Université de Montréal

Facial expressions of pain in cats: the development and validation of the Feline Grimace Scale

par Marina Cayetano Evangelista

Département de sciences cliniques

Faculté de médecine vétérinaire

Thèse présentée à la Faculté de médecine vétérinaire
en vue de l'obtention du grade de *Philosophiae Doctor* (Ph. D.)

en sciences vétérinaires

Août, 2020

Université de Montréal

Faculté de médecine vétérinaire

Cette thèse intitulée

Facial expressions of pain in cats: the development and validation of the Feline Grimace Scale

Présentée par

Marina Cayetano Evangelista

A été évaluée par un jury composé des personnes suivantes

M. Gustavo Zamberlam
Président-rapporteur

M. Paulo Steagall
Directeur de recherche

Mme Diane Frank
Membre du jury

Mme Kathy Murphy
Examinatrice externe

Résumé

L'évaluation de la douleur chez le chat est souvent un défi en raison de leur nature discrète et les changements de comportement potentiels dans des situations inhabituelles et stressantes, telles que l'environnement vétérinaire. Différents outils d'évaluation de la douleur (c.-à.-d. des échelles de douleur) basés sur l'observation des comportements ont été proposés pour les chats; cependant, la majorité de ces outils manque de tests de validité, de fiabilité et/ou de généralisabilité. De plus, les échelles de douleur sont peu utilisées dans la pratique clinique. Des outils simples, pratiques et fiables tels que les échelles de grimace (instruments d'évaluation de la douleur basés sur l'expression faciale), ont le potentiel de changer ce scénario. Elles ont été développées pour plusieurs espèces, excluant le chat. L'objectif général de cette thèse était de développer un nouvel instrument basé sur l'expression faciale pour l'évaluation de la douleur aiguë chez les chats, la « Feline Grimace Scale » (FGS) et d'explorer ses applications et ses limitations. Nos hypothèses étaient que la FGS permettrait l'identification de la douleur chez les chats avec précision (dans différentes conditions telles que la douleur d'origine naturelle et postopératoire); elle serait valide et fiable (parmi différents évaluateurs); elle serait capable de détecter la réponse aux analgésiques; et finalement, elle pourrait être appliquée en temps réel dans le contexte clinique.

La FGS a été développée et validée en utilisant une approche psychométrique pour détecter la douleur aiguë chez les chats. Cette échelle discriminait entre les chats en douleur de ceux qui ne le sont pas; détectait la réponse à différents analgésiques; et corrélait fortement avec un autre système de notation de la douleur. Une bonne fiabilité inter et intra-observateur a été démontrée, non seulement parmi les vétérinaires, mais aussi parmi les propriétaires de chats, les étudiants vétérinaires et les techniciens en santé animale. L'utilisation de la FGS en temps réel était aussi réalisable. D'autre part, nos résultats suggèrent que le genre de l'évaluateur influencerait l'évaluation de la douleur, car les évaluatrices attribuaient des scores plus élevés que les évaluateurs.

La FGS est un outil valide, fiable et pratique pour l'utilisation potentielle en recherche ou en clinique; en temps réel ou par l'évaluation des images. Elle pourrait être aussi applicable dans une large gamme de conditions douloureuses et par des évaluateurs avec différents niveaux d'expertise, et potentiellement aussi à la maison (par les propriétaires de chats). Cela représente un progrès substantiel dans l'identification et la gestion de la douleur féline, vers les plus hautes exigences en matière de soins vétérinaires.

Mots-clés : Chats, Douleur, Évaluation de la douleur; Expressions Faciales, Félin, Gestion de la douleur; Grimace.

Abstract

Pain assessment in cats is challenging due to a number of reasons, including their discrete nature and potential behavioral changes in unfamiliar and stressful situations, such as the veterinary environment. Different pain assessing instruments (i.e. pain scales) that rely on the observation of behaviors have been proposed for cats; however, the majority lack validity, reliability and/or generalizability testing. Additionally, the adherence to their use in clinical practice is low and warrants improvement. Simple, practical and reliable tools such as grimace scales (facial expression-based pain assessment instruments), have the potential of changing this scenario. They have been developed for several species, among which the cat was not included. The overall aim of this thesis was to develop a novel facial expression-based instrument for acute pain assessment in cats, the Feline Grimace Scale (FGS) and to explore its applications and limitations. Our hypotheses were that the FGS would be able to accurately identify pain in cats (in different conditions such as naturally-occurring or spontaneous and postoperative pain); it would be valid and reliable (among different raters); it would be able to detect the response to analgesics; and its application in real-time in the clinical context would be feasible.

The FGS was developed and validated using a comprehensive psychometric approach to detect acute pain in cats. It has demonstrated a high discriminative ability between painful and non-painful cats; it is capable of detecting the response to different analgesic drugs and it is strongly correlated with another pain scoring system. Furthermore, it demonstrated good inter- and intra-rater reliability, not only among veterinarians, but also among cat owners, veterinary students and nurses (technicians). Real-time scoring using the FGS was proven feasible. On the other hand, our results suggested that the rater gender may influence pain assessment, as female raters assigned higher scores than males.

The FGS is a valid, reliable and practical tool potentially for both research and clinical use in real-time or using image assessment; that may be applicable in a wide range of painful conditions, by raters with different degree of expertise, and potentially at home (by cat owners). This represents a substantial progress in feline pain management, towards the highest standards in veterinary care.

Keywords: Cats; Facial expression; Feline; Grimace; Pain assessment; Pain management.

Table of Contents

List of Tables

List of figures

List of abbreviations

AD: Action descriptor

AU: Action unit

COPS-C/F: Composite Oral and Maxillofacial Pain Scale-Canine/Feline

EQUUS-Donkey-FAP: Equine Utrecht University Scale for Donkey Facial Assessment of Pain

EQUUS-FAP: Equine Utrecht University Scale for Facial Assessment of Pain

FACS: Facial Action Coding System

FerGS: Ferret Grimace Scale

FGS: Feline Grimace Scale

G-CMPS: Glasgow Composite Measure Pain Scale for acute pain in cats

HGS: Horse Grimace Scale

ICC: Intraclass Correlation Coefficient

LoA: Limits of Agreement

MGS: Mouse Grimace Scale

NRS: Numeric Rating Scale

OVH: Ovariohysterectomy

PGS: Piglet Grimace Scale

RbtGS: Rabbit Grimace Scale

RGS: Rat Grimace Scale

SF-UBCPS: Short-form of the UNESP-Botucatu Composite Pain Scale for cats

SGS: Sheep Grimace Scale

SPFES: Sheep Pain Facial Expression Scale

UB-MCPS: UNESP-Botucatu Multidimensional Composite Pain Scale for assessing postoperative pain in cats

VAS: Visual Analog Scale

This thesis is dedicated to all cats. To the cats who contributed to this work, to those who might benefit from the results and to my most faithful furry companion (Panqueca). Thank you for the inspiration.

Acknowledgements

First, I would like to thank my director Paulo Steagall for the support, the mentorship and for believing in my potential. Thank you for this PhD opportunity, for all the knowledge and experiences shared, the advices and encouragement. I would also like to thank the members of the jury, Dr. Gustavo Zamberlam, Dr. Diane Frank and Dr. Kathy Murphy for their time and valuable contribution to this thesis.

Thank you to all the people who have been in any way involved with my education and training at the Université de Montréal, including my former co-director Dr. Daniel Pang, my advisory committee (comité conseil), Dr. Elizabeth O'Toole and Dr. Derek Boerboom, and my pre-doc examination jury (examen de synthèse), Dr. Éric Troncy, Dr. Bertrand Lussier and Dr. Jeff Mogil. I appreciate all the feedback received.

Thank you to my lab mates Hélène Ruel and Ryota Watanabe (and other members of the Feline Anesthesia and Analgesia Research Team) for the friendship, support and collaboration over these four years. Without all of you, this experience would not have been the same.

I also wish to thank the clinicians, technicians, residents, interns, students and staff in the emergency and critical care department at the Centre hospitalier universitaire vétérinarie (CHUV) for contributing in the recruitment of cats during part of this work.

I am grateful to my family and friends for all the love and support throughout this journey. I would not be here without their help, cheer and encouragement.

Finally, I would like to thank the cats who participated in this research.

Introduction

In the late 1800's, Charles Darwin researched the expression of emotions and he proposed that much like other traits found in animals, emotions also evolved and were adapted over time. In his work "The Expression of the Emotions in Man and Animals" (Darwin, 1872), he believed that expressions were unlearned and innate, therefore important for survival. His work looked at facial expressions in animals and humans, and also attempted to point out parallels between their behaviors, as exemplified by the following excerpts:

> "When animals suffer from an agony of pain, they generally writhe about with frightful contortions; and those which habitually use their voices utter piercing cries or groans. […] With man the mouth may be closely compressed, or more commonly the lips are retracted, with the teeth clenched or ground together […] and I have plainly heard the grinding of the molar teeth of a cow which was suffering acutely from inflammation of the bowels. With man the eyes stare wildly as in horrified astonishment, or the brows are heavily contracted […] The circulation and respiration are much affected. Hence the nostrils are generally dilated and often quiver […] The ears through their movements are highly expressive in many animals; but in some, such as man, the higher apes, and many ruminants, they fail in this respect. A slight difference in position serves to express in the plainest manner a different state of mind, as we may daily see in the dog […] The retraction of the ears may be seen in kittens fighting together in their play, and in full-grown cats when really savage. […] Although their ears are thus to a large extent protected when retracted, yet they often get much torn in old male cats during their mutual battles […]" (Darwin, 1872).

From an evolutionary perspective, emotions have evolved through their adaptive value in dealing with fundamental life tasks. Emotions are ways of relating to the environment (i.e. states of readiness for engaging, or not in interaction with that environment) (Oatley et al., 2006). They can be brief in duration and consist of a coordinated set of responses, which comprise different physiological, behavioral, and neural mechanisms (Fox, 2008). They have also been described as the result of evolution because they provided good solutions to ancient and recurring problems that ancestors faced (Ekman, 1992).

For many years, the sole argument-by-analogy was used to define that animals experience pain (Sherwin, 2001). This is based on the principle that if an animal responds to a stimulus in a similar way to humans, it is likely to have had an analogous experience. However, some critics of this theory suggest that inferences about animal emotions are biased by anthropomorphism; and much of the debate has been based on the difficulty in defining emotions and the cognitive requirements for the experience of emotions (Dawkins, 2000). The debate whether animals may or may not experience pain or suffering in the same way as humans dates back to the 17th century, when Descartes argued that animals lack consciousness, thus would not be capable of feeling pain (Cottingham, 2013). Further discussion was carried on around animal consciousness and ability to feel pain, involving theories on neural correlates, specific brain areas and cognitive processes (J. R. Anderson, 2005; Crick & Koch, 1990; Nani et al., 2019). However, only in 2012 was the Cambridge Declaration on Consciousness signed, stating that the presence or absence of specific brain areas does not appear to preclude an organism from experiencing affective states, and that non-human animals have the neuroanatomical, neurochemical, and neurophysiological substrates of conscious states along with the capacity to exhibit intentional behaviors (Low et al., 2012).

Many animals exhibit complex behavioral and physiological changes indicative of the ability to experience pain, which includes disruption in their normal behavior (appearance or disappearance of normal behaviors or adoption of unusual behavioral patterns), suppression of social behavior, emission of distress calls (vocalizations), respiratory and cardiovascular changes, as well as inflammatory responses and release of stress hormones (Sneddon et al., 2014). Proposed criteria that may indicate the potential of a species to feel pain include: 1) Evidence of a central processing of nociception (nervous system and sensory receptors); 2) Physiological changes to noxious stimuli; 3) Displays of protective motor reactions that might include reduced use of an affected area such as limping, rubbing, holding or autotomy; 4) Presence of opioid receptors and displays of reduced responses to noxious stimuli when given analgesics and/or local anesthetics; 5) Demonstrations of avoidance learning and other motivational requirements (revealing cognitive abilities) (Bateson, 1991; Sneddon et al., 2014). Nociceptors are found in most groups of vertebrates. However, the amount of sensory fibers and the complexity of a central processing of information about noxious stimuli may differ between groups (mammals, birds, reptiles, amphibians and fish) (Sneddon, 2014). Mammals, in particular, fill all the criteria cited above and stimuli that are considered painful in humans have been shown to induce similar physiological and behavioral changes in other non-human mammals (Allen et al., 2005; Sneddon et al., 2014), corroborating with Darwin’s ideas of the universality of emotions.

Nowadays, it is widely accepted that pain can be the result of injury, disease, or emotional distress as evidenced by physiological and behavioral changes. According to the updated definition by the International Society for the study of Pain (IASP), pain is *"An unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage"* (Raja et al., 2020)[1]. Additionally, the inability to communicate does not negate the possibility that a human or a non-human animal experiences pain.

Moreover, pain processing includes sensation and perception (Katz & Melzack, 1999). Sensation refers to the lower level of neural and biochemical components of encoding and processing noxious stimuli (nociception), whereas higher cognitive processing, such as interpretation, is associated with perception; however, pain signs vary between species, type of insult or stimulus, stage of development, among other factors (Bateson, 1991; Le Bars et al., 2001). Pain is also described as a complex multidimensional phenomenon that involves three key components. The *sensory-discriminative* component provides information on the onset, location, intensity, quality and duration of the stimulus. The *affective-motivational* component is associated with the autonomic nervous system's responses and disturbs the feeling of well-being in the individual (unpleasantness), thus triggering certain actions and urge to escape the unpleasantness. The affective or emotional components can also be thought as "how it makes the patient feel". Finally, the *cognitive-evaluative* component, which comprises the effects of prior experience, attention and conditioning (Broom, 2001; Melzack & Casey, 1968; D. D. Price, 2000).

As a multidimensional, complex and personal experience, pain has some important practical implications: others cannot see or measure it directly. The patient's self-report is considered the most valid measure of the experience - *"Pain is whatever the experiencing person says it is, existing whenever he says it does"* (McCaffery, 1968); however, in non-verbal humans (i.e. infants or cognitively impaired) and other animals, self-report is not possible. As a result, making inferences about pain in this context is more challenging and requires standardized methods of

[1] The term "aversive" instead of "unpleasant" had been proposed in 2019 in a version of the terminology released for comment; however, the term "aversive" was criticized as being not easily understood, especially by the lay public, and not readily translatable into many languages, for this reason, the IASP task force 2020 decided to keep the simpler term "unpleasant" (Raja et al., 2020).

behavior observations. The observation of pain behaviors (including changes in facial expressions) is considered a valid approach for pain assessment in the absence of self-report (Herr et al., 2011).

Recently, there has been an increased focus on determining and measuring specific pain-related behaviors in cats (Merola & Mills, 2016a, 2016b). Besides, the facial expressions of pain have been recognized in a wide variety of animal species and they were the subject of three recent reviews (Descovich et al., 2017; McLennan et al., 2019; Mogil et al., 2020). Similar features have been identified across different species and the subject has become critical in animal research. These expressions can be objectively assessed using a Facial Action Coding system (FACS) that measures the individual movements or "action units" (AU) of the face that comprise an expression (Ekman & Friesen, 1978). Grimace scales assess changes in facial expression during painful states based on a simplified approach to the FACS. An overview of the literature concerning these methods applied for assessing facial expressions is presented in a dedicated section of this thesis.

Finally, pain in companion animals is commonly undertreated, due to numerous factors including challenges in its recognition and assessment, fear of adverse effects (particularly linked to opioid analgesics), lack of specific training in the subject and limited availability of reliable instruments for pain assessment (i.e. validated pain scales) (Hewson et al., 2006b; Hugonnard et al., 2004; Hunt et al., 2015; Williams et al., 2005). Assessing pain in cats has always been deemed difficult due to their discrete nature and subtle expression of pain, especially in a clinical setting where it may be aggravated by the stress caused by hospitalization (Robertson, 2015).

Although the use of standardized pain scoring tools (pain scales) can decrease subjectivity and bias by observers (Steagall & Monteiro, 2019), studies regarding acute pain assessment instruments in cats are still scarce (Merola & Mills, 2016b). Two validated tools for acute pain in cats have been published in the past decade: the UNESP-Botucatu Multidimensional Composite Pain Scale for assessing postoperative pain in cats (UB-MCPS) (Brondani et al., 2013) and the Glasgow Composite Measure Pain Scale for acute pain in cats (G-CMPS) (Calvo et al., 2014). The latter was updated to include two features of facial expressions (Reid et al., 2017). Yet, several challenges still exist and the adherence to their use is low, with only 8 to 17% of veterinary practices reporting the use of a standardized or formal pain scoring system (Coleman & Slingsby, 2007; Dawson et al., 2017; Hunt et al., 2015). Also, discrepancies exist amongst individuals with different degree of expertise (i.e. pet owners, veterinary students, veterinary nurses/technicians, recently graduated veterinarians and board-certified specialists) regarding their attitudes, perceptions and ability to recognize pain in animals (Barletta et al., 2016; Coleman & Slingsby, 2007; Dohoo & Dohoo, 1998; Doodnaught et al., 2017; Simon et al., 2018; Steagall, Monteiro, Ruel, et al., 2017; Väisänen et al., 2008). Additionally, although validated, each of these tools has its own limitations such as time-consuming implementation, context of use limited to a specific pain stimulus, etc., which are discussed further in this thesis.

The challenges of pain assessment in cats, including the scarcity of validated pain assessment tools, the low adherence to their use and the degree of variability among observers with different levels of expertise inspired the subject of this thesis and consequently led to the development of a new simple and practical facial expression-based pain scoring system, the Feline Grimace Scale (FGS). In the first part of this thesis, a review of the literature regarding the methods of clinical pain assessment in cats (physiological and behavioral assessments); the challenges and limitations to their use; the psychometric principles in the process of validation of pain scales; and the use of facial expressions in pain research, including preliminary results of a systematic review on the psychometric properties of grimace scales is presented. The second part of this thesis presents the research studies performed in order to achieve the overall objectives of developing, validating and testing the applicability of the FGS.

1. Literature review

1.1. Why should we assess pain?

Effective pain management can only be achieved and maintained when the signs of pain can be assessed accurately and reliably (Dobromylskyj et al., 2000). Assessment is the foundation of pain management, and can ultimately impact in animal health and welfare. Pain assessment provides information regarding the severity of the condition. In addition, appropriate pain assessment is critical for guiding treatment decisions (i.e. choice of drugs, doses and intervals of administration). It also allows practitioners, clinicians and researchers to monitor the time course of the pain disorder and assess treatment effects, thus adapting it accordingly (Fillingim et al., 2016). Finally, comprehensive pain assessment can yield clues regarding the pathophysiological mechanisms underlying the pain condition, enabling a more personalized approach to its management (Hui & Bruera, 2014).

1.2. Why is it so difficult to assess pain in cats?

Most cats instinctively hide their pain as a survival mechanism (the cat is, by nature, considered both predator and prey). Their ability to mask signs of pain is possibly related to being a largely solitary species, a characteristic still conserved throughout the process of domestication (Driscoll, MacDonald, et al., 2009). Domestication is known to influence the morphology, behavior and cognitive abilities of a species (Driscoll et al., 2009, Montague et al., 2014), but in the case of the domestic cat, researchers are just starting to understand these modifications. The ancestor to the domestic cat (*Felis sylvestris lybica)* was solitary, nocturnal and intolerant to humans whereas the modern domestic cat seems to be facultatively social and more integrated into the human environment (Driscoll, MacDonald, et al., 2009; Macdonald et al., 2000). The domestication of cats took a different trajectory than that of other animals. For instance, dogs were the earliest domesticate (approximately >14,000 years ago during the hunter-gatherer nomadic period) and they proved useful as guards and as hunters (Morey, 2010). In contrast, cat domestication happened much later, after humans started building houses, farms, and settlements. Additionally, some authors believe that the wildcat would not be the first choice as a house pet by an early agricultural community, and researchers hypothesized that wildcats exploiting human environments were tolerated by people and, over time they gradually diverged from their "wild" relatives (Driscoll, Clutton-Brock, et al., 2009; Driscoll, MacDonald, et al., 2009). It is proposed that it took many years before the cat achieved a position in which selective breeding could help develop the behavioral characteristics desired in a domesticated animal (Beaver, 2003).

The particular behavior of the modern domestic cat has only recently begun to be explored and understood. Cats showed a remarkable ability to adjust their behavior throughout the years to deal with the constraints of being a pet. Their current living conditions provide them with shelter, food, and relief from most predators and disease, but constrain them in terms of the size and density of their living area, access to the outdoors (it may vary for indoor-only and indoor-outdoor cats), and the number and kinds of companions (or lack of it, considering that the removal of reproductive capability is routinely performed for most pet cats) (Bernstein, 2006).

Veterinary practices are continuously changing, now offering clinical behavior services or behavioral medicine, a reflection of the increasing knowledge of the cat as a house pet in the contemporary society. Cat ownership around the world is continuously growing, and in several countries cats outnumbered dogs as pets. In the province of Québec, the estimated number of cats is approximately 1.8 million (versus 1.2 million dogs) (AMVQ, 2020).

Still, veterinary consultations are considered as a source of stress and fear, and up to 58% of cat owners cited the stress associated with visiting the veterinarian as a reason for fewer consultations, mentioning that their cat "hates going to the vet" (Volk et al., 2011). Fear is an emotional response that stimulates an animal to avoid potentially dangerous situations and can impair the ability of assessing pain in a reliable manner. Pain assessment in a veterinary hospital context can be complicated by the fact that many pain-related behaviors, such as withdrawal, are also observed in response to fear (Mathews et al., 2014; Robertson, 2015; Rodan, 2010). It is thus not surprising that veterinarians consider the recognition of pain as one of the most significant barriers to effective pain management.

In a clinical context, hospitalized cats might experience pain due to recent surgery, trauma, or severe illness and it is important to adequately assess and treat these individuals with analgesics to improve their recovery. Studies have examined the prevalence of pain in emergency/intensive care unit showing percentages ranging from 22 to 56% in dogs (Moran & Hofmeister, 2013; Rousseau-Blass et al., 2020; Wiese et al., 2005) and up to 54% in cats (Wiese et al., 2005). Besides that, cats have historically received less analgesics than dogs (Dohoo & Dohoo, 1996b; Muir et al., 2004; Reimann et al., 2017; Simon et al., 2017; Williams et al., 2005). This scenario warrants improvement, which may be obtained by the incorporation of effective pain assessment tools in the clinical practice.

1.3. Methods of clinical pain assessment in cats

At its simplest, pain can be classified according to its duration as acute or chronic. Acute pain, mainly caused by injury, infection or inflammation, is often short-lived and responds well to treatment. Whereas chronic pain is long-lasting beyond the healing process and traditionally an arbitrary interval of time from onset of pain is used to define it (i.e. lasting more than 3 months) (Grichnik & Ferrante, 1991; Mathews et al., 2014; Turk & Okifuji, 2001). Recently, the terms 'adaptive' and 'maladaptive' have been suggested to better describe pain. Adaptive pain includes both nociceptive (protective) and inflammatory pain (Woolf, 2010). Maladaptive pain (also called pathological pain) is not protective and is primarily due to plastic changes in the pain processing system (Adrian et al., 2017; Epstein et al., 2015).

Many tests are used in both experimental pain models and for pain assessment. For example, pain can be assessed by changes in clinical parameters (i.e. heart rate, blood pressure, etc), neuroendocrine and neuroimmune mechanisms (i.e. reflected by changes in cortisol plasma concentrations), reflexive pain tests (i.e. evaluation of evoked responses after the application of thermal, mechanical or electrical stimuli), non-reflexive pain tests (i.e. systematic assessment of spontaneous pain behaviors), levels of activity, quality of life and function (Gregory et al., 2013).

This section provides a literature review on the physiological and behavioral assessment of clinical pain in cats. The focus of this thesis is feline acute pain, although some methods of chronic pain assessment are briefly mentioned.

1.3.1. Physiological assessment

Physiological changes, such as increases in heart rate, blood pressure, plasma cortisol and other chemical mediators, occur in response to sympathetic stimulation caused in part by pain (Smith et al., 1996). Objective clinical variables such as heart rate, systolic blood pressure, respiratory rate and rectal temperature are often unreliable or nonspecific, as similar changes may be observed in other conditions (National Research Council, 2009; Smith et al., 1999). Fear, stress, anesthesia, and pharmacologic interventions may also cause these parameters to change (Saritas et al., 2015; Smith et al., 1996). Moreover, consistent changes in these parameters in cats, dogs and horses expected to be in pain have not been demonstrated (Cambridge et al., 2000; Holton, Scott, Nolan, Reid, & Welsh, 1998; J. Price et al., 2003).

Stress could be defined as the physiological response to an internal or external stimulus that triggers the behavioral fight-or-flight response through the activation of the autonomic nervous system and the endocrine system (release of stress hormones) (Finestone et al., 2008; Ulrich-Lai & Herman, 2009). The acute stress response activates immediate physical reactions associated with preparation for demanding muscular action: increased heart and respiratory rate, bronchodilation, inhibition of gastric and intestinal motility, redirection of blood circulation from intestines to the muscles by vasoconstriction, increased metabolic activity by gluconeogenesis, mydriasis, increased muscular tension and downregulation of the immune system (McCorry, 2007; Oatley et al., 2006).

Even a routine clinical examination of cats at a veterinary clinic setting using low stress handling techniques was shown to increase stress responses (marked by higher blood glucose and more hiding behaviors), as opposed to the same manipulations performed at their home environment (Nibblett et al., 2015). Additionally, significant increases in blood pressure, heart rate, and respiratory rate between the home and veterinary hospital environments were also demonstrated (Quimby et al., 2011).

Given the wide range of factors that can alter heart and respiratory rate, it is not surprising that even handling can cause important changes in these parameters. Recently, more sophisticated analysis of heart rate variability has been proposed as adjunct to pain assessment (Arras et al., 2007; Rietmann et al., 2004). A device, that measures the heart rate variability has been described for use in veterinary patients, having as a fundament the nociception-antinociception index by measuring the activity of the parasympathetic tone in animals under anesthesia (Aguado et al., 2020; Mansour et al., 2017). A high value of the parasympathetic tone reflects an absence of nociception, while a lower value reflects a potentially harmful (or painful) stimulus (Aguado et al., 2020); though, its value in feline practice is yet to be determined.

Other stress and pain indicators include hormonal concentrations and measurement of chemical mediators. Examples of these are changes in plasma hormone concentrations like cortisol, β-endorphins and catecholamines; nevertheless, weak correlations with other parameters in cats (Smith et al., 1999) and conflicting results have been reported. While a few studies demonstrated that cortisol concentration increased in response to surgical stimulation and pain (Brondani et al., 2009; Case et al., 2015; Evangelista et al., 2014; Fox et al., 1998; Mollenhoff et al., 2005; Smith et al., 1999; Srithunyarat et al., 2017), others demonstrated that this increase was not specific to surgery (Benson et al., 1991; Cambridge et al., 2000; Glerum et al., 2001; Ueoka & Hikasa, 2015). The relationship between physiological stress and pain is complex; therefore, endocrine measures can reflect stress responses that may not always be related directly to pain or its severity. For example, anesthesia and bandaging alone have been shown to increase plasma cortisol and β-endorphin concentrations in control cats that did not undergo tenectomy or onychectomy (Cambridge et al., 2000). Anesthesia with ketamine, acepromazine, atropine in cats with or without a transdermal fentanyl patch have also been shown to increase cortisol levels (Glerum et al., 2001); and isoflurane anesthesia alone induced a transient increase in epinephrine concentration (Benson et al., 1991).

Pain is also associated with tissue injury and inflammation. Tissue damage results in inflammation, which is associated with increased mediators including tumor necrosis factors (TNF-α) and pro-inflammatory cytokines (interleukins, IL) such as IL-1β and IL-6 (Murtaugh et al., 1996; Omoigui, 2007; Zhou et al., 1993). Increased cytokines are used as markers of tissue injury and its intensity is directly related to the extent of surgical trauma (Cruickshank et al., 1990; Raeburn et al., 2002). A recent study revealed that some cytokines concentrations differed between cats with severe or minimal oral disease and were associated with the presence of tooth resorption and number of fractured and missing teeth (Watanabe et al., 2019). Additionally, certain cytokines such as interferon-γ, IL-4, IL-6 and IL-8 decreased after dental treatment in cats with severe oral disease (Watanabe et al., 2019). These markers were also associated with other painful conditions in cats, such as degenerative joint disease (Gruen et al., 2017), acute idiopathic cystitis (Parys et al., 2018), and other inflammatory and infectious diseases, including septic shock (O'Halloran et al., 2018; Paltrinieri, 2008; Troia et al., 2020). Although they lie outside the scope of this thesis, the role of cytokines for pain in cats seems promising and deserves to be further investigated along with other potential biomarkers.

1.3.2. Nociceptive thresholds

Nociception is a basic sensory ability that reflects the capacity to respond to potentially damaging stimuli (Dubin & Patapoutian, 2010). Nociceptive pathways connect with brain areas important for motivation, and animals are motivated to avoid the injurious stimulus (Bateson, 1991). These basic motor responses, such as withdrawal responses, occur in response to painful stimuli and are associated with protection from further tissue damage. However, simple withdrawal responses can occur consciously or unconsciously, in awake and in anesthetized animals (Bateson, 1991). Models of nociception date back to the late 19th century with the introduction of the von Frey hair aesthesiometer. They provide information about the sensory-discriminatory aspects of pain. Testing whether animals are able to respond to noxious stimuli is straightforward, and measures of withdrawal reflexes to noxious stimuli have been extensively described to examine mechanisms of pain (Gregory et al., 2013; Sneddon, 2018).

Quantitative sensory testing (QST) is a reliable way of assessing sensory nerve fiber function. In pain research, the somatosensory system can be measured by QST by gradually increasing the intensity of a stimulus such as mechanical pressure, electric current or heat applied to a certain area of the body (Backonja et al., 2009; Siao & Cros, 2003). These methods are considered semi-objective because they depend on the reporting and/or interpretation of the individual's reaction to the stimulus (Backonja et al., 2009). The subjectivity is amplified in animals, and the use of QST is considered more challenging because these methods rely solely on the observer's interpretation of the animal's reaction to the stimulus (Martinez-Arizala, 2003).

In cats, some modalities of QST (mechanical and thermal nociceptive thresholds) have been used for many years in the experimental setting with the goal of assessing opioid efficacy in healthy cats by testing different drugs, doses and routes of administration. (Dixon et al., 2002, 2007;

Ferreira et al., 2011; Lascelles & Robertson, 2004; Millette et al., 2008; Robertson et al., 2003; Steagall et al., 2006). However, studies involving these modalities in clinical studies for pain assessment purposes are scarce. In a clinical setting, mechanical nociceptive testing has been used in combination with subjective assessments of pain (pain scales) in cats undergoing surgery (Benito, Monteiro, Lavoie, et al., 2016; Brondani et al., 2009; Catbagan et al., 2011; Mollenhoff et al., 2005; Polson et al., 2012; Shah et al., 2019; Slingsby et al., 2001). In these studies, the mechanical nociceptive threshold was determined by a calibrated pressure algometer or von Frey filaments and it decreased after ovariohysterectomy or orthopedic surgery. A decrease in the threshold in relation to baseline indicates higher degree of pain or sensitivity, and this response was attenuated by the administration of analgesics (opioids and/or their combination with non-steroidal anti-inflammatory drugs and local anesthetics) (Benito, Monteiro, Lavoie, et al., 2016; Brondani et al., 2009; Catbagan et al., 2011; Mollenhoff et al., 2005; Polson et al., 2012; Slingsby et al., 2001).

On the other hand, QST modalities are often used in the assessment of chronic pain, especially in osteoarthritis (Addison & Clements, 2017; Guillot et al., 2013; Monteiro et al., 2017). These studies aimed to assess the differences between healthy cats and those with osteoarthritis as well as to assess the effect of treatments. Additionally, a recent systematic review and meta-analysis revealed that cats with osteoarthritis have lower punctate tactile threshold (measured by the use of von Frey filaments) and facilitated temporal summation (a modality of dynamic QST, that consists of the administration of a repetitive mechanical stimulus of constant intensity) of pain when compared with healthy cats (Monteiro et al., 2020).

1.3.3. Behavioral assessment

Classical pain tests (i.e. nociceptive withdrawal responses to calibrated stimuli) provide information about the sensory-discriminatory aspects of pain, but not the affective-motivational (emotional) and cognitive-evaluative components. The unavoidable lack of self-reporting in animals impairs the evaluation of the emotional components and makes the overall pain assessment more challenging. However, the lack of clear evidence relating to these features in non-human animals should not be interpreted as an absence of an emotional component (Merola & Mills, 2016b; Sneddon et al., 2014).

Both in non-verbal human patients and other mammals, observation of behavior is crucial, and specific behaviors or behavioral disturbances can be monitored as pain indicators (Hansen, 1997; Herr et al., 2011; Mathews, 2000; Merola & Mills, 2016a; Sneddon et al., 2014). Behaviors such as facial grimacing and guarding may indicate pain, as well as an increase or decrease in vocalizations, display of aggressive behaviors, hiding or escape behavior, changes in routine behavior patterns (i.e. grooming, sleeping), avoiding use of a limb and/or licking the painful region, changes in appetite and in mental status (Muir & Gaynor, 2008; Waran et al., 2007; Wiseman et al., 2001).

Systematic behavioral observation methods provide a valid and reliable way of assessing pain. These methods have been successfully applied with a wide range of pain conditions and are often used in clinical and research contexts (Keefe & Smith, 2002). These observations may be performed in the form of an ethogram, a catalogue or inventory of behaviors or actions exhibited by an animal (P. Martin & Bateson, 2007; Stanton et al., 2015). Previous studies looked at the frequencies and durations of certain behaviors shown by cats when in pain (postoperative pain related to: tenectomy/onychectomy, ovariohysterectomy, and dental extractions) (Cloutier et al., 2005; Waran et al., 2007; Watanabe, Frank, et al., 2020). These authors highlighted that some behaviors, such as shaking of the forepaws, crouching, pawing the face and difficulty in grasping dry food were expressed more often in painful cats than in non-painful ones. Furthermore, an expert consensus listed 25 signs that were considered sufficient to indicate pain in cats (Merola & Mills, 2016a). Among them, reaction to palpation, withdraw/hiding, overall activity decrease, hunched up posture, lower head posture, eyes closed and changes in appetite were also listed in frequently used pain scales such as the UB-MCPS and the G-CMPS (Brondani et al., 2013; Reid et al., 2017), reviewed later in this section.

It must be noted, however, that the intensity and frequency of the behavioral response varies between species; and measuring frequency alone may not discriminate between patients who exhibit different degrees of pain-behavior severity (von Korff et al., 1992). Therefore, a pain behavior measure that better reflects the whole pain experience would be a valuable clinical tool.

1.3.3.1. Pain Scales

Pain scales are often described as pain measurement instruments or tools, typically in the form of a questionnaire or behavior descriptors that receive individual scores; and the final scores enable the interpretation of the pain state (Reid et al., 2018). Interpretation of pain behaviors is an important part of a multidimensional assessment of pain (Fillingim et al., 2016). Several tools have been proposed to assess pain in dogs and cats, including categorical scales (i.e. Mild, Moderate, Severe), visual analog scales (VAS), numerical rating scales (NRS) (Figure 1), and well-validated multidimension composite measures scales. These methods provide simple, efficient, and minimally invasive measures of pain that have been used widely in research and clinical settings (Katz & Melzack, 1999; Mathews et al., 2014; Merola & Mills, 2016b). Facial expressions of pain have been recognized in a wide variety of animal species and the instruments used to assess them are appraised in a dedicated section within this thesis.

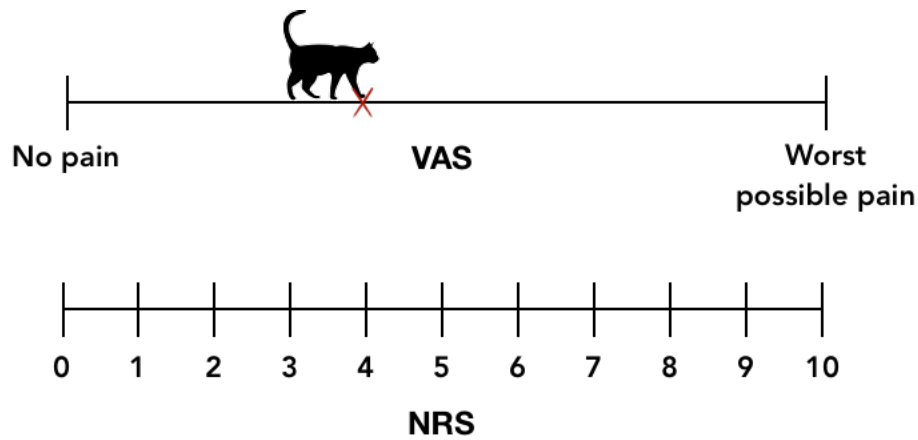


Figure 1. Visual Analogue Scale (VAS) and Numeric Rating Scale (NRS)

The VAS consists of a line, often 100 mm long, with 2 descriptors representing extremes of pain intensity (no pain and extreme pain or worst possible pain) at each end. Users make a mark somewhere along the line that represents the pain intensity, and the VAS is scored by measuring the distance from the "no pain" (0) and the mark. Dynamic and interactive visual analogue scales (DIVAS) are a form of modified VAS. The patient is observed at rest, then handled (stroked, palpated, stimulated to move, etc.) and then a mark is made on the linear scale. The NRS is represented by a series of integers on a line from 0 to 10 in which 0 is no pain and 10 is the worst possible pain.

While relatively easy and straightforward to use, unidimensional pain scales may not be very sensitive in distinguishing subtle changes in pain levels. The VAS has been widely used in human medicine and in several veterinary studies (Cambridge et al., 2000; Holton, Scott, Nolan, Reid, Welsh, et al., 1998; Hudson et al., 2004; Jensen et al., 2003; Slingsby et al., 2001). However, they only assess a limited part of the pain experienced, they lack validity and reliability, and an important shortcoming of these scoring system is that they rely on the experience and familiarity of the user (Holton, Scott, Nolan, Reid, Welsh, et al., 1998).

Methods based on behavior descriptors seem to have some advantage (Jensen & Karoly, 1992). Adjectival descriptors can be more readily employed in scales and have the potential for reducing response bias and increasing reliability (Gracely & Kwilosz, 1988). Descriptors may convey more subtle meanings of pain not readily communicable in VAS or NRS. This feature may be particularly useful in the measurement of the affective component of pain (Morley & Pallin, 1995).

A simple descriptive scale (SDS) usually ranges from 0 to 3 (or 4) and the scores are assigned based on a general observation of the cats' appearance. Pain may be classified into absent, mild, moderate and severe (Merola & Mills, 2016b). Each score may be accompanied by a brief description of the behaviors to score, sometimes with corresponding pictures. For example, (Slingsby et al., 1998) suggested the following scoring system based on both observation and interaction with the cat: 0 = No pain; 1 = Happy cat, purr and friendly, flinch with wound pressure but not with stroke over area; 2 = Happy cat, flinch on wound stroke; 3 = Looks uncomfortable but can touch wound; 4 = Worst possible pain, looks uncomfortable and cannot touch wound, growl and hiss. This system underwent some adaptation over the years by different authors (Polson et al., 2012; Steagall et al., 2009; Taylor et al., 2010).

Other types of descriptive rating scales include the Colorado State University (CSU) feline acute pain scale (Hellyer et al., 2006; Hellyer & Gaynor, 1998), a Composite Pain Scale (CPS) (Al-Gizawiy & Rudé, 2004), and a Total Clinical Score (TCS) (Kamata et al., 2012). The CSU pain scale includes psychological and behavioral components that are evaluated in combination with the cat's response to palpation of the site of surgery and its body tension (Hellyer et al., 2006). Al-Gizawiy & Rudé (2004) compiled and adapted a CPS using the CSU pain scale as the main template, and using other behaviors suggested as indicative of pain in cats and dogs (Firth & Haldane, 1999; Hansen, 1997; Hellyer & Gaynor, 1998; Mathews, 2000). This adapted scale involved observation of temperament (defined as friendly, confident, mildly aggressive or outwardly aggressive), appearance, body posture, comfort level, unprovoked and interactive behavior, movement and vocalization; the total score ranges from 0 to 21. Finally, the TCS, used the sum of scores for posture, behavior and pain on palpation, resulting in a maximum possible score of 9 (Kamata et al., 2012).

However, none of these tools mentioned so far were validated, and there is growing recognition of the importance of assessing the multiple dimensions of pain. Validated tools for acute pain in cats have been published during the past decade, including the UNESP-Botucatu Multidimensional Composite Pain Scale for assessing postoperative pain in cats (UB-MCPS) (Brondani et al., 2013) and the Glasgow Composite Measure Pain Scale for acute pain in cats (G-CMPS) (Calvo et al., 2014; Reid et al., 2017); also the partially validated Composite Oral and Maxillofacial Pain Scale-Canine/Feline (COPS-C/F) (Della Rocca et al., 2019).

The UB-MCPS is a scale used to assess postoperative pain levels in cats by assessing pain expression, psychomotor changes and physiological variables, which adds up to a score out of 30. The scale was originally devised in Brazilian Portuguese (Brondani et al., 2012), and its translations were further validated in English (Brondani et al., 2011, 2013), Spanish (Brondani et al., 2014), French (Steagall, Monteiro, Lavoie, et al., 2017) and Italian (Della Rocca et al., 2018). It has been thoroughly validated using psychometric approaches to confirm its validity and reliability (for the complete scale and each of the subscales). The scale consists of 10 questions on miscellaneous behaviors, reaction to palpation of the surgical wound, reaction to palpation of the abdomen/flank, vocalization, posture, comfort, activity, attitude, arterial blood pressure and appetite (0-3 each). Additionally, an intervention score (threshold for rescue analgesia) was defined (≥ 8 out of 30 or ≥ 7 out of 27, if the blood pressure is not assessed) (Brondani et al., 2013). A short version of this scale (Short-form of the UNESP-Botucatu Composite Pain Scale for cats; SF-UBCPS) was proposed and is currently being validated. It contains short descriptors structured as four items to evaluate the cats' posture, activity, attitude and reaction to touch and palpation of a painful site. Each item is scored from 0 to 3 adding up to a maximum score of 12 points. The threshold for rescue analgesia administration was determined at ≥ 4 out of 12 (Luna et al., 2020).

The G-CMPS in its first (prototype) version ranged from 0 to 22 points and was further revised (rCMPS-f) to comprise 6 questions regarding the cat's attitude, posture, attention to any wound or painful area, response to stroking and palpation and general impression. Each question had 2 (0-1) to 5 response options (0-4) and the final score added up to 16 points (Calvo et al., 2014). The intervention score that indicates the need for additional analgesia was defined as ≥ 4 out of 16. This scale has been once more updated to include two features of facial expressions (ears and muzzle, defined by Holden et al. (2014), increasing the discriminatory ability in its definitive version with misclassification rates decreasing from 26.7% to 17.6%) (Reid et al., 2017). The original tool was partially validated (construct and criterion validity) and provided some evidence for responsiveness. Its sensitivity was moderate (improved by the addition of the facial expressions); however, its reliability remains undetermined (Calvo et al., 2014; Reid et al., 2017). A practical approach to pain assessment in cats using these instruments is illustrated in Figure 2.

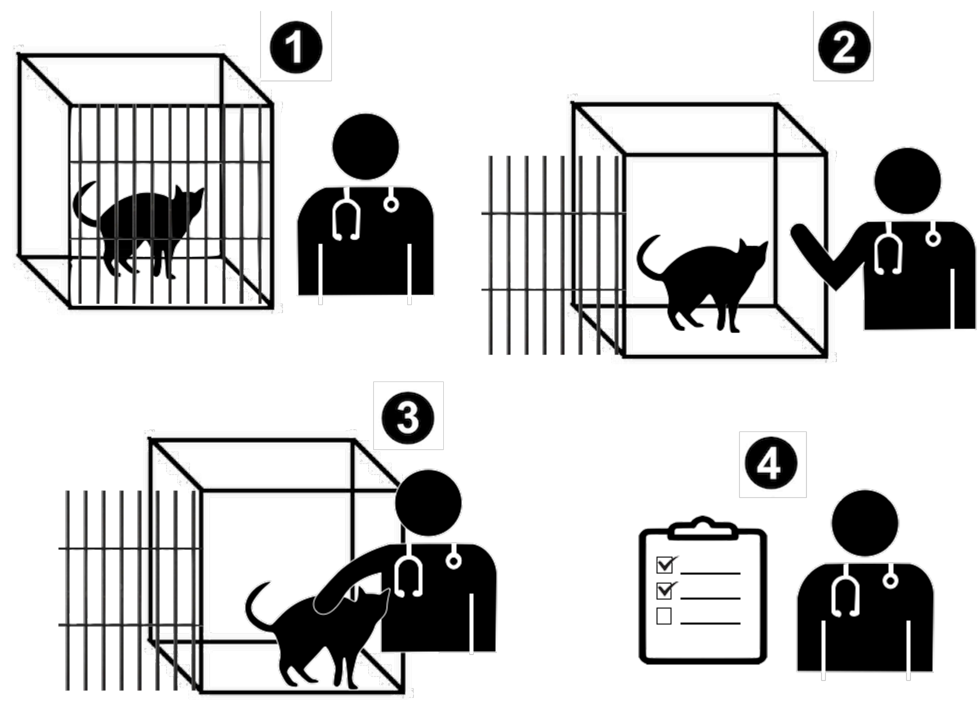


Figure 2. Practical approach to pain assessment in cats in a clinical setting

The general principle of utilization of composite pain scoring instruments is: 1) observation of the cat undisturbed in its cage; 2) approaching and opening the cage, calling the cat's name and stimulating it to move; 3) interaction - the cat is gently handled and stroked the cat (determining its receptiveness) and then the area around the wound or painful area is gently palpated; 4) the response is recorded (scores assigned) according to the instrument used [Steps adapted from (Epstein et al., 2015; Steagall & Monteiro, 2019)].

More recently, a tool specifically designed for assessing oral pain was developed. The Composite Oral and Maxillofacial Pain Scale-Canine/Feline (COPS-C/F) (Della Rocca et al., 2019) was initially developed in Italian and then translated to English. The COPS-C/F consists of two parts, one owner specific questionnaire containing 6 questions and another veterinary specific questionnaire with 4 questions, including examination of the oral cavity (Della Rocca et al., 2019).

In contrast to acute conditions, chronic pain is of longer duration and may also be present in the absence of ongoing clinical disease, persisting beyond the expected course of an acute disease process (Grichnik & Ferrante, 1991; Mathews et al., 2014). Designing instruments to measure chronic pain is considered more challenging than designing those for acute pain, because of the long-lasting painful condition interference in function, behavior and welfare (Reid, Wiseman-Orr, et al., 2013). Many of the instruments for assessment of chronic pain focus on measuring the effect of the pain on quality of life (QoL) (Epstein et al., 2015; Monteiro & Steagall, 2019; Reid, Wiseman-Orr, et al., 2013).

