation of all evaluators at the time of greatest pain (M2).	Non-hierarchical cluster analysis was employed, applying the "centroid" by using the "maximum" method ("dist" and "hclust" functions in the "stats" package) [93] followed by the Kruskal-Wallis test ("kruskal" function in the "agricolae" package ") to assess the difference between the groups formed by the cluster analysis.

Scales: numerical (NS), simple descriptive (SDS), visual analogue (VAS). Statistical analyzes were performed using R software in the RStudio integrated development environment [77]. For all analyzes, α of 5% was considered, MG - data of grouped moments (M1 + M2 + M3 + M4).

Results

A minimum sample size of 5 sheep [67] was estimated.

Behaviour data

With respect to the ethogram, from the description of the behaviors (S1 Table) and the time and percentages of each observed behavior (Table 2), comparing the moment of greatest pain (M2) with that of the supposed absence of pain (M1), the following differences were observed: at M2 there was a decrease in times of “normal interaction”, “normal locomotion”, and “head above the withers” and an increase in times of “reduced and absent locomotion” and “arch back”. After the analgesic rescue (M3), compared to M2, the times of “eat” and normal “interaction” increased, and “normal locomotion” and “head below the withers” decreased. The act of “eat” increased at M3, in relation to M2, because the animals were fasting at M1 and therefore it was not possible to compare M2 vs M1 for this variable.

Pain scale data

According to the inclusion/exclusion criteria cited for the refinement (S4 Table), the following sub-items were excluded: in the item “locomotion”, “walks backwards” and “walks in a circle”; in the item “posture”, “kicks and stamps limbs on the ground” and “extends one or more limbs”; in the item “miscellaneous behavior”, “body tremors” and “crawls in ventral recumbence, without getting up” were excluded. The two sub-items that remained in the “miscellaneous” item replaced those excluded from the “posture” item. The “miscellaneous” item continued with four sub-items and was renamed “posture”.

Next, the final version of the USAPS, containing six items (five with three sub-items and one with four sub-items) was validated (Table 3). The total score was based on the sum of each item, ranging from zero (without pain) to 12 (maximum pain).

Table 2. Median and range of the percentage of duration of behaviors of 48 sheep before and after laparoscopy.

The percentages of each behavioral category were calculated based on the total time of each evaluation moment (20 mins). Different letters express significant differences between moments (values in bold express differences at M2 compared to M1, or M3, specifically for “eating”) with $a>b>c$, according to the Friedman test

Moments	M1		M2		M3		M4	
Behavior category	Median	Range	Median	Range	Median	Range	Median	Range
Eat	0^c	0 – 0	11^b	0 – 82	31.25^a	0 – 86	34^{ab}	0 – 71
Ruminate	0^b	0 – 24	0^{ab}	0 – 41	4.04^a	0 – 20	4^a	0 – 44
Drink	0	0 – 7	0	0 – 26	0	0 – 16	0	0 – 2
Urinate	0	0 – 9	0	0 – 12	0	0 – 21	0	0 – 5
Defecate	0	0 – 0	0	0 – 19	0	0 – 0	0	0 – 0
Normal interaction	44^a	0 – 100	0^b	0 – 91	7.15^a	0 – 95	61^a	0 – 97
Reduced interaction	0 ^a	0 – 97	8 ^a	0 – 94	0 ^a	0 – 100	0 ^b	0 – 97
Absent interaction	0	0 – 100	0	0 – 100	0	0 – 30	0	0 – 88
Normal locomotion	23^a	0 – 57	0^b	0 – 68	0^c	0 – 47	14^a	0 – 50
Reduced/altered locomotion	0 ^b	0 – 21	0 ^a	0 – 25	0 ^b	0 – 40	0 ^a	0 – 47
Absent/abnormal locomotion	0 ^b	0 – 70	2 ^a	0 – 100	0 ^{ab}	0 – 60	0 ^b	0 – 30
Head above the withers	11^a	0 – 80	0^b	0 – 50	0^b	0 – 67	21^a	0 – 86
Head at the height of the withers	0	0 – 68	0	0 – 78	0	0 – 87	0	0 – 81
Head below the withers	10 ^a	0 – 70	13 ^a	0 – 81	6.46 ^b	0 – 60	5 ^b	0 – 25
Standing still in normal posture	61 ^b	33 – 89	67 ^{ab}	0 – 94	73.87 ^a	0 – 100	74 ^{ab}	10 – 92
Standing in altered posture	5 ^a	0 – 31	6 ^a	0 – 37	1.04 ^{ab}	0 – 60	0 ^b	0 – 8
Kick and stamp the limbs on the ground	0 ^b	0 – 0	0 ^{ab}	0 – 5	0 ^b	0 – 0	0 ^a	0 – 14
Lying down with extension of the head and neck and/or limb(s)	0 ^{ab}	0 – 31	0 ^a	0 – 38	0 ^{ab}	0 – 60	0 ^b	0 – 11
Lying down	0	0 – 31	0	0 – 90	0	0 – 35	0	0 – 44
Lying down with head turned back	0	0 – 0	0	0 – 0	0	0 – 0	0	0 – 0
Lying with head supported on or close to the ground	0	0 – 28	0	0 – 100	0	0 – 32	0	0 – 0
Look at affected area	0	0 – 0	0	0 – 0	0	0 – 0	0	0 – 0
Lick the affected area	0	0 – 0	0	0 – 0	0	0 – 0	0	0 – 0
Quick and repeated tail movements	0	0 – 40	0	0 – 0	0	0 – 0	0	0 – 0
Keep the tail straight	0	0 – 22	0	0 – 6	0	0 – 0	0	0 – 50
Arch the back	0^b	0 – 0	0^a	0 – 34	0^{ab}	0 – 13	0^{ab}	0 – 7
Body tremors	0	0 – 0	0	0 – 11	0	0 – 0	0	0 – 0

($p < 0.05$) [29,47]. M1: preoperative; M2 - postoperative, before analgesic rescue; M3 - postoperative, after analgesic rescue; M4 - 24h after surgery.

Table 3. Final validated Unesp-Botucatu sheep acute pain scale (USAPS).

Item	Sub-item (descriptors)	Score	Links to videos
Interaction	Active, attentive to the environment, interacts and/or follows other animals.	0	https://www.youtube.com/watch?v=4fOJWD-uNbg&t=9s
	Apathetic: may remain close to other animals, but interacts little.	1	https://www.youtube.com/watch?v=EEyMC_VIMpk
	Very apathetic: isolated or not interacting with other animals, not interested in the environment.	2	https://www.youtube.com/watch?v=5NsthhKoEP4
Activity	Moves normally	0	https://www.youtube.com/watch?v=dDx9FesiA2M
	Restless, moves more than normal or lies down and gets up frequently.	1	https://www.youtube.com/watch?v=3MjccV2yV74
	Moves less frequently or only when stimulated or does not move.	2	https://www.youtube.com/watch?v=EvLDBJo93jo
Locomotion	Moves about freely, without altered locomotion; when stopped, the pelvic limbs are parallel to the thoracic limbs.	0	https://www.youtube.com/watch?v=W0Hw2lbqbyk
	Moves about with restriction and/or short steps and/or pauses and/or lameness; when stopped, the thoracic or pelvic limbs may be more open and further back than normal.	1	https://www.youtube.com/watch?v=i8FxBj-yQhw
	Difficulty and/or reluctant to get up and/or not moving and/or walking abnormally and/or limping; leans against a surface.	2	https://www.youtube.com/watch?v=dPdT9VMJTio
Head position	Above the withers or eating.	0	https://www.youtube.com/watch?v=W8mi15l1dr8
	At the height of the withers.	1	https://www.youtube.com/watch?v=8xSUmoXaiZY
	Below the withers.	2	https://www.youtube.com/watch?v=YRxpWSTsgpw
Appetite	Normorexia and/or rumination present.	0	https://www.youtube.com/watch?v=no1VeiFglUE
	Hyporexia.	1	https://www.youtube.com/watch?v=aIEY1UkqQ-k
	Anorexia.	2	https://www.youtube.com/watch?v=YV40N-OHuNI
Posture	A. Arched back.		https://www.youtube.com/watch?v=gloa-38gTW8
	B. Extends the head and neck.		https://www.youtube.com/watch?v=rNh_aFePKAE
	C. Lying down with head resting on the ground or close to the ground		https://www.youtube.com/watch?v=LT6BJzhZO9E
	D. Moves the tail quickly (except when breastfeeding) and repeatedly and/or keeps the tail straight (except to defecate/urinate)		https://www.youtube.com/watch?v=91RbQMsa8Mg
	• Absence of these behaviors	0	
	• Presence of one of the related behaviors	1	
	• Presence of two or more of the related behaviors	2	

Intra-observer reliability

Repeatability ranged from reasonable to good for each item on the USAPS (with the exception of appetite for evaluator 3 which was poor) and from good to very good for their total score at all moments assessed (Table 4).

Table 4. Repeatability of USAPS, unidimensional scales and rescue analgesia indication in sheep.

Evaluator	1			2 (gold)			3			4		
Items	<i>k_w</i>	Min	Max	<i>k_w</i>	Min	Max	<i>k_w</i>	Min	Max	<i>k_w</i>	Min	Max
Interaction	0.64	0.55	0.74	0.66	0.56	0.77	0.52	0.41	0.64	0.65	0.56	0.74
Activity	0.54	0.42	0.66	0.65	0.54	0.76	0.46	0.34	0.58	0.47	0.35	0.60
Locomotion	0.71	0.63	0.80	0.66	0.55	0.77	0.48	0.36	0.61	0.61	0.52	0.71
Head position	0.67	0.56	0.77	0.71	0.62	0.80	0.48	0.35	0.61	0.59	0.48	0.69
Appetite	0.55	0.42	0.69	0.61	0.48	0.73	0.56	0.43	0.69	0.15	0.02	0.28
Posture	0.42	0.27	0.57	0.61	0.48	0.73	0.45	0.28	0.61	0.32	0.23	0.42
RA	0.67	0.56	0.77	0.75	0.65	0.85	0.53	0.41	0.65	0.53	0.40	0.65
NS	0.80	0.74	0.86	0.85	0.80	0.91	0.58	0.48	0.68	0.72	0.64	0.79
SDS	0.78	0.71	0.85	0.77	0.69	0.84	0.61	0.50	0.71	0.67	0.59	0.76
Scales	ICC	CI		ICC	CI		ICC	CI		ICC	CI	
USAPS	0.77	0.70	0.82	0.83	0.78	0.87	0.65	0.57	0.73	0.71	0.64	0.78
VAS	0.80	0.74	0.85	0.81	0.76	0.86	0.56	0.44	0.64	0.71	0.63	0.77

Scales: USAPS - Unesp-Botucatu sheep acute composite pain scale; NS – numerical; SDS - simple descriptive; VAS - visual analogue; RA - Rescue analgesia. Each item of the USAPS, the NS and SDS was calculated with kappa coefficient (*k_w*); Sum of the USAPS and the VAS was calculated using intraclass correlation coefficient (ICC consistency); CI - Confidence interval. Interpretation of the degree of reliability *k_w* or ICC (consistency): very good: 0.81 - 1.0; good: 0.61 - 0.80; moderate: 0.41 - 0.60; reasonable: 0.21 - 0.4; poor <0.2. Bold type corresponds to acceptable values > 50 [47,82,84].

Inter-observer reliability

Evaluator 2 was selected as the gold standard. Reproducibility between the gold standard and the other three observers for each item on the USAPS was moderate, with the exception of one evaluator that ranged from poor to moderate (Table 6).

Table 5. Reproducibility of USAPS, unidimensional scales and rescue analgesia indication between the gold standard and the other observers.