Several standardized instruments for chronic pain are available for cats. For example, general QoL instruments (questionnaires) such as the VetMetrica Health-related QoL instrument for cats with chronic pain (Noble et al., 2019); the Cat Health and Wellbeing (CHEW) instrument (Freeman et al., 2016) and the Feline QoL measure instrument (Tatlock et al., 2017). And others were developed for cats with specific diseases such as: diabetes (Niessen et al., 2010); cardiac disease (Freeman et al., 2012); chronic kidney disease (Bijsmans et al., 2016); skin disease (Noli et al., 2016) and cancer (Lynch et al., 2011). Additionally, several tools were developed specifically for the assessment of the progression and quality of life impairment caused by degenerative joint disease/osteoarthritis. Examples include the Feline Musculoskeletal Pain Index (FMPI) (Benito et al., 2012); Client-Specific Outcome Measures (CSOM) (Lascelles et al., 2007); the Montreal Instrument for Cat Arthritis Testing (MiCAT) (Klinck et al., 2015; Klinck, Gruen, et al., 2018; Klinck, Monteiro, et al., 2018); and the Feline Musculoskeletal Pain Screening Checklist (Feline MiPSC) (Enomoto et al., 2020).

Most of these instruments rely on owner rather than veterinarian assessment. One rare exception, the MiCAT(v) requires the veterinarian's assessment of the cat's body posture, gait, willingness and ease to jump, as well as other behaviors (Klinck et al., 2015; Klinck, Monteiro, et al., 2018). The caretaker version, MiCAT(c) takes into account the owner's evaluation of the cat's agility, social play and exploratory behaviors, self-maintenance and general physical condition (Klinck, Gruen, et al., 2018). The owner is considered the preferred proxy evaluator in cases of chronic pain because behavioral changes may be so subtle and gradual in onset that they may be apparent only to someone very familiar with the individual. These changes may not be obvious in a clinical setting where they may be masked by fear, excitement or anxiety associated with the unfamiliar environment (Reid et al., 2018). Additionally, the current pain state (during the veterinary consultation) may not accurately reflect the cat's overall pain experience over a given time period (i.e. in the past 24 hours or the past week).

1.4. Challenges in the applicability of pain scales in a clinical context

The use of pain scoring tools can decrease subjectivity and bias by observers (Steagall & Monteiro, 2019), resulting in more accurate pain assessment and more effective pain management, which leads to better patient care (Epstein et al., 2015); however, this is not without limitations. Pain assessment instruments have limitations to their effectiveness in assessing animal pain, including a lack of specificity in identifying pain over other negative internal states (i.e. fear, anxiety), limited context of use, extent or degree of validation, reliance on knowledge (and expertise) of species-specific behavior, inherent biases by observers, and in some cases being time consuming to develop and implement (Descovich et al., 2017; Reid, Scott, et al., 2013; Robertson, 2015).

This section provides an overview of the bias, limitations and challenges surrounding the use of pain scoring systems. First, limiting factors related to the instrument are listed, then animal-related sources of variability, user-related (observers or raters) sources of bias, and finally, other sources of variability are reviewed.

1.4.1. Instrument-related limitations

Some of the current acute pain scales were developed for a given context (i.e. postoperative or oral pain), thus they would not be suitable for other sources or types of pain. For example, the UB-MCPS was specifically designed for assessing abdominal pain related to ovariohysterectomy

(Brondani et al., 2013) and the COPS-C/F, for oral pain (Della Rocca et al., 2019); while the G-CMPS could be applied to a broader context of use (Reid et al., 2017). The UB-MCPS is considered to be more time-consuming (contains longer questions/descriptors) and is equipment reliant (for blood pressure monitoring) when all its sections are evaluated (Brondani et al., 2013). Its short version (SF-UBCPS) has been preliminary validated using video-based assessments; however, these results are awaiting publication. The G-CMPS in its original version was considered more accessible (i.e. received positive user feedback from the participants involved in its testing) and the time taken for the completion and calculation of the final score was short (Calvo et al., 2014). The COPS-C/F was initially developed and preliminarily validated in Italian, and its English version requires further testing in the clinical setting (Della Rocca et al., 2019).

While the UB-MCPS was considered the only thoroughly validated instrument, the G-CMPS's validity was only partially established by the time of the publication of a systematic review of the behavioral assessment of pain in cats (Merola & Mills, 2016b); and its inter-rater reliability still remains undetermined. Additionally, the decision to give additional analgesia (when the score exceeds the analgesic threshold) may change according to the instrument used. For example, in a previous study involving both the UB-MCPS and the G-CMPS for assessing postoperative pain in cats, one scale indicated the requirement for rescue analgesia for some animals while the other did not (Steagall et al., 2018).

1.4.2. Animal-related sources of variability

Cats often become anxious and/or fearful when introduced to a novel or unfamiliar environment (i.e. veterinary hospital) and new people, (Griffin & Hume, 2006; Rodan, 2010). These circumstances can impair pain assessment, despite of the choice of pain scoring instrument (Robertson, 2015). Likewise, pain assessment becomes more challenging in a stoic animal. It has been previously demonstrated that demeanor confounds pain assessment, as pain scores were falsely increased in shy, fearful or aggressive cats (classified according to their higher demeanor scores), even crossing the analgesic intervention thresholds (Buisman et al., 2017). This was observed using both the UB-MCPS and G-CMPS; though, the changes were only observed in the UB-MCPS psychomotor subscale and not in the UB-MCPS pain expression subscale (Buisman et al., 2017). In humans, changes in mood and emotional state have a significant impact on the resultant pain perception and ability to cope (Tracey & Mantyh, 2007). For example, anticipating and being anxious about pain can exacerbate the pain experienced (by enhancing pain perception) (Reicherts et al., 2017).

The presence of coping responses may also interfere with pain assessment (Keefe & Dunsmoret, 1992). For example, certain behaviors commonly considered as pain behaviors, such as rubbing of the painful area, may be present as a coping mechanism to modulate pain by means of descending inhibitory mechanisms (as explained by the gate control theory) (Melzack & Wall, 1965; Wall, 1978). Purring may be an example of coping behavior in cats. Purring is often considered to indicate a positive emotional state (Tavernier et al., 2020); however, cats sometimes purr when they are ill, tense, or experiencing traumatic or painful situations (Leyhausen, 1979; McComb et al., 2009). Moreover, some authors suggest that the low acoustic frequency of purrs (24-140 vibrations per minute) causes a series of related vibrations within their body that can serve as a form of pain relief and to speed up the healing process (von Muggenthaler, 2001). In practice, purring is evaluated differently in the G-CMPS and UB-MCPS. In the G-CMPS, purring is scored alongside with "silent" or "meowing" and receives a score of 0 (Reid et al., 2017), while in the UB-MCPS, if the cat is quiet and purring when stimulated, it receives a score of 0, but if the cat purrs spontaneously (without being stimulated or handled), it receives 1 point (Brondani et al., 2013). This may confound the observer, especially in the case of novice raters, leading to erroneous interpretation of the observed behavior.

1.4.3. User-related variability

In veterinary practice, the rater completes the pain scoring systems (or scales) after a few moments of observation and interaction with the cat (as exemplified in Figure 2). All scoring systems that depend on human observers have some intrinsic degree of subjectivity and leave room for error (i.e. under- or overestimating pain levels) (Robertson, 2015). This may bias the pain assessment in two ways, first, the presence of the observer in close proximity may change the patient reaction (observer effect) and second, the individual characteristics of the raters may influence their evaluation (observer bias).

The presence of the observer in close proximity may trigger a number of different responses. Rodents can detect the presence of humans by smell, and their behavior may be affected. Rabbits and guinea pigs may remain immobile in the presence of an observer (freezing response) (Leach et al., 2009; National Research Council, 2009). It has been also demonstrated that the presence of a male observer in the room produced stress-induced pain inhibition in mice, whereas the same was not observed with female observers (Sorge et al., 2014). Moreover, pain (after ovariohysterectomy) and analgesia have been shown to alter the non-interactive and interactive behaviors in bitches (i.e. in the absence of or presence of a caregiver); however, the observer effect has not been systematically assessed in these studies (S. M. Fox et al., 2000; Hardie et al., 1997; Kyles et al., 1998). In cats, it has been reported anecdotally that the presence of the observer may change the cat's behavior (Mathews et al., 2014; Robertson, 2015); however, studies are lacking to confirm these assumptions using the aforementioned pain assessment tools (UB-MCPS and G-CMPS). The observer effect (or lack of it in some situations) has been demonstrated using grimace scales and is discussed later in this thesis.

Observer bias has been studied more extensively. Individual characteristics of the observers (i.e. gender, empathy levels, degree of training, specialist status and year of graduation from veterinary school) have been shown to influence their perception and attitudes towards pain and use of analgesics in dogs and cats (Bartley & Fillingim, 2013; Bell et al., 2014; Beswick et al., 2016; Coleman & Slingsby, 2007; Dohoo & Dohoo, 1996a; Doodnaught et al., 2017; Hewson et al., 2006b; Hugonnard et al., 2004; Simon et al., 2018; Steagall, Monteiro, Ruel, et al., 2017; Väisänen et al., 2008; Williams et al., 2005).

Literature in humans is extensive regarding sex and gender differences in responses to pain (Bartley & Fillingim, 2013). A distinction has been made among the terms "sex" and "gender". Sex refers to a set of biological attributes in humans and animals, while gender refers to the socially constructed roles, behaviours, expressions and identities. Research on pain has typically only included masculinity and femininity. The gender spectrum considers gender as a continuum stretching from men to women and masculine to feminine, including androgyny. Studies considering trait gender rather than the dichotomous masculine/feminine classification are rare, and future research should explore these premises that sex, trait gender and state gender can play a role in pain experience (Fillingim, 2017; Keogh, 2014; R. M. Martin, 2019).

Research has shown that females are frequently more empathic towards the pain and distress of others, and often assign higher pain scores than males (Christov-Moore et al., 2014; Sadeghiyeh et al., 2012). Likewise, female veterinarians have been more likely to assess pain and administer analgesics for dogs and cats (Beswick et al., 2016; Dohoo & Dohoo, 1996b; Doodnaught et al., 2017; Hewson et al., 2006a; Williams et al., 2005).

Concerning the levels of expertise, discrepancies have been demonstrated between pet owners, veterinary students, nurses/technicians and veterinarians in regard to their attitudes, perceptions and ability to recognize pain in animals (Coleman & Slingsby, 2007; Dohoo & Dohoo, 1998; Doodnaught et al., 2017; Simon et al., 2018; Steagall, Monteiro, Ruel, et al., 2017; Väisänen et al., 2008). Pet owners frequently disagree with statements that pain assessment in animals is easy and that they had received sufficient information on appropriate methods of animal pain management (Steagall, Monteiro, Ruel, et al., 2017; Väisänen et al., 2008). This indicates a gap in communication between veterinarians and clients, and could definitely be improved by the use of simple and practical tools to teach owners how to recognize pain in their pets, thus leading to better patient care.

It has also been shown that veterinary students tend to overestimate pain scores when compared with experienced veterinarians and board-certified specialists in anaesthesia and analgesia (Barletta et al., 2016; Benito et al., 2017; Doodnaught et al., 2017). Also, veterinary students' knowledge of animal pain increases at later stages of their studies (Mich et al., 2010; Valros & Hänninen, 2018). Signs of pain vary greatly among individuals, and differences in pain expression may be subtle and difficult to evaluate without sufficient experience or knowledge. Similarly, veterinary nurses, generally assign higher pain scores than veterinarians (Coleman & Slingsby, 2007; Dohoo & Dohoo, 1998).

In general, health care professionals working at human hospitals tend to provide lower estimates of others' pain compared with lay people (Hadjistvropoulos et al., 1998; Prkachin et al., 2001). It has been recognized that estimation of others' pain tends to become more conservative with experience (Bartley & Fillingim, 2013; Miron-Shatz et al., 2020), possibly because clinicians recalibrate internal reference points as they become habituated to observing patients in pain through their clinical experience (Prkachin, 2011). The same phenomenon is likely to occur among veterinary health professionals.

1.4.4. Other sources of variability

Anesthetics and other sedatives produce confounding effects in pain assessment. For instance, the UB-MCPS scores are influenced by the administration of ketamine-based protocols (Buisman et al., 2016). More specifically, ketamine produces a confounding effect on the psychomotor subscale, falsely increasing pain scores, possibly leading to erroneous administration of rescue analgesia. However, it did not interfere with the pain expression subscale and similar changes in pain scores were not observed with the use of alfaxalone (Buisman et al., 2016).

Additionally, pain and dysphoria may be difficult to differentiate using these scales. Dysphoria has been described in cats after administration of opioid drugs. Dysphoric cats may resent being handled, may be restless, pace and vocalize (Robertson et al., 2009). In these cases, an analgesic trial and further reassessments may be indicated to help distinguishing pain and dysphoria (Steagall & Monteiro, 2009).

Studying these limitations and sources of variability will give interesting clues about the context of application of each instrument. This process of testing a scale performance in different and heterogenous settings is known as generalizability, and it is considered an important characteristic of ideal measurement instruments (Streiner & Norman, 2008).

1.5. Psychometric principles applied to the validation of pain scales

The role of measurement is to assign numerical values to the attribute of interest (McDowell, 2006). Subjective measures give insights into pain, that could not be assumed exclusively from physiological assessments or laboratory test results, as previously highlighted during this literature review. This section provides an overview of the measurement properties considered relevant for pain assessment instruments.

Psychometric principles are applied to describe the procedures used to assign numerical scores to subjective judgments, as an attempt to standardize these measurements using a scientific basis (McDowell, 2006; Streiner & Norman, 2008). Psychometric principles carry the ultimate goal of maximizing the quality of assessment through validity, reliability and responsiveness testing (also called measurement properties) (Kaplan & Saccuzzo, 2010; McDowell, 2006; Mokkink et al., 2010; Streiner & Norman, 2008). Part of the nomenclature diverges among authors (i.e. some authors consider responsiveness as part of construct validity, while others propose to assess it as a different quality domain); however, they agree on the importance of assessing these measurement properties (Mokkink et al., 2010; Streiner & Norman, 2008).

In brief, validity is often defined as the extent to which a test measures that which it is intended to measure. Traditionally, validity assessment involves the "three Cs": content, construct and criterion validity (Streiner & Norman, 2008). Content validity assesses whether an instrument contains all the items necessary to represent the concept being measured (i.e. pain). It may be established based on the opinion of a committee of experts in the field (McDowell, 2006; Streiner, 1993; Streiner & Norman, 2008). After developing a pool of items, experts are asked to classify and determine which ones are relevant. The UB-MCPS, G-CMPS and COPS-C/F used this approach to identify and sort the items (questions or descriptors) and have established content validity (Brondani et al., 2011, 2013; Calvo et al., 2014; Della Rocca et al., 2019; Merola & Mills, 2016b), as did other instruments for chronic pain assessment in cats (Benito et al., 2012; Klinck et al., 2015; Klinck, Gruen, et al., 2018; Klinck, Monteiro, et al., 2018; Reid et al., 2018; Zamprogno et al., 2010). The most common method for measuring content validity is the determination of the content validity index (CVI), which can be calculated for the whole instrument (Scale-level CVI - S-CVI) or for each item (Item-level CVI - I-CVI) (Polit & Beck, 2006).

Construct validity assess whether the tool is measuring something (a construct) that cannot be directly observed (i.e. pain or quality of life). There are several methods of assessing construct validity, such as convergent validity, discriminative validity, known-groups discrimination, etc. (McDowell, 2006; Streiner & Norman, 2008). One of the methods, known-groups discrimination determines whether the instrument is able to detect differences between groups (i.e. a group of painful individuals and a group of non-painful ones). This approach was used in the validation of the UB-MCPS (Brondani et al., 2013). Another approach used to confirm construct validity is by showing that pain scores increase or decrease expectedly over time following surgery or the administration of analgesics (in this case, overlapping with the responsiveness assessment, thus explaining the divergence in nomenclature by some authors). This method of testing was used in the validation of the G-CMPS and COPS-C/F (Calvo et al., 2014; Della Rocca et al., 2019).

Criterion validity reflects the correlation of a new instrument under development with some other measure of the trait or disorder under study, ideally a gold standard (Keszei et al., 2010; Streiner & Norman, 2008). However, in the case of pain there is no gold standard, so this property can be tested using concurrent criterion validation. This approach consists of correlating the new scale with another existing measure, both given at the same occasion or within a short period of time (Streiner, 1993). In this scenario, unidimensional pain scales (VAS, NRS, SDS) were widely used in the validation of other instruments, including the UB-MCPS, G-CMPS and COPS-C/F (Brondani et al., 2013; Calvo et al., 2014; Della Rocca et al., 2019; Firth & Haldane, 1999).

Finally, reliability reflects the overall consistency of a measure across time, patients, or observers, and the measurement error involved in it (Streiner & Norman, 2008). Methods of assessing reliability include inter- and intra-rater reliability (the consistency of the scores among different raters and from the same rater after repeated applications of the measure, respectively) and internal consistency (Mokkink et al., 2010). The latter evaluates the correlation between different items (or questions) of the instrument. Internal consistency is not reported as commonly as inter- and intra-rater reliability.

The usual approach to assess inter- and intra-rater reliability is to use video-recordings of cats in different situations and several raters (observers) blinded to the pain status of the cat and moment of video recording. The UB-MCPS showed overall good inter- and intra-rater reliability in its original version in Portuguese and in all its translations (Brondani et al., 2012, 2013, 2014; Della Rocca et al., 2018; Steagall, Monteiro, Lavoie, et al., 2017).

Developing standardized instruments that are robust and appropriate for their purpose is challenging, but may be achieved by incorporating the psychometric principles into the process. Besides the basic characteristics (validity, reliability, responsiveness) presented above, other qualities are desirable, such as the ability to discriminate intensity, frequency, duration and quality between different pain conditions; be simple; and require minimal training and instrumentation (Merola & Mills, 2016b; Reid et al., 2018). Moreover, the ease of application (feasibility) of the instrument in the intended context of use must be assured, and its usefulness for clinical and/or experimental contexts and generalizability should be confirmed (Prinsen et al., 2018). For example, an analgesic intervention threshold (i.e. score for rescue analgesia) increases the usefulness of a clinical tool, guiding the decision of analgesic administration and enabling treatment monitoring (Oliver et al., 2014; Reid et al., 2007). This threshold has been determined for the UB-MCPS (and SF-UBCPS) and G-CMPS (Brondani et al., 2013; Reid et al., 2017).

The process of development and validation of pain assessment instruments is constant, and validity and reliability are not fixed properties (Streiner & Norman, 2008). The instruments are refined and re-tested with new populations in new contexts and for new purposes continuously. Assessment tools that are valid, reliable, and practical are essential for effective pain assessment.

1.6. Facial expressions

Facial expressions have been extensively studied as a measure of the psychological and emotional experience; they are widespread with many facial movements conserved across mammal species (Darwin, 1872; Diogo et al., 2009; Ekman, 1993; Waller & Micheletta, 2013). Some theories suggest that facial expressions initially served a non-communicative adaptive function (Susskind & Anderson, 2008). For example, the widened eyes in the facial expression of fear have been shown to increase the visual field, which would help identifying and following threats; as well as the wrinkled nose and mouth of the facial expression of disgust could limit the intake of bothersome smells and possibly dangerous particles (Shariff & Tracy, 2011). However, throughout the years, the increasing role of the communicative function of facial expressions has been defined (Frith, 2009; Parkinson, 2005). This communicative function can influence the behavior of other members in the group. Primates, as the rhesus monkeys, or human infants can learn to fear potential dangers based on the facial expressions of fear of other group members or parents (Shariff & Tracy, 2011). Another example is the constriction of face openings (facial expression of disgust) leading to the reduction of dangerous inhalations, which can also serve as a communicative function of warning others of dangerous foods (Shariff & Tracy, 2011).

Research involving the facial expressions of emotions has evolved since Darwin's work, including clues on the neurobiological encoding of facial expressions and the creation of standardized systems of assessing facial muscular movements that constitutes the facial expressions (i.e. Facial Action Coding Systems, FACS) (Ekman & Friesen, 1978). Some evidence of similarities on the neural encoding of facial expressions and work on the universality of facial expressions corroborated with Darwin's previous inferences, showing that facial expressions are not only similar among mammals, but also across cultures in humans (Dolensek et al., 2020; Ekman, 1993; Ekman & Friesen, 1971).

This growing interest in facial expression has led to the development of practical bed side tools for assessing pain in non-verbal human patients like infants, individuals with cognitive impairment (Feldt et al., 1998; Krechel & Bildner, 1995; Prkachin, 2009). Additionally, the facial expressions of pain have been recognized in a wide variety of domestic species. In this section, the different contexts of use and the available instruments for pain assessment using facial expressions are reviewed.

1.6.1. Neurobiological encoding of facial expressions

Neuroimaging studies have been conducted to examine the neural base of the communication of affective states via facial expressions. Several studies in humans focused on the brain mechanisms involved in the perception of facial expressions by others (decoding) (Tsao & Livingstone, 2008). However, only a few others studied mechanisms underlying the encoding of affective states via facial expressions (Kunz et al., 2011). Functional magnetic resonance imaging (fMRI), is often used to study brain activity based on blood-oxygenation-level dependent (BOLD) signals, an indicator of neuronal activity (Logothetis, 2002). Certain brain imaging and lesion studies suggested that a network of both subcortical (basal ganglia) and cortical regions (motor-related structures such as the primary motor cortex and supplementary motor areas) contribute to the display of spontaneous and voluntary facial expressions (Kunz et al., 2011). Moreover, a meta-analysis concluded that each of the basic emotions (fear, anger, disgust, sadness, and happiness) was characterized by consistent neural correlates defined by reliable correlations with regional brain activations (Vytal & Hamann, 2010).

Such detailed information about the neurobiological signature in other species is not available; however, evidence is emerging. For example, the insular cortex (or insula) is activated in humans by pain with an important emotional component (Gogolla, 2017; Tracey & Mantyh, 2007). Human individuals with insular lesions can present with pain asymbolia (or pain dissociation), in which emotional responses to pain are reduced (lack of unpleasantness) but pain thresholds are often unaffected (Berthier et al., 1987). Likewise, lesions of the rostral anterior insula produced attenuation of the facial expressions of pain in mice, without affecting abdominal constriction behavior (Langford et al., 2010). Moreover, a recent study described the neuronal correlates of inferred emotion states in the insular cortex in mice. These emotion events (mimicking pleasure, disgust, pain, malaise and fear) were triggered by sweet tasting by administration of sucrose, bitter tasting - quinine, tail shocks, lithium chloride injections, escape and freezing reactions, respectively. The researchers measured the activity of individual neurons using two-photon calcium imaging microscopy and simultaneously recorded the facial expressions of the mouse, demonstrating that insular neuronal activity correlated with specific facial expressions (i.e. each individual neuron was linked to only one single emotion) (Dolensek et al., 2020). That brings interesting insights on the similarities of neural encoding of facial expressions and the pain processing by the central nervous system across mammals.

1.6.2. Facial Action Coding Systems

The Facial Action Coding System (FACS) is an anatomically based system for describing observable facial movements for every emotion (Ekman & Friesen, 1978). An update of this tool was published in the early 2000s (Ekman et al., 2002). Facial muscles are innervated by the facial nerve (CN7) and these muscles are characterized by their extensive connections to the superficial fascia and skin of the face (Diogo et al., 2009). The FACS are coding schemes, where the contraction of a particular muscle (or group of muscles) are called Action Units (AU) and receive a code (number). Action Descriptors (AD) are also used for more general facial movements where the muscular basis either cannot be identified or is the result of different muscle groups (i.e. head movements) (Ekman & Friesen, 1978). This creates a reliable system used to assess facial expressions in a more standardized way. Since its original version created for use with humans, FACS has been adapted for other species, such as chimpanzees (Vick et al., 2007), macaques (Parr et al., 2010), gibbons (Waller et al., 2012), orangutans (Caeiro et al., 2013), dogs (Waller et al., 2013), horses (Wathan et al., 2015), and cats (Caeiro et al., 2017). They are scientific observational tools for identifying and coding facial movements in cats, which according to their authors can be applied to investigate communication and emotion in animals. The available FACS are summarized in Table 1 with the number of AU and the similarities between the original FACS and each animal species. Furthermore, there are 12 AU that are similar or analogous between cats and humans. They are illustrated in Figure 3.

Table 1. Facial Action Coding Systems described for different animal species

Species / FACS	Reference	Listed AU / AD	Obs
Humans / FACS	Ekman & Friesen, 1978 / Ekman et al., 2002	33 AU and 25 AD	*In total, FACS describes 58 components in the human face
Chimpanzees / ChimpFACS	Vick et al., 2007	15 AU and 8 AD	12 AU similar or analogous to humans
Macaques / MaqFACS	Parr et al., 2010	15 AU, 1 AD, and 3 EAD	13 AU similar or analogous to humans
Gibbons and Siamangs / GibbonFACS	Waller et al., 2012	20 AUs and 5 AD	17 AU similar or analogous to humans
Orangutan / OrangFAC	Caeiro et al., 2013	17 AUs and 7 AD	17 AU similar or analogous to humans
Horses / EquiFACS	Wathan et al., 2015	17 AU	15 AU similar or analogous to humans
Dogs / DogFACS	Waller et al., 2013	16 AU and 5 EAD	12 AU similar or analogous to humans
Cats / CatFACS	Caeiro et al., 2017	15 AU, 7 AD and 7 EAD	12 AU similar or analogous to humans

AU: Action Units; AD: Action Descriptors; EAD: Ear Action Descriptors

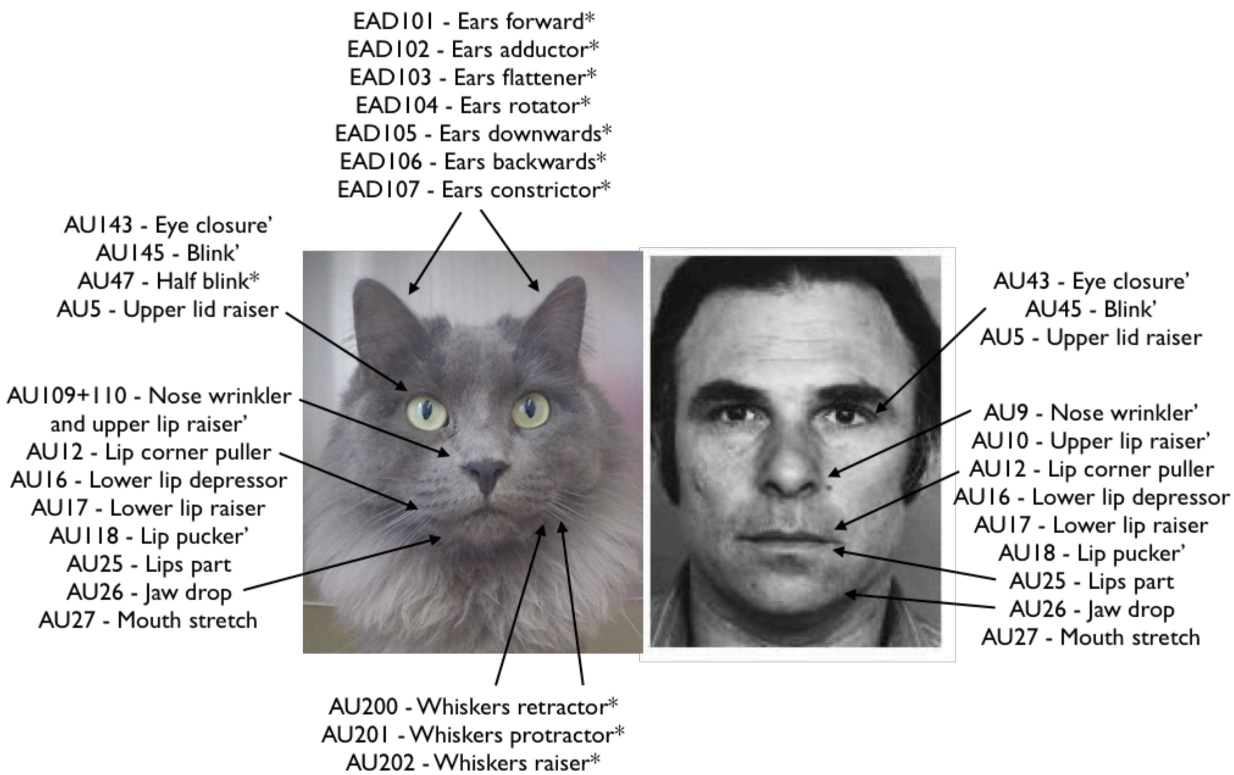


Figure 3. Action Units similar or analogous between cats and humans

AU: Action Unit; EAD: Ear Action Descriptor described within the CatFACS (Caeiro et al., 2017). ’AU similar between cats and humans; *AU or AD described in cats, but not in humans. [Cat image (left) - from Evangelista et al., 2019; man image (right) - from the FACS manual (Ekman & Friesen, 1978)].

The facial expression of pain is coded in terms of eight component facial actions based on FACS (Bartlett et al., 2014). Figure 4 highlights the similarities between the five AU related to pain identified in cats (Evangelista et al., 2019) and those described in humans (Bartlett et al., 2014).

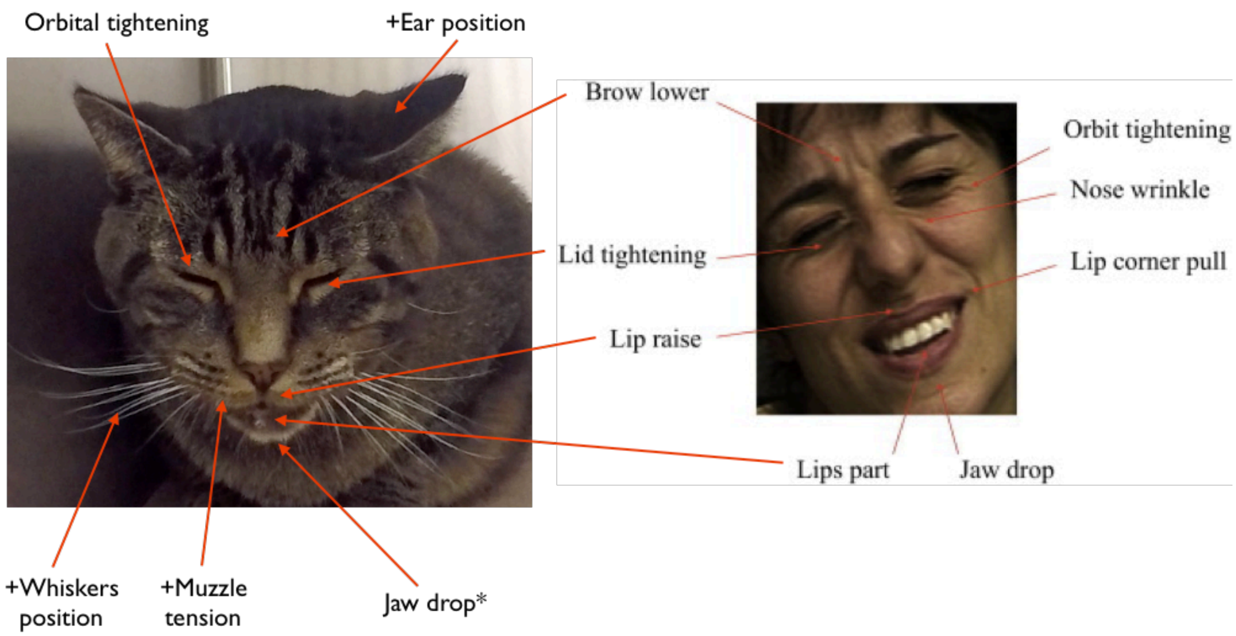


Figure 4. Action Units related to pain identified in cats and humans

Five AU related to pain were identified in cats (Ear position, Orbital tightening, Muzzle tension, Whisker position and Head position), while eight AU were described in this condition in humans. *Note that in these two pictures, it is possible to remark other similarities such as the brow lower (frowning - it may be difficult to see in long-haired cats), lip raise, lips part and jaw drop. [Cat image (left) - from Evangelista et al., 2019; woman's face image (right) - from Bartlett et al., 2014].

1.6.3. Grimace Scales

Grimace scales evaluate facial expressions based on a simplified version of the FACS by focusing on determined facial features and their changes during painful states. The observers (or raters) score the presence (and intensity) of AU on a 0-2 scale where 0 = the AU is absent, 1 = moderately present and 2 = obviously present (often performed using pictures). Grimace scales are effective methods of detecting pain that can be used rapidly and easily with minimal training (McLennan et al., 2019). Developed in the past decade, they have been tested in a range of painful situations (stimuli) and different species. Each grimace scale differs in length (number of AU or features to be assessed), level of accuracy, reliability, and degree correlation with other measures. This section points out their characteristics, the steps taken into their development and validation, and preliminary results of a systematic review that is underway concerning their measurement properties.

The first grimace scale was developed for laboratory mice (Mouse Grimace Scale, MGS) (Langford et al., 2010), then for rats (Rat Grimace Scale, RGS) (Sotocinal et al., 2011), rabbits (Rabbit Grimace Scale, RbtGS) (Keating et al., 2012), horses (Horse Grimace Scale, HGS) (Dalla Costa et al., 2014), adult sheep (Sheep Grimace Scale, SGS) (Hager et al., 2017) and lambs (Guesgen et al., 2016) (Lamb Grimace Scales, LGS), piglets (two distinct grimace scales with the same name were published for this species, a few months apart: Piglet Grimace Scales, here named PGS-a and PGS-b; though only the PGS-b was considered a standard grimace scale) (Di Giminiani et al., 2016; Viscardi et al., 2017) and ferrets (Ferret Grimace Scale, FerGS) (Reijgwart et al., 2017). Although they do not carry "grimace" in their name, the Sheep Pain Facial Expression Scale (SPFES) (McLennan et al., 2016) and the Equine Utrecht University Scale for Facial Assessment of Pain for horses (van Loon & van Dierendonck, 2015) and donkeys (van Dierendonck et al., 2020) (EQUUS-FAP and EQUUS-Donkey-FAP, respectively) are facial expression-based pain scales, and were considered altogether for the purpose of this literature review. They differ in length (number of AU), response options, ways of calculating the final score and method of scoring (i.e. using pictures, videos, live scoring). They are summarized in Table 2. The Feline Grimace Scale, subject of this thesis, was also included in the tables, enabling direct comparison among all the instruments.

Other instruments proposed to identify facial expressions of pain do not consist of a standard grimace scale, such as the ones mentioned above. They are the Facial expression tool for cats (Holden et al., 2014); Equine Pain Face (Gleerup, Forkman, et al., 2015); the Cow Pain Scale (which contains facial expression of dairy cow) (Gleerup, Andersen, et al., 2015); a Piglet Grimace Scale (PGS-a) (Di Giminiani et al., 2016), two other tools used for the identification of AU in seals (MacRae et al., 2018) and beef cattle (Müller et al., 2019), and a recently published Donkey Grimace Scale (DGS) (Orth et al., 2020). They are either not used alone (incorporated into another instrument), or they lack a final score calculation; which precludes direct comparison with the other classical grimace scales (Table 3).

Table 2. Summary of grimace scales for different species published to date

Species / scale	First reference	Pain stimulus used in development	Items (AU)	Response options	Range of scores (calculation)	Method of scoring	Obs
Mouse / MGS	Langford et al., 2010	Abdominal constriction test (0.9% acetic acid)	5 - **Ear position, Orbital tightening**, Nose bulge, Cheek bulge, Whisker change	0 - 1 - 2	0 - 2 (average)	Images (screenshots from videos)	RT scoring described (Miller et al., 2015)
Rat / RGS	Sotocinal et al., 2011	Intra plantar CFA, Kaolin/Carrageenan; Laparotomy	4 - **Ear changes, Orbital tightening**, Nose/cheek flattening, Whisker change	0 - 1 - 2	0 - 2 (average) AT: 0.67	Images (screenshots from videos)	RT scoring described (Leung et al., 2016)
Rabbit / RbtGS	Keating et al., 2012	Ear tattooing	5 - **Ear Position*, Orbital tightening**, Cheek flattening, Nose shape, Whisker position	0 - 1 - 2	0 - 2 (average)	Images (screenshots from videos)	RT scoring described (Banchi et al., 2020)
Horse / HGS	Dalla Costa et al., 2014	Castration	6 - **Stiffly backwards ears**, Tension above the eye area, **Orbital tightening,** Strained nostrils and flattening of profile, Prominent strained chewing muscles, Mouth strained and Pronounced chin	0 - 1 - 2	0 - 12 (sum)	Images (screenshots from videos)	RT scoring described (Coneglian et al., 2020)
Horse / EQUUS-FAP	van Loon & van Dierendonck, 2015	Colic (naturally-occurring)	9 - Head, **Eyelids**, Focus, Nostrils, Corners mouth/lips, Muscle tone, Flehmen/Yawning, Teeth grinding, **Ears**	0 - 1 - 2; Flehmen/ Yawning, Teeth grinding: 0 - 2	0 - 18 (sum)	Live (RT) scoring	
Sheep (adult) / SPFES	McLennan et al., 2016	Footrot and mastitis (naturally-occurring)	5 - **Abnormal ear position, Orbital tightening**, Abnormal nostril and philtrum shape, Cheek	0 - 1 - 2	0 - 10 (sum)	Images (photographs)	

			tightening, Abnormal lip and jaw profile				
Sheep (adult) / SGS	Hager et al., 2017	Tibial osteotomy	3 - **Orbital tightening, Ear** and head position, Flehmen	0 - 1 - 2; Flehmen: 0 - 1 - 3	0 - 10 (sum) AT: 5	Images (screenshots from videos)	
Sheep (lambs) / LGS	Guesgen et al., 2016	Tail docking	5 - **Ear position, Orbital tightening**, Nose changes, Cheek flattening, Mouth changes	0 - 1 - 2	0 - 2 (average)	Images (photographs)	
Piglet / PGS-b	Viscardi et al., 2017	Castration	3 - **Ears,** Cheek tightening/Nose bulge, **Orbital tightening**	0 - 1 - 2; Orbital tightening: 0 - 1	0 - 5 (sum)	Images (screenshots from videos)	
Ferret / FerGS	Reijgwart et al., 2017	Telemetry device implantation	5 - **Ear changes, Orbital tightening**, Nose bulging, Cheek bulging and Whisker retraction	0 - 1 - 2	0 - 10 (sum)	(Images) photographs	
Donkey / EQUUS-Donkey-FAP	van Dierendonck et al., 2020	Several (naturally-occurring)	9 - Head, **Eyelids,** Focus, Nostrils, Corners mouth/lips, Muscle tone, Flehmen/Yawning, **Ear position**, Ear response, Startle/Headshaking, Teeth grinding/moaning, Sweating behind the ears	0 - 1 - 2; Head, Flehmen, Ear position, Startle, Mouth, Teeth grinding and Sweating behind ears: 0 - 2	0 - 24 (sum)	Live (RT) scoring	
Cat / FGS	Evangelista et al., 2019	Several (naturally-occurring)	5 - **Ear position, Orbital tightening**, Muzzle tension, Whisker position, Head position	0 - 1 - 2	0.0 - 1.0 (sum divided by total) AT: 0.39	Images (screenshots from videos)	RT scoring described - *Presented in this thesis*

AU: Action Unit. AT: Analgesic threshold (determined based on sensitivity and specificity of the scale). RT: Real-time method of scoring. Note that AUs related to ear position and orbital tightening were listed for all scales. *Ear position was listed as AU in the RbtGS; however, it was not considered in further validation of the scale due to the pain stimulus used (ear tattooing).

Table 3. Other facial expression tools and identification of AU in different species

Species / scale	Reference	Pain stimulus	Items (AU)	Obs
Cat / Facial Expression Tool	Holden et al., 2014	Several (naturally-occurring)	2 - Ears and muzzle landmarks distances	Not used alone, incorporated into another pain scale (Reid et al., 2017)
Horse / Equine Pain Face	Gleerup, Forkman, et al., 2015	Tourniquet and topical capsaicin	6 - Ear position, Eye angle, Stare, Nostrils, Tension of the muzzle Tension of the mimic muscles	No final score calculated (Ethogram)
Dairy cattle / Cow Pain Scale	Gleerup, Andersen, et al., 2015	Several (naturally-occurring and post-surgical pain)	6 - Ears, Eyes, Tension of the muscles above the eyes, Tension of the facial muscles on the side of the head, Strained nostrils, Increased tonus of the lips	Part of a pain scale including other general behaviors
Piglet / PGS-a	DiGiaminiani et al., 2016	Tail docking and castration	7 - Initially, 10 AU were identified (Temporal tension, forehead, eyes, tension above the eyes, cheek, snout plate, snout angle, lip, jaw, nostril), 3 were excluded after refinement (upper lip, nostril, lower jaw profile)	No final score calculated (AU evaluated independently)
Seal / Identification of AU	MacRae et al., 2018	Flipper tagging and microchipping	4 - Eye change, Nose change, Whisker change, Mouth change	No final score calculated (Ethogram)
Beef cattle / Identification of AU	Müller et al., 2019	Branding with hot iron	5 - Orbital tightening, tension above eye, brow lowering, eye closure, and raised inner brow	No final score calculated
Donkey / DGS	Orth et al., 2020	Castration	9 - Ear frontal and side position, eye description (shape and tension), muzzle and nostril description (shape and tension), and stance	No final score calculated (preliminary validation of the scale was not based on grimace scores)

AU: Action Unit.

1.6.3.1. Considerations on the development of grimace scales

Study design

The majority of the scales were developed using a within-subject design (in randomized-controlled trials), in which the animal was evaluated before and after a painful stimulus (mice, rats, rabbits, horses, sheep/lambs, piglets and ferrets; MGS, RGS, RbtGS, HGS, SGS, LGS, PGS-b, FerGS, respectively) (Dalla Costa et al., 2014; Guesgen et al., 2016; Hager et al., 2017; Keating et al., 2012; Langford et al., 2010; Reijgwart et al., 2017; Sotocinal et al., 2011; Viscardi et al., 2017). The within-subject design enables the scores to be compared with a baseline value, better controlling for individual variability. However, in a clinical context, baseline values are not always available and important individual variability may be present. A between-subject design (case-control study) was performed to identify dissimilar facial features in footrot- or mastitis-affected sheep and a group of healthy, control sheep (SPFES) (McLennan et al., 2016). The EQUUS-FAP development study used colic as a model to study pain also in a case-control study, however the methods for the creation of the scale were not clearly reported (van Loon & van Dierendonck, 2015). The same occurred for the EQUUS-Donkey-FAP, which used both surgical and non-surgical cases, but the methods for the scale creation were not clearly reported (van Dierendonck et al., 2020). A recent review considered that ensuring there is a baseline for comparison during a scale development may yield better results and the authors included this in the best practice guidelines for facial expression scale development (McLennan et al., 2019).

In all these studies, one or more individuals, experienced on pain assessment or facial expression, visually compared images of painful vs. non-painful animals and described the features that changed between them (listing as AU). Additional quantitative measures (distances and/or angles) were performed in control and painful lamb faces (Guesgen et al., 2016), and the AU composing the FerGS were identified after an anatomical study of ferret faces using dissection of cadavers and only changes observed in >25% of the animals were listed as AU (Reijgwart et al., 2017).

Sex and strain/breed

Sex and strain differences in laboratory animals have been previously reported in pain research, hence the importance of including both sexes in the development of a pain scale (Mogil & Chanda, 2005). In addition, experimental models tend to comprise genetically homogeneous groups of young, male rodents in restricted and controlled environments. This may not reflect the population and condition of interest (i.e. chronic pain conditions in humans are generally more prevalent in women than in men), bringing the issue of reproducibility and failure in translational pain research (Bartley & Fillingim, 2013; Greenspan et al., 2007; Klinck et al., 2017; Shansky & Woolley, 2016).

Instruments devised for mice, rats, rabbits, horses, donkeys and sheep (MGS, RGS, RbtGS, EQUUS-FAP, EQUUS-Donkey-FAP and SPFES) included both males and females in their scale development studies (Keating et al., 2012; Langford et al., 2010; McLennan et al., 2016; Sotocinal et al., 2011; van Dierendonck et al., 2020; van Loon & van Dierendonck, 2015). Contrarily, another instrument for horses, the HGS, was developed using only male horses (castration was used as painful stimulus, therefore the choice of only males); and the SGS and FerGS only included female sheep and ferrets, respectively (Dalla Costa et al., 2014; Hager et al., 2017; Reijgwart et al., 2017). The development of PGS-a included both male and female piglets, however they were submitted to different painful stimuli (females - tail docking; males - castration), and PGS-b only used male piglets (DiGiaminiani et al., 2016; Viscardi et al., 2017).

The MGS was developed using CD-1 mice and tested using other 5 strains (C57BL/6, C3H/He, BALB/c, CBA and DBA/2) (Miller et al., 2015; Miller & Leach, 2015a). Some changes were reported in MGS scores across different strains of mice; however, the magnitude of the differences was small, and the study was conducted in healthy animals not receiving any painful stimulus (Miller & Leach, 2015a). The RGS was developed using Wistar rats and further tested on Sprague-Dawley, Holtzman Sprague-Dawley and Hooded Lister rats (Miller et al., 2016; Oliver et al., 2014; Philips et al., 2017). Direct comparisons between rat strains were not reported. Furthermore, animals of the same breed were used for the development of the RbtGS (New Zealand White rabbits), SGS (Blackface sheep), LGS (Romney cross lambs) and PGS-b (Yorkshire piglets) (Guesgen et al., 2016; Hager et al., 2017; Keating et al., 2012; Viscardi et al., 2017). Studies including several breeds for the development of the Grimace Scales were: HGS, EQUUS-FAP, EQUUS-Donkey-FAP, SPFES and FerGS (Dalla Costa et al., 2014; McLennan et al., 2016; Reijgwart et al., 2017; van Dierendonck et al., 2020; van Loon & van Dierendonck, 2015). It was reported that horses with dark brown or black coats were more difficult to score than grey and light brown coat (Dalla Costa et al., 2014); and in sheep, a main effect of breed on SPFES scores was reported, however, when performing contrasts, there were no significant differences on these scores between breeds (McLennan et al., 2016). To ensure the generalizability of the instrument in different contexts, it is important to include different breeds in the scale development and further testing.

Methods of scoring

Originally, the assessments of facial expression in laboratory animals were performed based on still images (screenshots obtained from video recordings) (Langford et al., 2010; Sotocinal et al., 2011). Similar methodology was adopted for the HGS, SGS, LGS and PGS-b (Dalla Costa et al., 2014; Guesgen et al., 2016; Hager et al., 2017; Viscardi et al., 2017). The duration of video-recordings ranged from 1 minute (Guesgen et al., 2016) up to 30 minutes (Sotocinal et al., 2011). On the other hand, for the SPFES and FerGS, photographs were used (McLennan et al., 2016; Reijgwart et al., 2017). For the development of the PGS-a (Di Giaminiani et al., 2016), both video-recordings and photographs were used to assess facial expressions, and the EQUUS-FAP and EQUUS-Donkey-FAP used live scoring (in real-time) (van Dierendonck et al., 2020; van Loon & van Dierendonck, 2015).