	Evaluator 2 (gold-standard) vs								
Evaluators	1			3			4		
Items	k_w	Min	Max	k_w	Min	Max	k_w	Min	Max
Interaction	0.52	0.42	0.62	0.60	0.49	0.70	0.45	0.35	0.55
Activity	0.58	0.47	0.69	0.55	0.43	0.67	0.50	0.35	0.55
Locomotion	0.56	0.46	0.66	0.48	0.35	0.60	0.49	0.37	0.60
Head position	0.54	0.42	0.65	0.48	0.36	0.61	0.43	0.31	0.54
Appetite	0.57	0.44	0.69	0.63	0.51	0.75	0.18	0.03	0.33
Posture	0.47	0.33	0.61	0.50	0.36	0.64	0.31	0.19	0.42
RA	0.53	0.41	0.64	0.63	0.52	0.74	0.46	0.35	0.57
NS	0.43	0.36	0.50	0.66	0.59	0.73	0.49	0.41	0.56
SDS	0.65	0.57	0.74	0.65	0.56	0.74	0.63	0.55	0.72
Scales	ICC	CI		ICC	CI		ICC	CI	
USAPS	0.72	0.65	0.78	0.76	0.70	0.82	0.63	0.53	0.70
VAS	0.30	0.16	0.42	0.59	0.49	0.68	0.44	0.32	0.55

Scales: USAPS - Unesp-Botucatu sheep acute composite pain scale; NS – numerical; SDS - simple descriptive; VAS - visual analogue; RA - Rescue analgesia. Each item of the USAPS, the NS and SDS was calculated with kappa coefficient (k_w); Sum of the USAPS and the VAS was calculated using intraclass correlation coefficient (ICC consistency); CI - Confidence interval. Interpretation of the degree of reliability k_w or ICC (consistency): very good: 0.81 - 1.0; good: 0.61 - 0.80; moderate: 0.41 - 0.60; reasonable: 0.21 - 0.4; poor <0.2. Bold type corresponds to acceptable values > 50 [47,82,84].

Criterion validity

Concurrent criterion validity

Reproducibility of the total score of USAPS according to matrix correlation was good among all observers (Table 6).

Table 6. Reproducibility matrix of the USAPS.

Evaluator	1		2 (gold)		3	
ICC	ICC	CI	ICC	CI	ICC	CI
1						
2	0.72	0.65 - 0.78				
3	0.71	0.64 - 0.78	0.76	0.70 - 0.82		
4	0.74	0.67 - 0.80	0.63	0.53 - 0.70	0.63	0.54 - 0.71

USAPS - Unesp-Botucatu sheep acute composite pain scale. Intra-class correlation coefficient (ICC consistency). Interpretation of the degree of reliability (ICC consistency): very good: 0.81 - 1.0; good: 0.61 - 0.80; moderate: 0.41 - 0.60; reasonable: 0.21 - 0.4; poor <0.2. Bold type corresponds to acceptable values > 50 [47,82,84].

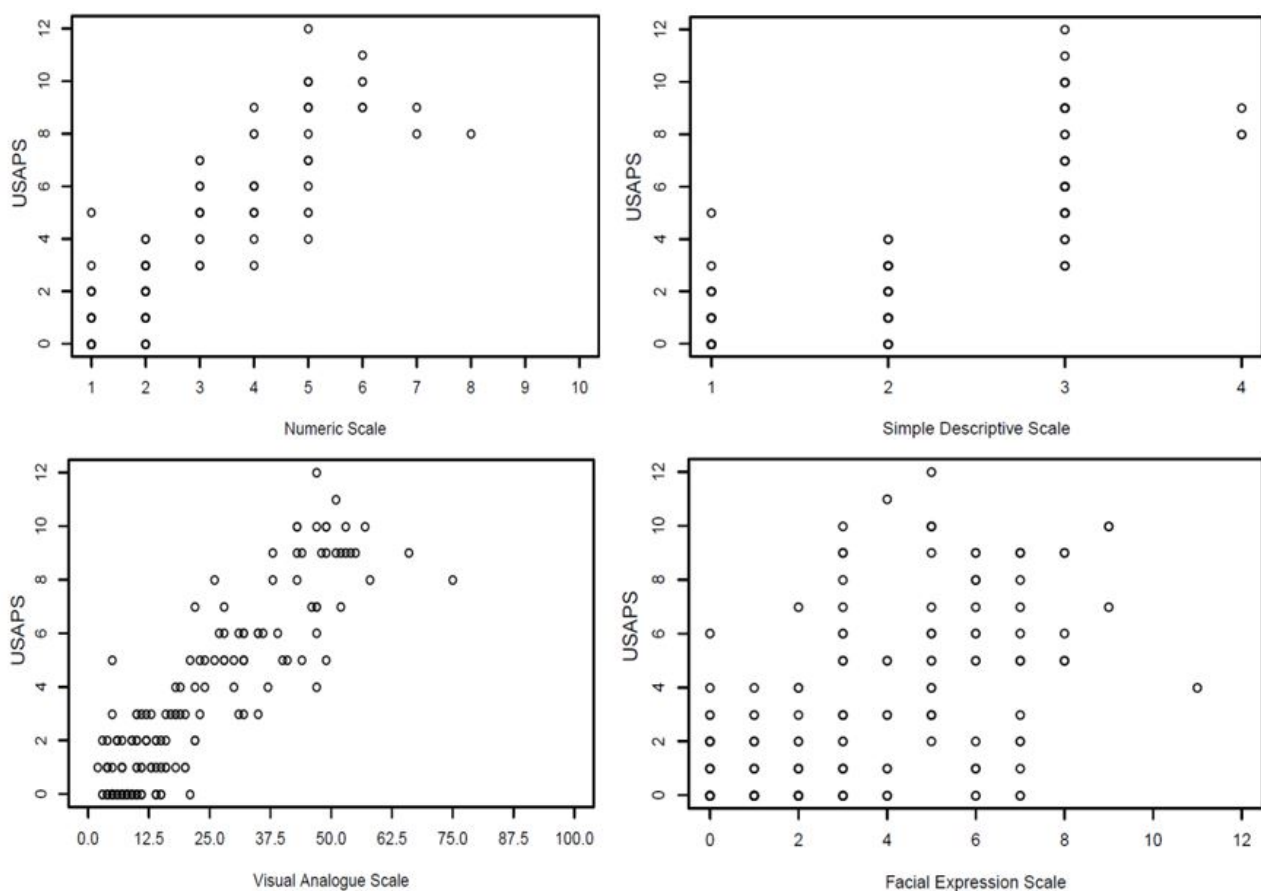
The concurrent criterion validation test registered a high correlation between pain scale scores and those of the NS ($r = 0.87$), SDS ($r = 0.86$), and VAS ($r = 0.81$), and moderate correlation with the facial scale ($r = 0.59$) for all moments (MG) evaluated by the four evaluators (Table 7; Fig

3).

Table 7. Correlation between the USAPS and the numerical, simple descriptive, visual analogue, and facial expression scales.

Scales / USAPS	MG
NS	0.87
SDS	0.86
VAS	0.81
Facial expression scale	0.59

Scales: USAPS - Unesp-Botucatu sheep acute composite pain scale; NS – numerical; SDS - simple descriptive; VAS - visual analogue; Interpretation of Spearman's correlation coefficient ($p < 0.05$): 0 - 0.35: low correlation; 0.35 - 0.7: moderate correlation; 0.7 - 1.0: high correlation [47]. Bold values correlation > 70 . MG - data of grouped moments (M1 + M2 + M3 + M4).

**Fig 3. Correlation between the scores of the USAPS (Unesp-Botucatu sheep acute composite pain scale) and those of the numerical, simple descriptive, visual analogue and facial expression scales for all moments.**

Predictive criterion validity

Considering the predictive criterion validity, grounded by the Youden index, between 83 and 98% of sheep (90% considering all the grouped evaluators) would receive rescue analgesia in the moment of most intense pain (M2) (Table 8). Unnecessary analgesia would be indicated in 8 to 23% of sheep at M1, based on the score given by each evaluator and 15% considering all the grouped evaluators.

This result demonstrated that the scale was sensitive in distinguishing pain, otherwise specific in distinguishing sheep not suffering pain.

There was no difference between indication of rescue analgesia according to clinical experience and according to the Youden index of the USAPS ($p = 0.74$ - Fisher's Exact Test).

Table 8. Percentage of sheep for which was indicated rescue analgesia according to clinical experience and according to the Youden index of the USAPS.

Evaluator	1		2 (gold)		3		4		All	
RA	Exp	YI	Exp	YI	Exp	YI	Exp	YI	Exp	YI
M1	13	13	8	8	23	15	19	25	16	15
M2	90	90	83	83	83	90	98	98	89	90

Calculation based on 48 sheep for each observer and 192 sheep for all evaluators. RA – indication of rescue analgesia according to clinical experience scored at the end of each video analysis (Exp) and according to the Youden index of the USAPS (score ≥ 4). USAPS - UNESP-Botucatu sheep composite acute pain scale. Gold - gold standard observer; M1 - preoperative; M2 - postoperative, before rescue analgesia; Youden index ≥ 4 is representative of the cut-off point for indication of rescue analgesia (see Table 12 for results of Youden index).

Construct validity (responsiveness)

The scores for all items and the sum of the scale were significantly higher at M2 than at M1, M3, and M4, demonstrating their responsiveness (construct validity). The differences between the total scores of the proposed scale, the NS, SDS, and VAS were as follows: $M2 > M3 > M4 > M1$ (Table 9; Fig 4). In the statistical model, the fixed effect of breeds did not have any significant influence on the total score of the final scale.

Table 9. Responsiveness of the USAPS, rescue analgesia and unidimensional pain scales, between the four perioperative moments.

Moments	M1		M2		M3		M4	
Items	Median	Amplit.	Median	Amplit.	Median	Amplit.	Median	Amplit.
Interaction	0 ^{bc}	0 - 2	1 ^a	0 - 2	0 ^b	0 - 2	0 ^c	0 - 1
Activity	0 ^{bc}	0 - 2	2 ^a	0 - 2	0 ^b	0 - 2	0 ^c	0 - 2
Locomotion	0 ^c	0 - 2	1 ^a	0 - 2	0.5 ^b	0 - 2	0 ^c	0 - 1
Head position	0 ^{bc}	0 - 2	1 ^a	0 - 2	0 ^b	0 - 2	0 ^c	0 - 1
Appetite	0 ^{bc}	0 - 0	0.5 ^a	0 - 2	0 ^b	0 - 2	0 ^b	0 - 2
Posture	0 ^{bc}	0 - 1	1 ^a	0 - 2	0 ^b	0 - 2	0 ^b	0 - 1
USAPS	2 ^c	0 - 8	7 ^a	0 - 12	2.5 ^b	0 - 10	0 ^c	0 - 7
RA	0 ^c	0-1	1 ^a	0-1	0 ^b	0-1	0 ^c	0-1
NS	1 ^c	1-4	5 ^a	1-8	2 ^b	1-5	1 ^c	1-5
SDS	1 ^c	1-3	3 ^a	1-3	2 ^b	1-3	1 ^c	1-3
VAS	8.5 ^{bc}	2-38	43 ^a	7-75	15 ^b	4-53	7 ^c	3-47

Each item and total score of the USAPS (0-12): Unesp-Botucatu sheep acute composite pain scale; RA - Rescue analgesia (0 no; 1 yes); NS (1-10), SDS (1-4) and VAS (0-100). Different letters express significant differences between moments where $a > b > c$, according to the Friedman test ($p < 0.05$) [29,47]. M1: preoperative; M2: postoperative, before rescue analgesia; M3: postoperative, after rescue analgesia; M4: 24h postoperative

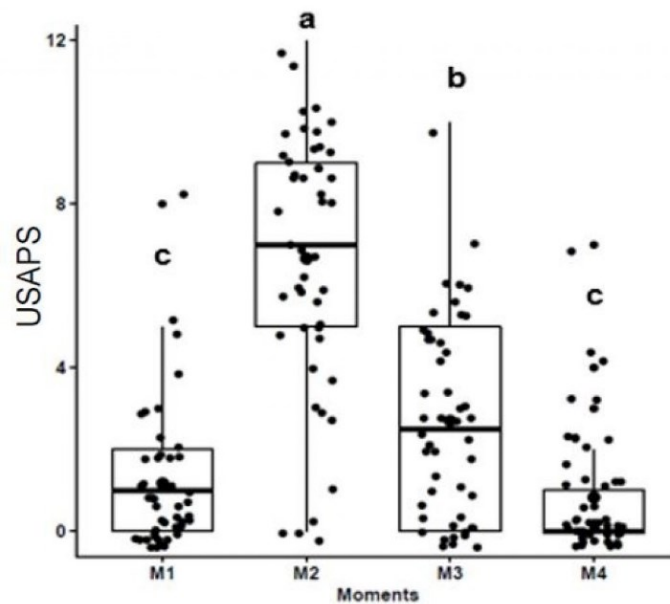


Fig 4. Box-plot of the scores (median/amplitude) of the USAPS, comparing the four perioperative moments.

Principal component analysis

The multiple association of the items of the scale with each other, at all moments, evaluated through principal component analysis selected the main component 1, corresponding to one representative dimension, the mathematical reason why the scale was considered unidimensional (Table 10; Fig 5).

Table 10. Load values, eigenvalues and variance of the USAPS items with each other after principal components analysis.