Real-time scoring was also described in mice and rats, in comparison with the standard (still image-based) methodology. Differences between real-time and image scoring method were reported for the MGS (Miller & Leach, 2015a), while a slight underestimation (with good agreement between methods) was observed for the RGS (Leung et al., 2016). Moreover, comparison of other methods were performed and there was no significant difference in HGS total scores obtained from still images and from 15-second video clips (Dalla Costa et al., 2016). The use of videos has the advantage of continuously recording the individual from a distance, whereas the presence of a camera (and observer) in close proximity to the animal's face may alter the behavior and facial expression.

Context of use

The grimace scales differ on the target population (or context) for which that the scale is intended. In the first paper published (MGS), it was proposed that grimace scales were not valid measures of chronic pain; however, further studies disproved this hypothesis, demonstrating the validity and applicability in different contexts (acute and chronic) of experimental and clinical pain (Liao et al., 2014; Schneider et al., 2017). As more studies using grimace scales have been published, defining the time course of grimacing is proving to be more complex (Mogil et al., 2020). This emphasizes the notion that, regardless of the setting in which the scale was developed, its validity should be tested in different contexts to ensure the generalizability of the instrument (Streiner & Norman, 2008).

The MGS was developed using a preclinical assay (0.9% acetic acid abdominal constriction test) and further tested on 14 other pain assays of different intensity and expected duration of nociception (acute nociceptive and neuropathic pain models, including tail flick, intra plantar injection of irritant substances, laparotomy, chronic constriction nerve injury, etc.) (Langford et al., 2010). The scale was also used in other contexts such as, vasectomy, ear notching, thoracotomy and myocardial infarction, tooth pulp injury, chronic constriction nerve injury, tendon injury and chronic gastrointestinal disease models (Akintola et al., 2017; Chartier et al., 2020; Faller et al., 2015; Leach et al., 2012; Matsumiya et al., 2012; Miller & Leach, 2015b; Moser et al., 2020; Rossi et al., 2020). The RGS was developed using inflammatory assays (intra plantar injection of Complete Freund's Adjuvant, CFA or Kaolin/Carrageenan) and laparotomy (Sotocinal et al., 2011). It was further tested using other inflammatory assays (i.e. intra plantar injection of irritant substances and plantar incision), surgical implantation of telemetry device, experimental tooth movement, models of trigeminal neuropathic pain, spinal cord injury, cervical radiculopathy, reserpine induced myalgia, acute and chronic colitis, low back pain, occlusal dental interference, temporomandibular joint loading (osteoarthritis) (Akintola et al., 2017; De Rantere et al., 2016; Leung et al., 2019; Liao et al., 2014; Nagakura et al., 2019; Oliver et al., 2014; Philips et al., 2017; Reed et al., 2020; Schneider et al., 2017; P. G. de B. Silva et al., 2020; Sperry et al., 2020), and also using a positive stimulus (tickling) (Finlayson et al., 2016). The RbtGS was developed using ear tattooing (Keating et al., 2012) and tested in calvarial bone surgery and naturally-occurring pain (from several sources) (Banchi et al., 2020; Raillard et al., 2019). The HGS was conceived using horses undergoing castration (Dalla Costa et al., 2014) and it was tested in laminitis, orthopedic pain and dental disorders (Coneglian et al., 2020; Dalla Costa et al., 2016; Samy et al., 2020). The EQUUS-FAP was tested in horses with colic, orofacial and orthopedic pain (van Dierendonck & van Loon, 2016; van Loon & van Dierendonck, 2015, 2017, 2019). The PGS-b was developed and tested using castration (and cryptorchidectomy) (Viscardi et al., 2017; Vullo et al., 2020). And finally, no further studies were published using the SPFES, HGS, LGS, FerGS and EQUUS-Donkey-FAP (Guesgen et al., 2016; Hager et al., 2017; McLennan et al., 2016; Reijgwart et al., 2017; van Dierendonck et al., 2020).

1.6.3.2. Psychometric properties applied to grimace scales

A systematic review on the psychometric properties of grimace scales using a validated checklist for the assessment of these properties for health-related outcome measures (Consensus-bases Standards for the Selection of Health Measurement Instruments - COSMIN) (Mokkink et al., 2010; Prinsen et al., 2018) is underway. The protocol was registered at the Systematic Review Facility (SyRF) from CAMARADES (http://syrf.org.uk) (#21-Nov-2019). The search strategy included seven databases and search items related to facial expression, grimace, pain and animals. A total of 3107 studies were retrieved (2213 duplicated were removed) and 894 were screened against title and abstract (808 were excluded). Finally, 86 studies (full-text) were assessed for eligibility and 41 were initially included (November-December 2019). The search was updated in June 2020 and 18 studies were added (total = 59 studies included for data extraction). Preliminary results demonstrated that 13 instruments for 9 species were described. Grimace scales showed considerable variability regarding their development and their reported level of validity and reliability. Additionally, the species with the highest number of studies from different research groups are those used in pre-clinical research (mice and rats); whereas for most of the other domestic species, information only by the original authors of the scale was available (Evangelista, Monteiro, et al., 2020). The preliminary assessment of the psychometric properties of different grimace scales is summarized in Table 4.

Content (face validity) and construct validity were reported for all scales. Criterion validity was properly assessed for six of them (MGS, RGS, HGS, SPFES, PGS-b and FGS). Various instruments were used as comparators in the assessment of criterion validity, including general behaviors scales, composite pain scales, lameness and lesion scores and clinical severity scores. The degree of correlation was commonly measured using Pearson's or Spearman's rank correlation coefficients. In some cases, other instruments were used to assess pain concomitantly; however, no direct comparisons (or correlations) were reported. It was the case for the MGS x general behaviors (Leach et al., 2012); RGS x withdrawal thresholds (Sperry et al., 2018); RGS x composite behavior score (Klune et al., 2019); SPFES x lameness and lesion score (McLennan et al., 2016); HGS x composite pain score and Obel grade for laminitis (Dalla Costa et al., 2014, 2016); PGS-b x active and inactive behaviors (Viscardi et al., 2017). The SGS was correlated with clinical severity score (Hager et al., 2017). It must be noted that in the latter study the measurements were not performed concomitantly or within a short period, thus the pain state may have changed between the assessments. Other methods of pain assessment were used in some of the studies, and sometimes the authors mentioned that the measures "coincided" or "corresponded" to the increase of the comparator; however, no direct measure of correlation was attempted. It was the case for MGS x welfare score (Faller et al., 2015); RGS x paw hypersensitivity and withdraw (De Rantere et al., 2016; Schneider et al., 2017; Nagakura et al., 2019); RGS x disease activity index, composite pain score and burrowing (Leung et al., 2019); and EQUUS-FAP x VAS (van Dierendonck et al., 2016).

The responsiveness of the MGS, RGS, RbtGS, HGS, SPFES, PGS-b and FGS was supported by the significant change in pain scores in response to treatment, including the following drug classes: opioids (i.e. morphine, buprenorphine, fentanyl, tramadol), non-steroidal anti-inflammatory drugs (i.e. meloxicam, ketoprofen, diclofenac, phenylbutazone), local anesthetics (i.e. topical EMLA cream, local infiltration of bupivacaine, intravenous regional analgesia with lidocaine or mepivacaine), anticonvulsants (i.e. gabapentin), antidepressants (i.e. duloxetine), and other treatment (monoclonal antibodies, tumor necrosis factor inhibitor - etanercept; and antibiotics) (Akintola et al., 2017; Dalla Costa et al., 2016; Keating et al., 2012; Langford et al., 2010; Leach et al., 2012; Leung et al., 2016; Matsumiya et al., 2012; McLennan et al., 2016; Nagakura et al., 2019; P. G. de B. Silva et al., 2020; Sotocinal et al., 2011; Sperry et al., 2020; Viscardi et al., 2017; Viscardi & Turner, 2018). Additionally, for the MGS, RGS and PGS-b, not only the responsiveness to analgesics was assessed in painful animals, but also the effect of opioids in non-painful animals (Leung et al., 2016; Miller et al., 2015; Oliver et al., 2014; Viscardi & Turner, 2018).

Internal consistency was reported for the MGS, RGS, HGS, SPFES, FerGS and FGS. Authors used the most common indicator, the Cronbach's alpha (Streiner & Normal, 2008) or Linear Discriminant Analysis to assess the weight of each AU in the composition of the final score (Dalla Costa et al., 2018; Langford et al., 2010; McLennan et al., 2016; Oliver et al., 2014; Philips et al., 2017; Reijgwart et al., 2017). Finally, inter-rater reliability was assessed for all scales, although the methodology was not appropriate for the EQUUS-FAP and EQUUS-Donkey-FAP, which used Pearson's correlation and Cronbach's alpha, respectively to assess this property (van Dierendonck et al., 2020; van Dierendonck & van Loon, 2016). Intra-rater reliability was properly assessed only for the MGS and RGS (Mittal et al., 2016; Oliver et al., 2014; Zhang et al., 2019). The methodology used for assessing this property for the FerGS was not appropriate (the raters underwent training between the scoring sessions, so reliability was expected to change) (Reijgwart et al., 2017) .

1.6.3.1. Limitations

Attention should be given to bias and confounding factors. The primary cause for some of the observed effects may be variables that have not been considered or measured (i.e. the effect of sedatives, anesthesia or stress induced by handling). They could be addressed by adding positive and negative controls and/or sham groups. These aspects were broadly addressed for the MGS (Matsumiya et al., 2012; Miller et al., 2015; Miller & Leach, 2016), RGS (Leung et al., 2016; Miller et al., 2016; Oliver et al., 2014) and RbtGS (Keating et al., 2012); and partially for the other studies. For example, the effect of handling or restraint was addressed for the SGS (sheep were housed in a suspension system and the impact of this system was considered before the surgical procedures) (Hager et al., 2017). In the development of the LGS and PGS-a, lambs and piglets were restrained while photographs were taken (Di Giminiani et al., 2016; Guesgen et al., 2016); and for the FerGS, ferrets were habituated to a handling procedure using a tube, however, the effect of this handling technique was not accounted (Reijgwart et al., 2017). Anesthesia and sedation may also cause confounding factor in pain scores. Effects of anesthesia on the Grimace Scales were addressed for the MGS, RGS, RbtGS and HGS using groups undergoing sham anesthesia (without surgery or painful stimuli) (Dalla Costa et al., 2014; Keating et al., 2012; Miller et al., 2015, 2016). And effects of opioids (which are known to cause sedation in some species) were assessed in non-painful animals using the MGS, RGS and PGS-b (Leung et al., 2016; Miller et al., 2015; Oliver et al., 2014; Viscardi & Turner, 2018).

The observer effect is particularly important given the potential advantages of real-time (live) scoring for the clinical applicability of grimace scales, compared to the retrospective method (still images obtained from videos). It has been previously demonstrated that the presence of a male observer could produce stress-induced analgesia (supported by reduced grimace scale scores) in mice (Sorge et al., 2014); but the presence of a female experimenter did not produce the same effect in rats habituated to this observer (Leung et al., 2016). In the majority of the studies, the gender of the observer was not reported, which precluded further inferences on the observer effect in other species.

Table 4. Psychometric properties of grimace scales

Species / scale	VALIDITY Content - Face validity	Criterion (comparator)	Construct	Responsiveness (treatments)	RELIABILITY Internal consistency	Inter-rater	Intra-rater	Obs
Mouse / MGS	X	X* (general behaviors)	X	X (OP, NSA*, LA*)	X	X	X*	*determined in follow-up studies
Rat / RGS	X	X* (withdrawal thresholds, behavior score)	X	X (OP, NSA*, LA*, AC*, AD*, MAB*)	X	X	X*	*determined in follow-up studies
Rabbit / RbtGS	X		X	X (LA)		X		
Horse / HGS	X	X (composite pain scale and Obel grade)	X	X (NSA, LA*)	X*	X		*determined in follow-up studies
Horse / EQUUS-FAP	X		X	X (NSA*)		?		Inter-rater reliability: method not appropriate
Sheep (adult) / SPFES	X	X (lameness and lesion score)	X	X (NSA, ATB)	X	X		
Sheep (adult) / SGS	X	? (clinical severity score)	X			X		Correlation: not performed at same time
Sheep (lambs) / LGS	X		X			X		
Piglet / PGS-b	X	X (general behaviors)	X	X (LA, OP*, NSA*)		X		*determined in follow-up studies
Ferret / FerGS	X		X		X	X	?	Intra-rater reliability: method not appropriate
Donkey / EQUUS-Donkey-FAP	X		X			?		Inter-rater reliability: method not appropriate
Cat / FGS	X	X (G-CMPS)	X	X (OP, NSA)	X	X	X	*Presented in this thesis*

Treatment drugs - OP: Opioids; NSA: Non-steroidal anti-inflammatories; LA: local anesthetics; AC: Anticonvulsants; AD: antidepressants; MAB: Monoclonal antibody; ATB: Antibiotics

1.6.4. The study of facial expressions in cats

Studies regarding facial expressions in cats are limited in comparison with other species. Some anecdotal information and exploratory studies on cats' emotion and facial expression have been published since Darwin's work (1872), but the role of facial expressions in the study of pain in this species has only been investigated in the past decade.

Social and predatory emotional states of the cat were described by Leyhausen (1979). Emotional states (offensive and defensive "moods" related to fear and degree of aggressiveness) were inferred from observations of the behavioral and facial expression, alongside with reference to brain stimulation studies. An ethogram of "offensive and defensive" moods in cats (including facial expressions) was described in his book (Leyhausen, 1979) and later reproduced in several articles, textbooks and guidelines on feline behavior (Bennett et al., 2017; Bowen & Heath, 2005; Overall et al., 2005; Wolski, 1982) (Figure 5).

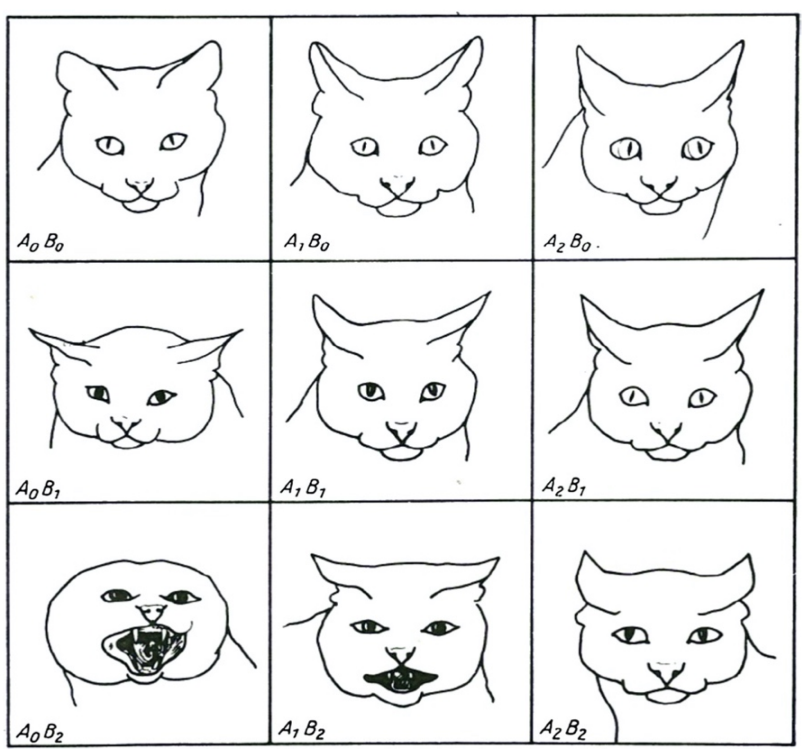


Figure 5. Facial expressions of offensive and defensive moods in cats

The images range from a "neutral face" (top left - A0B0) to "threat of attack" (bottom right - A2B2) and "readiness for defense" (bottom left - A0B2) [from Leyhausen, 1979].

The CatFACS was published in 2017 (previously contemplated in this literature review) (Caeiro et al., 2017) and a follow-up study aimed to develop an ethological description of the behavior and facial expression of cats according to their underlying emotional state (Bennett et al., 2017). The authors used the CatFACS to describe the relationship between behavior and facial expressions of cats in cages when they were filmed from a distance and when a human approached and interacted with them. The facial actions were associated with varying degrees of relaxed engagement, fear and frustration; and they were also compared to the descriptions of "offensive and defensive moods" by Leyhausen (1979) (Bennett et al., 2017).

Another recent study looked into the ability of participants in recognizing negative (i.e. fear; frustration) and positive states (i.e. being petted or given treats) in cats through facial expressions. Participants watched short videos featuring 20 cats and were asked whether the cat was in a positive state, a negative one, or if they were not sure. People identified the correct expression, on average 59% of the time. The authors also reported that 13% percent of the participants correctly scored the valence of cats' states in more than 75% of the cases (≥ 15 out of 20 videos) (Dawson et al., 2019). The variation observed was attributed to participant characteristics (gender and age, whether they had professional feline experience) and the activity levels of the cat in the video. Women were more successful at than men, and younger participants more successful than older, as were participants with professional feline experience (veterinarians or animal health technicians) (Dawson et al., 2019). Whether the ability to correctly interpret cats' facial expressions can indeed be learned with experience and/or improved via training should be further investigated. In this study, none of the cats showed extreme expressions of fear, such as bared teeth or flattened ears (which could also be pain-related); thus, the ability of people with different expertise to identify pain in cats through facial expressions remains unknown.

Much of the attention towards the facial expressions of pain in this species occurred during the past 6 years; including the publication a facial expression tool differentiating pain-free cats from those in acute pain (Holden et al., 2014) and a geometric morphometric approach to study facial expressions of pain in cats (Finka et al., 2019).

The first tool developed for assessing facial expression in cats differentiating pain-free cats from those in acute pain used anatomical landmarks and linear distances between points in the cats' faces (Holden et al., 2014). The authors used the results to construct a caricature face scale to complement the G-CMPS (Calvo et al., 2014; Reid et al., 2017). This tool includes two features: ear position and landmarks around the mouth (muzzle). Eye position and shape were not considered, due to the authors' concerns about the effects of opioids (11 out of the 28 cats considered painful in this study had received analgesia; eight of them had been treated with opioids).

More recently, a geometric morphometric approach has been described to study facial expressions of pain in cats (Finka et al., 2019). This approach has been proposed as basis for further application of machine learning algorithms for automated pain recognition in non-human animal species. This study considered images from 29 cats before and after ovariohysterectomy and assessed the changes in facial shape related to pain. According to the authors, it offers a powerful discriminatory tool for quantifying facial shape and its variation, regardless of their recognition as facial action units (Finka et al., 2019).

1.7. Justification, objectives and hypotheses

The literature review highlights the challenges surrounding pain assessment in cats, including the lack of simple, practical and validated tools easily applied in feline practice. It has also presented some of the limitations such as the substantial degree of variability among observers and the low adherence to use of standardized pain scoring systems. In the beginning of this PhD program, the definitive version of the G-CMPS (Reid et al., 2017) had not been published yet, neither the CatFACS (Caeiro et al., 2017; although brief mention of its existence was made at the AnimalFACS website). The only available tool for facial expression in cats was the one published by Holden et al. (2014) not designed for use as a sole instrument for pain assessment. Additionally, the role of facial expressions in pain research was increasingly recognized with the publication of several grimace scales for non-human mammals but did not include the feline species. Thus, the overall aim of this thesis was to develop a novel facial expression-based instrument for acute pain assessment in cats and to explore its applications and limitations.

This thesis had the following specific objectives: 1) to develop a facial expression-based instrument, a Feline Grimace Scale to detect acute pain associated with naturally-occurring conditions; 2) to validate the FGS, assessing its psychometric properties such as validity, reliability and responsiveness, including the derivation of a cut-off score indicating the need for rescue analgesia; 3) to evaluate the generalizability of the instrument and its clinical applicability in real-time, including the influence of sedation and surgery (ovariohysterectomy) on the FGS scores; 4) to explore if different individuals with varied levels of expertise (cat owners, veterinary students, technicians and veterinarians) were capable to use the instrument in a reliable manner; and finally 5) to investigate the impact of gender and other demographic characteristics on the FGS scores.

The hypotheses of this thesis were: 1) the FGS would be able to discriminate painful from non-painful cats in a clinical setting; 2) painful cats would score higher than non-painful cats (demonstrating its construct validity); the FGS would correlate with another pain scoring instrument (G-CMPS, supporting its criterion validity); the scores from different raters and from the same rater over time would be consistent (exhibiting good inter- and intra-rater reliability); the FGS scores would decrease after analgesic treatment (indicating its responsiveness); and a cut-off for rescue analgesia would be determined with high sensitivity and specificity. Furthermore; 3) FGS scores obtained in real-time and by image assessment would be comparable when evaluated by the same observer; FGS scores would not change after mild sedation with acepromazine and buprenorphine nor 24 h after extubation when compared with baseline values. Plus, 4) the FGS would be reliable among different groups of raters; and 5) that our findings would agree with the literature, in the sense that women would assign higher pain scores than men.

2. Publications

This thesis describes the steps taken to develop, validate and test the applicability of the FGS in different clinical contexts, exploring its strengths and limitations. The first and second objectives were to develop and validate a grimace scale to assess naturally occurring acute pain in cats. They were achieved with the first paper entitled “Facial expressions of pain in cats: The development and validation of a Feline Grimace Scale” by Marina C Evangelista, Ryota Watanabe, Vivian SY Leung, Beatriz Monteiro, Elizabeth O’Toole, Daniel SJ Pang and Paulo V Steagall; published in Scientific Reports (2019; 9: 19128). This study used a sound psychometric approach to validate the FGS, and demonstrated that the FGS is a simple, valid and reliable tool for acute pain assessment in cats. The contributions from the PhD candidate included design and conception of the study, data collection, image capture and selection, preparation of the online surveys, analysis of the results and drafting the manuscript.

The third objective was to evaluate the clinical applicability of the FGS by testing its use in real-time in cats undergoing ovariohysterectomy, including the assessment of the influence of sedation and surgery on the scores. This was achieved with the second paper entitled “Clinical applicability of the Feline Grimace Scale: real-time versus image scoring and the influence of sedation and surgery” by Marina C. Evangelista, Javier Benito, Beatriz P. Monteiro, Ryota Watanabe, Graeme M. Doodnaught, Daniel S. J. Pang and Paulo V. Steagall; published in PeerJ (2020; 8:e8967). This study compared the pain scores obtained in real-time from cats undergoing ovariohysterectomy with those obtained by image assessment and demonstrated good agreement between the methods; proving that real-time scoring using the FGS in a clinical postoperative setting is feasible. Additionally, the scores recorded after premedication with acepromazine-buprenorphine and 24 hours after surgery did not differ from baseline. The contributions from the PhD candidate included conception and design of the study, data collection (real-time pain assessments and image scoring), analysis of the results, preparation of figures and/or tables and drafting the manuscript.

The fourth and fifth objectives were to further investigate the applicability of the FGS, by exploring whether individuals with different levels of expertise were able to use the FGS in a reliable manner and the impact of demographic characteristic of the raters on FGS scores. These objectives were achieved with the third manuscript entitled "Agreement and reliability of the Feline Grimace Scale among cat owners, veterinarians, veterinary students and nurses" by Marina C. Evangelista and Paulo V. Steagall; submitted to Scientific Reports in August 2020. This study was carried out using images collected during the three aforementioned studies. It demonstrated that the FGS can be used in a reliable manner even by untrained individuals, and these results represent a substantial progress in feline pain assessment because the FGS is the first instrument for pain assessment that could be used by cat owners in the recognition of pain in their home environment. The contributions from the PhD candidate included study conception, recruitment of participants, image selection and preparation of the online surveys, data collection, analysis of the results and drafting the manuscript.

2.1. Facial expressions of pain in cats: the development and validation of a Feline Grimace Scale

Marina C Evangelista[1], Ryota Watanabe[1], Vivian SY Leung[1,2], Beatriz Monteiro[2,3], Elizabeth O'Toole[1], Daniel SJ Pang[1,2], Paulo V Steagall[1]

2.1.1. Article identifier

This article was published in Scientific Reports 9: 19128 (2019)

doi: 10.1038/s41598-019-55693-8

[1] Département de sciences cliniques, Faculté de médecine vétérinaire, Université de Montréal, 3200 rue Sicotte, Saint-Hyacinthe, Québec, J2S2M2, Canada
[2] Groupe de recherche en pharmacologie animale du Québec, Université de Montréal, 3200 rue Sicotte, Saint-Hyacinthe, Québec, J2S2M2, Canada
[3] Département de biomédecine vétérinaire, Faculté de médecine vétérinaire, Université de Montréal, 3200 rue Sicotte, Saint-Hyacinthe, Québec, J2S2M2, Canada.

2.1.2. Abstract

Grimace scales have been used for pain assessment in different species. This study aimed to develop and validate the Feline Grimace Scale (FGS) to detect naturally-occurring acute pain. Thirty-five client-owned and twenty control cats were video-recorded undisturbed in their cages in a prospective, case-control study. Painful cats received analgesic treatment and videos were repeated one hour later. Five action units (AU) were identified: ear position, orbital tightening, muzzle tension, whiskers change and head position. Four observers independently scored (0-2 for each AU) 110 images of control and painful cats. The FGS scores were higher in painful than in control cats; a very strong correlation with another validated instrument for pain assessment in cats was observed (rho = 0.86, $p < 0.001$) as well as good overall inter-rater reliability [ICC = 0.89 (95% CI: 0.85-0.92)], excellent intra-rater reliability (ICC > 0.91), and excellent internal consistency (Cronbach's alpha = 0.89). The FGS detected response to analgesic treatment (scores after analgesia were lower than before) and a cut-off score was determined (total pain score > 0.39 out of 1.0). The FGS is a valid and reliable tool for acute pain assessment in cats.

2.1.3. Introduction

Pain management is frequently overlooked in cats and they are prescribed less analgesic drugs when compared with dogs (Armitage et al., 2005; Hewson et al., 2006a; Hunt et al., 2015). This is due to challenges in feline pain recognition and assessment, lack of specific training in the subject and limited availability of pain assessment scoring tools in this species (Hewson et al., 2006b; Hunt et al., 2015).

Two validated behavior-based pain assessment instruments have been published, the UNESP-Botucatu multidimensional composite pain scale (UB-MCPS) (Brondani et al., 2013) and the Glasgow composite measure pain scale for cats (G-CMPS) (Calvo et al., 2014). The latter has been updated and the definitive version included two features of facial expression (ears and muzzle), improving its discriminatory ability (Reid et al., 2017). Although validated, each of these tools have their own limitations such as time-consuming implementation, validity tested for a single type of pain stimulus (i.e. ovariohysterectomy), and confounding effects of cats' demeanor and drugs on pain scores (Buisman et al., 2016, 2017; Steagall & Monteiro, 2019).

Along with the evaluation of behavioral changes, facial expressions have the potential to indicate emotional experiences in animals and provide valuable information regarding internal states (Darwin, 1872). Facial expressions can be a useful, valid and reliable tool for pain assessment in humans and other animals (McLennan et al., 2019). They can be objectively assessed using a facial action coding system (FACS) that measures the individual movements or 'action units' (AU) of the face that comprise an expression (Ekman & Friesen, 1978). This system assigns independent codes to activity of individual muscles or groups of muscles. A feline-specific coding scheme (CatFACS) has been developed by studying the facial musculature of the domestic cat (Caeiro et al., 2017).

Grimace Scales are simplified methods of assessing the facial expressions specifically related to pain. They were developed for mice (Langford et al., 2010), rats (Sotocinal et al., 2011), rabbits (Keating et al., 2012), horses (Dalla Costa et al., 2014), sheep (Hager et al., 2017; McLennan et al., 2016), lambs (Guesgen et al., 2016), piglets (Di Giminiani et al., 2016; Viscardi et al., 2017) and ferrets (Reijgwart et al., 2017). Most of these scales consider four to five AU rated as absent, partially present or markedly present. Action units such as orbital tightening and ear position are listed across all species, however other facial features and some specific changes are different (i.e. flattening of the nose and cheek regions are observed in rats, in contrast to bulging in mice) (Langford et al., 2010; Sotocinal et al., 2011). For this reason, it is important to develop species-specific grimace scales (McLennan et al., 2019).

In cats, methods to quantify facial changes, focusing on linear distances between specific facial landmarks (i.e. distances between ears and muzzle) allowed distinction between painful and pain-free animals (Holden et al., 2014). However, orbital tightening and whiskers position, that are commonly listed as action units in other species, were not included and a grimace scale for assessing pain in cats using facial expressions has not been published. More recently, a geometric morphometric approach has been described to study facial expressions of pain in cats. This approach has been proposed as basis for further application of machine learning algorithms for automated pain recognition (Finka et al., 2019). On the other hand, grimace scales are simple and readily applicable in a clinical context, and the development of a new tool could improve feline pain management.

The widespread adoption of pain assessment scales requires testing of its validity (the ability of the scale to identify pain), responsiveness (ability to detect clinically important changes, such as worsening pain or an improvement following analgesic administration), and reliability (the overall reproducibility of the scores between and within raters) (Streiner & Norman, 2008). Additionally, to be useful as a clinical tool, an analgesic intervention threshold (cut-off score) should be determined to guide when analgesics administration is warranted (Oliver et al., 2014; Reid et al., 2007).

This study aimed to develop and validate the Feline Grimace Scale (FGS) to detect acute pain associated with naturally-occurring conditions (i.e. diseases causing somatic or visceral pain). More specifically, we aimed 1) to identify the facial features associated with pain in cats during development of the FGS; 2) to assess construct validity (including responsiveness), criterion validity and reliability of the FGS; and 3) to derive a cut-off score for rescue analgesia.

We hypothesized that 1) the FGS would be able to discriminate painful from non-painful cats in a clinical setting; 2) painful cats would score higher than non-painful cats (construct validity); the FGS scores would decrease after analgesic treatment (responsiveness); the FGS would correlate with the G-CMPS (criterion validity); the scores from different observers and from the same observer over time would be reliable (inter- and intra-rater reliability); and 3) a cut-off for rescue analgesia would be determined with high sensitivity and specificity.

2.1.4. Methods

2.1.4.1. Ethical statement

The study protocol was reviewed and approved by the institutional animal care and use committee of the Faculty of Veterinary Medicine, Université de Montréal (protocol number 17-Rech-1863). Experiments were conducted in accordance with the Canadian Council on Animal Care guidelines.

2.1.4.2. Animals

Fifty cats of any age, breed and sex, admitted to the emergency and critical care unit of our veterinary teaching hospital (Centre hospitalier universitaire vétérinaire – CHUV) were recruited after owner's written and informed consent. Another twenty healthy cats from the teaching colony of our institution were included as controls (non-painful cats).

Cats were excluded if they presented diseases or conditions that could affect facial expressions (e.g. ophthalmic conditions, head trauma, etc.), excessively shy and feral behavior, or if they were administered sedatives and/or analgesics up to 24 hours before admission. Cats requiring immediate treatment (e.g. respiratory distress, severe hypovolemia, shock, active bleeding) were not recruited.

2.1.4.3. Pain assessment and video-recording

Client-owned cats were recruited between March and November 2017. Upon presentation, full physical examination was performed, and cats were placed individually in cages for a five-minute acclimation period. Pain was assessed by the main observer (MCE) according to the Glasgow composite measure pain scale for acute pain in cats (G-CMPS) (Calvo et al., 2014). Briefly, this scale contains six questions regarding the overall cat's appearance, response to interaction and palpation of a potentially painful area. Cats with scores of ≥ 4/16 were considered painful. Following the initial pain assessment, cats were filmed using a wide-angle high definition camera (GoPro Hero 5, GoPro, San Mateo, CA, USA) placed between the cage bars at the level of the eyes (Figure 6a). Video-recording of each animal was performed for six minutes for later image selection and assessment. Hospital staff were not present, and cats were left undisturbed during video-recording. After video-recording, analgesics were administered to painful cats (treatment choice at the discretion of the responsible clinician) and the same procedure for pain assessment using G-CMPS and video-recording was repeated one hour later.

Control cats were video-recorded between January and February 2017. The same video-recording procedure described above was performed with the control cats. A physical examination was performed, and pain was evaluated using the G-CMPS. These cats were filmed twice, with a one-hour interval (no analgesic treatment was given).

2.1.4.4. Image capture selection and development of the FGS

Image selection and assessment of facial expressions were performed by the main observer (MCE) using screenshots obtained from video-recordings (QuickTime Player, version 10.4, Apple Inc., Cupertino, CA, USA) when the cat was facing the camera, but not sleeping, grooming or vocalizing (Figure 6b). From each six-minute video, one image was selected from every two-minute interval and the best of three images (based on image quality) was chosen. The images were cropped to include only the cats' face and part of the shoulders. This observer was not blinded to the time point of the video (before or after analgesia or one-hour interval) during the image selection; however, the observer did not have access to their corresponding G-CMPS scores.

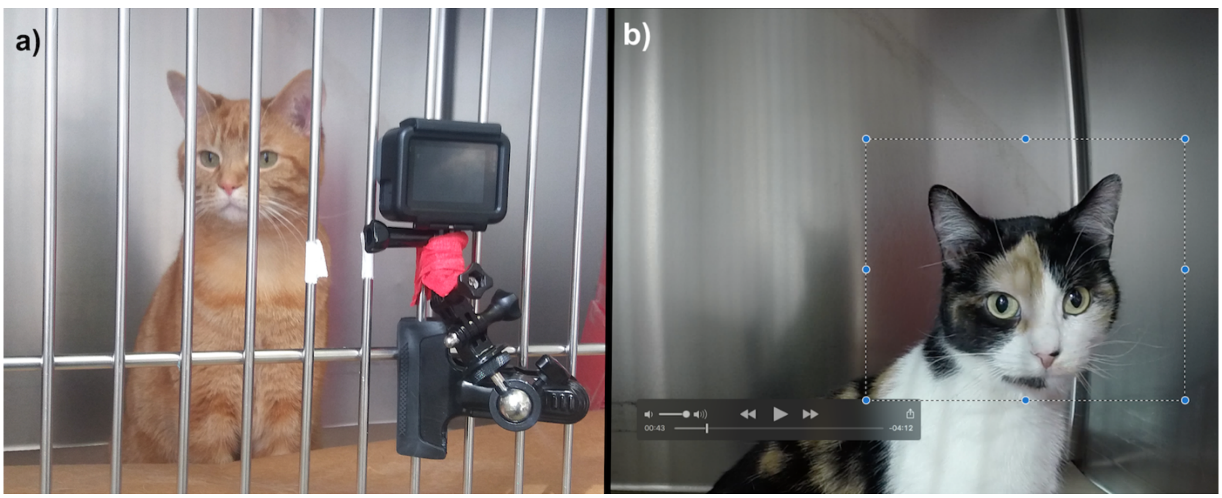


Figure 6. Video-recording and image capture

a) Video-recording of cats. A wide-angle lens camera was placed between the cage bars and cats were video-recorded undisturbed for 6 min. b) Image cropping and selection. Files were downloaded on to a computer and screenshots obtained from video-recordings.

Two individuals (MCE and RW) visually compared images (thumbnails and full-size pictures) of 20 controls and 31 client-owned painful cats (G-CMPS ≥ 4/16) before the administration of analgesics to identify any differences in facial features between these two groups. Features that were consistently different between both groups were listed as action units (AU) and used to create the FGS. The AU were illustrated and described in a manual that was later used for training purposes (Appendix A - FGS training manual). In order to corroborate these findings, distances between pairs of landmarks were measured (PixelStick, version 2.10.1, PlumAmazing Softwares, Princeville, HI, USA) and the ratios between two pairs were calculated (distances between the ear tips and ear bases, eye height and width, muzzle height and width; adapted from (Finlayson et al., 2016; Holden et al., 2014) (Figure 7a). Ear angles were also measured (Screen rulers for Mac, version 1.13.1, Ondesoft Softwares, Venice, CA, USA) (Figure 7b). These measurements were performed on 51 images (20 controls and 31 client-owned painful cats before analgesia) by the main observer (MCE), not blinded to the groups. A second observer (RW), blinded to the groups, independently repeated the measurements in a randomly chosen sample of one third of the images using a random number generator. The agreement between both observers was calculated.

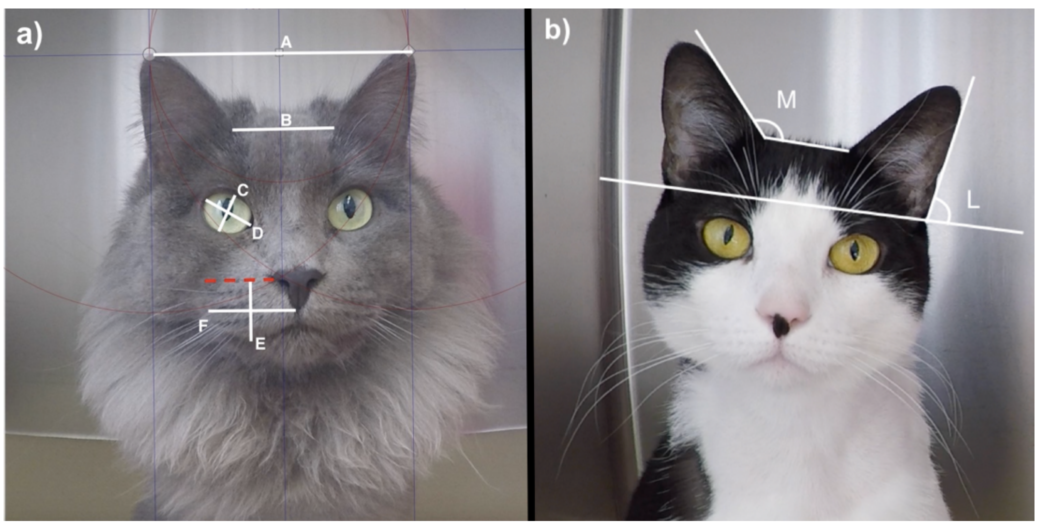


Figure 7. Distances and angles measured in the selected images

a) Measured distances for the ears, eyes and muzzle. A - Ears: tip to tip; B - Ears: base to base; C - Eye: height; D - Eye: width; E - Muzzle: height; F - Muzzle: width. Eyes and muzzle distances were measured bilaterally and averaged. b) Measured ear angles. M - Medial angle: between the medial border of the ear and top of the head; L - Lateral angle: between the lateral border of the ear and an imaginary line connecting both marginal cutaneous pouches. Angles were measured bilaterally and averaged.

2.1.4.5. Image scoring and validation of the FGS

Four observers, blinded to the groups and time when the images were obtained (VSYL, BM, PVS and DSJP; two PhD candidates and two board-certified veterinary anesthesiologists) were supplied with the training manual. They independently scored 110 images from 55 cats (20 controls and 35 client-owned, before and after analgesic treatment or one-hour interval), twice 30 days apart. The images were numerated using a random sequence generator (www.randomization.com) by one investigator (MCE) and a different sequence was generated for each round. An online questionnaire (SurveyMonkey, San Mateo, CA, USA) was built with one image per page and five questions regarding each of the AU: ears, eyes, muzzle, whiskers and head position. For each AU, the observers could select one of the four following options: 0 = AU is absent; 1 = moderate appearance of the AU, or uncertainty over its presence or absence; 2 = obvious appearance of the AU; or "not possible to score" (e.g. if the AU was not clearly visible). Image scoring was performed between February and April 2018.

A single total pain score per image was calculated as the sum of scores from each AU divided by the maximum possible score (e.g. 4/10 = 0.4), excluding the AU marked as "not possible to score". The final FGS score ranged from 0 to 1. Images receiving more than two "not possible to score" from the same rater were excluded from the analysis.

2.1.4.6. Statistical analysis

Statistical analyses were performed using the SPSS software (version 24.0.0.1, IBM SPSS Statistics, Armonk, NY, USA).

Development of the FGS

Bonferroni-corrected independent t-tests were used to compare linear measurements (mean distance ratios and angles) between controls and painful client-owned cats.

Intraclass correlation coefficients (ICC) were calculated based on single measures (ICC_{single}) and average of measures ($ICC_{average}$) when indicated, using two-way random effects model for absolute agreement between the linear measurements (distance ratios and angles) performed by the two observers.

Validation of the FGS

The effect of sex on FGS scores was assessed using a linear mixed model with group as between-subject factor and time as within-subject factor, using sex as a cofactor.

Construct validity (by known-groups discrimination) was assessed based on hypothesis testing. The hypothesis was that painful cats would score higher than non-painful ones and Mann-Whitney U test was used to compare the scores from both groups. Responsiveness, the sensitivity to change (as part of construct validity), was assessed based on the hypothesis that FGS scores for painful client-owned cats before analgesia would be higher than those assessed after analgesic treatment and the scores would not change in control group after the 1h interval. Wilcoxon signed rank tests were used to compare the scores within each group. The average of the scores given by four observers unaware of the groups and time points was used. A p value < 0.05 was considered significant.

Criterion validity, the correlation with a gold standard (G-CMPS), was assessed using Spearman's rank correlation between FGS and G-CMPS scores provided by the main observer (MCE) and interpreted as following: rho < 0.19 = very weak; 0.20-0.39 = weak; 0.4-0.59 = moderate; 0.6-0.79 = strong; 0.8-1.0 = very strong (Evans, 1996).

Reliability between raters (inter-rater) and by a single rater over time (intra-rater) were assessed with an ICC, calculated for each of the four observers independently assessing the same image and a single rater (comparing scores assigned to the same images on two rounds, 30 days apart). For inter-rater reliability, a two-way random effects ICC model for absolute agreement was used (calculated for both rounds 1 and 2). For intra-rater reliability, a two-way mixed effects ICC model, for absolute agreement was chosen. Interpretation was based on the ICC_{single} as following: ICC < 0.5 = poor, 0.5-0.75 = moderate, 0.75-0.9 = good, and > 0.90 = excellent reliability (Koo & Li, 2016).

The agreement between the scores from the main observer (MCE) and the average of the four raters was calculated using the Bland and Altman method (Bland & Altman, 1986).

Internal consistency was assessed with Cronbach's alpha coefficient calculated for the final FGS score and for each AU based on the scores of the main observer (MCE), recalculating the coefficient with each AU deleted. Interpretation was performed as following: alpha <0.65 =

unsatisfactory; 0.65-0.69 = fair; 0.7-0.74 = moderate; 0.75-0.79 = good; > 0.8 = excellent (Ponterotto & Ruckdeschel, 2007).

The analgesic threshold (cut-off for rescue analgesia) was determined with a receiver operating characteristics (ROC) curve analysis. The ability of the FGS to discriminate between absence (G-CMPS < 4 = no pain) and presence of pain (G-CMPS ≥ 4 = pain) was assessed by comparing the area under the curve (AUC) generated from the scores of the main observer (MCE) with an AUC of 0.5. An AUC between 0.50-0.70 represented low accuracy; between 0.70-0.90 = moderate accuracy; and for AUCs over 0.90 = high accuracy (Fischer et al., 2003).

2.1.5. Results

2.1.5.1. Development of the FGS

Client-owned cats: 50 adult cats of various breeds (24 females, 26 males) were enrolled in the study. Most cats presented with abdominal pain (diagnoses included hepatic lipidosis, cholangitis, pancreatitis, inflammatory bowel disease, suspected foreign body, lymphoma, constipation, urethral obstruction, urolithiasis and idiopathic cystitis). Eleven cats were excluded (Figure 3). Landmarks for measurements could not be identified in black cats (n = 2, comprised within poor image quality). Finally, 31 cats [domestic short haired (n = 17), domestic long-haired (n = 7), Siamese (n = 2), Cornish Rex (n = 1), Bengal (n = 1), Maine Coon (n = 1), Savannah (n = 2)] presented G-CMPS scores ≥ 4/16 [median (range): 5 (4-8)] and their images were included in the development of the FGS (14 females, 17 males; mean ± SD age 6.3 ± 3.6 years and weight 5.6 ± 1.9 kg) (Figure 8).

Control cats: All of the 20 cats [domestic short-haired (n = 17), domestic long-haired (n = 3)] presented G-CMPS scores = 0 [median (range): 0 (0-0)] and were included in image assessment (15 spayed females, 5 castrated males; mean ± SD age 3.1 ± 1.1 years and weight 3.8 ± 0.5 kg) (Figure 8).

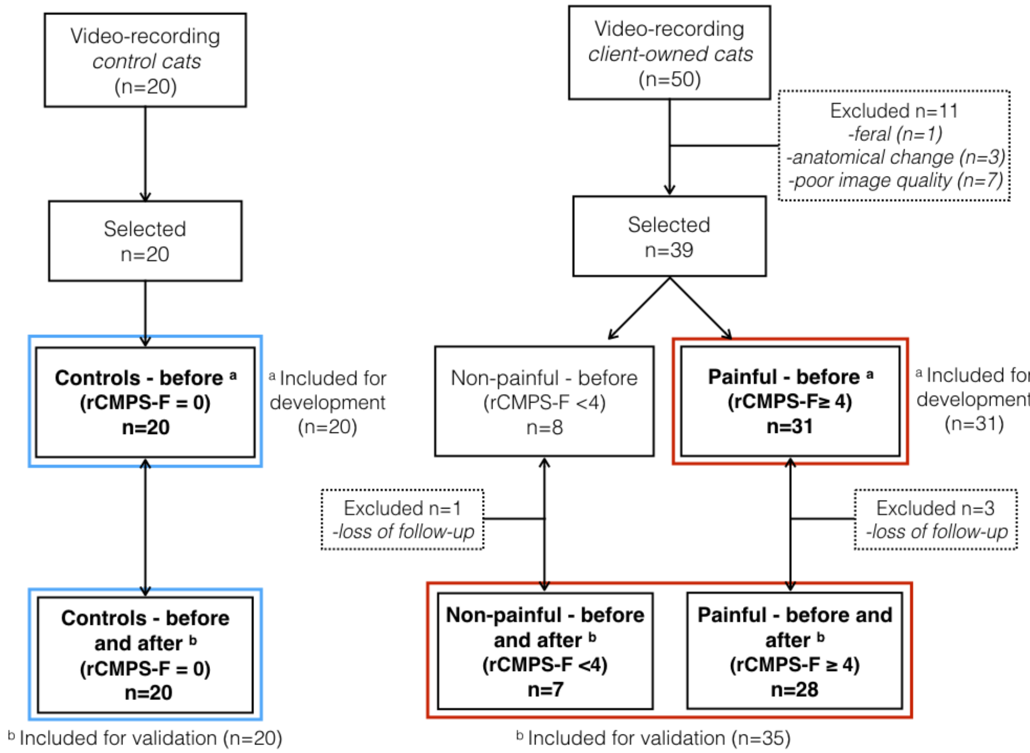


Figure 8. Flowchart of the animals included in the study

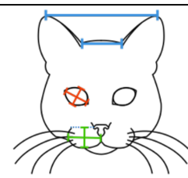
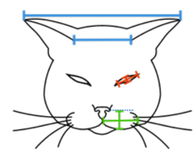
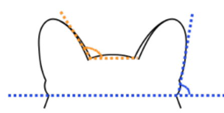
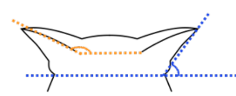
Blue boxes = control cats included for: [a] development and [b] validation of the Feline Grimace Scale (FGS). Red boxes = client-owned cats included for: [a] development and [b] validation of the FGS.

Both evaluators agreed on their visual assessment. Action units were defined as follows: 1) Ear position: refers to the tips of ears pulled apart (flattened) and rotated outwards; 2) Orbital tightening: narrowing of the orbital area, with a height between eyelids smaller than 50% of eyes width, or tightly closed eyelid (squinted eyes); 3) Muzzle tension: flattening and stretching of the muzzle from round to an elliptical shape (muzzle may be bulged); 4) Whiskers position: movement of whiskers forward (rostrally and away from the face), as if standing on end (spiked); 5) Head position (in relation to the shoulders): head below the shoulder line or tilted down (chin toward the chest) (Appendix A - FGS training manual).

Linear distance ratios and angles were significantly different between groups (Table 5). The agreement between the measurements of the two observers (MCE and RW) was good for ears ratio and medial ear angle, and excellent for the eyes ratio, muzzle ratio and lateral ear angle. Ears ratio

- ICC $_{single}$ = 0.76 (95% CI: 0.46 - 0.90); Eyes ratio - ICC $_{single}$ = 0.97 (95% CI: 0.92 - 0.99); Muzzle ratio - ICC $_{single}$ = 0.91 (95% CI: 0.78 - 0.97); Medial angle - ICC $_{single}$ = 0.77 (95%CI: 0.48 - 0.91); Lateral angle - ICC $_{single}$ = 0.91 (95% CI: 0.76 - 0.96).

Table 5. Calculated ratios for the ears, eyes and muzzle, and measured ear angles to discriminate control and client-owned painful cats

Measurements	Control (n = 20)	Painful (n = 31)	p value
			
Ear tips/bases ratio	2.85 ± 0.3	2.34 ± 0.3	p < 0.001
Eyes height/width ratio	0.79 ± 0.1	0.50 ± 0.2	p < 0.001
Muzzle height/width ratio	0.70 ± 0.1	0.50 ± 0.1	p < 0.001
			
Medial ear angles	126.5 ± 4.7˚	140.4 ± 6.5˚	p < 0.001
Lateral ear angles	78.9 ± 3.1˚	68.5 ± 5.9˚	p < 0.001

Data are presented as mean ± SD. Ratios were calculated between the distances of ear tips/bases, eyes height/width, muzzle height/width. Angles are presented in degrees. Excluding the distances between ear tips and bases, other measurements were performed bilaterally and averaged. Means across groups were compared using Bonferroni-corrected independent t-tests.

2.1.5.2. Validation of the FGS

Images from 35 client-owned [domestic short haired (n = 21), domestic long-haired (n = 8), Siamese (n = 2), Cornish Rex (n = 1), Bengal (n = 1), Maine Coon (n = 1), Savannah (n = 1)] and 20 control cats (before and after analgesic treatment or one-hour interval) were included for the validation of the FGS. Seven out of the 35 client-owned cats presented G-CMPS scores < 4/16 and 28 presented scores ≥ 4/16 (16 females, 19 males; mean ± SD age 6.8 ± 3.8 years and weight 5.8

± 2.2 kg) (Figure 3). Sex did not produce a significant effect on FGS scores (linear mixed model; p = 0.63).

Median (range) time to complete the survey was 78 (33 - 103) minutes and 63 (32 - 71) minutes for rounds 1 and 2, respectively. The AU whiskers position had the highest percentage of "not possible to score" answer, representing 10.2% of the images whereas the incidence of "not possible to score" selections for muzzle, head, ears and eyes was 3.6%, 2.7%, 0.22% and 0%, respectively.

Construct validity

Known-groups discrimination was used to assess construct validity. The FGS scores were higher in client-owned painful (before analgesia) than in control cats [median (range): 0.71 (0.18 - 0.98) and 0 (0 - 0.1), respectively (Mann-Whitney U test, p < 0.001; Figure 9)].

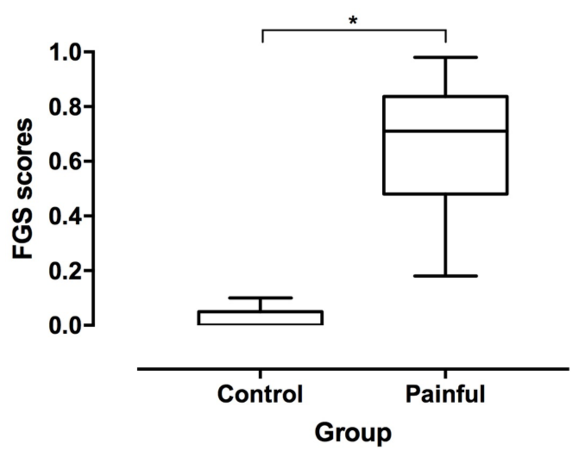


Figure 9. Construct validity of the FGS

Boxplot showing median (solid line) and interquartile ranges for Feline Grimace Scale (FGS) scores in control (n = 20) and painful cats (n = 28). The whiskers represent the range. The FGS scores were higher for painful than control cats (Mann-Whitney U test, *p < 0.001).

Responsiveness

For the responsiveness assessment, data from 15 painful cats before and after analgesia were available. Different analgesic drugs, doses and routes of administration were used:

butorphanol (n = 9), buprenorphine (n = 4), hydromorphone (n = 1), meloxicam (n = 1). The FGS scores did not change in the control group after the one-hour interval (without any treatment) [median (range): 0 (0-0.1) and 0 (0-0.16); Wilcoxon signed rank test, p = 0.342], but a significant decrease was observed in painful cats after analgesic treatment [0.72 (0.18-0.98) and 0.44 (0.11-0.93); Wilcoxon signed rank test, p = 0.003] (Figure 10). Similarly, G-CMPS scores did not change in the control group [median (range) before: 0 (0-0) and after: 0 (0-0), Wilcoxon signed rank test, p = 1.0] and significantly decreased in painful cats after treatment [before: 5 (4-8) and after 3 (0-10), Wilcoxon signed rank test, p = 0.006].

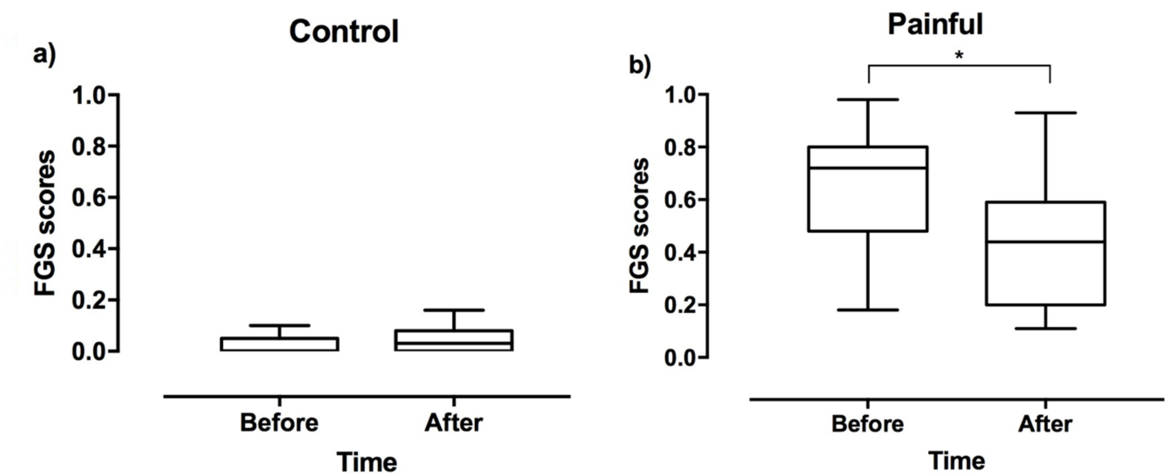


Figure 10. Responsiveness of the FGS

a) Boxplot showing median (solid line) and interquartile ranges for Feline Grimace Scale (FGS) scores in control cats (n = 20). The FGS scores did not change for control cats after 1h (without treatment; Wilcoxon signed rank test, p = 0.342). b) Boxplot showing median (solid line) and interquartile ranges for FGS scores in painful cats (n = 15). The FGS scores decreased in painful cats after analgesia (Wilcoxon signed rank test, *p = 0.003).

Criterion validity

A very strong correlation was observed between the G-CMPS and FGS (rho = 0.86; p < 0.001). A total of 110 images from both groups (control, n = 20 and client-owned, n = 35) and both time points (before and after) were considered for this analysis (Figure 11).

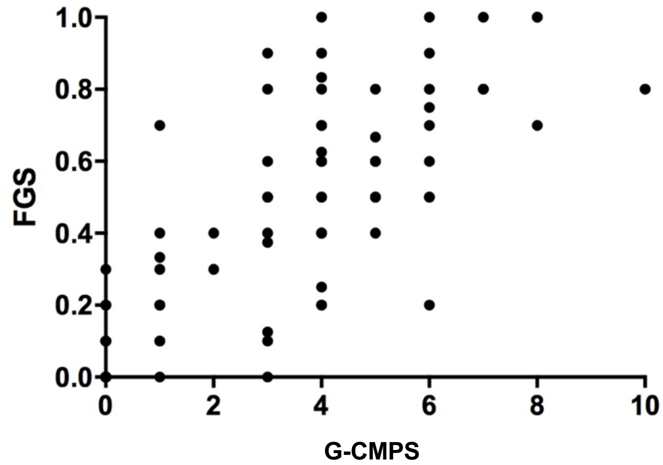


Figure 11. Criterion validity of the FGS

Scatterplot showing correlation between Feline Grimace Scale (FGS) and Glasgow composite measure pain scale for acute pain in cats (G-CMPS) scores (Spearman's rho = 0.86, $p < 0.001$) from control (n = 20) and client-owned cats (n = 35).

Reliability

Inter-rater reliability for FGS scoring was good during both round 1 and 2 [ICC_{single} = 0.89 (95% CI: 0.85 - 0.92) and ICC_{single} = 0.89 (95% CI: 0.84 - 0.92), respectively]. Inter-rater reliability of individual AU ranged from moderate (muzzle and whiskers) to good (ears, eyes and head position) (Table 6).

Intra-rater reliability was excellent for all observers 30 days after the first round of scoring. Observer 1: ICC_{single} = 0.91 (95% CI: 0.81 - 0.95); observer 2: ICC_{single} = 0.94 (95% CI: 0.91 - 0.96); observer 3: ICC_{single} = 0.95 (95% CI: 0.93 - 0.97); observer 4: ICC_{single} = 0.92 (95% CI: 0.88 - 0.94).

The scores from the main observer (MCE) and the average of the four raters had strong agreement, minimal bias (0.0047) and narrow limits of agreement (-0.18 – 0.19).

Table 6. Inter-rater reliability of the Feline Grimace Scale

Action unit		Round 1 ICC (95% CI)	Round 2 ICC (95% CI)
Ears	ICC $_{single}$	0.87 (0.82 - 0.90)	0.85 (0.81 - 0.89)
	ICC $_{average}$	0.96 (0.95 - 0.97)	0.96 (0.94 - 0.97)
Eyes	ICC $_{single}$	0.86 (0.81 - 0.89)	0.83 (0.77 - 0.88)
	ICC $_{average}$	0.96 (0.95 - 0.97)	0.95 (0.93 - 0.97)
Muzzle	ICC $_{single}$	0.63 (0.53 - 0.72)	0.67 (0.58 - 0.75)
	ICC $_{average}$	0.87 (0.82 - 0.91)	0.89 (0.85 - 0.92)
Whiskers	ICC $_{single}$	0.55 (0.43 - 0.66)	0.60 (0.48 - 0.70)
	ICC $_{average}$	0.83 (0.75 - 0.89)	0.86 (0.79 - 0.90)
Head position	ICC $_{single}$	0.90 (0.87 - 0.93)	0.88 (0.84 - 0.91)
	ICC $_{average}$	0.97 (0.96 - 0.98)	0.97 (0.96 - 0.98)
Final FGS score	ICC $_{single}$	0.89 (0.85 - 0.92)	0.89 (0.84 - 0.92)
	ICC $_{average}$	0.97 (0.96 - 0.98)	0.97 (0.96 - 0.98)

Intraclass correlation coefficient (ICC) estimates and their 95% confidence intervals (95% CI) were calculated based on single measures (ICC $_{single}$) and average (ICC $_{average}$) of measures (raters = 4), using two-way random effects model for absolute agreement. Interpretation was performed based on ICC $_{single}$ as following: ICC < 0.5 = poor reliability, between 0.5 and 0.75 = moderate reliability, between 0.75 and 0.9 = good reliability, and > 0.90 = excellent reliability (Koo & Li, 2016).

Internal consistency

Cronbach's alpha coefficient calculated for the final FGS score was 0.89, indicating excellent internal consistency. The recalculated coefficients by deleting each AU were: ears (alpha if deleted = 0.86), eyes (alpha if deleted = 0.90), muzzle (alpha if deleted = 0.85), whiskers (alpha if deleted = 0.88), head position (alpha if deleted = 0.86). The alpha was minimally affected by removing any single AU, thus all AU contributed similarly to the final score.

Analgesic threshold

The classification of 110 images according to the G-CMPS scores resulted in 43 considered as “presence of pain” and 67 as “absence of pain”. The ROC curve was originated by plotting the true positive rate (sensitivity) against false positive rate (1 – specificity). The AUC of 0.94 (95% CI: 0.89 - 0.98) with $p < 0.001$ indicated a high discriminative ability (high accuracy) for the FGS (Figure 12). The cut-off score of 0.39 (from a maximum of 1.0) was selected for representing an optimal balance between sensitivity (90.7%) and specificity (86.6%) (Table 7).

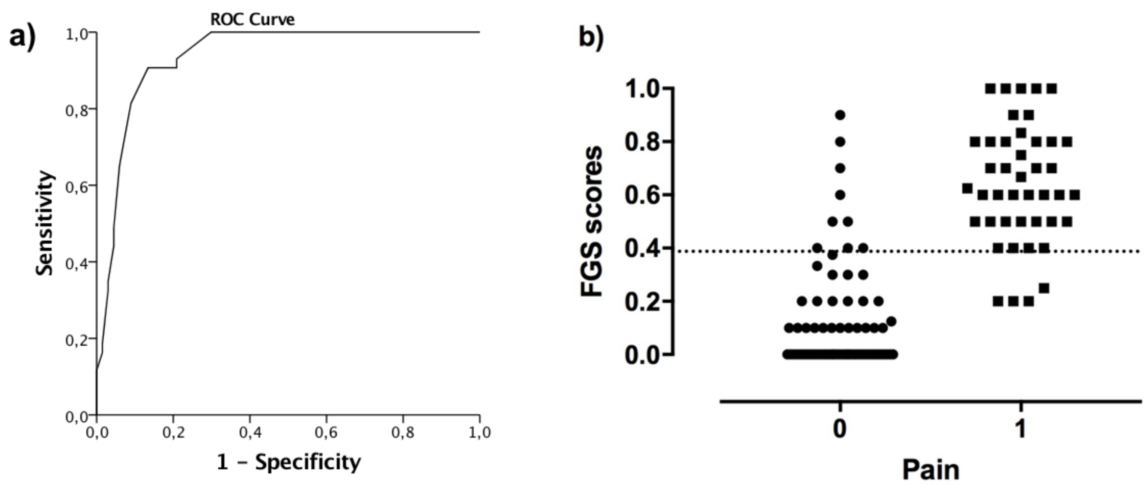


Figure 12. Analgesic threshold derivation

a) Receiver operating characteristics (ROC) curve showing the cut-off point > 0.39 for rescue analgesia, with sensitivity of 90.7%, and specificity of 86.6%. The area under the curve (AUC) of 0.94 (95% CI: 0.89 - 0.98), represents high accuracy of the Feline Grimace Scale (FGS). b) Scatterplots showing the FGS scores and cut-off point for rescue analgesia (0.39), identified from the ROC curve analysis. N = 110 images of cats' faces were classified as “absence of pain (0)” or “presence of pain (1)” according to Glasgow composite measure pain scale for acute pain in cats (G-CMPS) scores.

Table 7. Analgesic thresholds and their sensitivity and specificity

Cut-off score	Sensitivity (%)	Specificity (%)
0.05	100.0	50.7
0.11	100.0	68.7
0.16	100.0	70.1
0.23	93.0	79.1
0.28	90.7	79.1
0.32	90.7	83.6
0.35	90.7	85.1
0.39*	**90.7**	**86.6**
0.45	81.4	91.0
0.55	65.1	94.0
0.61	48.8	95.5
0.65	46.5	95.5
0.68	44.2	95.5
0.73	34.9	97.0
0.78	32.6	97.0
0.82	18.6	98.5
0.87	16.3	98.5
0.95	11.6	100.0

*The cut-off for rescue analgesia of >0.39 was selected based on the highest values of sensitivity (90.7%) and specificity (86.6%). All the other cut-off values are the averages of two consecutive ordered observed test values.

2.1.6. Discussion

This study reported the development and validation of the FGS in the clinical setting using image assessment. A diverse sample of cats was included, with cats presenting pain from different sources and intensities. The FGS was developed by comparing the facial features of control and painful cats, then its validity and reliability were tested using images obtained from the video-recordings.

Two distinct populations of cats (healthy control cats and client-owned painful cats) of various breeds were studied. There were more females than males in the control group, but both sexes were equally represented within the client-owned group, besides, there was no effect of sex on FGS scores. Sex and strain differences in laboratory animals have been previously reported in pain research, hence the importance of including both sexes in the development of a pain scale (Mogil & Chanda, 2005). Although different breeds were represented within our client-owned population, most of the cats were domestic short-haired. Brachycephalic breeds were not included. Indeed, one Persian and one Himalayan were initially recruited, however they were excluded from final analysis due to poor image quality. Morphological differences (round-shaped skulls and decrease in facial width) related to breed-specific features have been previously observed in brachycephalic cats (Kunzel et al., 2003). At this point it is not known if brachycephalic cats present the same AU related to pain as mesocephalic and dolichocephalic cats, and if these changes could present a source of bias in the FGS.

Five AU (ear position, orbital tightening, muzzle tension, whiskers position and head position) were identified. Similar AU have been previously described in mice, rats, rabbits, horses and ferrets (Dalla Costa et al., 2014; Keating et al., 2012; Langford et al., 2010; Reijgwart et al., 2017; Sotocinal et al., 2011), and head position has been described in sheep (Hager et al., 2017).

These AU described on the FGS are consistent with ear action descriptors: "ears flattener", "ears downwards", and action units: "eye closure", "nose wrinkler and upper lip raiser", "whiskers retractor" and "whiskers protractor" presented on CatFACS (Caeiro et al., 2017). They are also consistent with a geometric morphometric study that identified changes in the feline facial shape after a painful stimulus including "a more lateral and ventral positioning of the ears", "a slightly narrowed eye aperture" and "reduced distance between the cheeks, mouth and nose region" (Finka et al., 2019).

The visual comparison of the images for the development of the FGS was performed based on previous work on the development and validation of the Mouse, Rat, Rabbit, Horse, Sheep and Lamb Grimace Scales (Dalla Costa et al., 2014; Guesgen et al., 2016; Hager et al., 2017; Keating et al., 2012; Langford et al., 2010; McLennan et al., 2016; Sotocinal et al., 2011) where similar methodology of comparison of two distinct group of animals were described. In addition to the visual comparison of the images, distances and angles were measured as an additional quantitative outcome to corroborate our findings. These measurements were used in the description of the AU for the training manual. A second observer independently repeated the measurements in a randomly selected sample of one third of the images to avoid bias in measurements, and the agreement between their measures was good to excellent. The ratios between two distances rather than the actual measures were considered for group comparisons, to account for differences in the distance between the cat and the camera. According to our results, the eyes' height was approximately 80% of its width in control cats and approximately 50% in the painful ones (when eyes are partially closed or squinted). Similar results were observed in lambs undergoing tail docking (Guesgen et al., 2016). Likewise, the muzzle height decreased from nearly 70% of its width to 50% in the presence of pain. The medial ear angle increases and the lateral decreases as the ears flattens in painful cats. Distances between two pairs of landmarks in the cats' ears and muzzle were previously reported as significantly different between painful and pain-free cats (Holden et al., 2014).

The assessment of facial expressions in laboratory animals has been performed using still images (screenshots obtained from video recordings) (Langford et al., 2010; Sotocinal et al., 2011). We used similar methodology, where the cats were video-recorded undisturbed in their cages. Animals were free-ranging and able to express their natural behaviors. Other studies included photographs taken when the animals were restrained (Di Giminiani et al., 2016; Holden et al., 2014; McLennan et al., 2016; Reijgwart et al., 2017). However, physical restraint significantly affected facial expressions scores in lambs (Guesgen et al., 2016). Avoiding handling or close contact with the animal during image acquisition has been suggested, and the authors argue that leaving an animal to perform the behaviors in conditions that meet their needs is likely to yield the best results during the development of facial expression scales (McLennan et al., 2019).

In our study, black cats were excluded due to lacking image quality and difficulty in identifying landmarks in their faces. Similar difficulties were reported when scoring dark coated horses' faces (Dalla Costa et al., 2014). These issues may be resolved using high definition cameras or real-time scoring, which has been demonstrated to be possible in rats (Leung et al., 2016). In a recent study, our research group tested real-time scoring using the FGS and reported good agreement (small bias and narrow limits of agreement) with image assessment (Evangelista, Benito, et al., 2020).

The FGS showed high discriminative ability between painful and non-painful cats. Construct validity assesses whether the tool is measuring something (a construct) that cannot be directly observed (e.g. pain) (Streiner & Norman, 2008). Known-groups discrimination was the method chosen to confirm the construct validity through hypothesis testing. This approach is in agreement with the validation of the Mouse and Rat Grimace Scales (Langford et al., 2010; Sotocinal et al., 2011) and behavior-based feline pain scales (Brondani et al., 2013; Calvo et al., 2014). For responsiveness assessment, different analgesic drugs, doses and routes of administration were used. Even with such a diverse sample receiving different analgesic drugs and dosage regimens, the FGS detected the response to analgesic treatment in painful cats. Correspondingly, the scores in the control group did not change after one hour, however, these cats did not receive any sham analgesia or handling. This represents a limitation and the impact of the physical experience of drug administration was not accounted. In a follow-up study, it would be important to determine the effect of specific analgesic drugs using fixed dosage regimens (including sham analgesia) after a standardized painful stimulus to confirm these findings.

Criterion validity was tested using concurrent validation of a new scale and a 'gold standard' (Streiner, 1993; Streiner & Norman, 2008). In the absence of a gold standard when evaluating pain, a validated pain scale for cats (G-CMPS) (Calvo et al., 2014) was used. Similar approaches have been applied in laboratory rodents, where the criterion standard was mechanical hypersensitivity testing (De Rantere et al., 2016; Langford et al., 2010). The most recent version of the G-CMPS (Reid et al., 2017) was not used since it includes two features of facial expressions that could bias our results. Concurrent validation was shown by correlating the FGS with G-CMPS scores and a very strong correlation was observed. Additionally, the G-CMPS scores were used to determine the presence of pain (G-CMPS ≥ 4) or its absence (G-CMPS < 4) for the ROC curve analysis. The drawback of this approach is how to be sure that the animals were in pain and not stressed. To reduce the bias, excessively shy and feral cats were excluded, knowing that demeanor influences the scores of feline-specific pain scales (Buisman et al., 2017). Future studies should investigate how changes in demeanor impact the FGS.

The FGS showed good overall inter-rater reliability ($ICC_{single} = 0.89$) and excellent intra-rater reliability ($ICC_{single} > 0.91$). Our results are similar to those reported for mice, ICC = 0.90 (Langford et al., 2010); rats, ICC = 0.90 (Sotocinal et al., 2011) / $ICC_{single} = 0.85$ (Oliver et al., 2014) and horses, ICC = 0.92 (Dalla Costa et al., 2014). Interpretation was performed based on the ICC estimate for single measures (ICC_{single}). The choice of reporting the estimates based on a single measure or the average of k measures depends on how the scale will be applied in a clinical context (e.g. if a decision will be made based on the scores of a single observer or on the average of a number of observers). The ICC $_{average}$ is usually higher and ideally, both estimated should be reported along with their confidence interval 95%, as reported herein (Koo & Li, 2016).

A neglected area in grimace scale research is the role of rater training. It is currently unclear to what extent training is important as most papers do not describe if training has taken place (Zhang et al., 2019). The raters in our study have years of experience working with cats. It is unknown, and deserves further investigation, how reliability would be affected by novice raters.

The internal consistency of the FGS was excellent. This result agrees with those reported for the Mouse and Rat Grimace Scales (alpha = 0.89 and 0.84, respectively) (Langford et al., 2010; Oliver et al., 2014). The Cronbach's alpha must be interpreted with caution, as the value will be higher for longer scales (Streiner, 1993). However, interpretation was performed according to the guidelines for scales with 6 items or less and a sample size between 100 and 300 (Ponterotto & Ruckdeschel, 2007).

The clinical utility of a pain scale is improved when an objective cut-off (or interventional) score informs the need for analgesia. For the FGS, the score for rescue analgesia is >0.39 out of 1.0. The cut-off score is a guide to help with treatment decisions and other values can be adopted depending on desired sensitivity and specificity. Similar methodology based on the ROC curve analysis, was used for the validation of the UB-MCPS (Brondani et al., 2013), Rat Grimace Scale (Oliver et al., 2014) and a Sheep Pain Facial Expression Scale (McLennan et al., 2016). Further studies are warranted to corroborate this finding in the clinical setting, using real-time assessment.

This study has some limitations: 1) It was an observational study, the decision for analgesic treatment was made by the clinician in charge. The observer did not interfere with the clinical judgement. It would be pertinent to test the FGS's performance in a controlled interventional study, using the cut-off score for analgesic treatment determined by the ROC curve. 2) Image selection from the video recordings and the development of the scale were performed by the same observer (MCE). This observer was not blinded to the groups or time points of the video (before or after) during the image selection. Additionally, image selection took place three months before the beginning of the scoring sessions. Our research group is carrying out another study addressing this limitation, where image selection is performed by an independent observer. 3) No power analysis or sample size calculation was performed before the beginning of the study since it is not possible to estimate the percentage of client-owned cats that would be presented with pain. There is no consensus to define sample size in studies involving the development of pain scales with the same rigor as found in controlled trials (Anthoine et al., 2014). Some grimace scales were developed using a within-subject design. In this setting, fewer subjects are required and the animal is evaluated before and after the induction of a standardized painful stimulus, which was the case in mice, rats, horses, sheep and ferrets (Dalla Costa et al., 2014; Hager et al., 2017; Langford et al., 2010; Reijgwart et al., 2017; Sotocinal et al., 2011). In contrast, other studies used a between-subject design (case-control studies) to identify dissimilar facial features and behaviors associated with naturally-occurring painful diseases in cats and sheep (Calvo et al., 2014; Holden et al., 2014; McLennan et al., 2016). 4) The lack of a baseline from the same animal is a limitation, and within-, rather than between-subject design is preferable (McLennan et al., 2019), however, to account for the variability among individuals we included a larger number of animals, similar to the population size used in the development of a Sheep Pain Facial Expression Scale (McLennan et al., 2016).

In conclusion, the FGS demonstrated high discriminative ability, a very high correlation with another validated instrument for pain assessment in cats, good overall inter-rater reliability, excellent intra-rater reliability, and excellent internal consistency. Furthermore, the FGS detected the response to analgesic treatment and a cut-off score was determined, making this a potential practical tool in both research and clinical settings. The FGS is a valid and reliable tool for acute pain assessment in cats.

2.1.7. Acknowledgements

This study received an unrestricted grant by Zoetis and Fonds en santé des animaux de compagnie (FSAC), Faculté de médecine vétérinaire, Université de Montréal. Dr. Marina Cayetano Evangelista is a recipient of the International Veterinary Academy of Pain Management fellowship (2017) and a scholarship from the Merit scholarship program for foreign students of the Ministère de l'Éducation et de l'Enseignement Supérieur du Québec. Dr. Beatriz Monteiro is a recipient of the Vanier Canada Graduate Scholarship. The authors wish to thank Dr. Hélène Ruel for technical support.

2.1.8. Additional Information

2.1.8.1. Author Contributions Statement

MCE, DSJP, PVS conceived the study. MCE, RW, EOT performed data collection. BM, VL, DSJP, PVS rated the images. MCE, DSJP, PVS analyzed the results. All authors reviewed the manuscript.

2.1.8.2. Competing interests

Dr. Paulo Steagall has received speaker honoraria and provided consultancy services to Zoetis. Dr. Beatriz Monteiro has provided consultancy services to Zoetis. As a sponsor of the study, Zoetis reviewed the proposal and approved the study design, but it was not involved with manuscript publication. This does not alter the authors' adherence to Scientific Reports policies on sharing data and materials.

2.2. Clinical applicability of the Feline Grimace Scale: real-time versus image scoring and the influence of sedation and surgery

Marina C. Evangelista[1], Javier Benito[1], Beatriz P. Monteiro[1], Ryota Watanabe[1], Graeme M. Doodnaught[1], Daniel S. J. Pang[1,2], Paulo V. Steagall[1]

2.2.1. Article identifier

This article was published in PeerJ 8:e8967 (2020)

doi: 10.7717/peerj.8967

---

[1] Department of Clinical Sciences, Faculty of Veterinary Medicine, Université de Montréal, Saint-Hyacinthe, QC, Canada
[2] Veterinary Clinical and Diagnostic Sciences, Faculty of Veterinary Medicine, University of Calgary, Calgary, Alberta, Canada

2.2.2. Abstract

Background. The Feline Grimace Scale (FGS) is a facial expression-based scoring system for acute pain assessment in cats with reported validity using image assessment. The aims of this study were to investigate the clinical applicability of the FGS in real-time when compared with image assessment, and to evaluate the influence of sedation and surgery on FGS scores in cats.

Methods. Sixty-five female cats (age: 1.37 ± 0.9 years and body weight: 2.85 ± 0.76 kg) were included in a prospective, randomized, clinical trial. Cats were sedated with intramuscular acepromazine and buprenorphine. Following induction with propofol, anesthesia was maintained with isoflurane and cats underwent ovariohysterectomy (OVH). Pain was evaluated at baseline, 15 minutes after sedation, and at 0.5, 1, 2, 3, 4, 6, 8, 12 and 24 hours after extubation using the FGS in real-time (FGS-RT). Cats were video-recorded simultaneously at baseline, 15 minutes after sedation, and at 2, 6, 12, and 24 hours after extubation for subsequent image assessment (FGS-IMG), which was performed six months later by the same observer. The agreement between FGS-RT and FGS-IMG scores was calculated using the Bland & Altman method for repeated measures. The effects of sedation (baseline versus 15 minutes) and OVH (baseline versus 24 hours) were assessed using linear mixed models. Responsiveness to the administration of rescue analgesia (FGS scores before versus one hour after) was assessed using paired t-tests.

Results. Minimal bias (-0.057) and narrow limits of agreement (-0.351 to 0.237) were observed between the FGS-IMG and FGS-RT. Scores at baseline (FGS-RT: 0.16 ± 0.13 and FGS-IMG: 0.14 ± 0.13) were not different after sedation (FGS-RT: 0.2 ± 0.15, $p = 0.39$ and FGS-IMG: 0.16 ± 0.15, $p = 0.99$) nor at 24 hours after extubation (FGS-RT: 0.16 ± 0.12, $p = 0.99$ and FGS-IMG: 0.12 ± 0.12, $p = 0.96$). Thirteen cats required rescue analgesia; their FGS scores were lower one hour after analgesic administration (FGS-RT: 0.21 ± 0.18 and FGS-IMG: 0.18 ± 0.17) than before (FGS-RT: 0.47 ± 0.24, $p = 0.0005$ and FGS-IMG: 0.45 ± 0.19, $p = 0.015$).

Conclusions. Real-time assessment slightly overestimates image scoring; however, with minimal clinical impact. Sedation with acepromazine-buprenorphine and ovariohysterectomy using a balanced anesthetic protocol did not influence the FGS scores. Responsiveness to analgesic administration was observed with both the FGS-RT and FGS-IMG.

2.2.3. Introduction

Despite the advent of scoring systems and assessment tools, pain is still under recognized and under treated in cats worldwide (Hunt et al., 2015; Lorena et al., 2014; Morales-Vallecilla et al., 2019; Reimann et al., 2017). Facial expressions are considered a reliable method of pain assessment in non-verbal humans and other mammals (Dalla Costa et al., 2014; Finka et al., 2019; Langford et al., 2010; Prkachin, 2009; Sotocinal et al., 2011).

Linear distances between specific facial landmarks and the quantification of changes in facial shape for the study of pain in cats have been published (Finka et al., 2019; Holden et al., 2014). Recently, a facial expression-based scoring system has been proposed for assessing acute pain in cats, namely the Feline Grimace Scale (FGS). This instrument has been developed and validated in cats with different sources and intensities of pain produced by naturally-occurring conditions; additionally, its reliability, criterion and construct validity (including responsiveness) were assessed using image scoring (Evangelista et al., 2019).

The standard methodology for assessing facial expressions of pain using grimace scales usually relies on image scoring (Dalla Costa et al., 2014; Langford et al., 2010; Sotocinal et al., 2011). The images are commonly screenshots obtained from videos which represents a time-consuming procedure that takes place weeks or months after video-recordings. In clinical practice, pain must be promptly assessed to ensure appropriate and immediate treatment. Real-time scoring has been described using the Mouse Grimace Scale (MGS) (Miller & Leach, 2015) and the Rat Grimace Scale (RGS) (Leung et al., 2016), but not the FGS.

Elective sterilization of female companion animals (i.e. ovariohysterectomy, OVH) is one of the most common surgical procedures performed in veterinary medicine. The administration of sedatives, anesthetics and analgesics are required in the perioperative setting and drugs such as ketamine have been shown to produce confounding effects by increasing psychomotor scores with feline pain scales (Buisman et al., 2016). Indeed, the influence of anesthesia, surgery and opioids (i.e. buprenorphine) on grimace scale scores have been demonstrated in mice and rats (Leung et al., 2016; Miller et al., 2015). Changes in facial shape have been reported in cats after OVH (Finka et al., 2019). It is not known how sedation and OVH itself may affect FGS scores, more specifically, and if these represent potential limitations in the application of the FGS in clinical practice.

The objectives of this study were: 1) to investigate whether the FGS could be successfully implemented in a clinical setting in cats undergoing OVH, enabling real-time pain assessment; 2) to evaluate the influence of sedation and OVH on the FGS; and 3) to reassess the responsiveness to analgesic treatment of the FGS in real-time. Our hypotheses were: 1) FGS scores obtained in real-time and by image assessment would be comparable when evaluated by the same observer; 2) FGS scores would not change after sedation with acepromazine and buprenorphine or at 24 hours after extubation when compared with baseline values; and 3) the FGS scores in real-time would decrease after analgesic treatment.

2.2.4. Materials and Methods

2.2.4.1. Ethical statement

The protocol was approved by the local animal care committee, Comité d'éthique de l'utilisation des animaux - Université de Montréal (protocol number 18-Rech-1825).

2.2.4.2. Animals

Eighty-one adult female domestic cats from local shelters were admitted to the veterinary teaching hospital (*Centre Hospitalier Universitaire Vétérinaire*) of the Faculty of Veterinary Medicine, Université de Montréal for elective OVH between June and October 2018. The study was designed to evaluate the analgesic efficacy of intraperitoneal administration of bupivacaine alone or in combination with dexmedetomidine (Benito et al., 2019). The current study occurred in parallel to evaluate changes in the FGS. Written informed consent for participation in the study was obtained for each patient.

Inclusion criteria included healthy cats based on physical examination and hematology. Exclusion criteria included cats presenting with cardiac arrhythmias, body condition score > 7 or < 3 on a scale from 1 to 9, anemia (hematocrit < 25%), hypoproteinemia (total protein < 5.9 g/dL), cats with high pain scores at presentation (see pain assessment), or any clinical signs of diseases during physical examination such as upper respiratory tract disease and conjunctivitis. Cats were admitted the day before the procedures and discharged 24 hours after the surgery. They were housed individually in adjacent cages in a cat-only ward with free access to water and food, and a litter box. Environmental enrichment included hanging toys, blankets and a cardboard box that cats could use as a hiding spot or as an elevated surface.

2.2.4.3. Anesthesia and surgery

Anesthetic and surgical procedures are described thoroughly elsewhere (Benito et al., 2019). In brief, food was withheld for 8-12 hours before general anesthesia; cats were premedicated with acepromazine (0.02 mg/kg) and buprenorphine (0.02 mg/kg) intramuscularly; anesthesia was induced with propofol intravenously to allow endotracheal intubation (4-6 mg/kg) and maintained with isoflurane in oxygen. Cats were randomly assigned to receive one of the two following treatments: bupivacaine (2 mg/kg) or bupivacaine (same dose) with dexmedetomidine (1 µg/kg). All cats received an intraperitoneal administration of either bupivacaine alone or in combination with dexmedetomidine intraoperatively. The intraperitoneal sterile solution was instilled (splashed) before OVH by the veterinary surgeon, over the right and left ovarian pedicles, and the caudal aspect of the uterine body in three equal volumes. Ovariohysterectomy was performed approximately one minute after intraperitoneal administration using the pedicle tie technique. All cats were discharged the day after surgery and returned to their respective shelters for adoption.

2.2.4.4. Sedation and pain assessment

Sedation is the depression of the animal's awareness to the environment and reduction of its response to external stimuli. Knowing that it can produce confounding effects in pain scores, sedation was assessed using a 5-point simple descriptive scale, where 0 = no sedation, 1 = able to stand but is wobbly; 2 = in sternal recumbency; 3 = can lift its head; 4 = fast asleep/no response to hand clap (adapted from Steagall et al., 2009).

Pain was evaluated using the short-form UNESP-Botucatu composite pain scale (SF-UBCPS) and the FGS. The SF-UBCPS is a novel pain scoring system consisting of 4 items (each item is scored from 0 to 3 adding up to a maximum score of 12 points) to evaluate the cats' posture, activity, attitude and reaction to touch and palpation of a painful site. This scale has been validated using video-based assessments (Luna et al., 2020). Cats with a baseline pain score of ≥ 2 using the SF-UBCPS were not included in the study. High baseline scores could indicate a mild degree of pain before the surgery, or an influence of the cat's demeanor (shy, fearful or feral) (Buisman et al., 2017). The FGS consists of five action units (ear position, orbital tightening, muzzle tension, whiskers change and head position), each one is scored from 0 to 2 (0 = action unit is absent; 1 = moderate appearance of the action unit, or uncertainty over its presence; and 2 = obvious appearance of the action unit) (Figure 13). If one or more action units were not visible, the observer had the option of marking "not possible to score" and the weight was redistributed to the other action units. A total FGS score was calculated by the sum of the scores of the action units divided by the total possible score, excluding those marked as not possible to score (e.g. 4/8 = 0.5). Images with more than 2 "not possible to score" were excluded from final analysis.

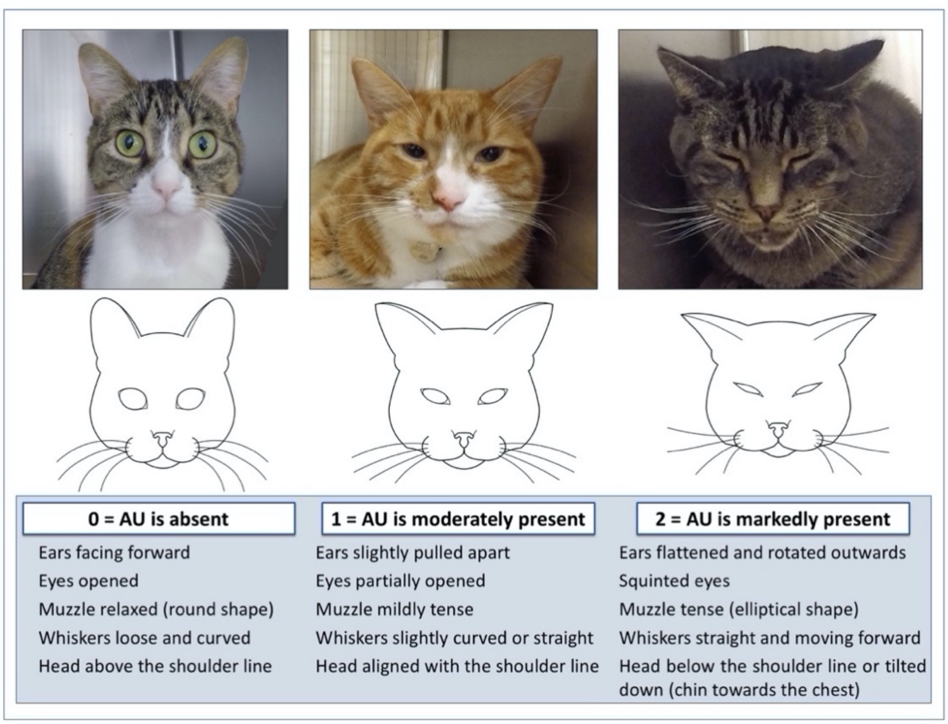


Figure 13. Illustration and description of the Feline Grimace Scale

The Feline Grimace Scale is composed of five action units (ear position, orbital tightening, muzzle tension, whiskers change and head position), each one is scored from 0 to 2 (0 = action unit is absent; 1 = moderate appearance of the action unit, or uncertainty over its presence; and 2 = obvious appearance of the action unit).

Sedation and pain assessments were performed by one observer (MCE) at baseline (morning before the surgery), 15 minutes after premedication, and at 0.5, 1, 2, 3, 4, 6, 8, 12 and 24 hours after extubation. Sedation was evaluated before pain assessment. The FGS scores were obtained in real time (FGS-RT) using instantaneous sampling during three minutes of observation without interacting with the cat. A stopwatch was set to control the time, and at the end of each minute (first, second and third minutes), one score per action unit was assigned (representing the whole minute of observation) and a total FGS-RT score was obtained. The average of those three scores was considered the FGS-RT score for that time-point [i.e. (0.5 + 0.4 + 0.5) / 3 = 0.47]. Three-minute videos were recorded simultaneously during the FGS-RT assessment using a high-definition video camera (GoPro Hero 5, GoPro Inc., San Mateo, CA, USA) at the following time-points: baseline, 15 minutes after premedication, and at 2, 6, 12 and 24 hours after extubation (Figure 14). The camera was placed between the cage bars at the level of the cats' eyes and set to record at 60 frames per second and a medium angle of view. Following the three-minute recording and FGS-RT scoring, the observer proceeded with the interaction and palpation of the wound/painful area to complete the SF-UBCPS.

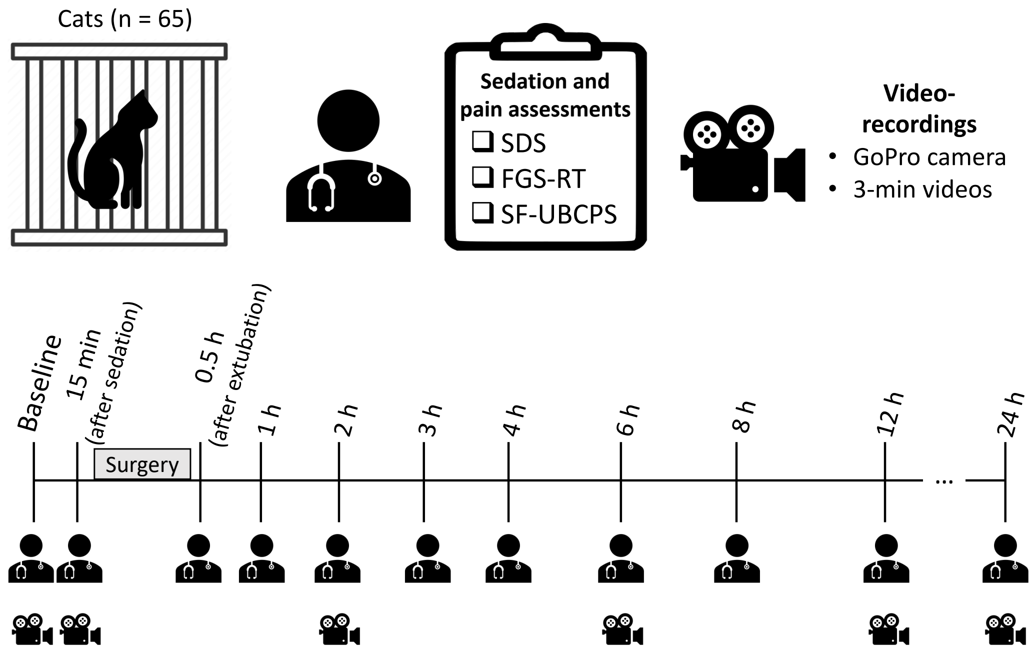


Figure 14. Timeline of the study and the time-points for sedation and pain assessments in real-time and video-recordings

Sedation was evaluated with a 5-point simple descriptive scale (SDS). Pain was evaluated using the Feline Grimace Scale (FGS) and the short-form UNESP-Botucatu composite pain scale (SF-UBCPS). FGS scores were obtained in real time (FGS-RT) during three minutes of observation of the cats undisturbed. Three-minute videos were recorded simultaneously to the FGS-RT assessment using a GoPro camera placed between the cage bars.

Rescue analgesia was provided with buprenorphine 0.02 mg/kg intravenously and meloxicam 0.2 mg/kg subcutaneously when SF-UBCPS ≥ 4 (the analgesic intervention threshold pre-determined for this scale). Additional rescue analgesia was administered if needed, with the same dose of buprenorphine. If a cat required the administration of rescue analgesia, additional videos (along with the pain evaluation using the FGS-RT and SF-UBCPS) were recorded before and one hour after its administration. If a cat did not require the administration of rescue analgesia before the 12h time-point, meloxicam was administered (same dose and route of administration as previously described for rescue analgesia). The choice of administering meloxicam at 12h for the cats that did not require rescue analgesia beforehand is explained by the time gap until the 24h evaluation and we wanted to make sure cats were not painful overnight.

2.2.4.5. Video handling, image selection and image scoring

The videos were downloaded into a computer and renamed after a random sequence of numbers (www.random.org). Videos were not recorded postoperatively if the cats had been previously spayed. In these cases, only videos recorded at baseline and 15 minutes after sedation were considered. Any identification of the cats' names or time-points were deleted for blinding purposes. Screenshots were obtained from video-recordings whenever the cat was facing the camera, but not when they were sleeping, grooming or vocalizing. The resulting images were identified as belonging to the first, second or third minute of the video. From each three-minute video, the single best image per minute was selected for later scoring. Images were not selected if quality was poor (i.e. blurred or dark images or when the cats' faces were partially/completely not visible).