Items	Dimension 1	Dimension 2
	Load value	Load value
Interaction	0.82	-0.06
Activity	0.80	-0.05
Locomotion	0.85	0.05
Head position	0.73	0.17
Appetite	0.59	-0.70
Posture	0.68	0.50
Eigenvalue	3.39	0.78
Variance	56.39	12.91

USAPS – Unesp-Botucatu sheep acute pain scale. The structure was determined considering items with a load value ≥ 0.50 or ≤ -0.50 , with representative dimension (eigenvalue > 1 and variance $> 20\%$). The load values in bold indicate the variables that contribute to each dimension and the respective accepted eigenvalue and variance [83].

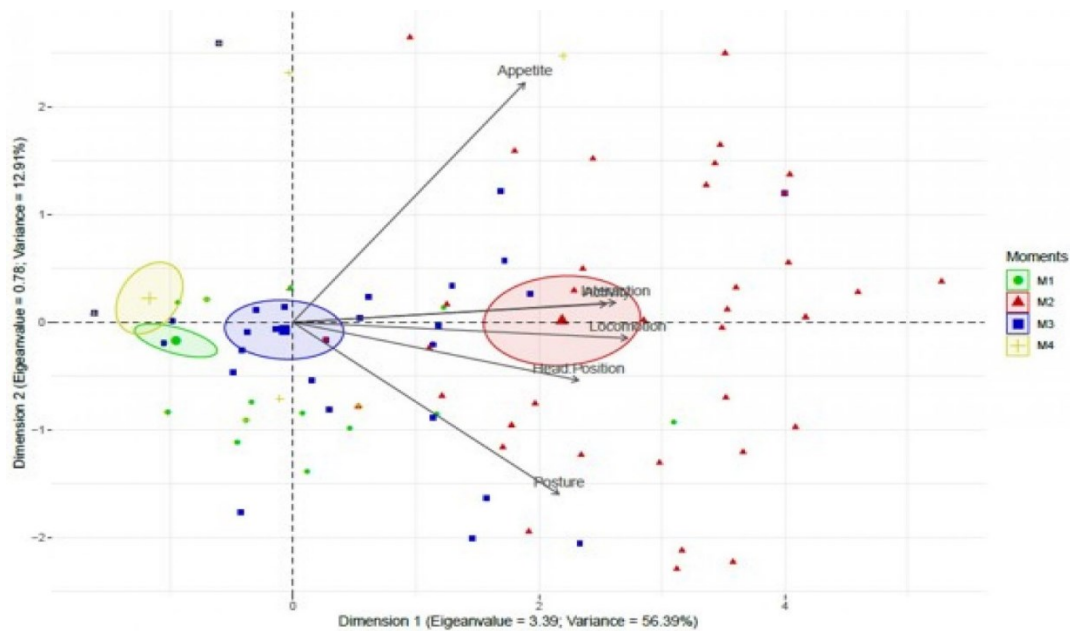


Fig 5. Biplot for the principal components analysis with the items of the USAPS.

USAPS – Unesp-Botucatu sheep acute pain scale. Confidence ellipses were built according to the moments of pain and animal assessment. The data were obtained from the analysis of the gold standard observer at M1 - preoperative; M2 - postoperative, before analgesic rescue; M3 - postoperative, after analgesic rescue; M4 - 24h after surgery. Ellipses were constructed according to the moments of pain assessment (M1 - green, M2 - red, M3 – blue, and M4 - yellow). The ellipse referring to the time when sheep were in severe pain (M2) was positioned at the right end of the graph, while on the opposite side were the times when the sheep were probably not in pain (M1 and M4). The moment of moderate pain (M3) is positioned in the middle. All items on the scale are influenced by moments of pain (M2 and M3) since their vectors are positioned in the direction of these ellipses.

Item-total correlation

Correlation coefficient of item score with total score (item-total score) ranged from 0.43 to 0.72 and, therefore, all items were accepted (Table 11).

Internal consistency

The Cronbach's α coefficient was 0.84 for all moments grouped, which indicates that the instrument presents excellent internal consistency and reinforces the possibility of using the full scale score to interpret the results obtained. Internal consistency was very good when locomotion (0.78) and interaction (0.79) were excluded and excellent when all other individual items were excluded, showing that all items contributed similarly and significantly to the total score (Table 11).

Specificity and sensitivity

All items on the USAPS showed moderate to excellent specificity. The items “interaction”, “activity” and “locomotion” presented moderate or good sensitivity. “Head position”, “appetite” and “posture” were not sensitive (Table 11).

Table 11. Item-total correlation, internal consistency, specificity and sensitivity of the USAPS.

Items	Item-total (Spearman)	Internal consistency (Cronbach's α)	Specificity (%)	Sensitivity (%)
Full scale		0.84		

	Excluding each item below			
Interaction	0.72	0.79	87	81
Activity	0.69	0.81	83	85
Locomotion	0.71	0.78	87	83
Head position	0.54	0.82	71	69
Appetite	0.43	0.84	100	50
Posture	0.53	0.83	79	67

USAPS: Unesp-Botucatu sheep acute composite pain scale. Interpretation of Spearman's rank correlation coefficient (r_s) - degree of correlation r_s : 0.3 - 0.7: acceptable values in bold [84]. Cronbach's α coefficient was calculated for the total score and excluding each item from the scale. Interpretation of the α coefficient values: 0.60-0.64: minimally acceptable; 0.65-0.69: acceptable; 0.70-0.74: good; 0.75-0.80: very good; > 0.80: excellent [84–86]; bold values > 70. Interpretation of specificity and sensitivity: excellent 95 - 100%; good 85 - 94.9%; moderate 70 - 84.9%; is not specific or sensitive <70%; bold values $\geq 70\%$ [84].

Distribution of scores

The distribution of scores “0”, “1” and “2” occurred as expected, according to the degree of pain. The score “0” predominated at moments M1, M3 and M4. Scores “1” and “2” were more frequent in M2 and decreased in M3 (Fig 6). Only the item “activity” of score “1” was not very representative when all moments were grouped. For calculation purposes, “appetite” was considered arbitrarily normal in M1.

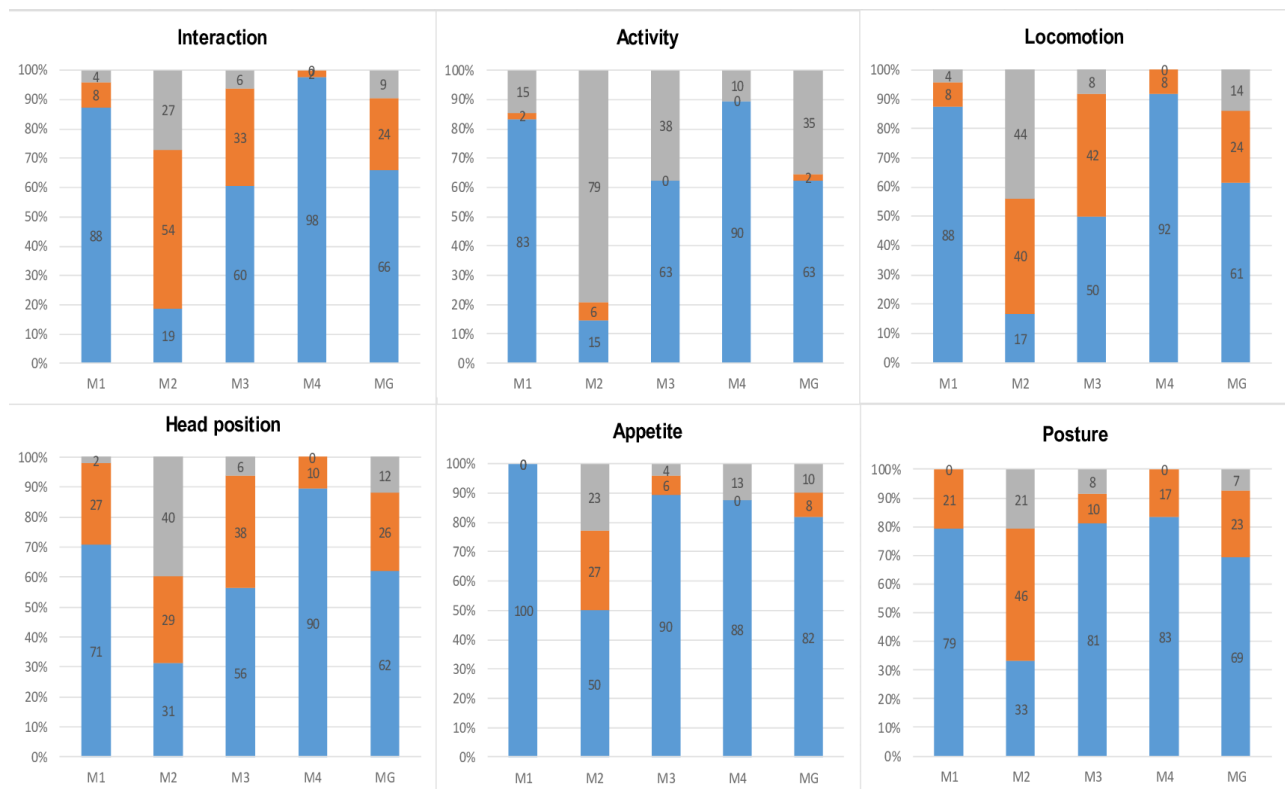


Fig 6. Distribution of the percentage of the presence of the **perioperative USAPS scores of each item.**

M1: preoperative; M2: postoperative, before rescue analgesia; M3: postoperative, after rescue analgesia; M4: 24h postoperative; MG - data of the grouped moments (M1 + M2 + M3 + M4).

ROC Curve, Youden index, cut-off point and diagnostic uncertainty zone of the USAPS

In the analysis of the ROC curve to determine the cut-off point for diagnosing pain and recommending analgesia, the Youden index was ≥ 3 for evaluators 1 and 3, ≥ 4 for evaluators 2 (gold standard) and 4, and again ≥ 4 of 12 for all grouped evaluators. The interval between the sensitivity and specificity values of 0.90 was between 3.3 and 3.7. Resampling confidence interval (Bootstrap) for Youden index was between 2.5 to 3.5. Based on the resampling result, which showed the highest interval between these two methods, the diagnostic uncertainty zone scores ranged from 2.5 to 3.5 for all grouped evaluators; therefore, ≤ 2.5 indicates truly negative pain (sheep without pain) and ≥ 3.5 indicates truly positive pain (sheep suffering pain). The individual results of the area under the curve were high (AUC > 0.95) for all evaluators, and 0.97 for grouped evaluators (0.96 - 0.98), indicating that the USAPS presents excellent discriminatory capacity (Fig 7).

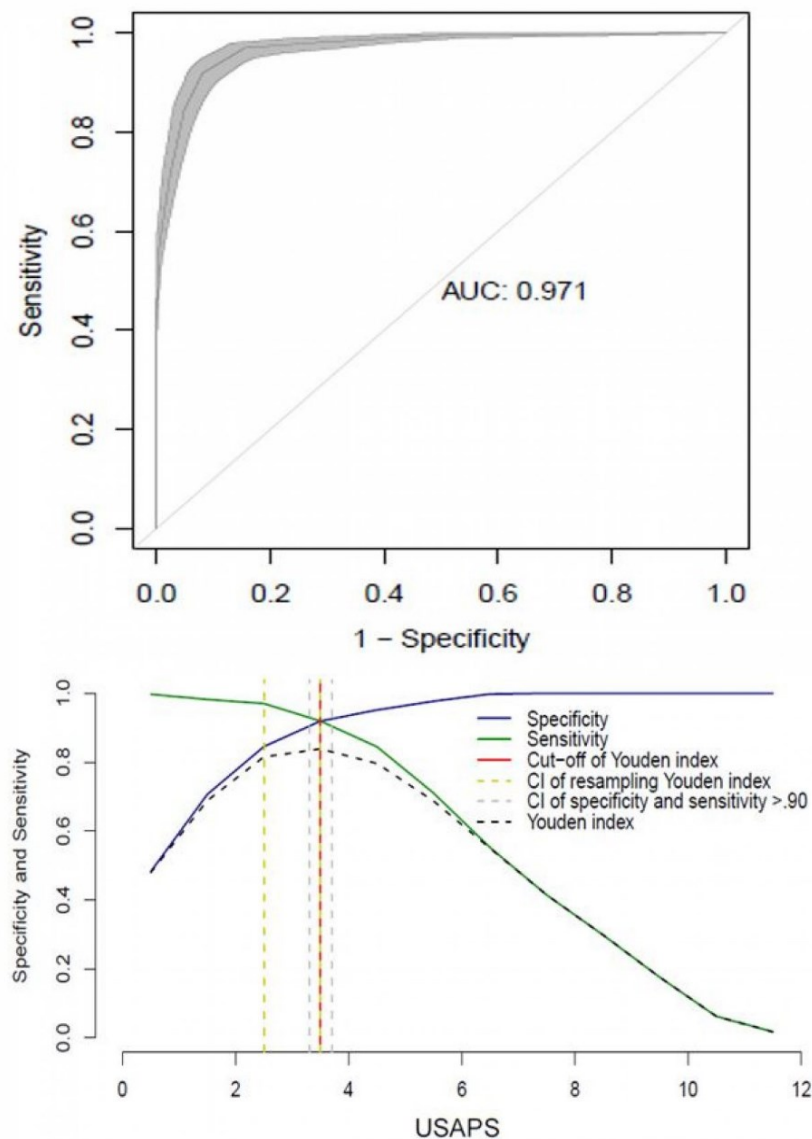


Fig 7. ROC curve and AUC [above] and ROC curve of two graphs with the diagnostic uncertainty zone for the USAPS [below].