The same observer (MCE) evaluated pain in real-time using the FGS followed by the SF-UBCPS, and scored the images using the FGS (FGS-IMG) in a blinded and randomized order, six months after the experimental study. The average score of the three images (each from the first, second and third minutes) was considered the final FGS-IMG score for that time-point.

2.2.4.6. Statistical analysis

Statistical analyses were performed with SAS v.9.3 (SAS Institute, Cary, NC, USA) and GraphPad Prism 7 (GraphPad software, San Diego, CA, USA).

The agreement between FGS-RT and FGS-IMG scores was assessed per minute of observation and for the average of three minutes using the Bland & Altman's method for repeated measures (Bland & Altman, 2007). This method describes the agreement between two quantitative measurements by evaluating a bias between the mean differences, and estimating an agreement interval, within which 95% of the differences fall.

Sedation scores (ordinal variable) over time were compared with baseline using a non-parametric approach (Friedman test and Dunn's multiple comparisons test). Sedation scores are presented as median (range).

Linear mixed models with individuals as random effect and time as fixed effect were used to evaluate the time course of the FGS-RT, and the influence of sedation (baseline versus 15 min after premedication) and OVH (baseline versus 24h after extubation) on both total FGS-RT and FGS-IMG scores (considered continuous variables). Scores obtained after the administration of rescue analgesia were not included in this analysis. FGS-IMG and FGS-RT scores are presented as mean ± standard deviation (SD). Only the FGS-RT scores were used for the time course evaluation because scores from all cats and all time-points were available. FGS-IMG scores were obtained at fewer time-points and they were sometimes missing due to the lack of acceptable images from some videos.

The Cochran-Mantel-Haenszel test for ordinal scores and repeated measures was used to assess the influence of sedation and OVH on each individual action unit (scores 0, 1 or 2). This approach takes the order into account, thus representing a more specific test and offering greater statistical power.

Responsiveness to rescue analgesia (i.e. construct validity) was assessed by comparing the scores before and one hour after the administration of buprenorphine and meloxicam. In this case, after confirming the normality of data distribution using the Shapiro Wilk test, FGS-RT and FGS-IMG scores were compared using paired t-tests. The SF-UBCPS scores (ordinal variable and not normally distributed) were compared using the Wilcoxon test. SF-UBCPS scores are presented as median (range).

P values were adjusted according to the number of comparisons for each analysis. Bonferroni-corrected values of $p < 0.05$ were considered significant.

2.2.5. Results

Sixty-five mixed-breed cats met the inclusion criteria (age: 1.37 ± 0.9 years and body weight: 2.85 ± 0.76 kg; domestic short-haired: n = 52 and domestic long-haired cats: n = 13).

Sixteen cats were excluded because of upper respiratory tract disease (n = 4), conjunctivitis (n = 1), facial hemiparalysis (n = 1), anemia and hypoproteinemia (n = 1), body condition score > 7/9 (n = 1) and < 3/9 (n = 1) or high baseline SF-UBCPS scores (n = 7).

A total of 394 videos and 1373 still images were recorded and obtained, respectively. The median (range) number of selected images per cat was 6 (0-20). The maximum number of possible still images selected per cat would be 18 if this cat did not require rescue analgesia, considering that it could be possible to select one image from each minute for each video of 3 minutes recorded at the six planned time-points (baseline, 15 min, 2h, 6h, 12h and 24) (3 x 6 = 18). Following the same reasoning, if the cat required rescue (i.e. at 4h), two additional videos would be recorded (at 4h and 5h) and the maximum number of possible images would be 20 (18 + 2 = 20).

The total number of images used for scoring was 540. It was not possible to obtain any image from 160 videos. The selected images (n = 540) were scored in four sessions of three hours maximum (135 images per day, every other day) to avoid observer fatigue.

2.2.5.1. Agreement FGS-RT and FGS-IMG

Minimal bias and narrow limits of agreement (LOA) were observed between the FGS-RT and FGS-IMG scores. The agreement was calculated for each minute of observation (Figure 15A-15C) and for the average of three minutes (final score per time-point; Figure 15D). The FGS-RT scores overestimated FGS-IMG scores ($Bias_{1st\ min}$ = -0.078, $Bias_{2nd\ min}$ = -0.054, $Bias_{3rd\ min}$ = -0.050 and $Bias_{final\ score}$ = -0.057).

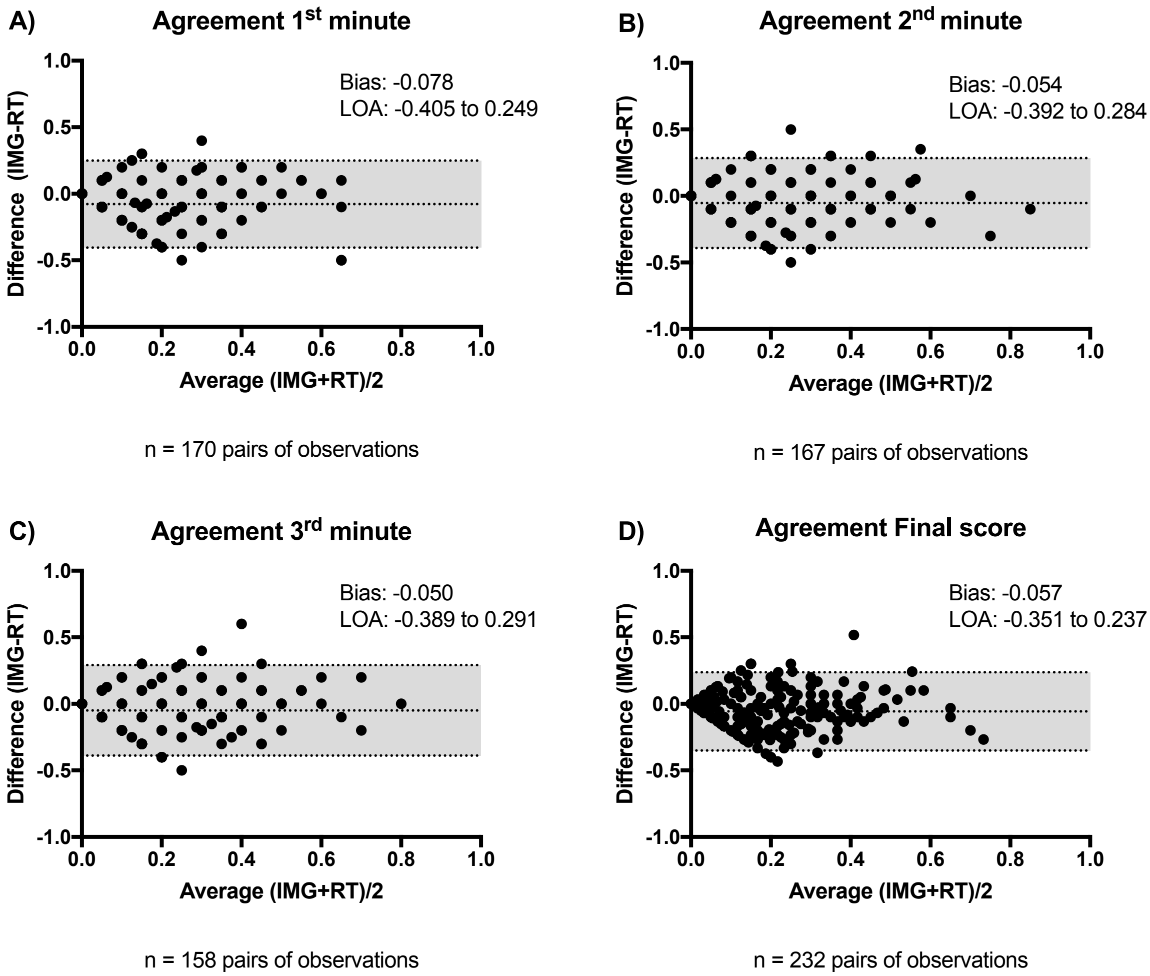


Figure 15. Agreement between Feline Grimace Scale scores obtained in real-time (FGS-RT) and by image assessment (FGS-IMG)

Bland & Altman's plots showing the bias and limits of agreement (LOA) between FGS-RT and FGS-IMG for the A) First minute of observation; B) Second minute of observation; C) Third minute of observation and D) Final score (average of the three minutes). Numbers of pairs of scores used for each analysis are indicated below the charts.

2.2.5.2. Sedation scores

Sedation scores were significantly higher than baseline at 15 min after sedation ($p = 0.012$) and 0.5h after extubation ($p < 0.001$) (Table 8).

Table 8. Sedation scores obtained pre- and postoperatively using a 5-point simple descriptive scale (adapted from Steagall et al., 2009)

Time-point	Sedation scores		Median of differences	95% CI of the differences	p value
	Median (range)	Interquartile range			
Baseline	0 (0-2)	[0-0]	-	-	-
15 min	1 (0-2)*	[0-2]	0	0 to 1	0.012
0.5h	1 (0-3)*	[1-1]	1	1 to 1	< 0.001
1h	1 (0-3)	[0-2]	0	0 to 1	0.053
2h	1 (0-3)	[0-2]	0	0 to 1	0.051
3h	0 (0-3)	[0-2]	0	0 to 0	0.161
4h	0 (0-3)	[0-2]	0	0 to 0	0.155
6h	0 (0-3)	[0-2]	0	0 to 0	> 0.999
8h	0 (0-3)	[0-2]	0	0 to 0	> 0.999
12h	0 (0-3)	[0-0]	0	0 to 0	> 0.999
24h	0 (0-2)	[0-0]	0	0 to 0	> 0.999

Scores over time were compared with baseline using Friedman test and Dunn's multiple comparisons test. Sedation scores are presented as median (range) and interquartile range. Median of the differences (in comparison with baseline) and their 95% confidence interval (CI) are presented. *Significantly different from baseline ($p < 0.05$).

2.2.5.3. Time course and influence of sedation and OVH on FGS scores

In comparison with baseline, mean ± SD FGS-RT scores were not different at 15 minutes after sedation, nor at 0.5h and 24h after extubation. The FGS-RT scores were significantly higher than baseline from 1 to 12h after extubation (Table 9 and Figure 16).

In comparison with baseline, mean ± SD FGS-IMG scores (0.14 ± 0.13) were not different at 15 minutes after sedation (0.16 ± 0.15, mean difference = 0.028, 95% CI of the differences: -0.0298 to 0.0864, p = 0.9911) nor at 24h after extubation (0.12 ± 0.12, mean difference = -0.033, 95% CI of the differences: -0.1 to 0.0339, p = 0.9569).

Table 9. Mean ± SD Feline Grimace Scale scores in real-time (FGS-RT) in cats undergoing ovariohysterectomy

Time-point	FGS-RT scores Mean ± SD	Mean difference	95% CI of the differences	*p* value
Baseline	0.16 ± 0.13	-	-	-
15 min	0.2 ± 0.15	0.04	0.0045 to 0.0755	0.3847
0.5h	0.21 ± 0.14	0.047	0.0028 to 0.0920	0.2624
1h	0.26 ± 0.17*	0.097	0.0462 to 0.1480	0.0003
2h	0.29 ± 0.2*	0.132	0.0824 to 0.1820	< 0.0001
3h	0.25 ± 0.18*	0.092	0.0434 to 0.1410	0.0003
4h	0.26 ± 0.16*	0.106	0.0605 to 0.1520	< 0.0001
6h	0.23 ± 0.18*	0.075	0.0277 to 0.1230	0.0095
8h	0.26 ± 0.14*	0.097	0.0454 to 0.1480	0.0001
12h	0.22 ± 0.14*	0.059	0.0096 to 0.1090	0.0316
24h	0.16 ± 0.12	<0.0001	-0.0434 to 0.0434	0.9998

Scores were obtained at baseline, 15 minutes after sedation with acepromazine-buprenorphine, and between 0.5 to 24 hours after extubation. Scores over time were compared with baseline using linear mixed models. The mean differences (in comparison with baseline) and their 95% confidence interval (CI) are presented. *Significantly different from baseline ($p < 0.05$).

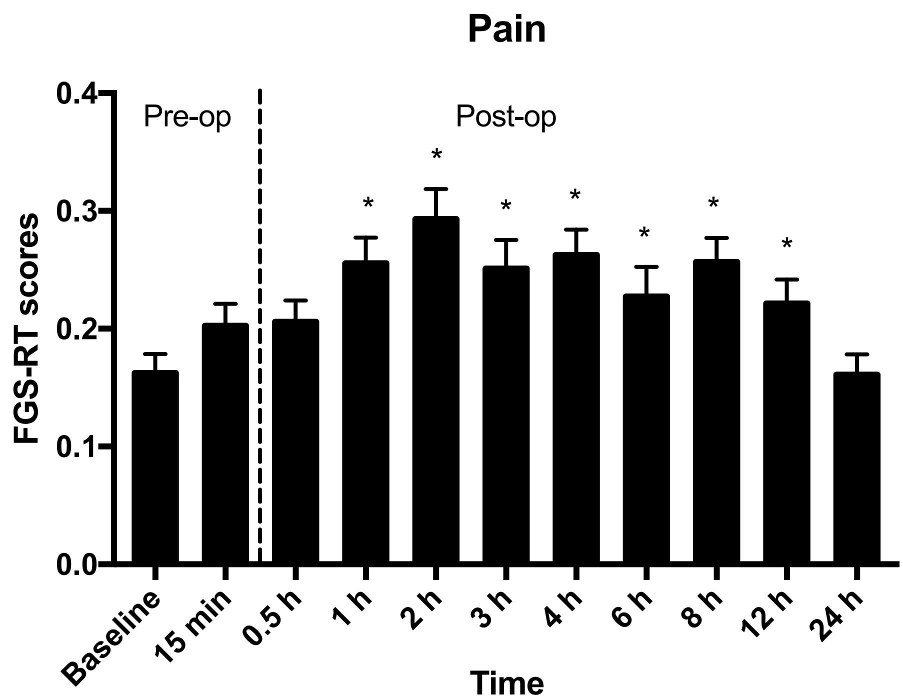


Figure 16. Mean ± SEM Feline Grimace Scale scores in real-time (FGS-RT) in cats undergoing ovariohysterectomy

Scores were obtained pre- and postoperatively (Pre-op and Post-op, respectively) at baseline, 15 minutes after sedation with acepromazine-buprenorphine, and between 0.5 to 24 hours after extubation. The average score of three minutes of observation of the cats' facial expression was considered the final FGS score per time-point. *Significantly different from baseline scores ($p < 0.05$).

2.2.5.4. Influence of sedation and OVH on individual action units

Sedation did not influence final FGS scores (both -RT and -IMG); however, orbital tightening scores obtained in real-time 15 minutes after sedation were higher than baseline ($p = 0.026$). This effect was not observed with any other action unit (ear position: $p = 0.72$; muzzle tension: $p = 0.17$; whiskers change: $p = 0.58$; head position: p = 0.22). Similarly, no significant differences in FGS-IMG scores were observed after sedation (ear position: $p = 0.1$; orbital tightening: $p = 0.32$; muzzle tension: $p = 0.14$; whiskers change: $p = 0.17$; head position: $p = 0.61$).

Ovariohysterectomy using a balanced anesthetic protocol did not influence any action unit scores obtained using either FGS-RT or FGS-IMG. All action units scores 24h after extubation were not significantly different from baseline values (FGS-RT - ear position: $p = 0.076$; orbital tightening: $p = 0.35$; muzzle tension: $p = 0.72$; whiskers change: $p = 0.62$; head position: $p = 0.35$ and FGS-IMG - ear position: $p = 0.16$; orbital tightening: $p = 0.22$; muzzle tension: $p = 0.12$; whiskers change: $p = 0.1$; head position: $p = 0.93$).

2.2.5.5. Rescue analgesia

Thirteen cats required rescue analgesia throughout the postoperative period. Mean ± SD scores before and after rescue analgesia were 0.47 ± 0.24 and 0.21 ± 0.18 (mean difference = -0.266, 95% CI of the differences: -0.389 to -0.143, $p = 0.0005$) respectively, using FGS-RT assessment; and 0.45 ± 0.19 and 0.18 ± 0.17 (mean difference = -0.274, 95% CI of the differences: -0.481 to -0.0681, $p = 0.0154$) respectively, using FGS-IMG assessment. Median (range) SF-UBCPS scores before and after analgesia were 7 (5-9) and 1 (0-3), respectively (median of differences: -3, 95% CI of the differences: -4 to -3, $p = 0.0005$).

2.2.6. Discussion

This study demonstrated the applicability of the Feline Grimace Scale in a clinical setting using real-time scoring. Minimal bias and narrow LOA were observed between scoring methods; the sedation protocol and surgery did not influence FGS scores and both FGS-RT and FGS-IMG detected responsiveness to analgesic administration.

The Bland & Altman for repeated measures method was used to verify the applicability of the FGS in real-time in a clinical setting. Real-time scoring using the FGS slightly overestimated image assessment ($Bias_{final\ score}$: -0.057), considered the standard scoring method. This could be explained by the clinical context of FGS-RT scoring and the three-minute observation of the cats' facial expression. Although the observer was focused on the action units to be scored, the body position and posture of the cat could be observed, and this might be another reason for the overestimation with the FGS-RT scores. Additionally, the FGS-IMG scores reflected the pain states using only one still image selected during that minute. It is possible that facial expressions might not be the exact same if an image was selected at the beginning or the end of each minute of observation, which may also explain the discrepancy observed with the FGS-RT when compared with FGS-

IMG. Real-time scoring using the RGS was also demonstrated feasible with good agreement between real-time and image scoring (Leung et al., 2016). In contrast to what has been observed in rats, real-time MGS scores were significantly lower than those obtained by image assessment in mice (Miller & Leach, 2015). In both rodent species, real-time scoring underestimated the scores obtained through image assessment. These differences might be related to the procedure for image capture. In rats, similar methodology was used; however, multiple intervals and punctual observations for RGS scoring methods were tested in real-time (Leung et al., 2016). In mice, photographs were taken to obtain the MGS scores (Miller & Leach, 2015). In the latter study, the authors speculated that the use of photographs may have resulted in an artificial elevation of scores by capturing specific behaviors or movements (such as blinking), and that some difficulty in real-time scoring could be expected by their constant (and fast) level of activity. For the FGS, screenshots were taken from the videos when the cat was facing the camera. It is possible that the screenshots were obtained when the cat was paying attention to the surroundings (eyes open and ears facing forward) thus decreasing the resulting FGS-IMG score. Additionally, during FGS-RT scoring, the observer was aware of the timing (pre- or postoperatively). Although the non-blind nature of real-time assessment is inevitable and could represent a source of bias, it represents the method by which pain is evaluated in the clinical setting. Moreover, the good agreement between IMG and RT scores implies that the fact of the observer being aware of the timing did not introduce substantial bias.

There is no consensus for the classification of the LOA for pain scores with the same degree of standardization that exists for physiological variables (e.g. blood pressure, cardiac output, blood gas measurements, etc.) when designing measurement comparison studies (Mantha et al., 2000). As in the case of blood pressure monitors, the American National Standards of the Association for the Advancement of Medical Instrumentation recommends a limit for the mean difference of the paired measurements (bias) of ± 5 mmHg or less between the test system and the comparison system (White et al., 1993). These criteria are not readily applicable for pain scores. Alternatively, the evaluation of the LOA in relation to the analgesic threshold has been proposed (Leung et al., 2016). The LOA observed in the present study, although narrow ($LOA_{final\ score}$: -0.351 to 0.237), spans the analgesic threshold of 0.39 out of 1.0 previously determined for the FGS (Evangelista et al., 2019). This would mean that FGS-RT scores that are close to the analgesic threshold (slightly lower or higher than 0.39) should be interpreted with caution since this threshold is a suggestion for the administration of rescue analgesia based on the probability of being painful above that score. However, this should not be a major clinical problem if the final decision for giving additional analgesics rely on clinical judgement while taking in consideration the FGS score, context and disease, and the potential reassessment of the cat after a brief period.

Considering the small average discrepancy between methods for bias (with consistent variability for most of the FGS score range) and narrow LOA, the methods (FGS-IMG and FGS-RT) would probably provide similar assessments in the clinical setting. It should be noted; however, that the authors did not define the LOA *a priori* to state that real time and image assessments are interchangeable.

Although the assessment can be difficult in patients under the effects of anesthetics and sedatives, pain must be evaluated in a valid and reliable manner to ensure adequate treatment. For example, it is known that ketamine-based protocols confound pain scores in cats using the UB-MCPS (Buisman et al., 2016). In this study, the premedication did not seem to affect FGS scores. Sedation scores increased after premedication with acepromazine-buprenorphine; however, it did not influence FGS scores using both methods of assessment (FGS-RT and FGS-IMG). Similarly, in mice and rats, the administration of buprenorphine alone had no impact on the MGS and RGS scores (Miller et al., 2015; Leung et al., 2016). Even if the real-time evaluation (sedation scores and FGS-RT) might have been biased by the fact that the observer was aware of premedication, this type of bias is not present when using FGS-IMG scores. It is very unlikely that an observer could memorize all cat faces given the number of subjects, timepoints and the long delay between real-time assessment and evaluation of images after blinding and randomization. The action unit "Orbital tightening" was influenced by sedation using the FGS-RT. This effect might be related to the effect of acepromazine causing enophthalmos leading to protrusion of the third eyelid (Hatch et al., 1984) and buprenorphine causing mydriasis (Steagall et al., 2014). This influence should be minimal in the clinical setting, since the final score was not affected by sedation and such changes were not detected using FGS-IMG.

The influence of OVH was determined by comparing the FGS scores 24 hours after extubation with baseline values, in view of the long duration of action of the drugs used herein. Under this condition, OVH did not affect FGS scores (both FGS-RT and FGS-IMG). At this time-point we also would not expect high pain scores considering the minimal degree of surgical manipulation and the use of multimodal analgesia. Mean FGS-RT scores were significantly higher than baseline from 1 to 12 hours after extubation; however, the mean scores remained below the analgesic threshold (≥ 0.39 out of 1.0) and the majority of the cats did not require rescue analgesia. Drugs such as acepromazine, buprenorphine, bupivacaine and meloxicam are long-acting drugs, thus their effects would still be present during the postoperative period. In cats, increased sedation scores have been reported for up to 4 hours after intramuscular administration of similar doses of acepromazine and buprenorphine (Hunt et al., 2013). Furthermore, the mean elimination half-life of bupivacaine was 4.8 ± 2.7 and 10.5 ± 10.3 hours after intraperitoneal administration of bupivacaine alone and in combination with dexmedetomidine, respectively, in cats (Benito et al., 2016, 2018). Significant differences in pain or sedation scores, and the prevalence of rescue analgesia were not found between cats receiving intraperitoneal bupivacaine alone or in combination with dexmedetomidine (Benito et al., 2019). For this reason, the FGS scores considered in the present study were analyzed together. Meloxicam was administered at the 12h time-point (except if a cat required rescue analgesia before it) and it has a serum half-life of approximately 24 hours (Lehr et al., 2009). Even though the effect of other drugs may have worn off, the results observed at 24 hours after extubation may have been influenced by the long duration of action of meloxicam.

Responsiveness to rescue analgesia, as part of construct validity testing, was previously assessed during the development and validation of the FGS using various analgesic protocols including different drugs, doses and routes of administration (Evangelista et al., 2019). In that study, FGS-IMG scores decreased after the administration of analgesics when compared with those at presentation (i.e. before interventional analgesia). In the present study, responsiveness was tested again since the study design included a standardized protocol for rescue analgesia and type of surgical stimulus (OVH). Additionally, responsiveness should also be assessed using FGS-RT. Both methods of pain assessment (FGS-RT and FGS-IMG) detected changes in pain scores corroborating our previous findings (Evangelista et al., 2019). Decreases in pain scores were also observed with the SF-UBCPS.

During real-time scoring the observer was present in front of the cage and was able to move around to look at the cat from different angles, while videos were recorded simultaneously from a single angle. The cats' faces were not always visible from the camera angle when screenshots were taken, thus explaining missing FGS-IMG scores in many cases. Indeed, it was not possible to obtain any image from 160 videos (approximately 40%) for the following reasons: movement of the cat either too close or far from the front of the cage) and out of view; when cats were facing the back of the cage or hiding behind the litter box, sleeping, grooming, or when the image was blurred. Perhaps the number of excluded images could be reduced by using two cameras, placed on either side of the cage, as reported in mice (Langford et al., 2010), rats (Sotocinal et al., 2011) and horses (Dalla Costa et al., 2014). The effect of the presence of the observer in front of the cage while the videos were recorded was not evaluated in this study; however, it is currently being investigated. A recent study in mice showed that the presence of a male observer in the room reduced MGS scores, whereas the same was not observed with female observers. These differences in MGS scores are likely a result of stress-induced analgesia (Sorge et al., 2014). Similarly, the presence of a female observer did not interfere with RGS scores in rats (Leung et al., 2016). In the present study, a female observer performed pain assessments and it remains unknown if the gender of the observer influences FGS scores in cats.

Some limitations of this study include the use of a single type of surgical painful stimulus and model of acute pain (OVH). Effects of sedation were assessed 15 minutes after premedication, before the surgery and any other painful stimulus and only one protocol for premedication was studied. Therefore, our findings are still limited to OVH involving premedication and mild sedation with acepromazine-buprenorphine. It is not clear how other sedative protocols producing moderate to profound sedation may affect the FGS scores. In addition, it was not possible to evaluate the effects of anesthesia and surgery separately on FGS scores since we did not include sham and negative control groups undergoing general anesthesia without the surgical procedure, or surgery without the administration of analgesics (the latter for ethical reasons). This is a clinical study involving surgery that requires the administration of anesthetics; therefore, it is an intrinsic limitation of the methodology. The effects of general anesthesia alone on pain scores were assessed in rats and horses. Higher RGS scores were observed in rats in the immediate time period, approximately 20 minutes after the discontinuation of short duration isoflurane anesthesia (Miller et al., 2016); however, Horse Grimace Scale scores were unchanged 8 hours after general anesthesia without the application of a nociceptive stimulus (Dalla Costa et al., 2014).

Real-time scoring using the FGS is feasible, although it slightly overestimates image assessment. The minimal bias and narrow limits of agreement between FGS-RT and FGS-IMG suggest minimal clinical impact. Sedation with acepromazine-buprenorphine and ovariohysterectomy using a multimodal analgesia and a balanced anesthetic protocol did not influence the FGS scores. Responsiveness to analgesic administration was detected both with FGS-RT and FGS-IMG.

2.2.7. Acknowledgements

This study was funded by an unrestricted grant provided by Zoetis. Funding was also provided by the Companion Animals Health Fund from the Faculty of Veterinary Medicine of the Université de Montréal, supported by Zoetis and a donation by Ms. Valeria Rosenbloom and M. Mike Rosenbloom. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

2.2.8. Additional Information

2.2.8.1. Author Contributions Statement

Marina C. Evangelista conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft. Javier Benito, Beatriz P. Monteiro, Ryota Watanabe and Graeme M. Doodnaught performed the experiments, authored or reviewed drafts of the paper, and approved the final draft. Daniel S.J. Pang analyzed the data, authored or reviewed drafts of the paper, and approved the final draft. Paulo V. Steagall conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.

2.2.8.2. Competing interests

Beatriz Monteiro and Paulo Steagall have provided consultancy services for Zoetis. This does not alter the authors' adherence to PeerJ policies on sharing data and materials.

2.3. Agreement and reliability of the Feline Grimace Scale among cat owners, veterinarians, veterinary students and nurses

Marina C. Evangelista[1] and Paulo V. Steagall[1]

2.3.1. Article identifier

This article was submitted to Scientific Reports in Aug 5th, 2020

Submission ID 859e666d-b2fb-464d-bf9b-e0aa06d6a87f (under revision)

[1] Département de sciences cliniques, Faculté de médecine vétérinaire, Université de Montréal, 3200 rue Sicotte, Saint-Hyacinthe, Québec, J2S2M2, Canada

2.3.2. Abstract

This study aimed to evaluate the agreement and reliability of the Feline Grimace Scale (FGS) among cat owners, veterinarians, veterinary students and nurses/technicians.

Raters (n = 5/group) scored 100 images using the FGS (ear position, orbital tightening, muzzle tension, whiskers position and head position). Intra-class correlation coefficients (ICC) were used to assess inter- and intra-rater reliability. Agreement between each group and the veterinarian group (gold-standard) was calculated using the Bland-Altman method. The effects of gender, age and number of cats owned on FGS scores were assessed using linear mixed models.

Inter-rater reliability was good for FGS final scores (ICC > 0.8). The muzzle and whiskers yielded lower reliability (ICC = 0.39 to 0.74). Intra-rater reliability was excellent for students and veterinarians (ICC = 0.91) and good for owners and nurses (ICC = 0.87 and 0.81, respectively). A very good agreement between all groups and veterinarians (bias < 0.1 and narrow limits of agreement) was observed. Female raters assigned higher FGS scores than males ($p = 0.006$); scores were not affected by age ($p = 0.93$) or number of cats owned ($p = 0.26$).

The FGS is reliable for feline acute pain assessment when used by individuals with different experience.

2.3.3. Introduction

The inherent subjectivity of pain has been broadly recognized, particularly with the updated definition of pain by the International Society for the Study of Pain (IASP): “An unpleasant sensory and emotional experience typically caused by, or resembling that caused by, actual or potential tissue injury” (Raja et al., 2020). Pain is always a subjective experience that is influenced by varying degrees of biological, psychological and social factors. Therefore, pain assessment represents a major challenge in paediatric patients and individuals with cognitive impairment who cannot self-report pain (Hadjistvropoulos et al., 1998; Prkachin, 2011), and in other non-human species. Indeed, veterinary health professionals must rely upon observations and pain scoring systems to assess pain in animals (Mathews et al., 2014; Merola & Mills, 2016a). In these cases, pain is estimated by observation of behaviours, posture, activity, along with an increasing role of facial expressions for pain assessment in the past decade (Steagall & Monteiro, 2019). Recently, a facial expression-based tool has been published for acute pain assessment in cats, the Feline Grimace Scale (FGS) (Evangelista et al., 2019). It comprises five action units (AU): ear position, orbital tightening, muzzle tension, whiskers position and head position. This instrument has reported validity and reliability in different painful conditions for use by veterinarians (Evangelista et al., 2019; Watanabe, Doodnaught, et al., 2020) using both image and real-time assessment (Evangelista, Benito, et al., 2020).

The incorporation of pain scales into feline practice allows the assessment of pain in a more rational and systematic means (Steagall & Monteiro, 2019), however the application of these instruments in the clinical setting is low, with under 10% of veterinary practices reporting the use of a standardized or formal pain scoring system (Coleman & Slingsby, 2007; Dawson et al., 2017). Additionally, discrepancies exist between pet owners, veterinary students, veterinary nurses/technicians and veterinarians regarding their attitudes, perceptions and ability to recognize pain in animals (Coleman & Slingsby, 2007; Dohoo & Dohoo, 1998; Doodnaught et al., 2017; Simon et al., 2018; Steagall et al., 2017; Väisänen et al., 2008). Pet owners frequently disagree with statements that pain assessment in animals is easy (Steagall et al., 2017; Väisänen et al., 2008) and, in general veterinary nurses assigned higher pain scores than veterinarians (Coleman & Slingsby, 2007; Dohoo & Dohoo, 1998). Moreover, studies suggested that pain assessment in cats and dogs may be affected by gender and previous experience of the observer (Beswick et al., 2016; Dohoo & Dohoo, 1996b; Doodnaught et al., 2017; Hewson et al., 2006a; Simon et al., 2017), and that veterinary students' knowledge of animal pain increases at later stages of their studies (Mich et al., 2010; Valros & Hänninen, 2018).

The inherent subjectivity of pain can be an issue in feline medicine and, for example, how the different expertise of individuals could affect FGS scores in cats. Thus, the objectives of this study were to assess the agreement and reliability of the FGS by groups of individuals with different expertise on pain assessment: cat owners, veterinary undergraduate students and veterinary nurses (animal health technicians). A secondary objective was to assess the effects of demographics (age, gender, number of cats owned, etc.) on the FGS scores. Our hypotheses were that there would be good reliability and agreement among FGS scores of different groups, and that our findings would agree with the literature, in the sense that women would assign higher pain scores than men.

2.3.4. Methods

2.3.4.1. Ethical statement

The study protocol was reviewed and approved by the "Comité d'éthique de la recherche en sciences et en santé (CERSES)" of the Université de Montréal (#CERSES-20-004-D).

2.3.4.2. Image selection

Face images of cats presenting with different levels of pain from three previous studies [Study A (Evangelista et al., 2019); Study B (Watanabe, Doodnaught, et al., 2020) and Study C (Evangelista, Benito, et al., 2020)] were included. These images were obtained from cats presenting pain associated with medical conditions (i.e. pancreatitis, cystitis, urethral obstruction, etc.) or surgery (i.e. multiple dental extractions and ovariohysterectomy). Briefly, the cats were filmed undisturbed in their cages before and after the painful stimulus/surgery and/or before and after analgesic treatment, if needed. Images (screenshots) were obtained from videos when the cats were facing the camera, but not sleeping, grooming, eating, playing or vocalizing (Evangelista et al., 2019).

Images were retrieved from our database, screened and pre-selected on two rounds by an independent investigator (MCE) who was not involved with image scoring. On the first round, images from each study were screened separately (Studies A, B and C). Poor quality images, those of cats showing signs of sedation (i.e. images obtained after premedication) and repeated images from the same cat at the same time point were excluded. On the second round of pre-selection, images from the three studies were pooled (A+B+C) and low-medium quality images, images where the head was not well aligned with the camera and similar images of the same cat (in slightly different position) at different time points were excluded to reach the goal of 100 images for evaluation (Figure 17). Approximately 40% of the images showcase cats presenting some degree of pain (mild/moderate/severe) and about 60%, no pain. This classification was based on pain scoring of cats in real-time during the original studies (Benito et al., 2019; Evangelista et al., 2019; Watanabe et al., 2019).

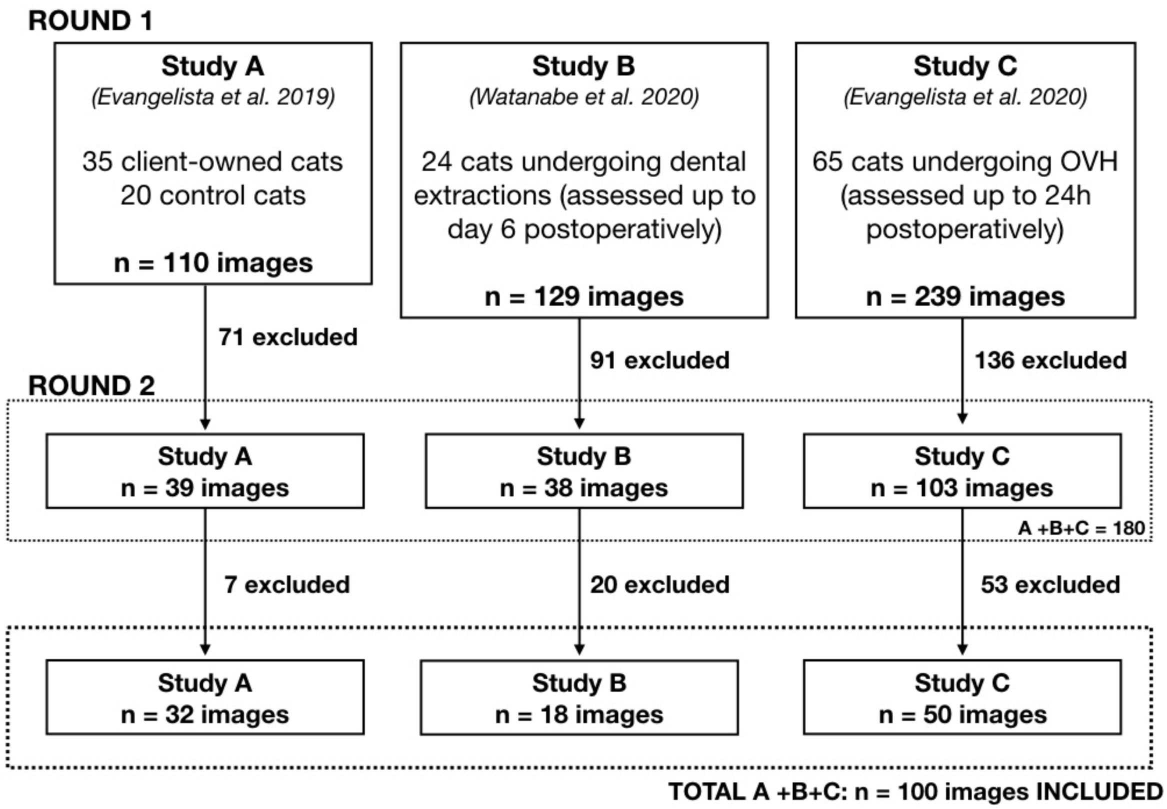


Figure 17. Flowchart of the screening and selection of images from three previous studies involving the Feline Grimace Scale

Images (screenshots) were obtained from video-recordings of cats undisturbed in their cages before and after the painful stimulus/surgery and/or before and after analgesic treatment. OVH: Ovariohysterectomy

2.3.4.3. Participant selection

Eligible participants had to: 1) be 21 years of age or older; 2) presently have or have had one (or more) cat(s) in the past (applicable to owners); 3) have access to a computer connected to the internet and the online questionnaire; 4) be committed to evaluating 110 images of cats in two sessions (one week apart).

The recruitment of participants took place from March 30th to April 17th, 2020. Emails were sent to six cat owners (investigators' personal contacts or acquaintances, who were not involved with any activity of our laboratory), approximately 400 veterinary students (mailing list of the undergraduate student association) and 90 veterinary nurses (animal health technicians) working at the Centre Hospitalier Universitaire Vétérinaire (CHUV) of the Université de Montréal (mailing list of the CHUV employees). The initial plan was to advertise and recruit cat owners randomly at the veterinary teaching hospital; however, this study happened amidst the COVID-19 pandemic, which impaired the recruitment in person and on-site.

The first five eligible individuals who responded to the recruiting emails were selected. Five veterinarians with previous experience in animal pain studies and/or with the FGS in our laboratory were selected to act as the gold-standard group of raters.

2.3.4.4. Image scoring - questionnaire

The selected images (n = 100) were numbered and randomised (using a random sequence generator available at: www.randomization.com) and uploaded into a two-part online questionnaire (SurveyMonkey). Ten images (selected using a random number generator) were repeated across the two parts of the questionnaire to assess intra-rater reliability. Each part contained 55 images to be completed one-week apart.

The participants were supplied with a training manual on how to use the Feline Grimace Scale (www.bit.ly/FGSmanual). They were asked to read it and contact the researchers in case of questions. A private link to the online questionnaire was sent by email to the selected participants. The first part was completed the week of April 20th and the second part, the week of April 27th, 2020. The first part of the questionnaire contained instructions on how to score the images, a consent form (in English or French, according to the raters' language preference), questions about the gender, age, number of cats at home, and years in veterinary school (for students) or years of experience working with cats (for veterinary nurses and veterinarians), and the images to be scored. The second part contained instructions on how to score the images and the remaining images to be scored. All participants gave written informed consent. Individuals were offered a reward for participation (20$ coffee shop gift card). Answers were anonymized during data analyses.

Participants were presented with one cat image at a time in order to score the five action units (AU) that comprise the FGS (ear position, orbital tightening, muzzle tension, whiskers position and head position in relation to the shoulders). Each AU was scored from 0 to 2, as following: 0 - the AU is absent, 1 - moderate appearance of the AU, 2 - obvious appearance of the AU, or N/A - not possible to score. The final FGS score was calculated as the sum of the scores assigned to each action unit divided by the total possible score, excluding those marked as not possible to score (i.e. 3/10 = 0.3 or 4/8 = 0.5) (Evangelista et al., 2019).

2.3.4.5. Statistical analysis

Intra-class correlation coefficients (ICC) were used to assess inter- and intra-rater reliability within groups. Inter-rater reliability was assessed for each of the AU and for the final FGS score using a two-way random effects ICC for absolute agreement. Intra-rater reliability was assessed using a two-way mixed effects ICC for absolute agreement. Estimates for single measures and average of measures, accompanied by their 95% confidence intervals (95% CI) are presented. Interpretation was based on the ICC single as following: ICC < 0.5 = poor, 0.5–0.75 = moderate, 0.75–0.9 = good, and > 0.90 = excellent reliability (Koo & Li, 2016).

Agreement between owners, students, nurses and the veterinarian group (considered as the gold standard) was calculated using the Bland and Altman method (Bland & Altman, 1999). Based on our previous publication (Evangelista, Benito, et al., 2020), the authors considered that a bias lower than 0.1 was considered acceptable, indicating very good agreement. A bias larger than 0.1 (more than 1 unit in the FGS score) was considered unacceptable, indicating poor agreement. The limits of agreement (LoA) were interpreted in relation to the analgesic threshold pre-determined for the FGS (Evangelista et al., 2019). The LoA should not span the analgesic threshold of 0.39 out of 1.0.

The effects of gender, age and number of cats on FGS scores were assessed using linear mixed models with the rater as a random effect and the gender, age and number of cats as fixed effects. There were few raters to assess the effect of years of work experience or years in school and these effects were not evaluated. Data for this analysis were transformed using the arcsine square root transformation to normalize the distributions. Values of $p < 0.05$ were considered significant.

2.3.5. Results

Seven cat owners (five out of the six invited owners responded to the recruitment emails and two others volunteered to participate), eight students, seven nurses and all five veterinarians responded to the recruitment within the acceptable deadline of three weeks. A total of 20 participants were included in the study (n = 5 in each group). Demographic information is presented in Table 10.

The mean ± SD time spent to complete each part of the questionnaire was 50.8 ± 21 minutes and 37.5 ± 28 minutes for parts 1 and 2, respectively.

Table 10. Demographic information of the included participants in a study involving the Feline Grimace Scale via image assessment

Group	Gender	Age range [years]	Number of cats owned [median (range)]	Years in school [median (range)]	Years of work experience
Cat owners	Female n = 3 Male n = 2	[21-29] n = 1 [40-49] n = 1 [50-59] n = 1 [60-69] n = 2	1 (1-2)	-	-
Veterinary Students	Female n = 5	[21-29] n = 5	2 (1-8)	4 (1-5)	-
Veterinary nurses	Female n = 5	[30-39] n = 3 [40-49] n = 2	3 (0-6)	-	[6-10] n = 2 [11 or more] n = 3
Veterinarians	Female n = 3 Male n = 2	[21-29] n = 1 [30-39] n = 3 [40-49] n = 1	2 (0-3)	-	[1-5] n = 2 [6-10] n = 1 [11 or more] n = 2

2.3.5.1. Reliability

Inter-rater reliability (final FGS score) was good for all groups [owners - ICC single = 0.80 (95% CI: 0.74 to 0.85); students - ICC single = 0.88 (95% CI: 0.85 to 0.91); nurses - ICC single = 0.83 (95% CI: 0.79 to 0.88); veterinarians - ICC single: 0.86 (95% CI: 0.81 to 0.90)]. The inter-rater reliability was good for the AU ears and eyes for all groups; poor to moderate for the AU muzzle (poor among owners and nurses; moderate among students and veterinarians) and whiskers (poor among owners; moderate among students, nurses and veterinarians), and moderate to good for the AU head position (moderate among owners and students; good among nurses and veterinarians) (Figure 18 and Table 11).

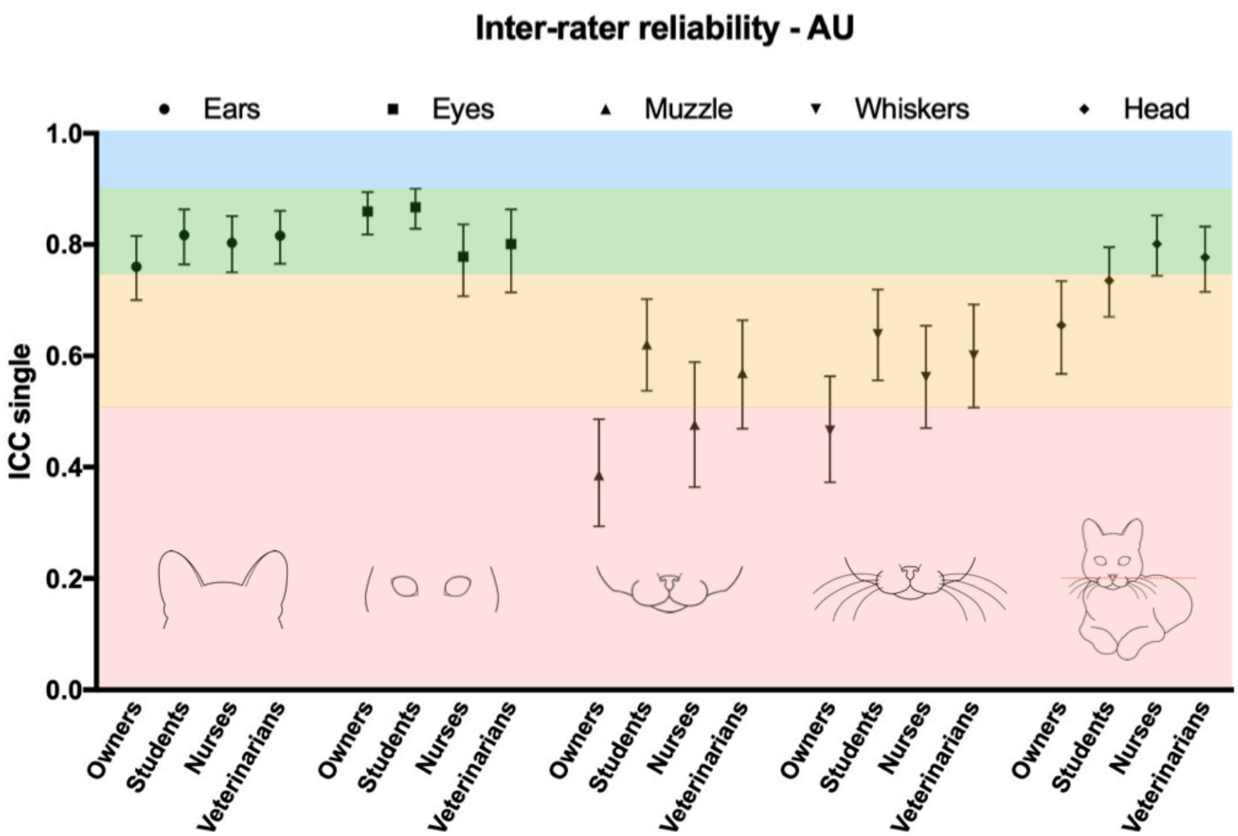


Figure 18. Inter-rater reliability of the action units (AU) composing the FGS by raters with different degrees of expertise in feline pain assessment

Inter-rater reliability was assessed using a two-way random effects intra-class correlation coefficient (ICC) for absolute agreement (raters: n = 5/group). Estimates for ICC single measures, accompanied by their 95% confidence intervals are presented. Interpretation was as following: ICC < 0.5 = poor (pink), 0.5-0.75 = moderate (yellow), 0.75-0.9 = good (green), and >0.90 = excellent reliability (blue)(Koo & Li, 2016).

Table 11. Inter-rater reliability of the FGS using 100 images assessed by raters with different degrees of expertise in feline pain assessment

	Group	ICC single (95% CI)	ICC average (95% CI)
FGS Final score	Owners	0.80 (0.74 to 0.85)	0.95 (0.93 to 0.97)
	Students	0.88 (0.85 to 0.91)	0.97 (0.97 to 0.98)
	Nurses	0.83 (0.79 to 0.88)	0.96 (0.95 to 0.97)
	Veterinarians	0.86 (0.81 to 0.90)	0.97 (0.95 to 0.98)
AU - Ears	Owners	0.76 (0.70 to 0.82)	0.94 (0.92 to 0.96)
	Students	0.82 (0.76 to 0.86)	0.96 (0.94 to 0.97)
	Nurses	0.80 (0.75 to 0.85)	0.95 (0.94 to 0.97)
	Veterinarians	0.82 (0.77 to 0.86)	0.96 (0.94 to 0.97)
AU - Eyes	Owners	0.86 (0.82 to 0.89)	0.97 (0.96 to 0.98)
	Students	0.87 (0.83 to 0.90)	0.97 (0.96 to 0.98)
	Nurses	0.78 (0.71 to 0.84)	0.95 (0.92 to 0.96)
	Veterinarians	0.80 (0.71 to 0.86)	0.95 (0.93 to 0.97)
AU - Muzzle	Owners	0.39 (0.29 to 0.49)	0.76 (0.68 to 0.83)
	Students	0.62 (0.54 to 0.70)	0.89 (0.85 to 0.92)
	Nurses	0.48 (0.36 to 0.59)	0.82 (0.74 to 0.88)
	Veterinarians	0.57 (0.47 to 0.66)	0.87 (0.82 to 0.91)
AU - Whiskers	Owners	0.47 (0.37 to 0.56)	0.81 (0.75 to 0.87)
	Students	0.64 (0.56 to 0.72)	0.90 (0.86 to 0.93)
	Nurses	0.56 (0.47 to 0.65)	0.87 (0.82 to 0.90)
	Veterinarians	0.60 (0.51 to 0.69)	0.88 (0.84 to 0.92)
AU - Head position	Owners	0.66 (0.57 to 0.73)	0.91 (0.87 to 0.93)
	Students	0.74 (0.67 to 0.80)	0.93 (0.91 to 0.95)
	Nurses	0.80 (0.74 to 0.85)	0.95 (0.94 to 0.97)
	Veterinarians	0.78 (0.72 to 0.83)	0.95 (0.93 to 0.96)

Intraclass correlation coefficient (ICC) estimates and their 95% confidence intervals (95% CI) were calculated based on single measures (ICC single) and average (ICC average) of measures (raters: n = 5/group), using two-way random effects model for absolute agreement.

Intra-rater reliability (final FGS score) was good for owners and nurses [ICC single = 0.87 (95% CI: 0.75 to 0.93) and 0.81 (95% CI: 0.69 to 0.89), respectively] and excellent for students and veterinarians [ICC single = 0.91 (95% CI: 0.84 to 0.95) and 0.91 (95% CI: 0.85 to 0.95), respectively]. The intra-rater reliability was good for the AU ears for all groups; for the AU eyes, it was good among students and nurses and excellent among owners and veterinarians. Intra-rater reliability was moderate for the AU muzzle and whiskers for all groups and moderate for AU head position for all groups, except owners, which presented good reliability (Figure 19 and Table 12).

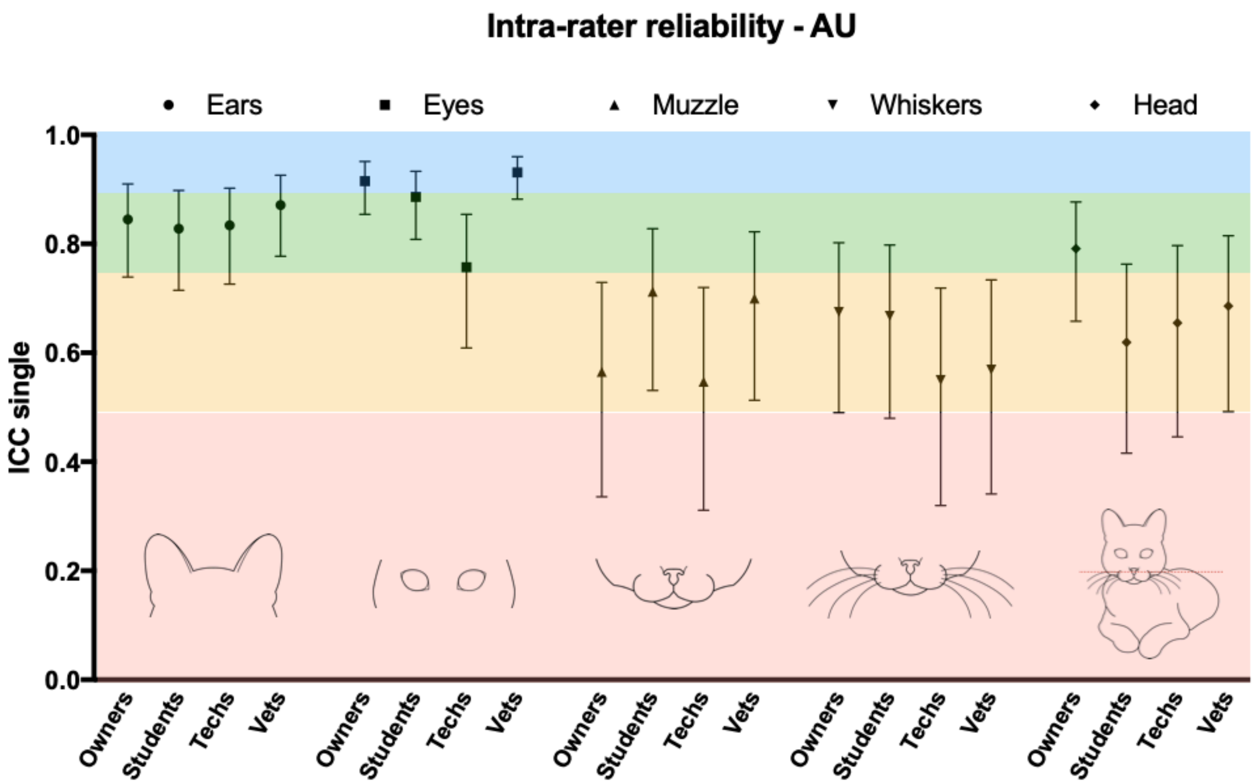


Figure 19. Intra-rater reliability of the action units (AU) composing the FGS by raters with different degrees of expertise in feline pain assessment

Intra-rater reliability was assessed using a two-way mixed effects intra-class correlation coefficient (ICC) for absolute agreement (raters: n = 5/group). Estimates for ICC single measures, accompanied by their 95% confidence intervals are presented. Interpretation was as following: ICC < 0.5 = poor (pink), 0.5–0.75 = moderate (yellow), 0.75–0.9 = good (green), and >0.90 = excellent reliability (blue)(Koo & Li, 2016).

Table 12. Intra-rater reliability of the FGS using 10 images (repeated across two scoring sessions, one week apart) assessed by raters with different degrees of expertise in feline pain assessment

	Group	ICC single (95% CI)	ICC average (95% CI)
FGS Final score	Owners	0.87 (0.75 to 0.93)	0.93 (0.86 to 0.97)
	Students	0.91 (0.84 to 0.95)	0.95 (0.91 to 0.97)
	Nurses	0.81 (0.69 to 0.89)	0.90 (0.82 to 0.94)
	Veterinarians	0.91 (0.85 to 0.95)	0.95 (0.92 to 0.97)
AU - Ears	Owners	0.85 (0.74 to 0.91)	0.92 (0.85 to 0.95)
	Students	0.83 (0.72 to 0.90)	0.91 (0.83 to 0.95)
	Nurses	0.83 (0.73 to 0.90)	0.91 (0.84 to 0.95)
	Veterinarians	0.87 (0.78 to 0.93)	0.93 (0.88 to 0.96)
AU - Eyes	Owners	0.92 (0.85 to 0.95)	0.96 (0.92 to 0.98)
	Students	0.89 (0.81 to 0.93)	0.94 (0.89 to 0.97)
	Nurses	0.76 (0.61 to 0.85)	0.86 (0.76 to 0.92)
	Veterinarians	0.93 (0.88 to 0.96)	0.96 (0.94 to 0.98)
AU - Muzzle	Owners	0.57 (0.34 to 0.73)	0.72 (0.50 to 0.84)
	Students	0.71 (0.53 to 0.83)	0.83 (0.69 to 0.91)
	Nurses	0.55 (0.31 to 0.72)	0.71 (0.48 to 0.84)
	Veterinarians	0.70 (0.51 to 0.82)	0.82 (0.68 to 0.90)
AU - Whiskers	Owners	0.68 (0.49 to 0.80)	0.81 (0.66 to 0.89)
	Students	0.67 (0.48 to 0.80)	0.80 (0.65 to 0.89)
	Nurses	0.55 (0.32 to 0.72)	0.71 (0.49 to 0.84)
	Veterinarians	0.57 (0.34 to 0.73)	0.73 (0.51 to 0.85)
AU - Head position	Owners	0.79 (0.66 to 0.88)	0.88 (0.79 to 0.93)
	Students	0.62 (0.42 to 0.76)	0.76 (0.59 to 0.87)
	Nurses	0.66 (0.45 to 0.80)	0.79 (0.62 to 0.89)
	Veterinarians	0.69 (0.49 to 0.82)	0.81 (0.66 to 0.90)

Intraclass correlation coefficient (ICC) estimates and their 95% confidence intervals (95% CI) were calculated based on single measures (ICC single) and average (ICC average) of measures (raters: n = 5/group), using two-way mixed effects model for absolute agreement.

2.3.5.2. Agreement

The agreement between groups was very good with minimal bias (-0.038 to -0.060) and narrow limits of agreement that did not span the analgesic threshold of the FGS (0.39). Owners, students and nurses tend to slightly overestimate the veterinarians' scores (bias owners X veterinarians = -0.041; bias students X veterinarians = -0.038 and bias nurses X veterinarians = -0.060). The limits of agreement were narrow in all situations (Figure 20).

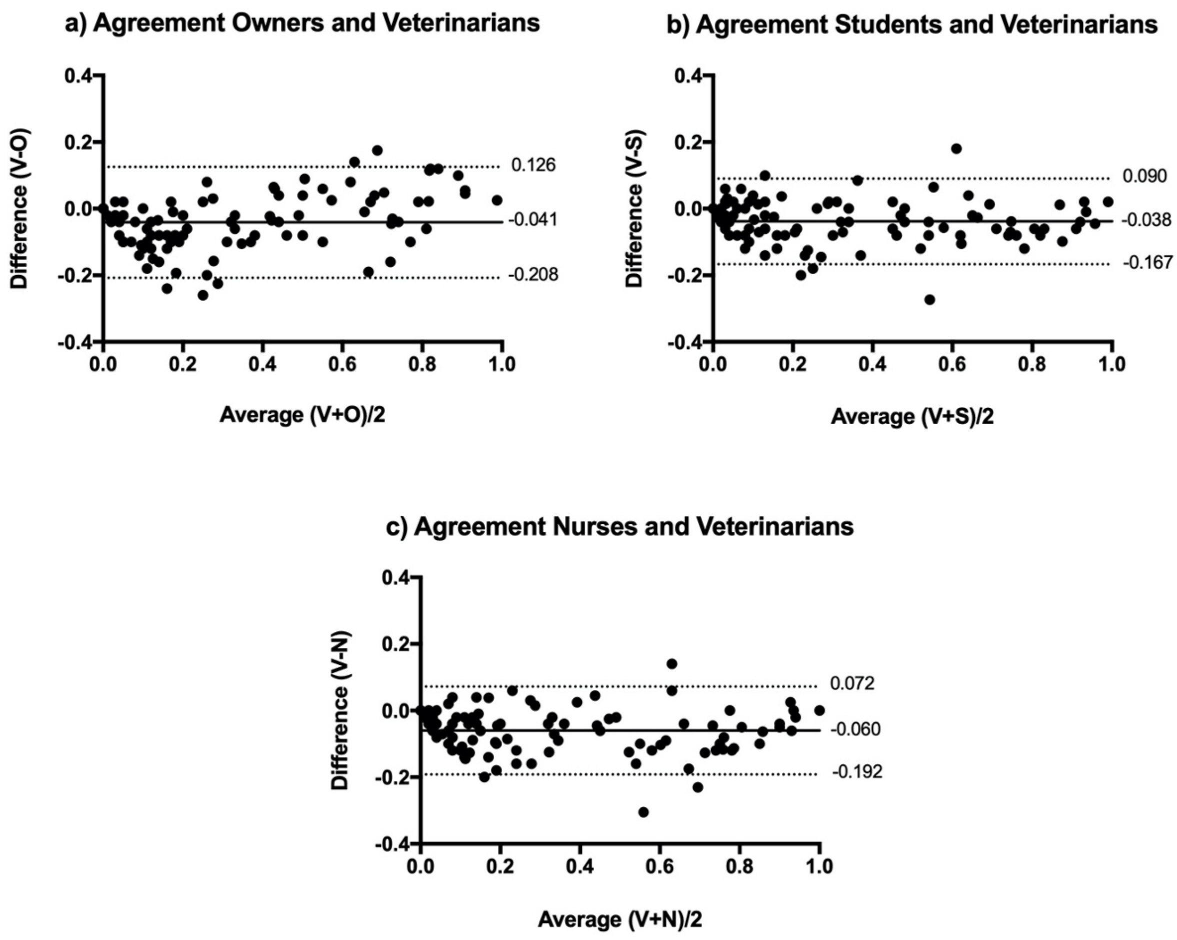


Figure 20. Bland and Altman plots showing the agreement of the FGS scores between each group and the veterinarian group (the gold standard)

a) Agreement between cat owners (O) and veterinarians (V). b) Agreement between veterinary students (S) and veterinarians (V). c) Agreement between veterinary nurses (N) and veterinarians (V). The bias (central continuous line) and limits of agreement (dotted lines) are indicated on each plot (raters: n = 5/group).

2.3.5.3. Effect of gender, age and number of cats owned on FGS scores

The model indicated that mean FGS scores were significantly larger in female than in male observers (transformed means ± standard error: 0.59 ± 0.014 and 0.491 ± 0.023, respectively; $p = 0.006$) but did not vary with age ($p = 0.93$). The mean FGS scores increased with the number of cats owned but this effect was not significant ($p = 0.26$).

2.3.6. Discussion

This study assessed the reliability and agreement of the FGS scores among four groups with different levels of expertise in feline pain assessment. It also demonstrated the effect of gender, age and the number of cats owned by raters on these scores. The overall inter-rater reliability was good for all groups (final FGS score), demonstrating that the FGS can be used reliably even by untrained individuals for pain assessment in cats. Previous studies revealed that owners frequently disagreed with statements that they had received sufficient information on appropriate methods of animal pain management (Väisänen et al., 2008). The FGS could change this paradigm and provide reliable means of feline pain assessment by cat owners. This could represent a significant advance for feline medicine. Pain in cats has been historically neglected and these animals receive less analgesics in comparison with dogs (Dohoo & Dohoo, 1996; Reimann et al., 2017; Simon et al., 2017; Williams et al., 2005). If cat owners can recognize pain in their pets, or that "something is wrong", a significant increase in veterinary consultations could result in better feline health and welfare, and potential early disease diagnosis.

This study demonstrated good reliability for the AU ears and eyes for all groups (ICC single > 0.75) and moderate to good for the AU head position [ICC single = 0.66 and 0.74 - moderate; owners and students, respectively and ICC single = 0.8 and 0.78 - good; nurses and veterinarians, respectively). The reliability was lower for the AUs muzzle (ICC single = 0.39, 0.62, 0.48 and 0.57 for owners, students, nurses and veterinarians, respectively) and whiskers (ICC single = 0.47, 0.74, 0.56 and 0.6 for owners, students, nurses and veterinarians, respectively), especially within the owners' group (ICC < 0.5 - poor), indicating more difficulty in scoring whiskers and muzzle using images. Likewise, moderate reliability was reported for the AU muzzle and whiskers by trained veterinarians using the FGS for assessing pain in cats with naturally occurring conditions and following dental extractions (ICC single = 0.63 and 0.56 for muzzle and ICC single = 0.55 and 0.64 for whiskers, respectively) (Evangelista et al., 2019; Watanabe, Doodnaught, et al., 2020). Overall, this shows that static and frontal image assessment of muzzle and whiskers can be challenging independently of training experience. Indeed, image assessment lacks three dimensional views that would help identifying the correct position of whiskers and muzzle. The latter two AUs are more subtle to be observed than the other three AUs, and may be impaired by the fur colour, background and the position of the cat in relation to the camera. Even with these potential limitations of muzzle and whiskers scoring, the clinical applicability of the FGS was previously demonstrated when comparing real-time versus image assessment (Evangelista, Benito, et al., 2020). Clinical experience in our laboratory also shows that scoring the muzzle and whiskers in real-time is commonly less challenging than image assessment. Finally, the poor to moderate reliability of whiskers and muzzle using the FGS across groups with different expertise may seem to be worse than similar AUs from other grimace scales [i.e. Ferret Grimace Scale - cheek bulging ICC = 0.86 and whiskers retraction ICC = 0.88 (Reijgwart et al., 2017); Rat Grimace Scale - nose/cheek flattening ICC = 0.86 (Sotocinal et al., 2011)]. However, one should bear in mind that some studies have reported ICC average without reporting ICC single, which normally produces higher estimates (Shrout & Fleiss, 1979). This could give the impression that muzzle and whiskers are highly reliable with these other grimace scales with superior reliability than the FGS. However, it may not be the case if the same approach for the reporting and interpretation of ICC were used. We have chosen the guidelines for interpretation of ICC proposed by (Koo & Li, 2016), which represents a more rigorous classification of the ICC. An alternative classification proposes that ICCs < 0.40 indicate poor reliability; between 0.40 and 0.59 = fair; between 0.60 and 0.74 = good; and > 0.75 = excellent (Cicchetti, 1994). Considering less rigorous classifications of ICC, our interpretation for the inter-rater reliability of the final FGS score would have been excellent for all groups, and as follows for each AU: ears and eyes - excellent for all groups; muzzle - poor for owners, fair for nurses and veterinarians, and good for students; whiskers - fair for all groups, except for students (good); and head position - good for owners and students, and excellent for nurses and veterinarians.

Our laboratory is currently studying the multiple association of AUs with each other and item-total correlation with the final score in order to understand the weigh-value of each AU in painful conditions. For example, if the Aus ears, eyes and head position are more reliable and "valuable" than muzzle and whiskers, this could means that evaluating the three Aus in the first place could already provide an overall idea of the painful status of a cat. The evaluation/scoring of the AUs muzzle and whiskers will perhaps become less important especially if the threshold for analgesic administration has been reached using the three other AUs (ears, eyes and head position).

The intra-rater reliability was calculated by comparing the scores of 10 images repeated across the two sessions one-week apart. The intra-rater reliability was good (owners and nurses) to excellent (students and veterinarians) with ICC values higher than 0.8, indicating that the final FGS scores across the two sessions were consistent. However, it is unknown if similar results would be obtained if a larger number of images had been used in the study. On the other hand, intra-rater reliability has also been previously reported for the FGS by veterinarians scoring images in two sessions, 30 days apart (excellent - ICC single = 0.91 to 0.95) (Evangelista et al., 2019).

The participants were supplied with the FGS training manual; however, no formal training or discussion on how to score the images was given. It has been demonstrated that training (discussion of ambiguous images between an experienced rater and trainees) improved the reliability of the Rat Grimace Scale (Zhang et al., 2019). One interesting finding of the present study was that students were as reliable (according to inter-rater reliability) and consistent in their scores as veterinarians, given that both groups demonstrated excellent intra-rater reliability. Veterinary students enrolled in the present study were mostly in the final years of training (1^{st} year, n = 1; 4^{th} year, n = 3 and 5^{th} year, n = 1). Students in the 4^{th} and 5^{th} year at the Université de Montréal had already completed courses in anaesthesia and pain management. Students receive 12 hours of training in pain management including the FGS (Simon et al., 2017), indicating some familiarity with the instrument and a certain degree of training on the subject. As shown previously, veterinary students at a later stage of their studies assigned higher pain perception scores to different animal species, which may reflect their increased knowledge acquired over time (Valros & Hänninen, 2018). This is an important finding and it shows that the FGS can be used reliably even by individuals under veterinary training.

The results for inter-rater and intra-rater reliability suggest that veterinary nurses could also potentially benefit from training before using the FGS. These professionals play a key, front-line role in pain assessment in the clinical setting. Previous studies in humans demonstrated that the inter-observer reliability on pain assessment was only moderate between experienced nurses working at a human emergency department (Hangaard et al., 2018). This is in agreement with previous findings from surveys which indicated that more than 90% of veterinary nurses and technicians in the UK and New Zealand considered that their knowledge of pain assessment could be improved (Coleman & Slingsby, 2007; Kongara et al., 2016), while 68% of Canadian veterinary nurses considered their knowledge related to the recognition and control of pain to be adequate (Dohoo & Dohoo, 1998). However, most of these surveys' outcomes come from studies performed before the publication of validated pain scales for cats and none of them included the recently published FGS, which can make study comparisons difficult. Additionally, our findings do not allow us to conclude if years of nurse training would also improve reliability using the FGS.

Although the ICC is an index of reliability that reflects both degree of correlation and agreement between measurements (Hallgren, 2012; Koo & Li, 2016), it does not reflect accuracy (i.e. the scores could be perfectly reliable and consistently wrong). For this reason, we calculated the agreement of the scores from each group with the veterinarian group (considered as the gold standard) using the Bland and Altman method. This method allows to evaluate the bias (mean of differences between two measurements) and the limits of agreement (which indicates the interval in which 95% of the differences lie) (Bland & Altman, 1999). Bias larger than 0.1 were considered unacceptable, indicating poor agreement between measures. This threshold was determined before the study had begun, based on the range of the final FGS scores (0.0 to 1.0) and considering that a difference of more than 1 unit in the FGS score would result in erroneous interpretation of the pain state. The LoA did not span the analgesic threshold previously established for the FGS (0.39 out of 1.0) (Evangelista et al., 2019) in any of the comparisons, showing that all groups agreed with the veterinarian group. We have used a similar approach to compare two methods of scoring using the FGS (real-time versus image assessment) (Evangelista, Benito, et al., 2020).The minimal bias observed (< 0.1) demonstrated a very good agreement for all groups (owners, students and nurses), with a slight overestimation of veterinarians' scores. This finding corresponds with a previous survey which demonstrated that nurses overestimated veterinarians' pain scores (Coleman & Slingsby, 2007) and with other studies showing that veterinary students tend to overestimate pain scores when compared with experienced veterinarians and board-certified specialists in anaesthesia and analgesia (Barletta et al., 2016; Benito et al., 2017; Doodnaught et al., 2017). In general, health care professionals tend to provide lower estimates of others' pain compared with lay people (Hadjistvropoulos et al., 1998; Prkachin et al., 2001). Discrepancies between self-report of pain and external observers' judgments of pain intensity occur among different raters (e.g., physicians, nurses, family members, aides, caretakers, etc.), with family members sometimes overestimating and care providers underestimating the intensity of pain experienced by the patient (Kappesser et al., 2006). Moreover, estimation of others' pain tends to become more conservative with experience (Kappesser et al., 2006; Miron-Shatz et al., 2020), possibly because clinicians recalibrate internal reference points as they become habituated to observing patients in pain through their clinical experience (Prkachin, 2011).

Another interesting finding of this study was the effect of the rater gender on FGS scores. The different responses by male and female observers might reflect that women may show higher empathy towards pain. Literature in humans is extensive regarding sex differences in responses to pain (Bartley & Fillingim, 2013). Females are frequently more empathic towards the pain and distress of others than males (Christov-Moore et al., 2014; Sadeghiyeh et al., 2012). Humans can also feel empathy, including assessment of pain towards non-human species (Prguda & Neumann, 2014). Among veterinary students and production animal practice-oriented veterinarians, empathy towards animals was greater than towards humans (Norring et al., 2014). Additionally, female veterinarians have been more likely to assess pain and administer analgesics for dogs and cats (Beswick et al., 2016; Dohoo & Dohoo, 1996b; Doodnaught et al., 2017; Hewson et al., 2006a; Williams et al., 2005).

This study has limitations. The number of female participants (n = 16) was greater than males (n = 4), and two of the groups (students and nurses) were composed exclusively by females. Moreover, there were few observers to examine the effect of years of work experience (nurses and veterinarians) or years in school (students). Students are normally part of a young population of individuals who could bias our results in terms of age effect. These findings could be further explored using a larger number of raters. On the other hand, our study aim was not to primarily assess the effect of training on the FGS scores. We wanted to compare the reliability of groups with different levels of expertise in pain assessment. Training programs have been shown to improve veterinary students' knowledge on pain and their ability to identify pain in animals (Mich et al., 2010). Training also improved the reliability of the Rat Grimace Scale (Zhang et al., 2019). Similar training programs designed for students and veterinary nurses could positively contribute to better pain management in the clinical setting including the application of the FGS. Additionally, such training modalities could also be valuable for pet owners when pain management is indicated at home.

In conclusion, our results show that all groups agreed with veterinarians' evaluations; that FGS can be used in a reliable manner even by untrained individuals for pain assessment in cats, although reliability could potentially be improved by training. These results represent a substantial progress in feline pain assessment if one considers that the FGS can be used reliably by groups with different expertise. We highlight that the FGS is the first instrument for pain assessment that could be used by cat owners in the recognition of pain in the home environment where cats tend to exhibit normal behaviours. This is important since fear, stress and anxiety during hospitalization could influence pain assessment by veterinary health professionals. Female observers assigned higher FGS scores than males, potentially indicating that when an image is ambiguous, female raters may have more empathy towards the identification of pain.

2.3.7. Acknowledgements

The authors would like to acknowledge the cat owners, veterinary students and nurses who participated in the study. The authors would also like to acknowledge Dr. Hélène Ruel, Dr. Beatriz Monteiro, Dr. Ryota Watanabe and Dr. Rubia Tomacheuski for participating in the study and M. Guy Beauchamp for the linear mixed model analysis.

2.3.8. Additional Information

2.3.8.1. Author Contributions Statement

MCE and PVS conceived the study, performed data collection, analysed the results and both authors reviewed the manuscript.

2.3.8.2. Competing interests

The authors declare no competing interests.

3. General Discussion

3.1. Overview and contributions to the field

This thesis aimed to fulfill the gaps in the knowledge regarding pain assessment using facial expression in cats by conceiving a new tool, the FGS. The FGS was shown to be a valid, reliable and practical tool potentially for both research and clinical use in real-time or using image assessment; and that it could be applicable in a wide range of painful conditions, by raters with different degree of expertise. However, the FGS is still a recent development and a wide range of its possible applications and limitations deserve further studying.

The studies presented herein described the development, validation and generalizability of the FGS. The first study (highlighting the development and validation of the FGS by Evangelista et al., 2019) used a sound psychometric approach, following the footsteps for the development of other instruments for cats and other species. This elevated the instrument to the high standards set by the MGS and RGS, with the FGS as the next best-investigated grimace scale (Evangelista, Monteiro, et al., 2020; Mogil et al., 2020). The second study by Evangelista, Benito, et al. (2020) demonstrated the real-time application of the FGS in a clinical setting in cats undergoing ovariohysterectomy. Elective sterilization of companion animals (spay / neuter) is one of the most common surgical procedures performed in veterinary medicine. Ovariohysterectomy has been used as a model to the study of visceral pain and the effectiveness of analgesics for many years (Brennan, 1999; Brondani et al., 2009; Evangelista et al., 2014; Gonzalez et al., 2000; Quarterone et al., 2017; Steagall et al., 2009). This study established the construct validity of the FGS for use in perioperative pain, reassessed responsiveness of the FGS, and demonstrated that real-time scoring is feasible, although it slightly overestimates image assessment, with good agreement between the methods of scoring.

Finally, the third study (agreement and reliability of the FGS among different observers by Evangelista & Steagall) answered the questions regarding the observer bias and the influence of the gender and other characteristics of the observers on FGS scores. The FGS was shown as a reliable tool even for untrained individuals, and these results represent a substantial progress in feline pain assessment because the FGS is the first instrument for acute pain assessment that could be used by cat owners in the recognition of pain in their home environment, where cats tend to exhibit normal behaviours. If cat owners can recognize pain in their pets reliably, that could lead to an increase in veterinary consultations and could potentially result in early diagnosis and better feline health and welfare. Hopefully, the FGS will also contribute to a change in the scenario of low adherence to the use of pain scoring instruments in the clinical setting (Coleman & Slingsby, 2007; Dawson et al., 2017; Hunt et al., 2015).

3.2. Development, validation and application of the FGS

The methodology used to develop the FGS followed the footsteps of the previously published grimace scales with the use of video-recording of the cats undisturbed in their cages and later image capture. The use of videos has the advantage of continuously recording the individual from a distance, whereas the presence of a camera (and observer) in close proximity to the animal may change the animal's response. Image caption obtained from the videos eliminates the need for physical restraint or handling the cats. Other studies included photographs taken when the animals were restrained (Di Giminiani et al., 2016; Holden et al., 2014; Reijgwart et al., 2017) and this physical interaction was shown to significantly alter facial expressions scores in lambs (Guesgen et al., 2016). However, taking screenshot from videos may also involve a subjective judgement towards the appropriate timing to capture a still image (as it is usually done with photographs).

After obtaining the still images from videos, the FGS was developed by comparing the facial features of control and painful cats. Five AU (ear position, orbital tightening, muzzle tension, whiskers position and head position) were identified and included in the FGS. Similar AU have been previously described in rodents, horses and ferrets (Dalla Costa et al., 2014; Keating et al., 2012; Langford et al., 2010; Reijgwart et al., 2017; Sotocinal et al., 2011). These features were summarized in the preliminary results of the systematic review on the measurement properties of grimace scales (Evangelista, Monteiro, et al., 2020). Ear position and orbital tightening were listed for all grimace scales. The other grimace scales also include changes around the nose/cheek and/or mouth area. One particularity of the cat is the extensive network of vibrissae with well-developed intrinsic musculature, which influences the range of facial expressions (Caeiro et al., 2017). Although not specifically related to the face per se, head position has also been listed as AU in horses, donkeys and sheep (Hager et al., 2017; van Dierendonck et al., 2020; van Loon & van Dierendonck, 2015) and AD related to general head movements are also scored using the FACS in humans (Ekman et al., 2002).

3.2.1. Validity and reliability

As validity should be paramount in the choice of a scale for use in the clinical or experimental setting, the fact that now there are three validated pain scales for acute pain assessment in cats (UB-MCPS, G-CMPS and FGS) should prompt practitioners to choose from, and implement them in feline practice. However, it should also be noted that no single pain assessment strategy, such as measurement of physiological parameters, interpretation of behaviors or estimates of pain by others, is considered sufficient by itself. Each instrument or approach to pain assessment has its own context and ease of application, its strengths and drawbacks. Having a greater selection of instruments to choose from, may overcome some of the challenges of pain assessment in animals, and possibly change the scenario of low adherence to their use in clinical practice. It also may lead to the ultimate goals of improving pain assessment, optimizing treatments and promoting better animal care.

In the recent decade, emphasis on validating pain scales has led to an increase in awareness of pain incidence and accuracy of measurements. The overall process of validating a pain scale involves testing of content validity (extent to which the scale contains the relevant items related to pain, subjectively viewed by knowledgeable individuals or experts in the area); construct validity (the ability of the instrument of measuring a construct, i.e. pain, that cannot be directly observed); criterion validity (degree of correlation with a gold standard or other measure of the phenomenon), reliability (inter- and intra-rater) and internal consistency (Keszei et al., 2010; McDowell, 2006; Streiner & Norman, 2008). The psychometric approach used herein to validate the FGS was in agreement with previously validated behavior-based pain scales (i.e. UB-MCPS and G-CMPS) and other grimace scales (MGS, RGS, HGS, among others) (Brondani et al., 2013; Evangelista, Monteiro, et al., 2020; Reid et al., 2017).

The construct validity was confirmed during the first study using the known-groups discrimination method in cats presenting with naturally-occurring pain (control versus painful cats) (Evangelista et al., 2019), and restated in the second study in cats undergoing ovariohysterectomy (before and after the painful stimulus; where the FGS were higher than baseline from 1 to 12h after extubation) (Evangelista, Benito, et al., 2020); demonstrating the FGS's high discriminative ability between painful and non-painful cats.

The criterion validity was also determined in the first study, by correlating the newly developed FGS with the revised version of the G-CMPS (which did not contain the facial expression features; Calvo et al., 2014). A very strong correlation was obtained, confirming the concurrent criterion validity (Evangelista et al., 2019). Likewise, a subsequent study employing the FGS in real-time to assess pain in cats undergoing ovariohysterectomy (using an opioid-free anesthetic protocol) found a positive and significant association between the FGS and the definitive version of the G-CMPS (containing facial expressions) (Diep et al., 2020). This demonstrated that both instruments measure the same construct (pain) and that they go in the same direction (the scores of both instruments were consistently elevated at postoperative timepoints; as pain intensity was expected to increase after surgery). Although a ketamine-based protocol (ketamine-dexmedetomidine-midazolam) was used in this study (Diep et al., 2020), the effect of these drugs on FGS scores in non-painful cats was not determined.

In addition to correlating the pain scores, it would be interesting to determine the percentage of agreement and disagreement regarding the decision to give rescue analgesia using different instruments. For example, the study by Steagall et al. (2018) demonstrated that despite being strongly correlated (rho > 0.7; p <0.0001), the percentage of disagreement between the UB-CMPS (French version, excluding the blood pressure and appetite, and considering the analgesic threshold of ≥6) (Steagall, Monteiro, Lavoie, et al., 2017) and G-CMPS (original revised version, considering the analgesic threshold of ≥4) (Calvo et al., 2014) was 7.8%. The decision for giving rescue analgesia would have changed if G-CMPS was used instead of UB-CMPS in 6.4% of cases in that study (Steagall et al., 2018). Still, this represents a low percentage of disagreement, which may not be of great importance in clinical practice when the whole context is taken into consideration; however, in a research setting, it could potentially change the outcome of a study comparing different treatments (especially when a small sample size is employed or in an underpowered study).

The internal consistency of the FGS was determined in the first study and considered excellent (Cronbach's alpha = 0.89) (Evangelista et al., 2019); and comparable to the other grimace scales (MGS = 0.89; RGS = 0.84 and 0.87) (Langford et al., 2010; Oliver et al., 2014; Philips et al., 2017). This measure is often used to check the homogeneity of the scale during its development phase (i.e. each item should correlate to one another and with the final score) (Streiner & Norman, 2008). The Cronbach's alpha coefficient gives an overall idea of this inter-relatedness of the items, and it may be recalculated by excluding each item (or AU in the case of grimace scales). If the alpha increases significantly when a specific item is dropped, it indicates that its exclusion would increase the homogeneity of the scale (Streiner & Norman, 2008). The Cronbach's alpha is considered the measure of choice for internal consistency, given its ease of calculation and the possibility of pooling the coefficient across studies when the results are sufficiently similar (Prinsen et al., 2018). Yet, other methods of measuring internal consistency have been described, such as the inter-item and item-total correlation, used in the development of the SPFES and RGS (McLennan et al., 2016; Sotocinal et al., 2011), and a Linear Discriminant Analysis (LDA), which is related to principal component analysis and factor analysis (Martinez & Kak, 2001). These sophisticated approaches may be used in multidimensional (or multifactor) scales and enables the assessment of the homogeneity of the whole instrument and its subscales, as it was the case for the factor analysis used in the validation of the UB-MCPS (Brondani et al., 2013). A principal component analysis was used in the development of a geometric morphometric approach to assess pain in cats (Finka et al., 2019). The LDA may also be used in a context of automation of the pain detection process (i.e. for pattern recognition or machine learning), to identify the combination of features (or items) that classifies or separates the objects or events (i.e. the presence or absence of pain). A classifier based on the weights of each AU, defined by LDA was proposed for the FerGS (Reijgwart et al., 2017), and for the MGS and HGS (Dalla Costa et al., 2018). Lastly, it would be interesting to use the LDA approach for the FGS to understand the the multiple association of AUs with each other and the final score, to understand the weight of each individual AU in painful conditions. Our laboratory is currently working in the study of these classifiers, which could potentially broaden the contexts of application of the FGS.

The FGS was also recognized as a reliable tool in different painful conditions. Its inter-rater reliability was tested in the first study (Evangelista et al., 2019) and reassessed in the third study (Evangelista & Steagall, submitted). Another study performed in our laboratory assessed the performance of the FGS in cats undergoing dental extractions (Watanabe, Doodnaught, et al., 2020) - Appendix B. This study confirmed the good reliability of the FGS for the assessment of pain in cats with oral pain and demonstrated that the presence of the observer (caregiver) did not affect FGS scores. These findings helped build evidence towards the use of the FGS in the clinical setting, by expanding its applications.

The raters enrolled in the aforementioned studies were experienced in animal pain assessment (PhD candidates and board-certified veterinary anesthesiologists) (Evangelista et al. 2019; Watanabe, Doodnaught, et al. 2020; Evangelista & Steagall, submitted). In all studies, the reliability was assessed using ICC and the interpretation of the coefficients was performed based on the ICC single measures estimates and according to the same guideline (Koo & Li, 2016). Both single and average estimates, along with their confidence intervals were provided in all three studies concerning the FGS reliability, which may facilitate comparisons with other studies (especially those that only display the ICC average, which yields always higher estimates). Additionally, the classification proposed by Koo & Li (2016) is more rigorous than other previously available (Cicchetti, 1994), which may give the false impression that one instrument is more or less reliable than others. According to the classification used in our studies (Koo & Li, 2016), the intra-rater reliability was good among owners and nurses, and excellent among students (Evangelista & Steagall, submitted) and veterinarians using the FGS (Evangelista et al., 2019; Evangelista & Steagall, submitted). Moreover, using the same classification, the FGS displayed good inter-rater reliability when used for assessing naturally-occurring and postsurgical (abdominal and oral) pain by all groups in the different studies (FGS final score ICC single: between 0.8 and 0.89; Evangelista et al., 2019; Evangelista & Steagall, submitted; Watanabe, Doodnaught et al., 2020). If the classification by Cicchetti (1994) had been used, or even the ICC estimate for the average of measures instead of single measures, the intra- and inter-rater reliability of the final FGS scores would have been considered excellent in all cases.

At the same time, the inter-rater reliability of each AU was variable and often the AU muzzle and whiskers yielded lower reliability estimates than ears, eyes and head position. In all occasions, reliability was assessed using images obtained from videos, and the quality of the images and lack of three-dimensional view were considered a shortcoming for the assessment of these two AU in all the studies assessing the reliability of the FGS (Evangelista et al., 2019; Evangelista & Steagall, submitted; Watanabe, Doodnaught et al., 2020). Our clinical experience showed that scoring whiskers and muzzle in real-time was less challenging. Similar challenges have been described when scoring the whiskers in mice; and some authors decided to simply remove this AU during their studies using the MGS (Miller et al., 2015; Miller & Leach, 2016). As mentioned previously, studying the weight of each AU in the final score (and their multiple associations) may help to understand the influence of skipping or not the assessment of a given item in the assessment of painful states.

3.2.2. Analgesic threshold

The determination of an analgesic intervention threshold has been considered to expand the usefulness of a clinical tool (Oliver et al., 2014; Reid et al., 2007). It may guide the decision of analgesic administration and enable treatment monitoring (by observing the trend of the scores). Ensuring the use of pain scoring systems with proposed analgesic threshold in clinical practice allows animal health professionals to assess if a patient is in pain and promptly determine if treatment is required.

Analgesic intervention thresholds have been proposed for the RGS, SPFES and FGS (Evangelista et al., 2019; McLennan et al., 2016; Oliver et al., 2014). In the case of the RGS and FGS, thresholds were determined based on the optimal balance between sensitivity and specificity for each scale (RGS: 0.67 out of 2 and 0.39 out of 1). In parallel, the analgesic threshold proposed for the SPFES (5 out of 10) was considered based only on the high specificity (nearly 100%), thus yielding a very low sensitivity (nearly 30%). The higher specificity relates to few false positives (i.e. SPFES scores above 5 result indicate a high probability of the presence of pain; therefore, animals receiving higher scores will likely benefit from rescue analgesia). The drawback of setting a threshold with high specificity and low sensitivity is failing to identify potentially painful animals. On the other hand, lowering the analgesic threshold (increasing sensitivity and decreasing specificity), could be more advantageous when the consequences of failing to identify potentially painful animals outweighs the risks of providing analgesics to potentially non-painful animals. According to the sensitivity and 1-specifity table provided in the study by McLennan et al. (2016); an alternative choice of analgesic threshold with balanced sensitivity and specificity would be between 2.5 and 3.5 (depending on the disease, footrot or mastitis). Similar tables (with potential analgesic thresholds and their values of sensitivity and 1-specifity) were provided along with the studies concerning the RGS and FGS (Evangelista et al., 2019; Oliver et al., 2014), hence allowing the user to change the threshold to increase or decrease sensitivity and specificity according to the clinical or experimental needs.

3.2.3. Real-time versus image assessment and observer effect

The second study described in this thesis (Evangelista, Benito, et al., 2020) demonstrated that the FGS can be a practical tool for clinical use with the application of real-time scoring. It represents a major advantage over the retrospective (and laborious) standard methodology of image-based scoring. The good agreement found between the methods means that in practice, the caretaker (veterinarian, veterinary student, nurse, animal care staff, etc) can simply observe the animal for a couple of minutes in real-time, enabling early pain identification and prompt analgesic intervention. This will hopefully encourage use of the FGS as a clinical tool, especially in busy practices, by allowing a quick and practical way to identify and assess pain. Real-time assessment has been previously demonstrated also using the MGS and RGS, with contrasting results (Leung et al., 2016; Miller & Leach, 2015a). Significantly lower MGS scores were obtained with real-time assessment in comparison to the standard method in mice (Miller & Leach, 2015a); while good agreement was demonstrated in rats (Leung et al., 2016). This difference in results may be because of species differences or due to disparities in their methodologies. For example, Miller & Leach (2015a) performed three 5-second observations over a 10-minute period and compared with image scores obtained from photographs using Mann-Whitney U tests. Leung et al., (2016) performed punctual and 15-seconds observations that were compared with still images captured from video-recordings using the Bland and Altman method (more appropriate in this case). Furthermore, to assess the potential clinical applicability of the HGS, researchers used short video-clips of horses with acute laminitis (Dalla Costa et al., 2016). In that study, the authors compared the scores obtained from 15s video-clips and those obtained from still images (captured from the same videos) using Spearman's rank correlation. The results demonstrated a positive correlation between the scoring methods; however, a high level of variation between the observers was noticed.

A recent review highlighted that novel approaches to the use of grimace scales (such as continuous video-recordings, real-time, etc.) should be encouraged; however, when scales are applied in novel ways, the authors strongly recommended that a comparison to the originally described method is included to allow the interpretation and immediate comparison of the results across studies (Mogil et al., 2020).

The method by which pain is evaluated in real-time is by nature non-blinded. The observer is often aware of the timing (i.e. pre- or postoperatively, or before/after analgesic treatment) and the general condition of the animal to be evaluated. Although it may be a potential source of bias, the similar results obtained using either method (RGS and FGS in real-time or image scoring) implies that the observer being aware of timing did not introduce substantial bias. For example, the significant differences between different treatment groups, and changes in scores before and after the painful stimulus were found using both real-time and image scores (Evangelista, Benito, et al., 2020; Leung et al., 2016).

Another point to be considered in the real-time application of grimace scales is the influence of observers' presence (observer effect). This was not assessed in the study testing the real-time application of the FGS (Evangelista, Benito, et al., 2020); however, it was considered in the follow-up study performed in our laboratory regarding the inter-rater reliability of the FGS in dental extractions (Watanabe, Doodnaught, et al., 2020). The results of said study demonstrated that the caregiver (male observer) presence did not alter FGS scores. Similarly, in the study by Leung et al. (2016), the presence of the observer did not alter RGS scores; however, it should be noted that the rats under observation were all habituated to the female observer. In contrast, another study remarked that the presence of a male observer produced stress-induced analgesia (supported by reduced grimace scale scores) (Sorge et al., 2014). In the majority of the studies involving pain assessment in animals using behavior-based pain scoring instruments or grimace scales, the gender of the observer was not reported, which prevented further inferences on the observer effect in other species.

3.2.4. Observer bias

The third study presented in this thesis (agreement and reliability of the FGS by different raters by Evangelista & Steagall) provided some insight into the observer bias. It was suggested that the rater gender may influence pain assessment, as female raters assigned scores in average 0.1 unit higher than males. This might reflect that women may show higher empathy towards pain; although, these findings should be interpreted with caution because in our study, the number of females was greater than male raters. On the other hand, the sex of the cats did not seem to influence FGS scores (as demonstrated in the first study; Evangelista et al., 2019).

It has been previously demonstrated that females are frequently more empathic towards the pain and distress of others than males (Christov-Moore et al., 2014; Sadeghiyeh et al., 2012). Sex and gender differences in responses to pain stimulus, responsivity to treatments and in perceiving pain from others have been widely studied (Boerner et al., 2018; Coll et al., 2012; Green et al., 2009; Mogil, 2020). Multiple biological and psychosocial processes are cogitated as contributing factors for these differences. For instance, factors such as genotype and endogenous opioid functioning reproductive hormones, psychosocial processes such as pain coping and early-life exposure to stress, in addition to stereotypical gender roles that may contribute to the disparities in pain expression (Bartley & Fillingim, 2013); while stereotypical gender roles, social constructs and empathy levels seem to affect how individuals perceive pain in others (Baez et al., 2017; Christov-Moore et al., 2014; Green et al., 2009; Han et al., 2008).

This is still poorly understood in veterinary; however, it has been demonstrated that humans can also feel empathy towards non-human species (Prguda & Neumann, 2014). Veterinary students and production animal veterinarians show greater empathy towards animals than towards humans (Norring et al., 2014). Additionally, female veterinarians have been more likely to assess pain and administer analgesics for dogs and cats than male veterinarians (Beswick et al., 2016; Dohoo & Dohoo, 1996b; Doodnaught et al., 2017; Hewson et al., 2006a; Williams et al., 2005).