ROC curve (receiver operating characteristic) with a 95% confidence interval (CI) [89] calculated from 1,001 replications and area under the curve (AUC) [above]. Interpretation of $AUC \geq 0.95$ - high discriminatory capacity [90]. ROC curve of two graphs, CI of 1,001 replications and sensitivity and specificity > 0.90 applied to estimate the diagnostic uncertainty zone of the cut-off point of each evaluator, according to the Youden index for the Unesp-Botucatu sheep acute composite pain scale (USAPS) [91–94]. Data of all grouped evaluators. The diagnostic uncertainty zone scores ranged from 2.5 to 3.5; therefore, ≤ 2.5 indicates truly negative pain (sheep without pain) and ≥ 3.5 indicates truly positive pain (sheep suffering pain). Youden index was ≥ 4 , representative of the cut-off point for indication of rescue analgesia.

For the unidimensional scales, the cut-off points for rescue analgesia defined by the ROC curve and the Youden index were ≥ 4 for SN, ≥ 3 for SDS and ≥ 27 for VAS (Table 12).

Table 12. Scores, specificity, sensitivity, and Youden index corresponding to rescue analgesia indication of the USAPS and unidimensional scales.

Scale	Score	Specificity	Sensitivity	Youden index
USAPS	4	0.92	0.92	0.84
NS	4	0.97	0.94	0.91
SDS	3	1	0.71	0.71
VAS	27	0.94	0.96	0.90

Scales: USAPS - Unesp-Botucatu sheep acute composite pain scale; NS – numerical; SDS - simple descriptive; VAS - visual analogue.

The percentage of animals present in the diagnostic uncertainty zone was low at all times for all evaluators (11%; 9 - 15%). At M2, this percentage for all evaluators grouped was 7% (0 to 13%), which ensures that 93% of the sheep were detected as suffering pain with confidence at this moment (Table 13).

Table 13. Percentage of sheep present in the diagnostic uncertainty zone according to the Youden index of the USAPS.

Evaluator	1	2 (gold)	3	4	All
M1	6	8	23	13	13
M2	13	10	4	0	7
M3	8	21	17	13	15
M4	10	8	15	15	12
MG	9	12	15	10	11

USAPS: Unesp-Botucatu sheep acute composite pain scale. Calculation based on 48 sheep for each evaluator and 192 sheep for all evaluators. M1: preoperative; M2: postoperative, before rescue analgesia; M3: postoperative, after rescue analgesia; M4: 24h postoperative; MG - data of grouped moments (M1 + M2 + M3 + M4). The diagnostic uncertainty zone scores ranged from 2.5 to 3.5; therefore, ≤ 2.5 indicates truly negative pain (sheep without pain) and ≥ 3.5 indicates truly positive pain (sheep suffering pain).

Determination of pain intensity

The total sums of the USAPS scores were divided into three groups (Fig 8) of pain intensity given by median, minimum and maximum and named according to the intensity: mild (1; 0 - 3); moderate (7; 4 - 8) and severe (10; 9 - 12) (Fig 9). From this classification, it was observed that in M2, the moment of greatest pain, 10% of sheep were classified as mild intensity, 47% as moderate intensity

and 43% as severe intensity

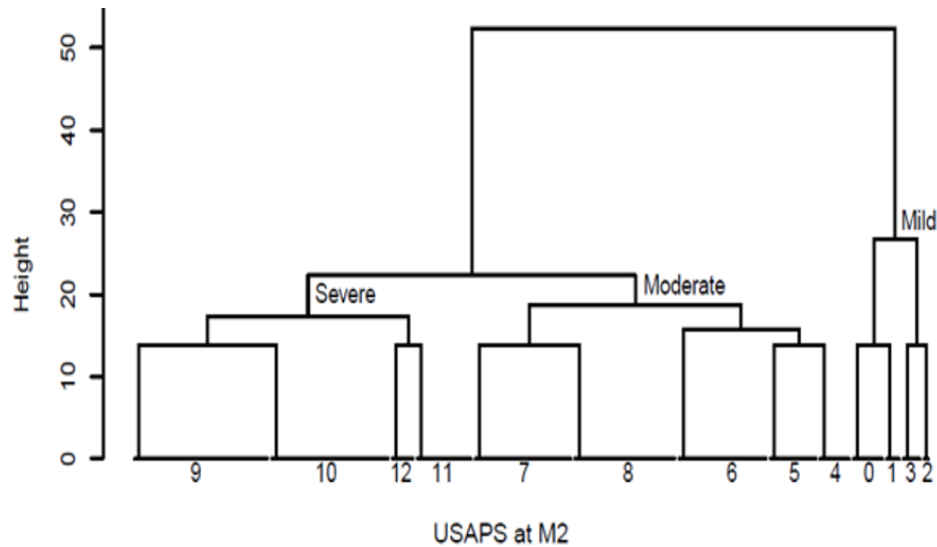


Fig 8. Dendrogram created by the non-hierarchical cluster analysis based on the total score of the USAPS.

USAPS: Unesp-Botucatu sheep acute composite pain scale. The scores were divided into 3 groups: mild (0 - 3); moderate (4 - 8) and severe (9 - 12) pain intensity.

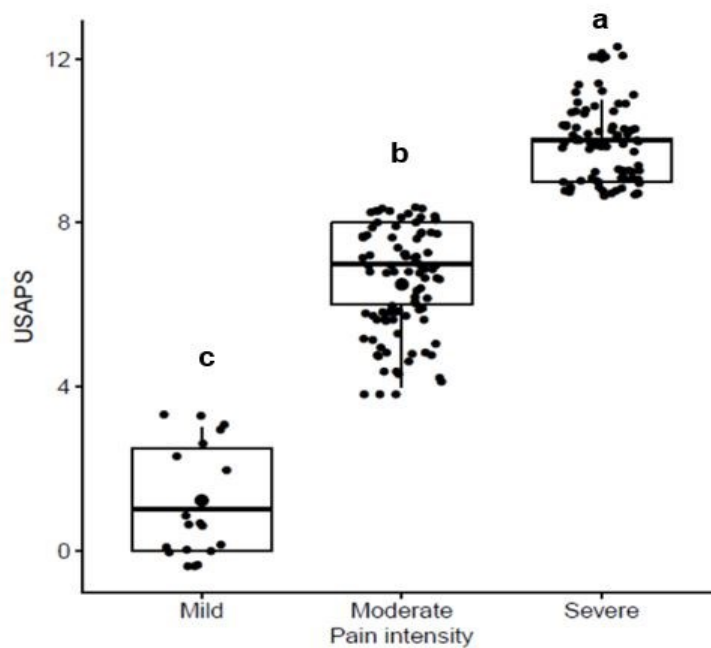


Fig 9. Box plot for the mild, moderate and severe pain intensity of the scores of the USAPS.

USAPS: Unesp-Botucatu sheep acute composite pain scale. Data were grouped at the time of greatest pain (M2). According to the median, minimum and maximum scores obtained by the non-hierarchical cluster analysis followed by the Kruskal-Wallis test, the scores were divided into 3 groups: mild (1; 0 - 3); moderate (7; 4 - 8) and (10; 9 - 12). Different letters indicate statistical difference by the Kruskal-Wallis test, where $a > b > c$.

Discussion

The creation of valid species-specific tools to assess pain is a prerequisite for recognizing the phenomenon and determining the need and effectiveness of analgesic treatment. From this

perspective, the behavioral pain scale proposed herein can be considered as a reliable, valid instrument with a defined analgesic intervention point, which can be used to assess postoperative abdominal pain in sheep. This instrument demonstrates potential clinical applicability, as it can guarantee that sheep benefit from analgesia when necessary, and potential experimental applicability, for example to perform studies that compare the analgesic efficacy of drugs, as well as translational studies [2,17].

The validation process of an instrument to assess pain is based on the investigation of behaviors and, when possible, of species-specific physiological data present during pain situations, followed by comparison of these changes with the state of normality [2]. This methodology was followed in the current study; an ethogram was constructed that covered the preoperative period, at which time the animals were supposedly devoid of pain, followed by the postoperative period, at which time the animals probably had severe pain, followed by rescue analgesia for pain reduction and assessment after 24 hours. Thus, the experimental design aimed to test the instrument at different pain intensities. The ethogram, together with the pain expression behaviors in sheep described in the literature, served as a basis for the construction of the scale and then, after content analysis by the committee of experienced researchers in the area, the first instrument was defined including relevant behaviors and excluding irrelevant ones, to make the instrument as simple and representative as possible.

Filming using video cameras adds value to the data as it enables the material to be archived for future research and minimizes the influence of the observer in the evaluation, avoiding possible behavioral changes that the animal may present which are inherent to the presence of the observer. On the other hand, it should be emphasized that in a clinical situation, it is often not possible to make a remote assessment, and the presence of the observer could interfere with the animals' behavior [94–97]. Thus, it is still necessary to validate the proposed scale in clinical situations and with the presence of the observer, to ensure that these results are reproducible.

According to the ethogram, sheep in pain ate less and consequently ruminated less in relation to the moments after rescue analgesia, which justified the introduction of appetite as one of the criteria to be evaluated on the scale. At the preoperative moment, the animals were fasting and there was no feed available, so, obviously this behavior was not present. Previous studies have also shown decreased appetite in sheep submitted to castration and laparoscopy [28,57,73,98], however, sheep in moderate pain may have appetite and rumination [66,99]. The improvement in appetite after analgesic therapy strengthens the need for postoperative analgesia in sheep [100], as in cattle [29].

As reported by the literature [31,62,99,101], in the current study, in sheep in pain, reduced interaction with the environment and activity, low head, and dorsal arching were observed, which confirms that laparoscopy leads to behavioral alterations indicative of acute pain [73]. Some of these behaviors only returned to normal 24 hours after the surgery.

The findings of the ethogram were similar to those in the literature. Lambs submitted to the

mulesing procedure were bent over and walked less in the postoperative period [101] and after orchietomy different behaviors and physiological indicators of postoperative pain were observed [102,103]. Orchietomy and tail cutting trigger behaviors related to various pain intensities. During severe pain, limb, tail, and head movements and postures with full extension of the pelvic limbs occur; during moderate pain vocalization, standing, sitting, and lying positions with partial extension of the pelvic limbs or with tremor are observed; and during mild pain or no pain, postures may be normal [28].

Vocalization could be a possible indicator of pain, as reported in other ruminants, such as cattle [29,104–106] and goats [106]. However, in sheep, this behavior is more related to social isolation, restraint, and the trans-operative period. Vocalization was not included in the ethogram in the current study, as it was not observed, except at times when feed was supplied. Thus, vocalization is not, in the sheep species, an indicator of postoperative pain, especially in adult animals, even after intense painful stimuli [28,107].

A differential of the current study, which presents advantages and disadvantages compared to others, is that the majority of studies that evaluated acute pain in sheep used lambs [8,28,31–33,62,98,101–103], which could limit extrapolation of the results to adult animals. Some behaviors more specific to lambs mentioned in the literature, such as “jumping like a rabbit”, did not occur, as they are more frequent in young animals, up to about 5 months of age, than in adult animals. A common behavior in cattle [29] and cited in lambs [28,91], looking at the flank and lick the painful area, was not observed in the adult sheep in the current work.

Regarding the validation of the scale, the instrument was subjected to a blinded and random methodology, through the same experimental design already used in cats [47], cattle [29] horses [38], and pigs [30], which presents recognized scientific robustness [71].