3.3. Exploring the limitations and future work

Applications and limitations of the FGS have only begun to be explored. Some questions remain unanswered and they are certainly worth considering as subject of future work. For example, the influence of cats' breed on the FGS; the effect of higher doses of opioids on FGS scores, as well as the effect of other sedatives; the quality of the pictures when using image scoring; the overlap between pain and other emotional states (i.e. fear, anxiety, stress, etc); use of grimace scales in chronic pain conditions etc. Some of these topics are currently being investigated by our laboratory.

In the first study (development and validation of the FGS), two distinct populations of cats (healthy control cats and client-owned painful cats) of various breeds were included (Evangelista et al., 2019). Similarly, in the following studies (Evangelista, Benito, et al., 2020; Watanabe, Doodnaught, et al., 2020), different breeds were represented and most of the cats were domestic short-haired, not including brachycephalic breeds (i.e. Persian, Himalayan, Exotic, etc). Cat faces are variable in form (i.e. dolichocephalic or brachycephalic breeds) and are generally covered in hair (with the exception of breeds such as the Sphinx). This can make distinguishing the nuances of facial expression in cats, in general, challenging. Morphological differences (round-shaped skulls and decrease in facial width) related to breed-specific features have been previously observed in brachycephalic cats (Kunzel et al., 2003) and was also raised in a study using the CatFACS to code cats faces. According to the authors, it was overcome by basing facial action coding on the observed movements, regardless of the starting reference point (Bennett et al., 2017). The same principle could potentially be applied to the FGS; yet it is still unknown if this instrument could be used as reliably in brachycephalic cats as in dolichocephalic ones.

The second study showed no effect of sedation (premedication with acepromazine and buprenorphine) on FGS scores; however, a low dose was used, and only mild sedation was observed (Evangelista, Benito et al., 2020). It is important to know how other protocols promoting deeper sedation (i.e. ketamine or dexmedetomidine) would affect these scores. In the study by Diep et al. (2020) performed in our laboratory using a ketamine-based opioid-free anesthetic protocol, the pain scores (FGS and G-CMPS) recorded at 0.5 and 1h postoperatively were not considered in the analyses because of purposeless movements and restlessness during anesthetic recovery. Based on these observations and the significant difference of sedation scores between baseline values and 0.5 and 1h, it might be possible that ketamine also impairs pain assessment using the FGS and G-CMPS, as it has been demonstrated to confound UB-MCPS scores (Buisman et al., 2016). This was a limitation of the study and there is still a need for understanding the effects of ketamine in other pain scoring systems, including the FGS. A new study is currently underway in our laboratory, looking into these effects.

The study presenting the reliability of the FGS in cats with dental pain (Watanabe, Doodnaught, et al., 2020) was an exploratory study and images of moderate to severely painful cats were underrepresented in the collection used for scoring, which was also considered a limitation in the second study presented in this thesis (Evangelista, Benito, et al., 2020). Additionally, image quality was an issue in these studies, and the height and angle of the video camera may have not been ideal to capture the frontal image of the cat for FGS scoring. Pose has been recognized as a main challenge in facial expression recognition along with variability related to lighting and background, even using automated systems (Finka et al., 2019; Li et al., 2019; McLennan & Mahmoud, 2019; Moore & Bowden, 2009). An alternative that may overcome this issue is the use of real-time scoring, in which the observer is able to move around and look at the cats' faces from different angles.

Another point to be considered is the potential overlap between pain and other states (stress, fear, anxiety, sleep, grooming). In mice, there may be an overlap between facial expressions displayed in situations of aggression (induced by the odor or close presence of an unfamiliar conspecific in a resident-intruder test of aggression), and those induced by pain, such as tightened eyes, flattened ears, nose and cheek bulging (Defensor et al., 2012). On the other hand, HGS scores did not change significantly in any of the four experimental conditions (novelty, grooming, anticipation food reward and fear) when compared to control, although fear seemed to have a greater impact than the other conditions in horses (mostly in the AUs stiffly backward ears and prominent strained chewing muscles) (Dalla Costa et al., 2017). Furthermore, another study looked into the changes in facial expression of rats subjected to a positive (tickling) and a mildly aversive contrast treatment (novel room and intermittent white noise). The authors found some changes in ear color and the ear angle was wider (ears more relaxed) in the positive emotional state; however, RGS scores did not change (Finlayson et al., 2016). In cats, facial expressions associated with negative affective states other than pain (i.e. fear, frustration and anxiety) have features similar to those described in the FGS, for example, the flattening of the ears (Bennett et al., 2017; Leyhausen, 1979). The direct impact of these emotional states on FGS scores have not yet been determined and deserves further study. For the time-being, a cat should not be evaluated when sleeping, grooming, eating or vocalizing.

Emotional states in mammals are typically recognized by analogy to human behaviors, and more recently by the involvement of homologous neuroanatomical structures involved in specific human emotions (D. J. Anderson & Adolphs, 2014). For example, the amygdala is involved in the processing of behaviors such as freezing in humans and rodents (Blanchard & Blanchard, 1972; Feinstein et al., 2011).

Previous studies on neuroimaging of pain processing in animals have described activations in areas seen in common with human studies (Borsook & Becerra, 2011; J. T. Da Silva & Seminowicz, 2019; Tracey, 2008; Tracey & Mantyh, 2007). Acute pain (induced by classical nociceptive models) in mice, rats and monkeys activated regions of the primary somatosensory cortex, insula, cingulate cortex, thalamus, retrosplenial cortex, and periaqueductal gray, all considered part of ascending and descending nociceptive pathways (J. T. Da Silva & Seminowicz, 2019; Tracey & Mantyh, 2007). A common pattern of nociception-evoked activations in these animals and the human brain has been described, however specific regions and their degree of activation may vary according to the species studied and the neuroimaging method (including limitations related to awake versus anesthetized animals, anesthetic protocol, induced stress, unknown effects of training, etc) (J. T. Da Silva & Seminowicz, 2019).

Recently, a fMRI study identified the specific brain regions activated by a noxious stimulus (tail clamp) in cats. Increased activity was detected in the regions corresponding to the human pain matrix (somatosensory area, the parietal association area, cingulate cortex and the cerebellum). Besides, activation of the hippocampus was also reported in these cats. (Nagakubo et al., 2017). This brain area is involved in cognition and memory. Hippocampal activity may be involved in the process of exacerbation of pain by anxiety in humans (Ploghaus et al., 2001). Also, its morphological changes are thought to be associated with persistent pain, emotional and cognitive changes (Yang & Chang, 2019). Neuroimaging techniques can yield valuable information on the brain areas and pathways involved in different emotional states, pain processing, modulation and analgesic targets, especially when correlated with behavioral responses and facial expressions. Multidisciplinary studies involving these techniques may potentially enable in the differentiation of specific behavioral and facial responses to different emotional states (i.e. stress, fear, anxiety) and pain in non-verbal humans and other mammals.

Formerly it was assumed that the grimace scales were only able to assess acute pain (Langford et al 2010); however, it was later demonstrated that rodent grimace scales (MGS and RGS) could also be used to assess neuropathic and chronic inflammatory pain of longer duration (Akintola et al., 2017; Leung et al., 2019; Liao et al., 2014). It is difficult to explain why researchers found these differences in grimacing duration. One possibility is that animals, especially prey species, would attempt to inhibit grimacing to hide it from potential predators or aggressive conspecifics (Mogil et al., 2020). Or still, another possibility is that animals with chronic pain would not display grimacing for a long period, unless the pain is exacerbated by certain activities or movements causing breakthrough pain. A third point may be raised, encompassing the issue of the lack of successful translation of these experimental pain models to the clinical setting. For example, the classical models (i.e. chronic constriction injury, spared nerve injury, joint loading, etc.) may not reflect the clinical chronic and/or neuropathic pain conditions (Klinck et al., 2017; Lascelles et al., 2017). Maybe testing the applicability of grimace scales in natural models of pain (such as osteoarthritis in cats) could give better insights on the usefulness of these tools in chronic pain. Nevertheless, this raises the possibility that grimace scales might indeed be useful in the assessment of chronic pain (or at least acute episodes on chronic pain) in animals and its application in other species (including cats) deserves further study.

3.4. Future perspectives in pain assessment

Some of the work presented in this thesis focused on determining the observer effect and observer bias during pain assessment. Additionally, the standard methodology of collecting images from video-recordings is laborious and retrospective. An alternative that could eliminate some bias and relieve the workload required to collect and score the images is the introduction of automated processes of pain detection.

Technologies for automated data collection have been described in rodents such as the HomeCageScan (a software that detects various behaviours of mice and records changes in their frequency and duration) (Roughan et al., 2009), the Rodent Face Finder (a software that automatically captures images from videos in rats for further manual scoring; Sotocinal et al., 2011) and the automated image classification for the MGS (aMGS) using a deep neural network which classifies images as "pain" or "no pain" (Tuttle et al., 2018). The aMGS was able to identify images into the two categories with similar accuracy to human raters (> 80%), but according to the authors, it functioned best when images had scores at the extreme ends of the scale (i.e. 0 or 2) and poorly when scores were in-between, demonstrating that its sensitivity could be improved to capture subtle facial changes. Another automated method of pain facial expression detection system has been described for sheep with high accuracy (McLennan & Mahmoud, 2019; Noor et al., 2020). This newly proposed framework combines concepts from SPFES with automatic facial expression analysis for the in-farm application of automated pain (and disease) assessment in sheep.

Improvements or development of new technologies integrating artificial intelligence is a promising approach that may help accurately identify facial expressions of pain in undisturbed animals (using still or moving images) and to detect subtle changes in pain states that may go unperceived even by experienced observers. On the other hand, in a clinical setting, there will always be the need for human observers to decide if an animal requires analgesic treatment or in the case of pre-clinical research, if it has reached a humane endpoint and intervention is required.

Steps towards the automation of pain detection and other applications of facial recognition in cats are emerging. Finka et al. (2019) used a geometric morphometrics approach to quantify changes in facial shape in cats before and after OVH. This approach has been proposed as a basis for further application of machine learning algorithms for automated pain recognition.

On the other hand, facial recognition in cats has been applied for other purposes such as identifying missing pets,[1] for automatically opening a pet door to let in the correct cat into the house[2] or even incorporated into smart feeders to correctly release the right type and/or amount of food for each cat.[3]

Maybe in the future it will be possible that a mobile application enables pointing the cell phone camera to the cats' face and the software will correctly identify its pain state. Meanwhile, one can only hope that new simple and practical tools, such as the FGS are routinely applied in clinics, improving pain assessment and management.

---

[1] The application Finding Rover uses facial recognition to help owners retrieve their lost pets www.findingrover.com
[2] Microsoft has incorporated a facial recognition software in a pet door with a webcam - https://www.windowscentral.com/microsoft-delivers-facial-recognition-pet-doors-windows-10-iot-core
[3] The Mookkie pet feeder uses facial recognition to distinguish between cats for individual feeding - www.mookkie.com

Conclusion

Pain assessment in cats is challenging due to a number of reasons, including their discrete nature (possibly linked to their particular domestication process) and potential behavioral changes in unfamiliar and stressful situations, such as the veterinary clinical context. Different pain assessing instruments have been proposed for cats; however, the majority lack validity, reliability and/or generalizability testing, which is a time-consuming and laborious process. Still, although thoroughly tested, each instrument has its own limitations and the adherence to their use in clinical practice needs to be improved. Alternative or complementary simple, practical and efficient tools such as the Feline Grimace Scale have the potential of changing this scenario.

The FGS was developed and validated using a comprehensive psychometric approach to detect acute pain in cats. It has demonstrated a high discriminative ability between painful and non-painful cats. It is capable of detecting the response to analgesic treatment and it is strongly correlated with another pain scoring system. Furthermore, its reliability has been broadly studied, demonstrating good inter- and intra-rater reliability, not only among veterinarians, but also among cat owners, veterinary students and nurses assessing pain from different sources and intensities in cats (naturally-occurring somatic and visceral pain, and postoperative abdominal pain). The determination of an analgesic threshold with high sensitivity and specificity, along with the evaluation of other aspects of generalizability support its clinical use. Real-time scoring using the FGS was proven feasible and good agreement has been reported between this method and the standard image-based one, although with a slight overestimation by real-time scores. Mild sedation and surgery did not influence FGS. The scores were not different from baseline after premedication, nor 24h after ovariohysterectomy. Furthermore, the FGS scores were not affected by the number of cats owned, nor by the age of the observer. It was however suggested that the rater gender may influence pain assessment, as female raters assigned higher scores than males.

The FGS is a valid, reliable and practical tool potentially for both research and clinical use in real-time or using image assessment. It may be applicable in a wide range of painful conditions, by raters with different degree of expertise. This represents a substantial progress in feline pain assessment and management, towards the highest standards in veterinary care.

Bibliography

Addison, E. S., & Clements, D. N. (2017). Repeatability of quantitative sensory testing in healthy cats in a clinical setting with comparison to cats with osteoarthritis. *Journal of Feline Medicine and Surgery*, *19*(12), 1274–1282. https://doi.org/10.1177/1098612X17690653

Adrian, D., Papich, M., Baynes, R., Murrell, J., & Lascelles, B. D. X. (2017). Chronic maladaptive pain in cats: A review of current and future drug treatment options. *Veterinary Journal*, *230*, 52–61. https://doi.org/10.1016/j.tvjl.2017.08.006

Aguado, D., Bustamante, R., García-Sanz, V., González-Blanco, P., & Gómez de Segura, I. A. (2020). Efficacy of the Parasympathetic Tone Activity monitor to assess nociception in healthy dogs anaesthetized with propofol and sevoflurane. *Veterinary Anaesthesia and Analgesia*, *47*(1), 103–110. https://doi.org/10.1016/j.vaa.2019.05.014

Akintola, T., Raver, C., Studlack, P., Uddin, O., Masri, R., & Keller, A. (2017). The grimace scale reliably assesses chronic pain in a rodent model of trigeminal neuropathic pain. *Neurobiology of Pain*, *2*(October), 13–17. https://doi.org/10.1016/j.ynpai.2017.10.001

Al-Gizawiy, M. M., & Rudé, E. P. (2004). Comparison of preoperative carprofen and postoperative butorphanol as postsurgical analgesics in cats undergoing ovariohysterectomy. *Veterinary Anaesthesia and Analgesia*, *31*(3), 164–174. https://doi.org/10.1111/j.1467-2987.2004.00180.x

Allen, C., Fuchs, P. N., Shriver, A., & Wilson, H. D. (2005). Deciphering Animal Pain. In M. Aydede (Ed.), *Pain: New essays on its nature and the methodology of its study* (pp. 351–366). The MIT Press. https://psycnet.apa.org/record/2006-01574-020

AMVQ. (2020). *Ils sont 3 millions au Québec*. Association Des Médecins Vétérinaires Du Québec. https://www.amvq.quebec/fr/nouvelles/ils-sont-3-millions-au-quebec

Anderson, D. J., & Adolphs, R. (2014). A framework for studying emotions across species. In *Cell* (Vol. 157, Issue 1, pp. 187–200). Cell Press. https://doi.org/10.1016/j.cell.2014.03.003

Anderson, J. R. (2005). *Cognitive psychology and its implications*. Worth Publishers.

Anthoine, E., Moret, L., Regnault, A., Sébille, V., & Hardouin, J.-B. (2014). Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health and Quality of Life Outcomes*, *12*, 176. https://doi.org/10.1186/s12955-014-0176-2

Armitage, E. A., Wetmore, L. A., Chan, D. L., & Lindsey, J. C. (2005). Evaluation of compliance among nursing staff in administration of prescribed analgesic drugs to critically ill dogs and cats. *Journal of the American Veterinary Medical Association*, *227*(3), 425–429.

Arras, M., Rettich, A., Cinelli, P., Kasermann, H. P., & Burki, K. (2007). Assessment of post-laparotomy pain in laboratory mice by telemetric recording of heart rate and heart rate variability. *BMC Veterinary Research*, *3*, 16. https://doi.org/10.1186/1746-6148-3-16

Backonja, M.-M., Walk, D., Edwards, R. R., Sehgal, N., Moeller-Bertram, T., Wasan, A., Irving, G., Argoff, C., & Wallace, M. (2009). Quantitative sensory testing in measurement of neuropathic pain phenomena and other sensory abnormalities. *The Clinical Journal of Pain*, *25*(7), 641–647. https://doi.org/10.1097/AJP.0b013e3181a68c7e

Baez, S., Flichtentrei, D., Prats, M., Mastandueno, R., García, A. M., Cetkovich, M., & Ibáñez, A. (2017). Men, women…who cares? A population-based study on sex differences and gender roles in empathy and moral cognition. *PLOS ONE*, *12*(6), e0179336. https://doi.org/10.1371/journal.pone.0179336

Banchi, P., Quaranta, G., Ricci, A., & Von Degerfeld, M. M. (2020). Reliability and construct validity of a composite pain scale for rabbit (CANCRS) in a clinical environment. *PLoS ONE*, *15*(4). https://doi.org/10.1371/journal.pone.0221377

Barletta, M., Young, C. N., Quandt, J. E., & Hofmeister, E. H. (2016). Agreement between veterinary students and anesthesiologists regarding postoperative pain assessment in dogs. *Veterinary Anaesthesia and Analgesia*, *43*(1), 91–98. https://doi.org/10.1111/vaa.12269

Bartlett, M. S., Littlewort, G. C., Frank, M. G., & Lee, K. (2014). Automatic Decoding of Facial Movements Reveals Deceptive Pain Expressions. *Current Biology*, *24*(7), 738–743. https://doi.org/10.1016/J.CUB.2014.02.009

Bartley, E. J., & Fillingim, R. B. (2013). Sex differences in pain: A brief review of clinical and experimental findings. *British Journal of Anaesthesia*, *111*(1), 52–58.

https://doi.org/10.1093/bja/aet127

Bateson, P. (1991). Assessment of pain in animals. *Animal Behaviour*, *42*, 827–839.

Beaver, B. V. (2003). Introduction to Feline Behavior. In *Feline Behavior: A Guide for Veterinarians* (2nd ed., pp. 1–41). Saunders. https://doi.org/10.1016/b0-72-169498-5/50002-2

Bell, A., Helm, J., & Reid, J. (2014). Veterinarians' attitudes to chronic pain in dogs. *The Veterinary Record*, *175*(17), 428. https://doi.org/10.1136/vr.102352

Benito, J., Evangelista, M. C., Doodnaught, G. M., Watanabe, R., Beauchamp, G., Monteiro, B. P., & Steagall, P. V. (2019). Analgesic Efficacy of Bupivacaine or Bupivacaine-Dexmedetomidine After Intraperitoneal Administration in Cats: A Randomized, Blinded, Clinical Trial. *Frontiers in Veterinary Science*, *6*, 307. https://doi.org/10.3389/fvets.2019.00307

Benito, J., Gruen, M. E., Thomson, A., Simpson, W., & Lascelles, B. D. X. (2012). Owner-assessed indices of quality of life in cats and the relationship to the presence of degenerative joint disease. *Journal of Feline Medicine and Surgery*, *14*(12), 863–870. https://doi.org/10.1177/1098612X12453904

Benito, J., Monteiro, B., Beaudry, F., & Steagall, P. V. (2018). Efficacy and pharmacokinetics of bupivacaine with epinephrine or dexmedetomidine after intraperitoneal administration in cats undergoing ovariohysterectomy. *Canadian Journal of Veterinary Research*, *82*, 124–130.

Benito, J., Monteiro, B. P., Beauchamp, G., Lascelles, B. D. X., & Steagall, P. V. (2017). Evaluation of interobserver agreement for postoperative pain and sedation assessment in cats. *Journal of the American Veterinary Medical Association*, *251*(5), 544–551. https://doi.org/10.2460/javma.251.5.544

Benito, J., Monteiro, B. P., Beaudry, F., Lavoie, A. M., Lascelles, B. D. X., & Steagall, P. V. (2016). Pharmacokinetics of bupivacaine after intraperitoneal administration to cats undergoing ovariohysterectomy. *American Journal of Veterinary Research*, *77*(6), 641–645.

Benito, J., Monteiro, B. P., Lavoie, A. M., Beauchamp, G., Lascelles, B. D. X., & Steagall, P. V. (2016). Analgesic efficacy of intraperitoneal administration of bupivacaine in cats. *Journal of*

Feline Medicine and Surgery, *18*(11), 906–912. https://doi.org/doi: 10.1177/1098612X15610162

Bennett, V., Gourkow, N., & Mills, D. S. (2017). Facial correlates of emotional behaviour in the domestic cat (Felis catus). *Behavioural Processes*, *141*, 342–350. https://doi.org/10.1016/j.beproc.2017.03.011

Benson, G. J., Wheaton, L. G., Thurmon, J. C., Tranquilli, W. J., Olson, W. A., & Davis, C. A. (1991). Postoperative catecholamine response to onychectomy in isoflurane-anesthetized cats. Effects of analgesics. *Veterinary Surgery : VS*, *20*(3), 222–225. https://doi.org/10.1111/j.1532-950x.1991.tb00340.x

Bernstein, P. L. (2006). Behavior of Single Cats and Groups in the Home. In J. R. August (Ed.), *Consultations in Feline Internal Medicine* (5th ed., pp. 675–685). Saunders. https://doi.org/10.1016/B0-72-160423-4/50074-3

Berthier, M., Starkstein, S., & Leiguarda, R. (1987). Behavioral Effects of Damage to the Right Insula and Surrounding Regions. *Cortex*, *23*(4), 673–678. https://doi.org/10.1016/S0010-9452(87)80057-6

Beswick, A., Dewey, C., Johnson, R., Dowsett-Cooper, J., & Niel, L. (2016). Survey of Ontario veterinarians' knowledge and attitudes on pain in dogs and cats in 2012. *Canadian Veterinary Journal*, *57*(12), 1274–1280.

Bijsmans, E. S., Jepson, R. E., Syme, H. M., Elliott, J., & Niessen, S. J. M. (2016). Psychometric Validation of a General Health Quality of Life Tool for Cats Used to Compare Healthy Cats and Cats with Chronic Kidney Disease. *Journal of Veterinary Internal Medicine*, *30*(1), 183–191. https://doi.org/10.1111/jvim.13656

Blanchard, D. C., & Blanchard, R. J. (1972). Innate and conditioned reactions to threat in rats with amygdaloid lesions. *Journal of Comparative and Physiological Psychology*, *81*(2), 281–290. https://doi.org/10.1037/h0033521

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, *1*(8476), 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*(2), 135–160. https://doi.org/10.1191/096228099673819272

Bland, J. M., & Altman, D. G. (2007). Agreement between methods of measurement with multiple observations per individual. *Journal of Pharmaceutical Statistics*, *17*(4), 571–582. https://doi.org/10.1080/10543400701329422

Boerner, K. E., Chambers, C. T., Gahagan, J., Keogh, E., Fillingim, R. B., & Mogil, J. S. (2018). Conceptual complexity of gender and its relevance to pain. *Pain*, *159*(11), 2137–2141. https://doi.org/10.1097/j.pain.0000000000001275

Borsook, D., & Becerra, L. (2011). CNS animal fMRI in pain and analgesia. In *Neuroscience and Biobehavioral Reviews* (Vol. 35, Issue 5, pp. 1125–1143). Neurosci Biobehav Rev. https://doi.org/10.1016/j.neubiorev.2010.11.005

Bowen, J., & Heath, S. (2005). *Behaviour Problems in Small Animals: Practical advice for the Veterinary Team*. Elsevier Health Sciences. https://books.google.ca/books?id=yX7OCgAAQBAJ&dq=leyhausen

Brennan, T. J. (1999). Postoperative Models of Nociception. *ILAR Journal*, *40*(3), 129–136. https://doi.org/10.1093/ilar.40.3.129

Brondani, J. T., Luna, S. P. L., Beier, S. L., Minto, B. W., & Padovani, C. R. (2009). Analgesic efficacy of perioperative use of vedaprofen, tramadol or their combination in cats undergoing ovariohysterectomy. *Journal of Feline Medicine and Surgery*, *11*(6), 420–429. https://doi.org/10.1016/j.jfms.2008.10.002

Brondani, J. T., Luna, S. P. L., Crosignani, N., Redondo, J. I., Granados, M. M., Bustamante, H., Palacios, C., & Otero, P. (2014). Validez y confiabilidad de la versión en español de la escala multidimensional de la UNESP-Botucatu para evaluar el dolor postoperatorio en gatos. *Archivos de Medicina Veterinaria*, *46*(3), 477–486. https://doi.org/10.4067/S0301-732X2014000300020

Brondani, J. T., Luna, S. P. L., Minto, B. W., Santos, B. P. R., Beier, S. L., Matsubara, L. M., & Padovani, C. R. (2012). Validade e responsividade de uma escala multidimensional para avaliação de dor pós-operatória em gatos. *Arquivo Brasileiro de Medicina Veterinária e*

Zootecnia, *64*(6), 1529–1538. https://doi.org/10.1590/S0102-09352012000600019

Brondani, J. T., Luna, S. P. L., & Padovani, C. R. (2011). Refinement and initial validation of a multidimensional composite scale for use in assessing acute postoperative pain in cats. *American Journal of Veterinary Research*, *72*(2), 174–183. https://doi.org/10.2460/ajvr.72.2.174

Brondani, J. T., Mama, K. R., Luna, S. P. L., Wright, B. D., Niyom, S., Ambrosio, J., Vogel, P. R., & Padovani, C. R. (2013). Validation of the English version of the UNESP-Botucatu multidimensional composite pain scale for assessing postoperative pain in cats. *BMC Veterinary Research*, *9*(1), 143. https://doi.org/10.1186/1746-6148-9-143

Broom, D. M. (2001). The evolution of pain. *Vlaams Diergeneeskunding Tijdschrift*, *70*, 17–21.

Buisman, M., Hasiuk, M. M. M., Gunn, M., & Pang, D. S. J. (2017). The influence of demeanor on scores from two validated feline pain assessment scales during the perioperative period. *Veterinary Anaesthesia and Analgesia*, 1–10. https://doi.org/10.1016/j.vaa.2016.09.001

Buisman, M., Wagner, M. C., Hasiuk, M. M., Prebble, M., Law, L., & Pang, D. S. (2016). Effects of ketamine and alfaxalone on application of a feline pain assessment scale. *Journal of Feline Medicine and Surgery*, *18*(8), 643–651. https://doi.org/10.1177/1098612X15591590

Caeiro, C. C., Burrows, A. M., & Waller, B. M. (2017). Development and application of CatFACS: Are human cat adopters influenced by cat facial expressions? *Applied Animal Behaviour Science*, *189*, 66–78. https://doi.org/10.1016/j.applanim.2017.01.005

Caeiro, C. C., Waller, B. M., Zimmermann, E., Burrows, A. M., & Davila-Ross, M. (2013). OrangFACS: A Muscle-Based Facial Movement Coding System for Orangutans (Pongo spp.). *International Journal of Primatology*, *34*(1), 115–129. https://doi.org/10.1007/s10764-012-9652-x

Calvo, G., Holden, E., Reid, J., Scott, E. M., Firth, A., Bell, A., Robertson, S., & Nolan, A. M. (2014). Development of a behaviour-based intervention level for assessing acute pain in cats. *Journal of Small Animal Practice*, *55*, 622–629. https://doi.org/10.1111/jsap.12280

Cambridge, A. J., Tobias, K. M., Newberry, R. C., & Sarkar, D. K. (2000). Subjective and objective measurements of postoperative pain in cats. *Journal of the American Veterinary Medical*

Association, *217*(5), 685–690. https://doi.org/10.2460/javma.2000.217.685

Case, J. B., Boscan, P. L., Monnet, E. L., Niyom, S. R., Imhoff, D. J., Wallace, M. L., & Smeak, D. D. (2015). Comparison of surgical variables and pain in cats undergoing ovariohysterectomy, laparoscopic-assisted ovariohysterectomy, and laparoscopic ovariectomy. *Journal of the American Animal Hospital Association*, *51*(1), 1–7. https://doi.org/10.5326/JAAHA-MS-5886

Catbagan, D. L., Quimby, J. M., Mama, K. R., Rychel, J. K., & Mich, P. M. (2011). Comparison of the efficacy and adverse effects of sustained-release buprenorphine hydrochloride following subcutaneous administration and buprenorphine hydrochloride following oral transmucosal administration in cats undergoing ovariohysterectomy. *American Journal of Veterinary Research*, *72*(4), 461–466. https://doi.org/10.2460/ajvr.72.4.461

Chartier, L. C., Hebart, M. L., Howarth, G. S., Whittaker, A. L., & Mashtoub, S. (2020). Affective state determination in a mouse model of colitis-associated colorectal cancer. *PLOS ONE*, *15*(1), e0228413. https://doi.org/10.1371/journal.pone.0228413

Christov-Moore, L., Simpson, E. A., Coudé, G., Grigaityte, K., Iacoboni, M., & Ferrari, P. F. (2014). Empathy: gender effects in brain and behavior. *Neuroscience and Biobehavioral Reviews*, *46 Pt 4*(Pt 4), 604–627. https://doi.org/10.1016/j.neubiorev.2014.09.001

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Cloutier, S., Newberry, R. C., Cambridge, A. J., & Tobias, K. M. (2005). Behavioural signs of postoperative pain in cats following onychectomy or tenectomy surgery. *Applied Animal Behaviour Science*, *92*(4), 325–335. https://doi.org/10.1016/J.APPLANIM.2004.11.009

Coleman, D. L., & Slingsby, L. S. (2007). Attitudes of veterinary nurses to the assessment of pain and the use of pain scales. *Veterinary Record*, *160*(16), 541–544. https://doi.org/10.1136/vr.160.16.541

Coll, M. P., Budell, L., Rainville, P., Decety, J., & Jackson, P. L. (2012). The role of gender in the interaction between self-pain and the perception of pain in others. *Journal of Pain*, *13*(7), 695–703. https://doi.org/10.1016/j.jpain.2012.04.009

Coneglian, M. M., Borges, T. D., Weber, S. H., Bertagnon, H. G., & Michelotto, P. V. (2020). Use of the horse grimace scale to identify and quantify pain due to dental disorders in horses. *Applied Animal Behaviour Science*, *225*, 104970. https://doi.org/10.1016/j.applanim.2020.104970

Cottingham, J. (2013). *René Descartes: Meditations on first philosophy: With selections from the objections and replies: A Latin-English Edition* (J. Cottingham (ed.)). Cambridge University Press.

Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in Neuroscience*, *2*, 263–275.

Cruickshank, A. M., Fraser, W. D., Burns, H. J., Van Damme, J., & Shenkin, A. (1990). Response of serum interleukin-6 in patients undergoing elective surgery of varying severity. *Clinical Science (London, England : 1979)*, *79*(2), 161–165. https://doi.org/10.1042/cs0790161

Da Silva, J. T., & Seminowicz, D. A. (2019). Neuroimaging of pain in animal models. *PAIN Reports*, *4*(4), e732. https://doi.org/10.1097/pr9.0000000000000732

Dalla Costa, E., Bracci, D., Dai, F., Lebelt, D., & Minero, M. (2017). Do Different Emotional States Affect the Horse Grimace Scale Score? A Pilot Study. *Journal of Equine Veterinary Science*, *54*, 114–117. https://doi.org/10.1016/j.jevs.2017.03.221

Dalla Costa, E., Minero, M., Lebelt, D., Stucke, D., Canali, E., & Leach, M. C. (2014). Development of the Horse Grimace Scale (HGS) as a pain assessment tool in horses undergoing routine castration. *PLoS ONE*, *9*(3), e92281. https://doi.org/10.1371/journal.pone.0092281

Dalla Costa, E., Pascuzzo, R., Leach, M. C., Dai, F., Lebelt, D., Vantini, S., & Minero, M. (2018). Can grimace scales estimate the pain status in horses and mice? A statistical approach to identify a classifier. *PLoS ONE*, *13*(8). https://doi.org/10.1371/journal.pone.0200339

Dalla Costa, E., Stucke, D., Dai, F., Minero, M., Leach, M. C., & Lebelt, D. (2016). Using the horse grimace scale (HGS) to assess pain associated with acute laminitis in horses (Equus caballus). *Animals*, *6*(8). https://doi.org/10.3390/ani6080047

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. J. Murray.

Dawkins, M. S. (2000). Animal Minds and Animal Emotions. *American Zoologist*, *40*(6), 883–888.

Dawson, L. C., Cheal, J., Niel, L., & Mason, G. (2019). Humans can identify cats' affective states from subtle facial expressions. *Animal Welfare*, *28*(4), 519–531. https://doi.org/10.7120/09627286.28.4.519

Dawson, L. C., Dewey, C. E., Stone, E. A., Mosley, C. I., Guerin, M. T., & Niel, L. (2017). Evaluation of a welfare assessment tool to examine practices for preventing, recognizing, and managing pain at companion-animal veterinary clinics. *Canadian Journal of Veterinary Research*, *81*(4), 270–279.

De Rantere, D., Schuster, C. J., Reimer, J. N., & Pang, D. S. J. (2016). The relationship between the Rat Grimace Scale and mechanical hypersensitivity testing in three experimental pain models. *European Journal of Pain*, *20*(3), 417–426. https://doi.org/10.1002/ejp.742

Della Rocca, G., Catanzaro, A., Conti, M. B., Bufalari, A., De Monte, V., Di Salvo, A., Tabarelli Brondani, J., & Pacca Loureiro Luna, S. (2018). Validation of the Italian version of the UNESP-Botucatu multidimensional composite pain scale for the assessment of postoperative pain in cats. *Veterinaria Italiana*, *54*(1), 49–61. https://doi.org/10.12834/VetIt.567.2704.22

Della Rocca, G., Di Salvo, A., Marenzoni, M. L., Bellezza, E., Pastorino, G., Monteiro, B. P., & Steagall, P. V. (2019). Development, Preliminary Validation, and Refinement of the Composite Oral and Maxillofacial Pain Scale-Canine/Feline (COPS-C/F). *Frontiers in Veterinary Science*, *6*, 274. https://doi.org/10.3389/fvets.2019.00274

Descovich, K. A., Wathan, J., Leach, M. C., Buchanan-smith, H. M., Flecknell, P., Farningham, D., & Vick, S. (2017). Facial expression : An under-utilised tool for the assessment of welfare in mammals. *ALTEX*, 1–21. https://doi.org/10.14573/altex.1607161

Di Giminiani, P., Brierley, V. L. M. H., Scollo, A., Gottardo, F., Malcolm, E. M., Edwards, S. A., & Leach, M. C. (2016). The Assessment of Facial Expressions in Piglets Undergoing Tail Docking and Castration: Toward the Development of the Piglet Grimace Scale. *Frontiers in Veterinary Science*, *3*, 100. https://doi.org/10.3389/fvets.2016.00100

Diep, T. N., Monteiro, B. P., Evangelista, M. C., Balleydier, A., Watanabe, R., Ruel, H. L. M., Doodnaught, G. M., Le Quang, T., & Steagall, P. V. (2020). Anesthetic and analgesic effects of an opioid-free, injectable protocol in cats undergoing ovariohysterectomy: A prospective, blinded, randomized clinical trial. *Canadian Veterinary Journal*, *61*(6), 621–628.

Diogo, R., Wood, B. A., Aziz, M. A., & Burrows, A. (2009). On the origin, homologies and evolution of primate facial muscles, with a particular focus on hominoids and a suggested unifying nomenclature for the facial muscles of the Mammalia. *Journal of Anatomy*, *215*(3), 300–319. https://doi.org/10.1111/j.1469-7580.2009.01111.x

Dixon, M. J., Robertson, S. A., & Taylor, P. M. (2002). A thermal threshold testing device for evaluation of analgesics in cats. *Research in Veterinary Science*, *72*(3), 205–210. https://doi.org/10.1053/rvsc.2001.0543

Dixon, M. J., Taylor, P. M., Steagall, P. V., Brondani, J. T., & Luna, S. P. L. (2007). Development of a pressure nociceptive threshold testing device for evaluation of analgesics in cats. *Research in Veterinary Science*, *82*(1), 85–92. https://doi.org/10.1016/j.rvsc.2006.03.010

Dobromylskyj, P., Flecknell, P. A., Lascelles, B. D., Livingston, A., Taylor, P., & Waterman-Pearson, A. (2000). Pain Assessment. In P. A. Flecknell & A. Waterman-Pearson (Eds.), *Pain Management in Animals* (pp. 53–79). W.B. Saunders. https://doi.org/https://doi.org/10.1016/B978-0-7020-1767-4.X5001-2

Dohoo, S. E., & Dohoo, I. R. (1996a). Factors influencing the postoperative use of analgesics in dogs and cats by Canadian veterinarians. *Canadian Veterinary Journal*, *37*, 552–556.

Dohoo, S. E., & Dohoo, I. R. (1996b). Perioperative use of analgesics in dogs and cats by Canadian veterinarians. *Canadian Veterinary Journal*, *37*, 546–551.

Dohoo, S. E., & Dohoo, I. R. (1998). Attitudes and concerns of Canadian animal health technologists toward postoperative pain management in dogs and cats. *Canadian Veterinary Journal*, *39*(8), 491–496.

Dolensek, N., Gehrlach, D. A., Klein, A. S., & Gogolla, N. (2020). Facial expressions of emotion states and their neuronal correlates in mice. *Science*, *368*(6486), 89–94. https://doi.org/10.1126/science.aaz9468

Doodnaught, G. M., Benito, J., Monteiro, B. P., Beauchamp, G., Grasso, S. C., & Steagall, P. V. (2017). Agreement among undergraduate and graduate veterinary students and veterinary anesthesiologists on pain assessment in cats and dogs: A preliminary study. *Canadian Veterinary Journal*, *58*(8), 805–808.

Driscoll, C. A., Clutton-Brock, J., Kitchener, A. C., & O'Brien, S. J. (2009). The Taming of the cat. *Scientific American*, *300*(6), 68–75. http://www.ncbi.nlm.nih.gov/pubmed/19485091

Driscoll, C. A., MacDonald, D. W., & O'Brien, S. J. (2009). From Wild Animals to Domestic Pets, an Evolutionary View of Domestication. In J. C. Avise & F. J. Ayala (Eds.), *In the Light of Evolution: Volume III: Two Centuries of Darwin*. National Academy of Sciences Press (US).

Dubin, A. E., & Patapoutian, A. (2010). Nociceptors: the sensors of the pain pathway. *The Journal of Clinical Investigation*, *120*(11), 3760–3772. https://doi.org/10.1172/JCI42843

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, *48*(4), 384–392. https://doi.org/10.1037/0003-066X.48.4.384

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124–129.

Ekman, P., Friesen, W. V., & Hager, J. (2002). *Facial Action Coding System*. Research Nexus.

Ekman, P., & Friesen, W. (1978). *Facial Action Coding System: A technique for the measurement of facial movement*. Consulting Psychologists Press.

Enomoto, M., Lascelles, B. D. X., & Gruen, M. E. (2020). Development of a checklist for the detection of degenerative joint disease-associated pain in cats. *Journal of Feline Medicine and Surgery*. https://doi.org/10.1177/1098612X20907424

Epstein, M. E., Rodanm, I., Griffenhagen, G., Kadrlik, J., Petty, M. C., Robertson, S. A., & Simpson, W. (2015). 2015 AAHA/AAFP Pain Management Guidelines for Dogs and Cats. *Journal of Feline Medicine and Surgery*, *17*(3), 251–272. https://doi.org/10.1177/1098612X15572062

Evangelista, M. C., Benito, J., Monteiro, B. P., Watanabe, R., Doodnaught, G. M., Pang, D. S. J., & Steagall, P. V. (2020). Clinical applicability of the Feline Grimace Scale: Real-time versus image scoring and the influence of sedation and surgery. *PeerJ*, *8*, e8967.

https://doi.org/10.7717/peerj.8967

Evangelista, M. C., Monteiro, B. P., & Steagall, P. V. (2020). Systematic review on the measurement properties of grimace scales for pain asessment - preliminary results. *In: Association of Veterinary Anaesthetists Spring Meeting. Dublin, Ireland*, March 11-13th.

Evangelista, M. C., Silva, R. A., Cardozo, L. B., Kahvegian, M. A. P., Rossetto, T. C., Matera, J. M., & Fantoni, D. T. (2014). Comparison of preoperative tramadol and pethidine on postoperative pain in cats undergoing ovariohysterectomy. *BMC Veterinary Research*, *10*, 252–260. https://doi.org/10.1186/s12917-014-0252-1

Evangelista, M. C., Watanabe, R., Leung, V. S. Y., Monteiro, B. P., O'Toole, E., Pang, D. S. J., & Steagall, P. V. (2019). Facial expressions of pain in cats: the development and validation of a Feline Grimace Scale. *Scientific Reports*, *9*, 19128. https://doi.org/10.1038/s41598-019-55693-8

Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Brooks/Cole Pub Co.

Faller, K. M. E., Mcandrew, D. J., Schneider, J. E., & Lygate, C. A. (2015). Refinement of analgesia following thoracotomy and experimental myocardial infarction using the Mouse Grimace Scale. *Experimental Physiology*, *100*(2), 164–172. https://doi.org/10.1113/expphysiol.2014.083139

Feinstein, J. S., Adolphs, R., Damasio, A., & Tranel, D. (2011). The human amygdala and the induction and experience of fear. *Current Biology*, *21*(1), 34–38. https://doi.org/10.1016/j.cub.2010.11.042

Feldt, K. S., Warne, M. A., & Ryden, M. B. (1998). Examining pain in aggressive cognitively impaired older adults. *Journal of Gerontological Nursing*, *24*(11), 14–22. https://doi.org/10.3928/0098-9134-19981101-07

Ferreira, T. H., Rezende, M. L., Mama, K. R., Hudachek, S. F., & Aguiar, A. J. a. (2011). Plasma concentrations and behavioral antinociceptive, and physiologic effects of methadone after intravenous and oral transmucosal administration in cats. *American Journal of Veterinary Research*, *72*(6), 764–771. https://doi.org/10.2460/ajvr.72.6.764

Fillingim, R. B. (2017). Sex, Gender, and Pain. In *Principles of Gender-Specific Medicine: Gender* *in the Genomic Era: Third Edition*. Elsevier Inc. https://doi.org/10.1016/B978-0-12-803506-1.00038-3

Fillingim, R. B., Loeser, J. D., Baron, R., & Edwards, R. R. (2016). Assessment of Chronic Pain: Domains, Methods, and Mechanisms. *The Journal of Pain : Official Journal of the American Pain Society*, *17*(9 Suppl), T10-20. https://doi.org/10.1016/j.jpain.2015.08.010

Finestone, H. M., Alfeeli, A., & Fisher, W. A. (2008). Stress-induced physiologic changes as a basis for the biopsychosocial model of chronic musculoskeletal pain: a new theory? *The Clinical Journal of Pain*, *24*(9), 767–775. https://doi.org/10.1097/AJP.0b013e3181790342

Finka, L. R., Luna, S. P., Brondani, J. T., Tzimiropoulos, Y., McDonagh, J., Farnworth, M. J., Ruta, M., & Mills, D. S. (2019). Geometric morphometrics for the study of facial expressions in non-human animals, using the domestic cat as an exemplar. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-46330-5

Finlayson, K., Lampe, J. F., Hintze, S., Würbel, H., & Melotti, L. (2016). Facial indicators of positive emotions in rats. *PLoS ONE*, *11*(11), e0166446. https://doi.org/10.1371/journal.pone.0166446

Firth, A. M., & Haldane, S. L. (1999). Development of a scale to evaluate postoperative pain in dogs. *Journal of the American Veterinary Medical Association*, *214*(5), 651–659. https://pubmed.ncbi.nlm.nih.gov/10088012/

Fischer, J. E., Bachmann, L. M., & Jaeschke, R. (2003). A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. *Intensive Care Medicine*, *29*(7), 1043–1051. https://doi.org/10.1007/s00134-003-1761-8

Fox, E. (2008). *Emotion science : cognitive and neuroscientific approaches to understanding human emotions*. Palgrave Macmillan.

Fox, S. M., Mellor, D. J., Lawoko, C. R. O., Hodge, H., & Firth, E. C. (1998). Changes in plasma cortisol concentrations in bitches in response to different combinations of halothane and butorphanol, with or without ovariohysterectomy. *Research in Veterinary Science*, *65*(2), 125–133. https://doi.org/10.1016/S0034-5288(98)90163-1

Fox, S. M., Mellor, D. J., Stafford, K. J., Lowoko, C. R. O., & Hodge, H. (2000). The effects of ovariohysterectomy plus different combinations of halothane anaesthesia and butorphanol analgesia on behaviour in the bitch. *Research in Veterinary Science*, *68*(3), 265–274. https://doi.org/10.1053/rvsc.2000.0375

Freeman, L. M., Rodenberg, C., Narayanan, A., Olding, J., Gooding, M. A., & Koochaki, P. E. (2016). Development and initial validation of the Cat HEalth and Wellbeing (CHEW) Questionnaire: a generic health-related quality of life instrument for cats. *Journal of Feline Medicine and Surgery*, *18*(9), 689–701. https://doi.org/10.1177/1098612X16657386

Freeman, L. M., Rush, J. E., Oyama, M. A., MacDonald, K. A., Cunningham, S. M., Bulmer, B., MacGregor, J. M., Laste, N. J., Malakoff, R. L., Hall, D. J., & Trafny, D. J. (2012). Development and evaluation of a questionnaire for assessment of health-related quality of life in cats with cardiac disease. *Journal of the American Veterinary Medical Association*, *240*(10), 1188–1193. https://doi.org/10.2460/javma.240.10.1188

Frith, C. (2009). Role of facial expressions in social interactions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1535), 3453–3458. https://doi.org/10.1098/rstb.2009.0142

Gleerup, K. B., Andersen, P. H., Munksgaard, L., & Forkman, B. (2015). Pain evaluation in dairy cattle. *Applied Animal Behaviour Science*, *171*, 25–32. https://doi.org/10.1016/j.applanim.2015.08.023

Gleerup, K. B., Forkman, B., Lindegaard, C., & Andersen, P. H. (2015). An equine pain face. *Veterinary Anaesthesia and Analgesia*, *42*(1), 103–114. https://doi.org/10.1111/vaa.12212

Glerum, L. E., Egger, C. M., Allen, S. W., & Haag, M. (2001). Analgesic effect of the transdermal fentanyl patch during and after feline ovariohysterectomy. *Veterinary Surgery*, *30*(4), ajvet0300351. https://doi.org/10.1053/jvet.2001.24387

Gogolla, N. (2017). The insular cortex. In *Current Biology* (Vol. 27, Issue 12, pp. R580–R586). Cell Press. https://doi.org/10.1016/j.cub.2017.05.010

Gonzalez, M. I., Field, M. J., Bramwell, S., McCleary, S., & Singh, L. (2000). Ovariohysterectomy in the rat: a model of surgical pain for evaluation of pre-emptive analgesia? *Pain*, *88*(1), 79–88. https://doi.org/10.1016/S0304-3959(00)00309-2

Gracely, R. H., & Kwilosz, D. M. (1988). The Descriptor Differential Scale: applying psychophysical principles to clinical pain assessment. *Pain*, *35*(3), 279–288. https://doi.org/10.1016/0304-3959(88)90138-8

Green, A. D., Tripp, D. A., Sullivan, M. J. L., & Davidson, M. (2009). The relationship between empathy and estimates of observed pain. *Pain Medicine*, *10*(2), 381–392. https://doi.org/10.1111/j.1526-4637.2009.00563.x

Greenspan, J. D., Craft, R. M., LeResche, L., Arendt-Nielsen, L., Berkley, K. J., Fillingim, R. B., Gold, M. S., Holdcroft, A., Lautenbacher, S., Mayer, E. A., Mogil, J. S., Murphy, A. Z., & Traub, R. J. (2007). Studying sex and gender differences in pain and analgesia: A consensus report. In *Pain* (Vol. 132, Issue SUPPL. 1, pp. S26-45). https://doi.org/10.1016/j.pain.2007.10.014

Gregory, N. S., Harris, A. L., Robinson, C. R., Dougherty, P. M., Fuchs, P. N., & Sluka, K. A. (2013). An overview of animal models of pain: Disease models and outcome measures. *Journal of Pain*, *14*(11), 1255–1269. https://doi.org/10.1016/j.jpain.2013.06.008

Grichnik, K. P., & Ferrante, F. M. (1991). The difference between acute and chronic pain. *Mount Sinai Journal of Medicine*, *58*(3), 217–220.