Validity and reliability are essential attributes for an instrument to identify and quantify pain in animals. Observers familiar with the behavior of ruminants validated the content of the scale by assessing the representativeness of each item. This analysis measures the extent to which the instrument reflects the phenomenon of interest, in this case, pain [40,47,108,109]. The evaluators were selected for their experience, which improves reliability, that is, the accuracy of the tool based on repeatability and reproducibility [110].

The initially proposed acute pain scale contained 33 variables including items, sub-items and their sub-divisions (S3 Table), which were excluded according to the criteria of the statistical tests (Table 1). This process of refinement of the composite scale contained several steps and identified the 12 most relevant and appropriate variables to measure pain in sheep. The refinement of the pain scale through the analysis of its items was essential to improve the quality of the validated final scale, giving rise to a simpler and more objective version [111].

In a validation of the acute pain scale in lambs, the principal component analysis generated two principal components [28]. Unidimensional scales are not as satisfactory as those with more than one dimension, as they only assess the intensity of pain. However, considering their simplicity

and practicality they are easily applicable. In the current study, it is premature to conclude about the dimension of the proposed pain scale, since only one type of pain was evaluated (abdominal), in soft tissues, using only one statistical model. The scale proposed for sheep includes several biological aspects of pain, such as physiological (appetite), sensory or motor (posture, activity), emotional (interaction with other animals and attention in the environment), and temporal (response to analgesia). Future studies addressing different pain models, such as orthopedic, and other statistical approaches, may confirm the number of dimensions of this scale.

Every new instrument needs to be compared with another already established and validated tool, considered the gold standard [71]. As this is not the case in sheep, since there is no validated scale with robust statistics to assess postoperative pain, the methodology based on the agreement between the pain scores assigned by the evaluators and a gold standard evaluator was adopted, following the same criteria applied in cats [47], cattle [29] horses [38], and pigs [30].

The intra- and inter-observer reliability for each item and for the total score of the sheep scale was similar to that of cattle [29] and pigs [30]. These results were lower than in cats [47] and higher than in horses [38]. When compared with other instruments developed in the sheep species, the scale proposed here presented reliability similar to the scale reported in lambs subjected to acute pain (73% intra and 79% inter-observer reliability) [28]. Compared to a sheep locomotion scale (with a score ranging from 0 to 6) that demonstrated very good intra (91%) and inter-observer (93%) reliability [112], the proposed instrument showed lower results, in which the item "locomotion" presented good intra-observer reliability and only moderate inter-observer reliability. Another study, in which 10 veterinarians and 10 sheep farmers scored a locomotion scale, obtained very good and good values for intra- and inter-observer reliability, however, the reliability for individual locomotion scores varied from reasonable to moderate [113].

Validity indicates that the instrument can accurately measure what is proposed. There are three types of validity. Criterion validity assesses the measuring efficiency of a scale. Two methods may be used: concurrent and predictive criterion validity. Concurrent validity compares the instrument to existing validated scales [43], by evaluating the instrument and the criterion simultaneously and predictive validity evaluates the criterion after the test. Both methods were used in this study [114]. As mentioned earlier, there is no validated instrument to assess postoperative pain in sheep that can be used as a comparison, therefore the proposed instrument was compared with the NS, SDS, and VAS, when a high correlation was observed. In the sheep species, the correlation of a claudication scale with the NS and VAS was good [115]. Otherwise when comparing the proposed scale with the facial scale in sheep [69], which, although not fully validated, was considered for comparison, the correlation was moderate. The agreement between the gold standard and the other observers was good and this was a second method to assess concurrent criterion validity, as reported previously in other species [29,30,38].

Predictive criterion validity was confirmed by the number of sheep that should receive rescue analgesia based on the Youden Index after surgery (M1). Consistent with this analysis rescue

analgesia was indicated in 93% of sheep; therefore the tool would foresee well that sheep were undergoing pain and then be treated, guaranteeing the animals' welfare.

Construct validity reflects the responsiveness of the scale and examines whether the instrument detects predictable differences between groups, or moments [36]. The method tests the hypothesis that time and surgical and analgesic intervention should alter pain scores [47] and has been used to validate scales in veterinary medicine [29,30,38,41,47]. In this study, the differences observed in the pain scores between the moments, and especially at the expected moment of greatest pain compared to the other moments, confirm that the proposed scale is responsive both to identify intense degrees of pain, as well as moderate degrees, which occurred after rescue analgesia, or even mild pain, which occurred 24 hours after surgery. In cattle, the alterations between scores ($M2 > M4 > M3 > M1$) [29] were different from sheep, where the pain at M4 decreased after M3 ($M2 > M3 > M4 > M1$), however, the surgical, anesthetic and analgesic protocols were different in the two species of ruminants. The increase and decrease in pain scores after surgery and rescue analgesia, respectively, also occurred in cats [47], horses [38] and pigs [30]. As the current study was carried out with three different breeds of sheep and there was no effect of breed on the scores, apparently the instrument can be used in different sheep breeds.

Principal component analysis relates the variables of the tool in a grouped manner [116] and calculates the number of dimensions determined by different variables [47], to establish the extension or dimensionality of the scale [85]. These variables are related so that the items that define specific parts of the construct are grouped by means of multiple association [117]. According to the Kaiser criterion, one main component was selected and the scale was considered as unidimensional [79]. The scales developed in cattle [29] and pigs [30] are also unidimensional and the scale in cats is multidimensional since it generated three dimensions [45-47]. An instrument is multidimensional when in addition to pain intensity, it includes qualitative and temporal characteristics, such as sensory, motor, emotional, and cognitive dimensions, which have a high correlation in the experience of pain [47,116,117].

All items of the proposed scale presented an acceptable item-total correlation, as in pigs [30], which demonstrates their individual relevance and ensures the homogeneity of the tool. The internal consistency of the proposed scale was excellent, with very similar results to cats - 0.86 [47], cattle - 0.87 [29] and pigs - 0.82 [30], which ensures that the scores of the items that compose the scale can be added and the total score will be representative of the pain intensity [47]. The similarity of the values when excluding each item demonstrates that they have a similar tendency and importance [85].

The scale was specific for all items, however, it was sensitive for only three. Of these three items, the head position and posture values were very close to those considered as good sensitivity (0.70). Only appetite did not present adequate sensitivity, so sheep, even when in pain, can feed, as described in other species [29,30].

The analysis of score distribution, by providing an overview of the occurrence of each score

at each moment, indicates the importance of each score. The results were as expected, since the score "0", related to the absence of pain, prevailed before and 24h after the surgery, the scores "1" and "2" occurred more in the postoperative period and after the rescue analgesia, and the score "2" was more evident, especially after surgery, suggesting a greater intensity of pain. The results of each item generally followed the results of the sum of the scale. Only the score "1" of the "activity" item was not so evident, which demonstrates that it is rare for sheep to move about more than normal or to lie down and get up frequently, as occurs in other species [29,30,38,47].

The analysis of the ROC curve [90] estimated the cut-off point for analgesic intervention in sheep as was the case in pain scales in cats [47], cattle [29] and pigs [30]. The determination of scores indicative of the need to use analgesics helps professionals in clinical decisions, confirms the efficacy of analgesic treatment [47], and prevents unnecessary suffering in animals. The diagnostic uncertainty zone of all evaluators ensures that sheep with a score of 3.5 out of a total of 12 points (≥ 4) are really in pain, while those with a score ≤ 2.5 do not have pain. The low percentage of animals within the zone of diagnostic uncertainty ensures good reliability in making decisions about the indication for rescue analgesia in animals that actually present pain and, therefore, should receive analgesia. Thus, the proposed scale presents excellent diagnostic accuracy. Although the definition of the score referring to the analgesic intervention point is a good tool, it is emphasized that even if the scores are < 4 , in some cases additional analgesia may be necessary according to the clinical evaluation, at the discretion of the observer. The cut-off point was $> 4/10$ in cattle [29], $> 4/18$ in pigs [30], and for the subscale "expression of pain" in cats it was $> 2/12$ [47]. In a recent empirical study, pain in sheep after cardiac surgery was scored; for scores of 0-2/25 there was no intervention, 3-9/25 rescue analgesia was performed, and $\geq 10/25$ multimodal analgesia was performed [118]. Another study subjected 18 sheep to laparoscopy and pain was assessed for 24 hours based on three items: decreased appetite, limited mobility, and back arching, but the pain scale ranging from 0 to 6 was insensitive, with 90% of animals with a "0" score and 10% with "1" [98].

In line with the low percentage of animals within the diagnostic uncertainty zone, the high areas under the curve observed in this study (> 0.95) indicate that the scale has high discriminatory capacity, that is, it correctly classifies individuals with or without pain [88], results that resembled cattle [29], pigs [30] and the subscale "expression of pain" in cats [47],

To our knowledge, the scales that assess acute pain in various animal species do not classify pain intensity based on their scores, except in an empirical way [44]. In this study, the zone of diagnostic uncertainty (2.5 - 3.5) corresponded to the lower limit of moderate pain scores (4), insuring that sheep suffering from moderate pain would be treated according to the cut-off point.

Limitations

The current study had some limitations. The main one is that the behavioral scale was validated only for a specific type of soft tissue surgery (abdominal - laparoscopy) and in females. Further studies are needed to test this tool in different procedures, such as orthopedic surgery and in clinical

circumstances, to ensure its versatility. To establish that the instrument is valid under field conditions, clinical validation with less experienced observers is also required. Since the majority of the studies that evaluated acute pain in sheep were in lambs, this can limit the collation of data, which means the instrument needs to be tested in lambs.

Another limitation is that video analysis does not necessarily equate to presential analysis. Video observation has the disadvantage of lacking some details observed in real time, while, as an advantage it can be reviewed. According to the previous experience of the authors, video analysis provides material to develop the ethogram and is an important step to validate a scale, as reported in cats [47], cattle [29] horses [38], and pigs [30], especially to assess intra-observer reliability. According to a study by our group in cats, the results of the video analysis were also reproducible in a clinical setting [45–47].

Some limitations relate specifically to the videos. Although the study was blinded, some videos may have suggested the moment they were taken: at baseline, the sheep were fasting, with no available feed, hence it was difficult for the observers to interpret if the animals did not eat due to lack of food or if they really had anorexia. To overcome this problem and avoid that appetite would be excluded according to the refinement criteria, appetite was arbitrarily considered normal before surgery regardless of the scores attributed by the observers. Around 21% of the videos at M3 were filmed at night with artificial light, which could suggest that they corresponded to M3; variations in the circadian cycle could alter some behaviors such as activity, so the reduction in activity observed at night may not be related to pain or discomfort, but to the natural reduction in activity at night [119]; given the small difference in the size of the stalls, the density of animals varied slightly, which could influence interaction and locomotion behaviors; the dark wool of some animals may also have made it difficult to evaluate some items in the videos/photos, making the analysis less accurate, especially on the facial scale.

To improve data reliability, the authors suggest that observers attend a training period, as, at least in laboratory animals, instruction and training have improved pain recognition [120].

Conclusion

It is concluded that, after refinement of the originally proposed scale, the Unesp-Botucatu composite scale to assess acute postoperative pain in sheep (USAPS) is a valid, reliable, specific and sensitive instrument, with excellent internal consistency and discriminatory capacity. The well-defined cut-off point for rescue analgesia and the classification of pain intensity supports the indication and type of analgesic therapy. To assess the clinical and experimental applicability of the scale and ensure its versatility, it is recommended that it be evaluated in other surgical procedures and in lambs.

Acknowledgements

Capes scholarship and FAPESP thematic research project number 2017/12815-0.

References

1. Huxley JN, Whay HR. Current attitudes of cattle practitioners to pain and the use of analgesics in cattle. *Vet Rec.* 2006;159: 662–668. doi:10.1136/vr.159.20.662
2. Flecknell P. Analgesia from a veterinary perspective. *Br J Anaesth.* 2008;101: 121–124. doi:10.1093/bja/aen087
3. Lizarraga I, Chambers JP. Use of analgesic drugs for pain management in sheep. *N Z Vet J.* 2012;60: 87–94. doi:10.1080/00480169.2011.642772
4. Lorena SERS, Luna SPL, Lascelles BD, Corrente JE. Attitude of brazilian veterinarians in the recognition and treatment of pain in horses and cattle. *Vet Anaesth Analg.* 2013;40: 410–418. doi:10.1111/vaa.12025
5. Raekallio M, Heinonen KM, Kuussaari J, Vainio O. Pain alleviation in animals: attitudes and practices of finnish veterinarians. *Vet J.* 2003;165: 131–135.
6. Anil L, Anil SS, Deen J. Detection P. Pain detection and amelioration in animals on the farm: issues and options. *J Appl Anim Welf Sci.* 2005;8(4):261-78. doi: 10.1207/s15327604jaws0804_3
7. Hewson CJ, Dohoo IR, Lemke KA, Barkema HW. Canadian veterinarians' use of analgesics in cattle, pigs, and horses in 2004 and 2005. *Can Vet J = La Rev Vet Can.* 2007;48: 155–164. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/17334029>
8. Futro A, Masłowska K, Dwyer CM. Ewes direct most maternal attention towards lambs that show the greatest pain-related behavioural responses. *PLoS One.* 2015;10: 1–15. doi:10.1371/journal.pone.0134024
9. West D, Bruere N, Ridler A. The sheep: health disease and production. 4th ed. New Zealand: Massey University Press; 2018.
10. Dolan S, Field LC, Nolan AM. The role of nitric oxide and prostaglandin signaling pathways in spinal nociceptive processing in chronic inflammation. *Pain.* 2000;86: 311–320.
11. Kaler J, Medley GF, Grogono-Thomas R, Wellington EMH, Calvo-Bado LA, Wassink GJ, et al. Factors associated with changes of state of foot conformation and lameness in a flock of sheep. *Prev Vet Med.* 2010;97: 237–244. doi:10.1016/j.prevetmed.2010.09.019
12. Ley SJ, Livingston A, Waterman AE. The effect of chronic clinical pain on thermal and mechanical thresholds in sheep. *Pain.* 1989;39: 353–357. doi:10.1016/0304-3959(89)90049-3
13. Ley SJ, Waterman AE, Livingston A. A field study of the effect of lameness on mechanical nociceptive thresholds in sheep. *Vet Rec.* 1995;137: 85–87. doi:10.1136/vr.137.4.85
14. Dolan S, Kelly JG, Monteiro AM, Nolan AM. Up-regulation of metabotropic glutamate receptor subtypes 3 and 5 in spinal cord in a clinical model of persistent inflammation and hyperalgesia. *Pain.* 2003;106: 501–512.
15. Taylor K, Rego L. EU statistics on animal experiments for 2014. *Altern to Anim Exp.* 2016;33. doi:10.14573/altex.1609291
16. Gigliuto C, De Gregori M, Malafoglia V, Raffaelli W, Compagnone C, Visai L, et al. Pain assessment in animal models: do we need further studies? *J Pain Res.* 2014;7: 227–236. doi:10.2147/JPR.S59161
17. Dias IR, Camassa JA, Bordelo JA, Babo PS, Viegas CA, Dourado N, et al. Preclinical and translational studies in small ruminants (sheep and goat) as models for osteoporosis research. *Curr Osteoporos Rep.* 2018;16: 182–197. doi:10.1007/s11914-018-0431-2

18. Sartoretto SC, Uzeda MJ, Miguel FB, Nascimento JR, Ascoli F, Calasans-Maia MD. Sheep as an experimental model for biomaterial implant evaluation. *Acta Ortop Bras.* 2016;24: 262–266. doi:10.1590/1413-785220162405161949
19. Dolan S, Gunn MD, Crossan C, Nolan AM. Activation of metabotropic glutamate receptor 7 in spinal cord inhibits pain and hyperalgesia in a novel formalin model in sheep. *Behav Pharmacol.* 2011;22: 582–588. doi:10.1097/FBP.0b013e3283478802
20. Wilkes D, Li G, Angeles CF, Patterson JT, Huang LYM. A large animal neuropathic pain model in sheep: A strategy for improving the predictability of preclinical models for therapeutic development. *J Pain Res.* 2012;5: 415–424. doi:10.2147/JPR.S34977
21. Stubbsj  en SM, Bohlin J, Skjerve E, Valle PS, Zanella AJ. Applying fractal analysis to heart rate time series of sheep experiencing pain. *Physiol Behav.* 2010;101: 74–80. doi:10.1016/j.physbeh.2010.04.018
22. Ong RM, Morris JP, O'Dwyer JK, Barnett JL, Hemsworth PH, Clarke IJ. Behavioural and EEG changes in sheep in response to painful acute electrical stimuli. *Aust Vet J.* 1997;75: 189–193. doi:10.1111/j.1751-0813.1997.tb10064.x
23. Mather LE, Cousins MJ, Huang YF, Pryor ME, Barratt SM. Lack of secondary hyperalgesia and central sensitization in an acute sheep model. *Reg Anesth Pain Med.* 2000;25: 174–180. doi:10.1097/00115550-200004000-00008
24. McLennan KM, Skillings EA, Rebelo CJB, Corke MJ, Pires Moreira MA, Morton AJ, et al. Technical note: Validation of an automatic recording system to assess behavioural activity level in sheep (*Ovis aries*). *Small Rumin Res.* 2015;127: 92–96. doi:10.1016/j.smallrumres.2015.04.002
25. Stubbsj  en SM, Bohlin J, Dahl E, Knappe-Poindecker M, Fjeldaas T, Lepschy M, et al. Assessment of chronic stress in sheep (part I): The use of cortisol and cortisone in hair as non-invasive biological markers. *Small Rumin Res.* 2015;132: 25–31. doi:10.1016/j.smallrumres.2015.09.015
26. Stubbsj  en SM, Knappe-Poindecker M, Langbein J, Fjeldaas T, Bohlin J. Assessment of chronic stress in sheep (part II): Exploring heart rate variability as a non-invasive measure to evaluate cardiac regulation. *Small Rumin Res.* 2015;133: 30–35. doi:10.1016/j.smallrumres.2015.10.026
27. Jongman EC, Morris JP, Barnett JL, Hemsworth PH. EEG changes in 4-week-old lambs in response to castration, tail docking and mulesing. *Aust Vet J.* 2000;78: 339–343. doi:10.1111/j.1751-0813.2000.tb11789.x
28. Molony V, Kent JE, McKendrick IJ. Validation of a method for assessment of an acute pain in lambs. *Appl Anim Behav Sci.* 2002;76: 215–238. doi:10.1016/S0168-1591(02)00014-X
29. de Oliveira FA, Luna SPL, do Amaral JB, Rodrigues KA, Sant'Anna AC, Daolio M, et al. Validation of the UNESP-Botucatu unidimensional composite pain scale for assessing postoperative pain in cattle. *BMC Vet Res.* 2014;10: 1–14. doi:10.1186/s12917-014-0200-0
30. Luna, S; Araujo, A; Neto, P; Oliveira, F; Brondani, J; Telles, F; Azeredo I. Validation of the UNESP-Botucatu unidimensional composite pain scale for assessing postoperative pain in pigs. 12th World Congress of Veterinary Anesthesiology. Kyoto; 2015.
31. Molony V, Kent JE. Assessment of acute pain in farm animals using behavioral and physiological measurements. *J Anim Sci.* 1997;75: 266–272. doi:10.2527/1997.751266x
32. Grant C. Behavioural responses of lambs to common painful husbandry procedures. *Appl Anim Behav Sci.* 2004;87: 255–273. doi:https://doi.org/10.1016/j.applanim.2004.01.011
33. Guesgen MJ, Beausoleil NJ, Stewart M. Effects of early human handling on the pain sensitivity of young lambs. *Vet Anaesth Analg.* 2013;40: 55–62. doi:10.1111/j.1467-2995.2012.00746.x

34. Morton CM, Reid J, Scott EM, Holton LL, Nolan AM. Application of a scaling model to establish and validate an interval level pain scale for assessment of acute pain in dogs. *Am J Vet Res.* 2005;66: 2154–2166. doi:10.2460/ajvr.2005.66.2154
35. McDowell I. *Measuring health: A guide to rating scales and questionnaires* 3rd ed. New York: Oxford University Press; 2006. doi: 10.1093/acprof:oso/9780195165678.001.0001
36. Crellin D, Sullivan TP, Babl FE, O'Sullivan R, Hutchinson A. Analysis of the validation of existing behavioral pain and distress scales for use in the procedural setting. *Paediatr Anaesth.* 2007;17: 720–733. doi:10.1111/j.1460-9592.2007.02218.x
37. Beyer JE, Wells N. The assessment of pain in children. *Pediatr Clin North Am.* 1989;36: 837–854.
38. Taffarel MO, Luna SPL, de Oliveira FA, Cardoso GS, Alonso JM, Pantoja JC, et al. Refinement and partial validation of the UNESP-Botucatu multidimensional composite pain scale for assessing postoperative pain in horses. *BMC Vet Res.* 2015;11. doi:10.1186/s12917-015-0395-8
39. von Baeyer CL, Spagrud LJ. Systematic review of observational (behavioral) measures of pain for children and adolescents aged 3 to 18 years. *Pain.* 2007;127: 140–150. doi:10.1016/j.pain.2006.08.014
40. Bussieres G, Jacques C, Lainay O, Beauchamp G, Leblond A, Cadore J-L, et al. Development of a composite orthopaedic pain scale in horses. *Res Vet Sci.* 2008;85: 294–306. doi:10.1016/j.rvsc.2007.10.011
41. Brondani JT, Luna SPL, Minto BW, Santos BPR, Beier SL, Matsubara LM, et al. Validade e responsividade de uma escala multidimensional para avaliação de dor pós-operatória em gatos. *Arq Bras Med Veterinária e Zootec.* 2012;64: 1529–1538. doi:10.1590/S0102-09352012000600019
42. Firth AM, Haldane SL. Development of a scale to evaluate postoperative pain in dogs. *J Am Vet Med Assoc.* 1999;214: 651–659.
43. Holton L, Reid J, Scott EM, Pawson P, Nolan A. Development of a behaviour-based scale to measure acute pain in dogs. *Vet Rec.* 2001;148: 525–531. doi:10.1136/vr.148.17.525
44. Reid J, Nolan AM, Hughes JML, Lascelles D, Pawson P, Scott EM. Development of the short-form Glasgow composite measure pain scale (CMPS-SF) and derivation of an analgesic intervention score. *Anim Welf.* 2007;16: 97–104.
45. Brondani JT, Luna SPL, Padovani CR. Refinement and initial validation of a multidimensional composite scale for use in assessing acute postoperative pain in cats. *Am J Vet Res.* 2011;72: 174–183. doi:10.2460/ajvr.72.2.174
46. Brondani JT, Luna S, Padovani CR. Assessing acute postoperative pain in cats. *Am J Vet Res.* 2011;72: 174–183. doi:10.2460/ajvr.72.2.174
47. Brondani JT, Mama KR, Luna SPL, Wright BD, Niyom S, Ambrosio J, et al. Validation of the english version of the UNESP-Botucatu multidimensional composite pain scale for assessing postoperative pain in cats. *BMC Vet Res.* 2013;9: 1. doi:10.1186/1746-6148-9-143
48. Graubner C, Gerber V, Doherr M, Spadavecchia C. Clinical application and reliability of a post abdominal surgery pain assessment scale (PASPAS) in horses. *Vet J.* 2011;188: 178–183. doi:10.1016/j.tvjl.2010.04.029
49. Sutton GA, Paltiel O, Soffer M, Turner D. Validation of two behaviour-based pain scales for horses with acute colic. *Vet J.* 2013;197: 646–650. doi:10.1016/j.tvjl.2013.04.007
50. Sutton GA, Dahan R, Turner D, Paltiel O. A behaviour-based pain scale for horses with acute colic: scale construction. *Vet J.* 2013;196: 394–401. doi:10.1016/j.tvjl.2012.10.008
51. Banks EM. Behavioral research to answer questions about animal welfare. *J Anim Sci.* 1982;54: 434–446. doi:10.2527/jas1982.542434x

52. Thornton PD, Waterman-Pearson AE. Quantification of the pain and distress responses to castration in young lambs. *Res Vet Sci.* 1999;66: 107–118. doi:10.1053/rvsc.1998.0252
53. Dwyer CM, Lawrence AB. Maternal behaviour in domestic sheep (*ovis aries*): Constancy and change with maternal experience. *Behaviour.* 2000;137: 1391–1413. doi:10.1163/156853900501999
54. Thornton PD, Waterman-Pearson AE. Behavioural responses to castration in lambs. *Anim Welf.* 2002;11: 203–212.
55. Mellema SC, Doherr MG, Wechsler B, Thueer S, Steiner A. Influence of local anaesthesia on pain and distress induced by two bloodless castration methods in young lambs. *Vet J.* 2006;172: 274–283. doi:10.1016/j.tvjl.2005.06.002
56. Melches S, Mellema SC, Doherr MG, Wechsler B, Steiner A. Castration of lambs: a welfare comparison of different castration techniques in lambs over 10 weeks of age. *Vet J.* 2007;173: 554–563. doi:10.1016/j.tvjl.2006.01.006
57. Molony V, Kent JE, Viñuela-Fernández I, Anderson C, Dwyer CM. Pain in lambs castrated at 2 days using novel smaller and tighter rubber rings without and with local anaesthetic. *Vet J.* 2012;193: 81–86. doi:10.1016/j.tvjl.2011.09.030
58. Guesgen MJ, Beausoleil NJ, Minot EO, Stewart M, Stafford KJ, Morel PCH. Lambs show changes in ear posture when experiencing pain. *Anim Welf.* 2016;25: 171–177. doi:10.7120/09627286.25.2.171
59. Small AH, Belson S, Holm M, Colditz IG. Efficacy of a buccal meloxicam formulation for pain relief in Merino lambs undergoing knife castration and tail docking in a randomised field trial. *Aust Vet J.* 2014;92: 381–388. doi:10.1111/avj.12241
60. McGlone JJ, Stobart RH. A quantitative ethogram of behavior of yearling ewes during two hours post-parturition. *Appl Anim Behav Sci.* 1986;16: 157–164. doi:10.1016/0168-1591(86)90108-5
61. Mellor DJ, Murray L. Effects of tail docking and castration on behaviour and plasma cortisol concentrations in young lambs. *Res Vet Sci.* 1989;46: 387–391.
62. Molony V, Kent JE, Robertson IS. Behavioural responses of lambs of three ages in the first three hours after three methods of castration and tail docking. *Res Vet Sci.* 1993;55: 236–245.
63. Kent JE, Molony V, Robertson IS. Comparison of the burdizzo and rubber ring methods for castrating and tail docking lambs. *Vet Rec.* 1995;136: 192–196. doi:10.1136/vr.136.8.192
64. Lester SJ, Mellor DJ, Holmes RJ, Ward RN, Stafford KJ. Behavioural and cortisol responses of lambs to castration and tailing using different methods. *N Z Vet J.* 1996;44: 45–54. doi:10.1080/00480169.1996.35933
65. Graham MJ, Kent JE, Molony V. Effects of four analgesic treatments on the behavioural and cortisol responses of 3-week-old lambs to tail docking. *Vet J.* 1997;153: 87–97.
66. Kent JE, Molony V, Graham MJ. Comparison of methods for the reduction of acute pain produced by rubber ring castration or tail docking of week-old lambs. *Vet J.* 1998;155: 39–51. doi:10.1016/S1090-0233(98)80033-6
67. Taylor KD, Mills DS. The development and assessment of temperament tests for adult companion dogs. *J Vet Behav.* 2006;1: 94–108. doi: 10.1016/j.jveb.2006.09.002
68. Belshaw Z, Asher L, Harvey ND, Dean RS. Quality of life assessment in domestic dogs: an evidence-based rapid review. *Vet J.* 2015;206: 203–212. doi:10.1016/j.tvjl.2015.07.016
69. McLennan KM, Rebelo CJB, Corke MJ, Holmes MA, Leach MC, Constantino-Casas F. Development of a facial expression scale using footrot and mastitis as models of pain in sheep. *Appl Anim Behav Sci.* 2016;176: 19–26. doi:10.1016/j.applanim.2016.01.007

70. Häger C, Biernot S, Buettner M, Glage S, Keubler LM, Held N, et al. The Sheep Grimace Scale as an indicator of post-operative distress and pain in laboratory sheep. *PLoS One*. 2017;12: e0175839. doi:10.1371/journal.pone.0175839
71. Merola I, Mills DS. Systematic review of the behavioural assessment of pain in cats. *J Feline Med Surg*. 2016;18: 60–76. doi:10.1177/1098612X15578725
72. Kilkeny C, Browne W, Cuthill IC, Emerson M, Altman DG. Animal research: reporting in vivo experiments: the ARRIVE guidelines. *Br J Pharmacol*. 2010;160: 1577–1579. doi:10.1111/j.1476-5381.2010.00872.x
73. Teixeira PPMM, Padilha LC, Motheo TF, Silva MAMM, Oliveira MEFF, da Silva ASLL, et al. Ovariectomy by laparotomy, a video-assisted approach or a complete laparoscopic technique in Santa Ines sheep. *Small Rumin Res*. 2011;99: 199–202. doi:https://doi.org/10.1016/j.smallrumres.2011.04.008
74. Rodriguez M, Ambrogi M, Dayane V, Maciel G, Feliciano M, Teixeira P. Aspição folicular por laparoscopia em ovinos. *Rev Investig*. 2015;14: 55–60. doi:10.26843/investigacao.v14i2.878
75. de Araujo EAB, de Oliveira SN, de Faria Tabet A, Bittencourt RF, Chalhoub M, Filho ALR. Aspição folicular videolaparoscópica comparativa em ovelhas Dorper e Santa Inês. *Cienc Anim Bras*. 2016;17: 98–104. doi:10.1590/1089-6891v17i128347
76. Martin P, Bateson PPG. *Measuring behaviour: An introductory guide*, 2nd ed. New York, NY, US: Cambridge University Press; 1993. doi:10.1017/CBO9781139168342
77. RstudioTeam. RStudio Team. 2016. Available: <http://www.rstudio.com/>
78. Suraseranivongse S, Santawat U, Kraiprasit K, Petcharatana S, Prakkamodom S, Muntraporn N. Cross-validation of a composite pain scale for preschool children within 24 hours of surgery. *BJA Br J Anaesth*. 2001;87: 400–405. doi:10.1093/bja/87.3.400
79. Schuster C. A note on the interpretation of weighted kappa and its relations to other rater agreement statistics for metric scales. *Educ Psychol Meas*. 2004;64: 243–253. doi:10.1177/0013164403260197
80. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33: 159–174.
81. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull*. 1968;70: 213–220.
82. Altman DG. *Practical statistics for medical research*. 1st ed. London: Chapman and Hall; 1991. doi: 10.1002/sim.4780101015
83. Kaiser HF. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*. 1958;23: 187–200. doi:10.1007/BF02289233
84. Streiner D, Norman G, Cairney J. *Health measurement scales: A practical guide to their development and use*. 5th ed. New York: Oxford University Press Inc; 2015. doi:10.1093/acprof:oso/9780199231881.003.0006
85. Jensen MP. Questionnaire validation: a brief guide for readers of the research literature. *Clin J Pain*. 2003;19: 345–352.
86. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J Pers Assess*. 2003;80: 99–103. doi:10.1207/S15327752JPA8001_18
87. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures statistics and strategies for evaluation. *Control Clin Trials*. 1991;12: 142–158. doi:10.1016/S0197-2456(05)80019-4

88. Streiner DL, Cairney J. What's under the ROC? An introduction to receiver operating characteristics curves. *Can J Psychiatry*. 2007;52: 121–128. doi:10.1177/070674370705200210
89. Mallat J, Meddour M, Durville E, Lemyze M, Pepy F, Temime J, et al. Decrease in pulse pressure and stroke volume variations after mini-fluid challenge accurately predicts fluid responsiveness. *Br J Anaesth*. 2015;115: 449–456. doi:10.1093/bja/aev222
90. Cannesson M, Le Manach Y, Hofer CK, Goarin JP, Lehot J-J, Vallet B, et al. Assessing the diagnostic accuracy of pulse pressure variations for the prediction of fluid responsiveness: a “gray zone” approach. *Anesthesiology*. 2011;115: 231–241. doi:10.1097/ALN.0b013e318225b80a
91. Gall O, Champigneulle B, Schweitzer B, Deram T, Maupain O, Montmayer Verchere J, et al. Postoperative pain assessment in children: a pilot study of the usefulness of the analgesia nociception index. *Br J Anaesth*. 2015;115: 890–895. doi:10.1093/bja/aev361
92. Celeita-Rodríguez N, Teixeira-Neto FJ, Garofalo NA, Dalmagro TL, Giroto CH, Oliveira GCV, et al. Comparison of the diagnostic accuracy of dynamic and static preload indexes to predict fluid responsiveness in mechanically ventilated, isoflurane anesthetized dogs. *Vet Anaesth Analg*. 2019;46: 276–288. doi:10.1016/j.vaa.2018.12.004
93. Trindade PHE, Hartmann E, Keeling LJ, Andersen PH, Ferraz GC, da Costa MJRP. Effect of work on body language of ranch horses in Brazil. Llor JJ, editor. *PLoS One*. 2020;15: e0228130. doi:10.1371/journal.pone.0228130
94. Pinho RH, Leach MC, Rocha FDL, Armani DJC, Ferreira-Souza G, Dissenha A, et al. Influence of the presence or absence of an observer on pain assessment in rabbits submitted to orthopaedic surgery. *AVA autumn meeting*. Ghent; 2019. p. 120.
95. Sorge RE, Martin LJ, Isbester KA, Sotocinal SG, Rosen S, Tuttle AH, et al. Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nat Methods*. 2014;11:629–32. doi:10.1038/nmeth.2935
96. Leach MC, Allweiler S, Richardson C, Roughan J V, Narbe R, Flecknell PA. Research in veterinary science behavioural effects of ovariohysterectomy and oral administration of meloxicam in laboratory housed rabbits. *Res Vet Sci*. 2009;87: 336–347. doi:10.1016/j.rvsc.2009.02.001
97. Lester LS, Fanselow MS. Exposure to a Cat Produces Opioid Analgesia in Rats. *Behav Neurosci*. 1985;99: 756–759.
98. Adams J. Assessment and management of pain in small ruminants. *Livestock*. 2017;22: 324–328. doi:10.12968/live.2017.22.6.324
99. Manteca X, Temple D, Mainau E, Llonch P. Évaluation de la douleur chez les ovins. Farm animal welfare education centre. janvier 2017; 1–2. Available from: https://www.fawec.org/media/com_lazy/pdf/pdf/Fiche_Technique_FAWEC_n17_fr.pdf
100. Teixeira PPM, Padilha LC, Oliveira MEF, Motheo TF, da Silva ASL, Barros FFPC, et al. Laparoscopic ovum collection in sheep: gross and microscopic evaluation of the ovary and influence on oocyte production. *Anim Reprod Sci*. 2011;127: 169–175. doi:10.1016/j.anireprosci.2011.08.001
101. Paull DR, Lee C, Atkinson SJ, Fisher AD. Effects of meloxicam or tolfenamic acid administration on the pain and stress responses of Merino lambs to mulesing. *Aust Vet J*. 2008;86: 303–311. doi:10.1111/j.1751-0813.2008.00325.x
102. Mellor DJ, Stafford KJ, Todd SE, Lowe TE, Gregory NG, Bruce RA, et al. A comparison of catecholamine and cortisol responses of young lambs and calves to painful husbandry procedures. *Aust Vet J*. 2002;80: 228–233. doi:10.1111/j.1751-0813.2002.tb10820.x

103. Dinniss AS, Stafford KJ, Mellor DJ, Bruce RA, Ward RN. The behaviour pattern of lambs after castration using a rubber ring and/or castrating clamp with or without local anaesthetic. *N Z Vet J*. 1999;47: 198–203. doi:10.1080/00480169.1999.36143
104. Currah JM, Hendrick SH, Stookey JM. The behavioral assessment and alleviation of pain associated with castration in beef calves treated with flunixin meglumine and caudal lidocaine epidural anesthesia with epinephrine. *Can Vet J = La Rev Vet Can*. 2009;50: 375–382.
105. Millman ST. Behavioral responses of cattle to pain and implications for diagnosis, management, and animal welfare. *Vet Clin North Am Food Anim Pract*. 2013;29: 47–58. doi:10.1016/j.cvfa.2012.11.007
106. Watts, Stookey. Vocal behaviour in cattle: the animal's commentary on its biological processes and welfare. *Appl Anim Behav Sci*. 2000;67: 15–33. doi:10.1016/s0168-1591(99)00108-2
107. Ferguson DM, Lee C, Fisher A. *Advances in sheep welfare*. 1st ed. Melbourne: Woodhead publishing; 2017. doi:10.1016/C2015-0-00983-2
108. Ison SH, Clutton RE, Di Giminiani P, Rutherford KMD. A review of pain assessment in pigs. *Front Vet Sci*. 2016. p.108. doi: 10.3389/fvets.2016.00108
109. Murrell JC, Psatha EP, Scott EM, Reid J, Hellebrekers LJ. Application of a modified form of the Glasgow pain scale in a veterinary teaching centre in the Netherlands. *Vet Rec*. 2008;162: 403–408. doi:10.1136/vr.162.13.403
110. DeVon HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ, Lazzara DJ, et al. A psychometric toolbox for testing validity and reliability. *J Nurs Scholarsh*. 2007;39: 155–164. doi:10.1111/j.1547-5069.2007.00161.x
111. Boateng GO, Neilands TB, Frongillo EA. Best practices for developing and validating scales for health, social, and behavioral research: a primer. 2018;6: 1–18. doi:10.3389/fpubh.2018.00149
112. Kaler J, Wassink GJ, Green LE. The inter- and intra-observer reliability of a locomotion scoring scale for sheep. *1129 Vet J*. 2009;180(2):189–94.
113. Angell JW, Cripps PJ, Grove-White DH, Duncan JS. A practical tool for locomotion scoring in sheep: reliability when used by veterinary surgeons and sheep farmers. *Vet Rec*. 2015;176: 521. doi:10.1136/vr.102882
114. Pasquali L. Psychometrics. *Ver. esc. enferm.USP*. 2009;43: 992–999. doi:10.1590/S0080-62342009000500002
115. Welsh EM, Gettinby G, Nolan AM. Comparison of a visual analogue scale and a numerical rating scale for assessment of lameness, using sheep as a model. *Am J Vet Res*. 1993;54: 976–983.
116. Gracely RH. Evaluation of multi-dimensional pain scales. *Pain*. 1992; 48: 297–300. doi:10.1016/0304-3959(92)90076-n
117. Netemeyer R, Bearden W, Sharma S. *Scaling procedures: issues and applications*. London: Sage Publications; 2003. doi:10.4135/9781412985772
118. Izer JM, LaFleur RA, Weiss WJ, Wilson RP. Development of a pain scoring system for use in sheep surgically implanted with ventricular assist devices. *J Investig Surg*. 2018;1939: 1–10. doi:10.1080/08941939.2018.1457191
119. Wyse CA, Zhang X, McLaughlin M, Biello SM, Hough D, Bellingham M, et al. Circadian rhythms of melatonin and behaviour in juvenile sheep in field conditions: Effects of photoperiod, environment and weaning. *Physiol Behav*. 2018;194: 362–370. doi:10.1016/j.physbeh.2018.06.001
120. Roughan J V, Flecknell PA. Evaluation of a short duration behaviour-based post-operative pain scoring system in rats. *Eur J Pain*. 2003;7: 397–406. doi:10.1016/S1090-3801(02)00140-4

Supporting information

S1 Table. Ethogram with the description of the behaviors analyzed in 48 sheep submitted to laparoscopy [8,28,31,32,52–66].

Behavioral category	Description
Eat	Act of eating feed
Ruminate	Act of ruminating
Drink	Act of drinking water
Urinate	Act of urinating
Defecate	Act of defecating
Normal interaction	Active, attentive to the environment, interacts with and/or follows other animals.
Reduced interaction	Apathetic: may remain close to other animals, but interacts little.
Absent interaction	Very apathetic: is isolated or does not follow other animals, is not interested in the environment.
Normal locomotion	Moves about freely, without altered locomotion; when stopped, the pelvic limbs are parallel to the thoracic limbs.
Reduced/altered locomotion	Moves about with restriction and/or short steps and/or pauses and/or lameness with limb support; when stopped.
Absent/abnormal locomotion	Reluctant to get up or gets up with difficulty; does not move or demonstrates unstable or rigid walking and/or limps with little or no limb support; may walk backwards or walk in a circle or leans against a surface or fall.
Head above the withers or eat	Raised head position or eats.
Head at the height of the withers	Head is positioned at the same level of the withers
Head below the withers	Head down
Standing still in normal posture	Quadrupedal position when not moving
Standing in altered posture	Quadrupedal position with arched back and pelvic members rigid and/or caudally extended (may be more open and further back than normal).
Kick and stamp the limbs on the ground	Kicks the limbs while standing
Lying down with extension of the head and neck and/or limb(s)	Stretches the head and neck and thoracic or pelvic limbs when in recumbence
Lying down	Lying in ventral or lateral recumbence.
Lying down with head turned back	Lying down with head turning back
Lying with head supported on or close to the ground	Lying down with head close or supported on the ground.
Quick and repeated tail movements	Energetic tail wagging (except when breastfeeding)
Keep the tail straight (not urinating or defecating).	Tail stretched (not urinating or defecating)
Arch the back	Back arching.
Body tremors	Back and abdomen (mostly) trembling.
Body rotation	Rotates body partially or totally, without getting up
Attention to the affected area	Does not look
	Turns the head and looks
	Licks or tries to lick or avoids contact of the area with surfaces or other animals

S2 Table. Criteria used to select the behaviors included in the pre-refinement Unesp Botucatu scale used for video analysis (S3 Table), based on content validity and behaviors reported in the literature.

Variable	Criterion	Ev.1	Ev.2	Ev.3	Score	Ethogram	Literature
Interaction	Active, attentive to the environment, interacts and/or follows other animals	1	1	1	1	x	[8,28,57]
	Apathetic: may remain close to other animals, but interacts little	1	1	1	1	x	[28,57]
	Very apathetic: isolated or does not follow other animals, not interested in the environment	1	1	1	1	x	[28,57]
Activity	Moves normally	1	1	1	1		[8,28,55,59]
	Restless, moves more than normal; or lies down and gets up frequently.	1	1	1	1		[8,28,55,57]
	Moves less frequently or only when stimulated or does not move.	1	1	1	1		[28,57]
Locomotion	Moves about freely, without altered locomotion; when stopped, the pelvic limbs are parallel to the thoracic limbs	1	1	1	1	x	[8,28,55,57]
	Moves about with restriction and/or short steps and/or pauses and/or lameness; when stopped, the thoracic or pelvic limbs may be more open and further back than normal.	1	1	1	1	x	[8,28,55,57,59]
	Reluctant to get up or gets up with difficulty; does not move or has an unstable or rigid gait and/or limps with little or no limb support; walks backwards or on knees; jumps like a rabbit, walks in a circle, leans against a surface or fall	1	0	1	0.66	x	[8,28,55,57,59]
Head position	Above the withers or eating	1	0	1	0.66	x	[28,55]
	At the height of the withers	1	0	1	0.66	x	[28,55]
	Below the withers.	1	0	1	0.66	x	[28,55]
Posture	Kicks and stamps the limbs on the ground	1	1	1	0.33	x	[28,55,57,59]
	Extends the head and neck and/or one or more limbs	1	0	1	0.66	x	[28,55]
	Body tremors (without considering the ears)	1	1	0	0	x	[28,57]
	The pelvic limbs may be more open and further back than normal	1	1	1	1	x	[8,28,57]
Posture in recumbence	Lying down, relaxed	1	1	1	1	x	[28,57]
	Extends neck and body forward	1	1	0	0.66	x	[8,28,57]
	Lying down with the head resting on the ground or close to the ground.	1	1	1	1		[8,28,55,57]
Appetite	Normorexia and/or rumination present.	1	1	1	1	x	[8,28,57]
	Hyporexia.	1	0	1	0.66		[8,28,57]
	Anorexia.	1	1	1	1		[8,28,57]
Miscellaneous behaviors	Moves the tail quickly and repeatedly (except when breastfeeding) or keeps the tail straight (except when defecating or urinating)	1	0	1	0.66	x	[8,28,57]
	Arched back	1	1	1	1	x	[8,28,55,57]
	Rotates body partially or totally, without getting up	1	0	0	0	x	[28,57]
Attention to the affected area	Does not look	1	1	1	1	x	[28,57]
	Turns the head and looks	1	1	1	1	x	[8,28,57]
	Licks or tries to lick or avoids contact of the area with surfaces or other animals	1	1	1	1	x	[8,28,57]

S3 Table. Pre-refinement Unesp-Botucatu scale to assess postoperative pain in sheep submitted to video analysis after content validation.

Principal item	Sub-items (with sub-divisions in some sub-items)	video 1	video 2	video 3	video 4
Interaction	A. Active, attentive to the environment, interacts and/or follows other animals.				
	B. Apathetic: may remain close to other animals, but interacts little.				
	C. Very apathetic: isolated or does not follow other animals, not interested in the environment.				
Activity	A. Moves normally.				
	B. Restless, moves more than normal; or lies down and gets up often.				
	C. Moves less frequently or only when stimulated or does not move.				
Locomotion	A. Moves about freely, without altered locomotion; when stopped, the pelvic limbs are parallel to the thoracic limbs.				
	B. Moves with restriction and/or short steps and/or pauses and/or lameness; when stopped, the pelvic limbs may be more open and further back than normal.				
	C. 1.1 Reluctant to get up and/or				
	1.2 Gets up with difficulty.				
	2.1 Does not move and/or				
	2.2 Walks abnormally and/or				
	2.3 Lameness.				
	3.1 Walks backwards and/or				
	3.2 Walks in circles and/or				
	3.3 Leans against a surface.				
Head position	A. Above the withers.				
	B. At the level of the withers.				
	C. Below the withers.				
Posture	1. Kicks or stamps one or more limbs on the ground.				
	2.1. Extends the head and neck and/or				
	2.2. Extends one or more limbs.				
	3. Lying down with the head resting on the ground or close to the ground.				
Appetite	A1. Normorexia and/or				
	A2. Rumination present.				
	B. Hyporexia.				
	C. Anorexia.				
Miscellaneous behaviors	1.1. Moves the tail quickly and repeatedly (except when breastfeeding) and/or				
	1.2. Keeps the tail straight, except when defecating or urinating.				
	2. Arched back.				
	3. Body tremors (without considering the ears).				
	4. Crawls in ventral recumbence, without getting up.				

Ev. = Evaluator; each item was classified as relevant (+1), not known (0) or irrelevant (-1), the values were added and divided by the number of experts. Score: final of content validation. In bold are the items approved in content validation, because mean score was ≥ 0.5 and because the specific pain-related behavior had been reported in the ethogram (x) and literature [78]; in italics are the behaviors that remained on the scale after refinement; “-” = behaviors that did not occur at any moment.

S4 Table. Refinement process for inclusion and exclusion of items and sub-items on the Unesp-Botucatu scale to assess acute postoperative pain

Items / Test	Sum	Etho. M2	CV	%M2 ≥ 15	PCA	Intra	Inter	F (M1xM2)	F (M2xM3)	F (M2xM4)	ITC	IC	Sp/S
Interaction	10		1		1	1	1	1	1	1	1	1	1
Activity	9		1		1		1	1	1	1	1	1	1
Locomotion	10		1		1	1	1	1	1	1	1	1	1
Locomotion 1	3			1		1	1						
Locomotion 2	3			1		1	1						
Locomotion 3	3			1		1	1						
Head position	7		1		1	1	1				1	1	1
Posture	2					1					1		
Posture A	2			1			1						
Posture B1	3	1	1	1									
Posture B2	2		1	1									
Posture C	4		1	1		1	1						
Appetite	9		1		1	1	1	1	1	1	1	1	
Miscellaneous behaviors	4						1	1			1	1	
Miscellaneous A	3		1	1			1						
Miscellaneous B	3	1	1	1									
Miscellaneous C	0												
Miscellaneous D	2					1	1						

Statistical tests: Etho. M2 - highest score of behavior in the Ethogram according the Friedman test at M2 vs the other moments; % M2 ≥ 15 - at least 15% frequency of occurrence of items/sub-items at M2; CV - content validation; Intra – intra-observer reliability; Inter – inter-observer reliability; PCA - Principal component analysis; F (M) - Friedman test between moments; ITC - item-total correlation; IC - Internal consistency; Sp - Specificity; S – Sensitivity. Sum – sum of the 10 statistical tests (Friedman test was divided into 3 tests). Locomotion: 1 - Difficulty and/or reluctance to get up; 2 - Does not move and/or walks abnormally and/or with a limp and/or when standing, the limbs may be more open and further back than normal; 3 - Leans against a surface; Posture: A - Kicks or stamps one or more limbs on the ground; B1 - Extends the head and neck; B2 - Extends one or more limbs; C – In ventral decubitus, the head rests on the ground or is close to the ground in column line; Miscellaneous behaviors: A - Moves the tail quickly and repeatedly and/or keeps the tail straight; B - Arched back; C - Body tremors (without considering the ears); D. Crawls in ventral decubitus, without getting up. Moments - M1: preoperative; M2: postoperative, before rescue analgesia; M3: postoperative, after rescue analgesia; M4: 24h postoperative. 1 - item included according to the criteria of each test. In bold the Items and subitems included in final scale after refinement; The main items were submitted to 12 tests and a minimum of 7 were included in the final scale (Table 3); sub-items were subjected to 5 tests, and a minimum of 3 were included in the final scale (Table 3).