Griffin, B., & Hume, K. R. (2006). Recognition and Management of Stress in Housed Cats. In J. R. August (Ed.), *Consultations in Feline Internal Medicine* (5th ed., pp. 717–734). Saunders. https://doi.org/10.1016/B0-72-160423-4/50079-2

Gruen, M. E., Messenger, K. M., Thomson, A. E., Griffith, E. H., Aldrich, L. A., Vaden, S., & Lascelles, B. D. X. (2017). Evaluation of serum cytokines in cats with and without degenerative joint disease and associated pain. *Veterinary Immunology and Immunopathology*, *183*, 49–59. https://doi.org/10.1016/j.vetimm.2016.12.007

Guesgen, M. J., Beausoleil, N. J., Leach, M. C., Minot, E. O., Stewart, M., & Stafford, K. J. (2016). Coding and quantification of a facial expression for pain in lambs. *Behavioural Processes*, *132*, 49–56. https://doi.org/10.1016/j.beproc.2016.09.010

Guillot, M., Moreau, M., Heit, M., Martel-Pelletier, J., Pelletier, J. P., & Troncy, E. (2013). Characterization of osteoarthritis in cats and meloxicam efficacy using objective chronic pain evaluation tools. *Veterinary Journal*, *196*(3), 360–367.

https://doi.org/10.1016/j.tvjl.2013.01.009

Hadjistvropoulos, T., LaChapelle, D., MacLeod, F., Hale, C., O'Rourke, N., & Craig, K. D. (1998). Cognitive Functioning and Pain Reactions in Hospitalized Elders. *Pain Research and Management*, *3*(3), 145–151. https://doi.org/10.1155/1998/621580

Hager, C., Biernot, S., Buettner, M., Glage, S., Keubler, L. M., Held, N., Bleich, E. M., Otto, K., Muller, C. W., Decker, S., Talbot, S. R., & Bleich, A. (2017). The Sheep Grimace Scale as an indicator of post-operative distress and pain in laboratory sheep. *PLoS ONE*, *12*(4), e0175839. https://doi.org/10.1371/journal.pone.0175839

Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34. https://doi.org/10.20982/tqmp.08.1.p023

Han, S., Fan, Y., & Mao, L. (2008). Gender difference in empathy for pain: An electrophysiological investigation. *Brain Research*, *1196*, 85–93. https://doi.org/10.1016/j.brainres.2007.12.062

Hangaard, M. H., Malling, B., & Mogensen, C. B. (2018). High inter-observer agreement of observer-perceived pain assessment in the emergency department. *BMC Emergency Medicine*, *18*(1), 1–5. https://doi.org/10.1186/s12873-018-0159-4

Hansen, B. (1997). Through a glass darkly: using behavior to assess pain. *Seminars in Veterinary Medicine and Surgery (Small Animal)*, *12*(2). https://doi.org/10.1016/S1096-2867(97)80003-5

Hardie, E. M., Hansen, B. D., & Carroll, G. S. (1997). Behavior after ovariohysterectomy in the dog: What's normal? *Applied Animal Behaviour Science*, *51*(1–2), 111–128. https://doi.org/10.1016/S0168-1591(96)01078-7

Hatch, R. C., Zahner, J. M., & Booth, N. H. (1984). Meperidine-acepromazine-pentobarbital anesthesia in cats: reversal by 4-aminopyridine and yohimbine. *American Journal of Veterinary Research*, *45*(12), 2658–2662.

Hellyer, P. W., & Gaynor, J. S. (1998). Acute post-surgical pain in dogs and cats. *Compendium of Continuing Education for the Practicing Veterinarian (Small Animals)*, *20*, 140–153.

Hellyer, P. W., Uhrig, S. R., & Robinson, N. G. (2006). *Feline acute pain scale*. Colorado State

University Veterinary Medical Center. http://csu-cvmbs.colostate.edu/Documents/anesthesia-pain-management-pain-score-feline.pdf

Herr, K., Coyne, P. J., McCaffery, M., Manworren, R., & Merkel, S. (2011). Pain Assessment in the Patient Unable to Self-Report: Position Statement with Clinical Practice Recommendations. *Pain Management Nursing*, *12*(4), 230–250. https://doi.org/10.1016/j.pmn.2011.10.002

Hewson, C. J., Dohoo, I. R., & Lemke, K. a. (2006a). Perioperative use of analgesics in dogs and cats by Canadian veterinarians in 2001. *Canadian Veterinary Journal*, *47*(4), 352–359.

Hewson, C. J., Dohoo, I. R., & Lemke, K. A. (2006b). Factors affecting the use of postincisional analgesics in dogs and cats by Canadian veterinarians in 2001. *Canadian Veterinary Journal*, *47*(5), 453–459.

Holden, E., Calvo, G., Collins, M., Bell, A., Reid, J., Scott, E. M., & Nolan, A. M. (2014). Evaluation of facial expression in acute pain in cats. *Journal of Small Animal Practice*, *55*(12), 615–621. https://doi.org/10.1111/jsap.12283

Holton, L. L., Scott, E. M., Nolan, A. M., Reid, J., & Welsh, E. (1998). Relationship between physiological factors and clinical pain in dogs scored using a numerical rating scale. *The Journal of Small Animal Practice*, *39*(10), 469–474. http://www.ncbi.nlm.nih.gov/pubmed/9816569

Holton, L. L., Scott, E. M., Nolan, A. M., Reid, J., Welsh, E., & Flaherty, D. (1998). Comparison of three methods used for assessment of pain in dogs. *Journal of the American Veterinary Medical Association*, *212*(1), 61–66. https://pubmed.ncbi.nlm.nih.gov/9426779/

Hudson, J. T., Slater, M. R., Taylor, L., Scott, H. M., & Kerwin, S. C. (2004). Assessing repeatability and validity of a visual analogue scale questionnaire for use in assessing pain and lameness in dogs. *American Journal of Veterinary Research*, *65*(12), 1634–1643. https://doi.org/10.2460/ajvr.2004.65.1634

Hugonnard, M., Leblond, A., Keroack, S., Cadoré, J., & Troncy, E. (2004). Attitudes and concerns of French veterinarians towards pain and analgesia in dogs and cats. *Veterinary Anaesthesia and Analgesia*, *31*(3), 154–163. https://doi.org/10.1111/J.1467-2987.2004.00175.X

Hui, D., & Bruera, E. (2014). A personalized approach to assessing and managing pain in patients with cancer. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, *32*(16), 1640–1646. https://doi.org/10.1200/JCO.2013.52.2508

Hunt, J. R., Grint, N. J., Taylor, P. M., & Murrell, J. C. (2013). Sedative and analgesic effects of buprenorphine, combined with either acepromazine or dexmedetomidine, for premedication prior to elective surgery in cats and dogs. *Veterinary Anaesthesia and Analgesia*, *40*(3), 297–307. https://doi.org/10.1111/vaa.12003

Hunt, J. R., Knowles, T. G., Lascelles, B. D. X. X., & Murrell, J. C. (2015). Prescription of perioperative analgesics by UK small animal veterinary surgeons in 2013. *Veterinary Record*, *176*(19), 493–493. https://doi.org/10.1136/vr.102834

Jensen, M. P., Chen, C., & Brugger, A. M. (2003). Interpretation of visual analog scale ratings and change scores: A reanalysis of two clinical trials of postoperative pain. *Journal of Pain*, *4*(7), 407–414. https://doi.org/10.1016/S1526-5900(03)00716-8

Jensen, M. P., & Karoly, P. (1992). Self-report scales and procedures for assessing pain in adults. In D. C. Turk & R. Melzack (Eds.), *Handbook of Pain Assessment* (pp. 135–151). Guilford Press.

Kamata, M., King, J. N., Seewald, W., Sakakibara, N., Yamashita, K., & Nishimura, R. (2012). Comparison of injectable robenacoxib versus meloxicam for peri-operative use in cats: Results of a randomised clinical trial. *The Veterinary Journal*, *193*(1), 114–118. https://doi.org/10.1016/J.TVJL.2011.11.026

Kaplan, R. M., & Saccuzzo, D. P. (2010). *Psychological Testing: Principles, Application and Issues* (8th ed.). Wadsworth, Cengage Learning.

Kappesser, J., Williams, A. C. C., & Prkachin, K. M. (2006). Testing two accounts of pain underestimation. *Pain*, *124*(1), 109–116. https://doi.org/10.1016/j.pain.2006.04.003

Katz, J., & Melzack, R. (1999). Measurement of pain. *Surgical Clinics of North America*, *7*(2), 231–252. https://doi.org/10.1016/s0039-6109(05)70381-9

Keating, S. C. J., Thomas, A. A., Flecknell, P. A., & Leach, M. C. (2012). Evaluation of EMLA Cream for Preventing Pain during Tattooing of Rabbits: Changes in Physiological, Behavioural and Facial Expression Responses. *PLoS ONE*, *7*(9), e44437. https://doi.org/10.1371/journal.pone.0044437

Keefe, F. J., & Dunsmoret, J. (1992). Pain behavior concepts and controversies. *APS Journal*, *1*(2), 92–100. https://doi.org/10.1016/1058-9139(92)90035-B

Keefe, F. J., & Smith, S. (2002). The assessment of pain behavior: Implications for applied psychophysiology and future research directions. *Applied Psychophysiology Biofeedback*, *27*(2), 117–127. https://doi.org/10.1023/A:1016240126437

Keogh, E. (2014). Gender differences in the nonverbal communication of pain: A new direction for sex, gender, and pain research? *Pain*, *155*, 1927–1931. https://doi.org/10.1016/j.pain.2014.06.024

Keszei, A. P., Novak, M., & Streiner, D. L. (2010). Introduction to health measurement scales. *Journal of Psychosomatic Research*, *68*(4), 319–323. https://doi.org/10.1016/j.jpsychores.2010.01.006

Klinck, M. P., Gruen, M. E., del Castillo, J. R. E., Guillot, M., Thomson, A. E., Heit, M., Lascelles, B. D. X., & Troncy, E. (2018). Development and preliminary validity and reliability of the montreal instrument for cat arthritis testing, for use by caretaker/owner, MI-CAT(C), via a randomised clinical trial. *Applied Animal Behaviour Science*, *200*, 96–105. https://doi.org/10.1016/J.APPLANIM.2017.11.013

Klinck, M. P., Mogil, J. S., Moreau, M., Lascelles, B. D. X., Flecknell, P. A., Poitte, T., & Troncy, E. (2017). Translational pain assessment: Could natural animal models be the missing link? *Pain*, *158*(9), 1633–1646. https://doi.org/10.1097/j.pain.0000000000000978

Klinck, M. P., Monteiro, B. P., Lussier, B., Guillot, M., Moreau, M., Otis, C., Steagall, P. V. M., Frank, D., Martel-Pelletier, J., Pelletier, J.-P. P., del Castillo, J. R. R. E., & Troncy, E. (2018). Refinement of the Montreal Instrument for Cat Arthritis Testing, for Use by Veterinarians: detection of naturally occurring osteoarthritis in laboratory cats. *Journal of Feline Medicine and Surgery*, *20*(8), 728–740. https://doi.org/10.1177/1098612X17730172

Klinck, M. P., Rialland, P., Guillot, M., Moreau, M., Frank, D., & Troncy, E. (2015). Preliminary validation and reliability testing of the montreal instrument for cat arthritis testing, for use by veterinarians, in a colony of laboratory cats. *Animals*, *5*(4), 1252–1267.

https://doi.org/10.3390/ani5040410

Klune, C. B., Larkin, A. E., Leung, V. S. Y., & Pang, D. S. J. (2019). Comparing the Rat Grimace Scale and a composite behaviour score in rats. *PLoS ONE*, *14*(5), e0209467. https://doi.org/10.1371/journal.pone.0209467

Kongara, K., Squance, H., Topham, I., & Bridges, J. (2016). Attitudes and perceptions of veterinary paraprofessionals in New Zealand to postoperative pain in dogs and cats. *New Zealand Veterinary Journal*, *64*(2), 112–116. https://doi.org/10.1080/00480169.2015.1111172

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Krechel, S. W., & Bildner, J. (1995). CRIES: a new neonatal postoperative pain measurement score. Initial testing of validity and reliability. *Paediatric Anaesthesia*, *5*(1), 53–61. https://doi.org/10.1111/j.1460-9592.1995.tb00242.x

Kunz, M., Chen, J. I., Lautenbacher, S., Vachon-Presseau, E., & Rainville, P. (2011). Cerebral regulation of facial expressions of Pain. *Journal of Neuroscience*, *31*(24), 8730–8738. https://doi.org/10.1523/JNEUROSCI.0217-11.2011

Kunzel, W., Breit, S., & Oppel, M. (2003). Morphometric Investigations of Breed-Specific Features in Feline Skulls and Considerations on their Functional Implications. *Anatomia, Histologia, Embryologia: Journal of Veterinary Medicine Series C*, *32*(4), 218–223. https://doi.org/10.1046/j.1439-0264.2003.00448.x

Kyles, A. E., Hardie, E. M., Hansen, B. D., & Papich, M. G. (1998). Comparison of transdermal fentanyl and intramuscular oxymorphone on post-operative behaviour after ovariohysterectomy in dogs. *Research in Veterinary Science*, *65*(3), 245–251. https://doi.org/10.1016/S0034-5288(98)90151-5

Langford, D. J., Bailey, A. L., Chanda, M. L., Clarke, S. E., Drummond, T. E., Echols, S., Glick, S., Ingrao, J., Klassen-Ross, T., Lacroix-Fralish, M. L., Matsumiya, L., Sorge, R. E., Sotocinal, S. G., Tabaka, J. M., Wong, D., van den Maagdenberg, A. M., Ferrari, M. D., Craig, K. D., & Mogil, J. S. (2010). Coding of facial expressions of pain in the laboratory mouse. *Nat Methods*, *7*(6), 447–449. https://doi.org/10.1038/nmeth.1455

Lascelles, B. D. X., Brown, D. C., Maixner, W., & Mogil, J. S. (2017). Spontaneous painful disease in companion animals can facilitate the development of chronic pain therapies for humans. *Osteoarthritis and Cartilage*, *26*, 175–183. https://doi.org/10.1016/j.joca.2017.11.011

Lascelles, B. D. X., Hansen, B. D., Roe, S., DePuy, V., Thomson, A., Pierce, C. C., Smith, E. S., & Rowinski, E. (2007). Evaluation of client-specific outcome measures and activity monitoring to measure pain relief in cats with osteoarthritis. *Journal of Veterinary Internal Medicine*, *21*(3). https://doi.org/10.1892/0891-6640(2007)21[410:EOCOMA]2.0.CO;2

Lascelles, B. D. X., & Robertson, S. A. (2004). Use of thermal threshold response to evaluate the antinociceptive effects of butorphanol in cats. *American Journal of Veterinary Research*, *65*(8), 1085–1089. https://doi.org/10.2460/ajvr.2004.65.1085

Le Bars, D., Gozariu, M., & Cadden, S. W. (2001). Animal models of nociception. *Pharmacological Reviews*, *53*(4), 597–652.

Leach, M. C., Allweiler, S., Richardson, C., Roughan, J. V., Narbe, R., & Flecknell, P. A. (2009). Behavioural effects of ovariohysterectomy and oral administration of meloxicam in laboratory housed rabbits. *Research in Veterinary Science*, *87*(2), 336–347. https://doi.org/10.1016/J.RVSC.2009.02.001

Leach, M. C., Klaus, K., Miller, A. L., Scotto di Perrotolo, M., Sotocinal, S. G., & Flecknell, P. A. (2012). The Assessment of Post-Vasectomy Pain in Mice Using Behaviour and the Mouse Grimace Scale. *PLoS ONE*, *7*(4), e35656. https://doi.org/10.1371/journal.pone.0035656

Lehr, T., Narbe, R., Jons, O., Kloft, C., & Staab, A. (2009). Population pharmacokinetic modelling and simulation of single and multiple dose administration of meloxicam in cats. *Journal of Veterinary Pharmacology and Therapeutics*, *33*(3), 277–286. https://doi.org/10.1111/j.1365-2885.2009.01134.x

Leung, V. S. Y., Benoit-Biancamano, M. O., & Pang, D. S. J. (2019). Performance of behavioral assays: The Rat Grimace Scale, burrowing activity and a composite behavior score to identify visceral pain in an acute and chronic colitis model. *Pain Reports*, *4*(2), e712. https://doi.org/10.1097/PR9.0000000000000712

Leung, V. S. Y., Zhang, E. Q., & Pang, D. S. J. (2016). Real-time application of the Rat Grimace Scale as a welfare refinement in laboratory rats. *Scientific Reports*, *6*, 31667.

https://doi.org/10.1038/srep31667

Leyhausen, P. (1979). *Cat behavior : the predatory and social behavior of domestic and wild cats*. Garland STPM Press.

Li, D., Li, Z., Luo, R., Deng, J., & Sun, S. (2019). Multi-Pose Facial Expression Recognition Based on Generative Adversarial Network. *IEEE Access*, *7*, 143980–143989. https://doi.org/10.1109/ACCESS.2019.2945423

Liao, L., Long, H., Zhang, L., Chen, H., Zhou, Y., Ye, N., & Lai, W. (2014). Evaluation of pain in rats through facial expression following experimental tooth movement. *European Journal of Oral Sciences*, *122*(2), 121–124. https://doi.org/10.1111/eos.12110

Logothetis, N. K. (2002). The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *357*(1424), 1003–1037. https://doi.org/10.1098/rstb.2002.1114

Lorena, S. E., Luna, S. P., Lascelles, B. D. X., & Corrente, J. E. (2014). Current attitudes regarding the use of perioperative analgesics in dogs and cats by Brazilian veterinarians. *Veterinary Anaesthesia and Analgesia*, *41*(1), 82–89. https://doi.org/10.1111/vaa.12104

Low, P., Edelman, D., & Koch, C. (2012). *The Cambridge Declaration on Consciousness Publicly proclaimed in Cambridge, UK, on July 7th, 2012, at the Francis Crick Memorial Conference on Consciousness in Human and non-Human Animals*.

Luna, S. P. L., Trindade, P. H. E., Della Rocca, G., Kronen, P. W., Crosignani, N., Monteiro, B. P., Ruel, H. L. M., Tseng, J., Yamashita, K., Teixeira, L. G., Brondani, J. T., & Steagall, P. V. (2020). Content validity, inter and intra-rater reliability of the novel short form UNESP-Botucatu cat pain scale (UCAPS) in eight languages. *In: American Association of Feline Practitioners - Online Conference*, October 3-4th.

Lynch, S., Savary-Bataille, K., Leeuw, B., & Argyle, D. J. (2011). Development of a questionnaire assessing health-related quality-of-life in dogs and cats with cancer. *Veterinary and Comparative Oncology*, *9*(3), 172–182. https://doi.org/10.1111/j.1476-5829.2010.00244.x

Macdonald, D. W., Yamaguchi, N., & Kerby, G. (2000). Group-living in the domestic cat: its sociobiology and epidemiology. In D. C. Turner & P. Bateson (Eds.), *The Domestic Cat* (2nd ed., pp. 95–118). Cambridge University Press.

MacRae, A. M., Joanna Makowska, I., & Fraser, D. (2018). Initial evaluation of facial expressions and behaviours of harbour seal pups (Phoca vitulina) in response to tagging and microchipping. *Applied Animal Behaviour Science*, *205*(April), 167–174. https://doi.org/10.1016/j.applanim.2018.05.001

Mansour, C., Merlin, T., Bonnet-Garin, J. M., Chaaya, R., Mocci, R., Ruiz, C. C., Allaouchiche, B., Boselli, E., & Junot, S. (2017). Evaluation of the Parasympathetic Tone Activity (PTA) index to assess the analgesia/nociception balance in anaesthetised dogs. *Research in Veterinary Science*, *115*. https://doi.org/10.1016/J.RVSC.2017.05.009

Mantha, S., Roizen, M. F., Fleisher, L. A., Thisted, R., & Foss, J. (2000). Comparing methods of clinical measurement: Reporting standards for bland and altman analysis. *Anesthesia and Analgesia*, *90*(3), 593–602. https://doi.org/10.1097/00000539-200003000-00018

Martin, P., & Bateson, P. (2007). Think before you measure. In P. Martin & P. Bateson (Eds.), *Measuring Behaviour: An Introductory Guide* (3rd ed., pp. 10–24). Cambridge University Press. https://doi.org/10.1017/CBO9780511810893.003

Martin, R. M. (2019). Influence of biological sex, trait gender, and state gender on pain threshold, pain tolerance, and ratings of pain severity. *Personality and Individual Differences*, *138*(July 2018), 183–187. https://doi.org/10.1016/j.paid.2018.09.038

Martinez-Arizala, A. (2003). Methods to measure sensory function in humans versus animals. *Journal of Rehabilitation Research and Development*, *40*(4 SUPPL. 1), S35–S39. https://doi.org/10.1682/JRRD.2003.08.0035

Martinez, A. M., & Kak, A. C. (2001). PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(2), 228–233. https://doi.org/10.1109/34.908974

Mathews, K. A. (2000). Pain assessment and general approach to management. *The Veterinary Clinics of North America. Small Animal Practice*, *30*(4), 729–755, v. https://doi.org/10.1016/s0195-5616(08)70004-4

Mathews, K. A., Kronen, P. W., Lascelles, D., Nolan, A., Robertson, S. A., Steagall, P. V. M., Wright, B., & Yamashita, K. (2014). Guidelines for Recognition, Assessment and Treatment of Pain. *Journal of Small Animal Practice*, *55*(6), E10–E68.

Matsumiya, L. C., Sorge, R. E., Sotocinal, S. G., Tabaka, J. M., Wieskopf, J. S., Zaloum, A., King, O. D., & Mogil, J. S. (2012). Using the Mouse Grimace Scale to reevaluate the efficacy of postoperative analgesics in laboratory mice. *Journal of the American Association for Laboratory Animal Science : JAALAS*, *51*(1), 42–49.

McCaffery, M. (1968). *Nursing practice theories related to cognition, bodily pain, and man-environment interactions*. University of California.

McComb, K., Taylor, A. M., Wilson, C., & Charlton, B. D. (2009). The cry embedded within the purr. *Current Biology : CB*, *19*(13), R507-8. https://doi.org/10.1016/j.cub.2009.05.033

McCorry, L. K. (2007). Physiology of the autonomic nervous system. *American Journal of Pharmaceutical Education*, *71*(4), 78. https://doi.org/10.5688/aj710478

McDowell, I. (2006). *Measuring Health: A guide to rating scales and questionnaires* (3rd ed.). Oxford University Press.

McLennan, K. M., & Mahmoud, M. (2019). Development of an Automated Pain Facial Expression Detection System for Sheep (Ovis Aries). *Animals*, *9*(4), 196. https://doi.org/10.3390/ani9040196

McLennan, K. M., Miller, A. L., Dalla Costa, E., Stucke, D., Corke, M. J., Broom, D. M., & Leach, M. C. (2019). Conceptual and methodological issues relating to pain assessment in mammals: The development and utilisation of pain facial expression scales. *Applied Animal Behaviour Science*, *217*, 1–15. https://doi.org/10.1016/j.applanim.2019.06.001

McLennan, K. M., Rebelo, C. J. B., Corke, M. J., Holmes, M. A., Leach, M. C., & Constantino-Casas, F. (2016). Development of a facial expression scale using footrot and mastitis as models of pain in sheep. *Applied Animal Behaviour Science*, *176*, 19–26. https://doi.org/10.1016/j.applanim.2016.01.007

Melzack, R., & Casey, K. L. (1968). Sensory, motivational and central control determinants of pain: A new conceptual model. In D. Kenshalo (Ed.), *The Skin Senses* (pp. 423–443). Thomas.

Melzack, R., & Wall, P. D. (1965). Pain mechanisms: A new theory. *Science*, *150*(3699), 971–979. https://doi.org/10.1126/science.150.3699.971

Merola, I., & Mills, D. S. (2016a). Behavioural Signs of Pain in Cats: An Expert Consensus. *PloS One*, *11*(2), e0150040. https://doi.org/10.1371/journal.pone.0150040

Merola, I., & Mills, D. S. (2016b). Systematic review of the behavioural assessment of pain in cats. *Journal of Feline Medicine and Surgery*, *18*(2), 60–76. https://doi.org/10.1177/1098612X15578725

Mich, P. M., Hellyer, P. W., Kogan, L., & Schoenfeld-Tacher, R. (2010). Effects of a pilot training program on veterinary students' pain knowledge, attitude, and assessment skills. *Journal of Veterinary Medical Education*, *37*(4), 358–368. https://doi.org/10.3138/jvme.37.4.358

Miller, A. L., Golledge, H. D. R., & Leach, M. C. (2016). The Influence of Isoflurane Anaesthesia on the Rat Grimace Scale. *PloS One*, *11*(11), e0166652. https://doi.org/10.1371/journal.pone.0166652

Miller, A. L., Kitson, G., Skalkoyannis, B., & Leach, M. C. (2015). The effect of isoflurane anaesthesia and buprenorphine on the mouse grimace scale and behaviour in CBA and DBA/2 mice. *Applied Animal Behaviour Science*, *172*, 58–62. https://doi.org/10.1016/j.applanim.2015.08.038

Miller, A. L., & Leach, M. C. (2015a). The mouse grimace scale: A clinically useful tool? *PLoS ONE*, *10*(9), e0136000. https://doi.org/10.1371/journal.pone.0136000

Miller, A. L., & Leach, M. C. (2015b). Using the mouse grimace scale to assess pain associated with routine ear notching and the effect of analgesia in laboratory mice. *Laboratory Animals*, *49*(2), 117–120. https://doi.org/10.1177/0023677214559084

Miller, A. L., & Leach, M. C. (2016). The effect of handling method on the mouse grimace scale in two strains of laboratory mice. *Laboratory Animals*, *50*(4), 305–307. https://doi.org/10.1177/0023677215622144

Millette, V. M., Steagall, P. V. M., Duke-Novakovski, T., & Livingston, A. J. (2008). Effects of meperidine or saline on thermal, mechanical and electrical nociceptive thresholds in cats. *Veterinary Anaesthesia and Analgesia*, *35*(6), 543–547.

Miron-Shatz, T., Ormianer, M., Rabinowitz, J., Hanoch, Y., & Tsafrir, A. (2020). Physician experience is associated with greater underestimation of patient pain. *Patient Education and*

Counseling, *103*(2), 405–409. https://doi.org/10.1016/J.PEC.2019.08.040

Mittal, A., Gupta, M., Lamarre, Y., Jahagirdar, B., & Gupta, K. (2016). Quantification of pain in sickle mice using facial expressions and body measurements. *Blood Cells, Molecules, and Diseases*, *57*, 58–66. https://doi.org/10.1016/j.bcmd.2015.12.006

Mogil, J. S. (2020). Qualitative sex differences in pain processing: emerging evidence of a biased literature. *Nature Reviews Neuroscience*, *21*(7), 353–365. https://doi.org/10.1038/s41583-020-0310-6

Mogil, J. S., & Chanda, M. L. (2005). The case for the inclusion of female subjects in basic science studies of pain. *Pain*, *117*(1), 1–5. https://doi.org/10.1016/j.pain.2005.06.020

Mogil, J. S., Pang, D. S. J., Silva Dutra, G. G., & Chambers, C. T. (2020). The development and use of facial grimace scales for pain measurement in animals. *Neuroscience & Biobehavioral Reviews*, *116*, 480–493. https://doi.org/10.1016/j.neubiorev.2020.07.013

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & De Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, *63*(7), 737–745. https://doi.org/10.1016/j.jclinepi.2010.02.006

Mollenhoff, A., Nolte, I., & Kramer, S. (2005). Anti-nociceptive Efficacy of Carprofen, Levomethadone and Buprenorphine for Pain Relief in Cats following Major Orthopaedic Surgery. *Journal of Veterinary Medicine Series A*, *52*(4), 186–198. https://doi.org/10.1111/j.1439-0442.2005.00710.x

Monteiro, B. P., Klinck, M. P., Moreau, M., Guillot, M., Steagall, P. V. M., Pelletier, J. P., Martel-Pelletier, J., Gauvin, D., Del Castillo, J. R. E., & Troncy, E. (2017). Analgesic efficacy of tramadol in cats with naturally occurring osteoarthritis. *PloS One*, *12*(4), e0175565. https://doi.org/10.1371/journal.pone.0175565

Monteiro, B. P., Otis, C., del Castillo, J. R. E., Nitulescu, R., Brown, K., Arendt-Nielsen, L., & Troncy, E. (2020). Quantitative sensory testing in feline osteoarthritic pain – a systematic review and meta-analysis. *Osteoarthritis and Cartilage*, *28*(7), 885–896. https://doi.org/10.1016/J.JOCA.2020.04.006

Monteiro, B. P., & Steagall, P. V. (2019). Chronic pain in cats: Recent advances in clinical assessment. *Journal of Feline Medicine and Surgery*, *21*(7), 601–614. https://doi.org/10.1177/1098612X19856179

Moore, S., & Bowden, R. (2009). The effects of Pose on Facial Expression Recognition. In A. Cavallaro, A. Prince, & D. Alexander (Eds.), *Procedings of the British Machine Vision Conference 2009* (pp. 79.1-79.11). British Machine Vision Association Press. https://doi.org/10.5244/C.23.79

Morales-Vallecilla, C., Ramírez, N., Villar, D., Díaz, M. C., Bustamante, S., & Ferguson, D. (2019). Survey of pain knowledge and analgesia in dogs and cats by Colombian veterinarians. *Veterinary Sciences*, *6*, 6. https://doi.org/10.3390/VETSCI6010006

Moran, C. E., & Hofmeister, E. H. (2013). Prevalence of pain in a university veterinary intensive care unit. *Journal of Veterinary Emergency and Critical Care*, *23*(1), 29–36. https://doi.org/10.1111/vec.12010

Morey, D. F. (2010). Domestication of Dogs and Other Organisms. In D. F. Morey (Ed.), *Dogs: Domestication and the Development of a Social Bond* (pp. 57–85). Cambridge University Press. https://doi.org/10.1017/CBO9780511778360.006

Morley, S., & Pallin, V. (1995). Scaling the affective domain of pain: a study of the dimensionality of verbal descriptors. *Pain*, *62*(1), 39–49. https://doi.org/10.1016/0304-3959(94)00226-5

Moser, H. L., Abraham, A. C., Howell, K., Laudier, D., Zumstein, M. A., Galatz, L. M., & Huang, A. H. (2020). Cell lineage tracing and functional assessment of supraspinatus tendon healing in an acute repair murine model. *Journal of Orthopaedic Research*, jor.24769. https://doi.org/10.1002/jor.24769

Muir, W. W., & Gaynor, J. S. (2008). Pain Behaviors. In J. S. Gaynor & W. W. Muir (Eds.), *Handbook of Veterinary Pain Management* (2nd ed., p. 672). Mosby.

Muir, W. W., Wiese, A. J., & Wittum, T. E. (2004). Prevalence and characteristics of pain in dogs and cats examined as outpatients at a veterinary teaching hospital. *Journal of the American Veterinary Medical Association*, *224*(9), 1459–1463. https://doi.org/10.2460/javma.2004.224.1459

Müller, B. R., Soriano, V. S., Bellio, J. C. B., & Molento, C. F. M. (2019). Facial expression of pain in Nellore and crossbred beef cattle. *Journal of Veterinary Behavior*, *34*, 60–65. https://doi.org/10.1016/j.jveb.2019.07.007

Murtaugh, M. P., Baarsch, M. J., Zhou, Y., Scamurra, R. W., & Lin, G. (1996). Inflammatory cytokines in animal health and disease. *Veterinary Immunology and Immunopathology*, *54*(1–4), 45–55. https://doi.org/10.1016/S0165-2427(96)05698-X

Nagakubo, D., Hamamoto, Y., Hasegawa, D., Kamata, M., Iizuka, T., Muta, K., Fujita, N., Nakagawa, T., & Nishimura, R. (2017). Functional MRI-based identification of brain regions activated by mechanical noxious stimulation and modulatory effect of remifentanil in cats. *Research in Veterinary Science*, *114*, 444–449. https://doi.org/10.1016/j.rvsc.2017.07.024

Nagakura, Y., Miwa, M., Yoshida, M., Miura, R., Tanei, S., Tsuji, M., & Takeda, H. (2019). Spontaneous pain-associated facial expression and efficacy of clinically used drugs in the reserpine-induced rat model of fibromyalgia. *European Journal of Pharmacology*, *864*. https://doi.org/10.1016/j.ejphar.2019.172716

Nani, A., Manuello, J., Mancuso, L., Liloia, D., Costa, T., & Cauda, F. (2019). The Neural Correlates of Consciousness and Attention: Two Sister Processes of the Brain. *Frontiers in Neuroscience*, *13*, 1169. https://doi.org/10.3389/fnins.2019.01169

National Research Council. (2009). *Recognition and Alleviation of Pain in Laboratory Animals*. National Academies Press (US). https://doi.org/https://doi.org/10.17226/12526.

Nibblett, B. M., Ketzis, J. K., & Grigg, E. K. (2015). Comparison of stress exhibited by cats examined in a clinic versus a home setting. *Applied Animal Behaviour Science*, *173*, 68–75. https://doi.org/10.1016/J.APPLANIM.2014.10.005

Niessen, S. J. M., Powney, S., Guitian, J., Niessen, A. P. M., Pion, P. D., Shaw, J. A. M., & Church, D. B. (2010). Evaluation of a quality-of-life tool for cats with diabetes mellitus. *Journal of Veterinary Internal Medicine*, *24*(5), 1098–1105. https://doi.org/10.1111/j.1939-1676.2010.0579.x

Noble, C. E., Wiseman-Orr, L. M., Scott, M. E., Nolan, A. M., & Reid, J. (2019). Development, initial validation and reliability testing of a web-based, generic feline health-related quality-of-life instrument. *Journal of Feline Medicine and Surgery*, *21*(2), 84–94.

https://doi.org/10.1177/1098612X18758176

Noli, C., Borio, S., Varina, A., & Schievano, C. (2016). Development and validation of a questionnaire to evaluate the Quality of Life of cats with skin disease and their owners, and its use in 185 cats with skin disease. *Veterinary Dermatology*, *27*(4), 247-e58. https://doi.org/10.1111/vde.12341

Noor, A., Zhao, Y., Koubaa, A., Wu, L., Khan, R., & Abdalla, F. Y. O. (2020). Automated sheep facial expression classification using deep transfer learning. *Computers and Electronics in Agriculture*, *175*, 105528. https://doi.org/10.1016/j.compag.2020.105528

Norring, M., Wikman, I., Hokkanen, A. H., Kujala, M. V., & Hänninen, L. (2014). Empathic veterinarians score cattle pain higher. *Veterinary Journal*, *200*(1), 186–190. https://doi.org/10.1016/j.tvjl.2014.02.005

O'Halloran, C., McCulloch, L., Rentoul, L., Alexander, J., Hope, J. C., & Gunn-Moore, D. A. (2018). Cytokine and Chemokine Concentrations as Biomarkers of Feline Mycobacteriosis. *Scientific Reports*, *8*(1), 17314. https://doi.org/10.1038/s41598-018-35571-5

Oatley, K., Keltner, D., & Jenkins, J. M. (2006). *Understanding emotions* (2nd ed.). Balckwell Publishing. https://doi.org/10.1016/0191-6599(94)90151-1

Oliver, V., De Rantere, D., Ritchie, R., Chisholm, J., Hecker, K. G., & Pang, D. S. J. (2014). Psychometric assessment of the rat grimace scale and development of an analgesic intervention score. *PLoS ONE*, *9*(5), e97882. https://doi.org/10.1371/journal.pone.0097882

Omoigui, S. (2007). The biochemical origin of pain: the origin of all pain is inflammation and the inflammatory response. Part 2 of 3 - inflammatory profile of pain syndromes. *Medical Hypotheses*, *69*(6), 1169–1178. https://doi.org/10.1016/j.mehy.2007.06.033

Orth, E. K., Navas González, F. J., Iglesias Pastrana, C., Berger, J. M., Jeune, S. S. le, Davis, E. W., & McLean, A. K. (2020). Development of a Donkey Grimace Scale to Recognize Pain in Donkeys (Equus asinus) Post Castration. *Animals*, *10*(8), 1411. https://doi.org/10.3390/ani10081411

Overall, K. L., Rodan, I., Beaver, B. V., Carney, H., Crowell-Davis, S., Hird, N., Kudrak, S., & Wexler-Mitchel, H. (2005). Feline Behavior Guidelines from the American Association of

Feline Practitioners. *Journal of the American Veterinary Medical Association*, *227*(1), 70–84. https://doi.org/10.2460/javma.2005.227.70

Paltrinieri, S. (2008). The feline acute phase reaction. *Veterinary Journal (London, England : 1997)*, *177*(1), 26–35. https://doi.org/10.1016/j.tvjl.2007.06.005

Parkinson, B. (2005). Do facial movements express emotions or communicate motives? *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *9*(4), 278–311. https://doi.org/10.1207/s15327957pspr0904_1

Parr, L. A., Waller, B. M., Burrows, A. M., Gothard, K. M., & Vick, S. J. (2010). Brief communication: MaqFACS: A muscle-based facial movement coding system for the rhesus macaque. *American Journal of Physical Anthropology*, *143*(4), 625–630. https://doi.org/10.1002/ajpa.21401

Parys, M., Yuzbasiyan-Gurkan, V., & Kruger, J. M. (2018). Serum Cytokine Profiling in Cats with Acute Idiopathic Cystitis. *Journal of Veterinary Internal Medicine*, *32*(1), 274–279. https://doi.org/10.1111/jvim.15032

Philips, B. H., Weisshaar, C. L., & Winkelstein, B. A. (2017). Use of the rat grimace scale to evaluate neuropathic pain in a model of cervical radiculopathy. *Comparative Medicine*, *67*(1), 34–42.

Ploghaus, A., Narain, C., Beckmann, C. F., Clare, S., Bantick, S., Wise, R., Matthews, P. M., Nicholas P Rawlins, J., & Tracey, I. (2001). Exacerbation of pain by anxiety is associated with activity in a hippocampal network. *Journal of Neuroscience*, *21*(24), 9896–9903. https://doi.org/10.1523/jneurosci.21-24-09896.2001

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? critique and recommendations. *Research in Nursing & Health*, *29*(5), 489–497. https://doi.org/10.1002/nur.20147

Polson, S., Taylor, P. M., & Yates, D. (2012). Analgesia after feline ovariohysterectomy under midazolam-medetomidine-ketamine anaesthesia with buprenorphine or butorphanol, and carprofen or meloxicam: a prospective, randomised clinical trial. *Journal of Feline Medicine and Surgery*, *14*(8), 553–559. https://doi.org/10.1177/1098612X12444743

Ponterotto, J. G., & Ruckdeschel, D. E. (2007). An overview of coefficient alpha and a reliability matrix for estimating adequacy of internal consistency coefficients with psychological research measures. *Perceptual and Motor Skills*, *105*, 997–1014.

Prguda, E., & Neumann, D. L. (2014). Inter-human and animal-directed empathy: A test for evolutionary biases in empathetic responding. *Behavioural Processes*, *108*, 80–86. https://doi.org/10.1016/j.beproc.2014.09.012

Price, D. D. (2000). Psychological and neural mechanisms of the affective dimension of pain. *Science*, *288*(5472), 1769–1772. https://doi.org/10.1126/science.288.5472.1769

Price, J., Catriona, S., Welsh, E. M., & Waran, N. K. (2003). Preliminary evaluation of a behaviour-based system for assessment of post-operative pain in horses following arthroscopic surgery. *Veterinary Anaesthesia and Analgesia*, *30*(3), 124–137. https://doi.org/10.1046/j.1467-2995.2003.00139.x

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, *27*(5), 1147–1157. https://doi.org/10.1007/s11136-018-1798-3

Prkachin, K. M. (2009). Assessing pain by facial expression: Facial expression as nexus. *Pain Research and Management*, *14*(1), 53–58. https://doi.org/10.1155/2009/542964

Prkachin, K. M. (2011). Facial pain expression. *Pain Management*, *1*(4), 367–376.

Prkachin, K. M., Solomon, P., Hwang, T., & Mercer, S. R. (2001). Does experience influence judgments of pain behaviour? Evidence from relatives of pain patients and therapists. *Pain Research and Management*, *6*(2), 105–112. https://doi.org/10.1155/2001/108098

Quarterone, C., Luna, S. P. L., Crosignani, N., De Oliveira, F. A., Lopes, C., Da Maia Lima, A. F., & De Araujo Aguiar, A. J. (2017). Ovariohysterectomy requires more post-operative analgesia than orchiectomy in dogs and cats. *Canadian Veterinary Journal*, *58*(11), 1191–1194. /pmc/articles/PMC5640292/?report=abstract

Quimby, J. M., Smith, M. L., & Lunn, K. F. (2011). Evaluation of the effects of hospital visit stress on physiologic parameters in the cat. *Journal of Feline Medicine and Surgery*, *13*(10), 733–

737. https://doi.org/10.1016/j.jfms.2011.07.003

Raeburn, C. D., Sheppard, F., Barsness, K. A., Arya, J., & Harken, A. H. (2002). Cytokines for surgeons. *American Journal of Surgery*, *183*(3), 268–273. https://doi.org/10.1016/s0002-9610(02)00781-x

Raillard, M., Detotto, C., Grepper, S., Beslac, O., Fujioka-Kobayashi, M., Schaller, B., & Saulacic, N. (2019). Anaesthetic and perioperative management of 14 male new zealand white rabbits for calvarial bone surgery. *Animals*, *9*(11), 896. https://doi.org/10.3390/ani9110896

Raja, S. N., Carr, D. B., Cohen, M., Finnerup, N. B., Flor, H., Gibson, S., Keefe, F. J., Mogil, J. S., Ringkamp, M., Sluka, K. A., Song, X.-J., Stevens, B., Sullivan, M. D., Tutelman, P. R., Ushida, T., & Vader, K. (2020). The revised International Association for the Study of Pain definition of pain. *Pain*. https://doi.org/10.1097/j.pain.0000000000001939

Reed, W. R., Little, J. W., Lima, C. R., Sorge, R. E., Yarar-Fisher, C., Eraslan, M., Hurt, C. P., Ness, T. J., Gu, J. G., Martins, D. F., & Li, P. (2020). Spinal Mobilization Prevents NGF-Induced Trunk Mechanical Hyperalgesia and Attenuates Expression of CGRP. *Frontiers in Neuroscience*, *14*, 385. https://doi.org/10.3389/fnins.2020.00385

Reicherts, P., Wiemer, J., Gerdes, A. B. M., Schulz, S. M., Pauli, P., & Wieser, M. J. (2017). Anxious anticipation and pain: The influence of instructed vs conditioned threat on pain. *Social Cognitive and Affective Neuroscience*, *12*(4), 544–554. https://doi.org/10.1093/scan/nsw181

Reid, J., Nolan, A., Hughes, J. M. L., Lascelles, B. D. X., & Pawson, P. (2007). Development of the short-form Glasgow Composite Measure Pain Scale (CMPS-SF) and derivation of an analgesic intervention score. *Animal Welfare*, *16*, 97–104.

Reid, J., Nolan, A. M., & Scott, E. M. (2018). Measuring pain in dogs and cats using structured behavioural observation. *Veterinary Journal*, *236*, 72–79. https://doi.org/10.1016/j.tvjl.2018.04.013

Reid, J., Scott, E. M., Calvo, G., & Nolan, A. M. (2017). Definitive Glasgow acute pain scale for cats: validation and intervention level. *Veterinary Record*, *180*(18), 449. https://doi.org/10.1136/vr.104208

Reid, J., Scott, M., Nolan, A., & Wiseman-Orr, L. (2013). Pain assessment in animals. *In Practice*, *34*(34), 3–5. https://doi.org/10.1136/inp

Reid, J., Wiseman-Orr, M. L., Scott, E. M., & Nolan, A. M. (2013). Development, validation and reliability of a web-based questionnaire to measure health-related quality of life in dogs. *Journal of Small Animal Practice*, *54*(5), 227–233. https://doi.org/10.1111/jsap.12059

Reijgwart, M. L., Schoemaker, N. J., Pascuzzo, R., Leach, M. C., Stodel, M., De Nies, L., Hendriksen, C. F. M., Van Der Meer, M., Vinke, C. M., & Van Zeeland, Y. R. A. (2017). The composition and initial evaluation of a grimace scale in ferrets after surgical implantation of a telemetry probe. *PLoS ONE*, *12*(11), e0187986. https://doi.org/10.1371/journal.pone.0187986

Reimann, J., Dewey, C., Bateman, S. W., Kerr, C., & Johnson, R. (2017). Perioperative analgesic use by Ontario veterinarians, 2012. *Canadian Veterinary Journal*, *58*(2), 149–156.

Rietmann, T. R., Stauffacher, M., Bernasconi, P., Auer, J. A., & Weishaupt, M. A. (2004). The association between heart rate, heart rate variability, endocrine and behavioural pain measures in horses suffering from laminitis. *Journal of Veterinary Medicine. A, Physiology, Pathology, Clinical Medicine*, *51*(5). https://doi.org/10.1111/J.1439-0442.2004.00627.X

Robertson, S. A. (2015). Acute Pain and Behavior. In I. Rodan & S. Heath (Eds.), *Feline Behavioral Health and Welfare* (pp. 162–183). Saunders.

Robertson, S. A., Taylor, P. M., Lascelles, B. D. X., & Dixon, M. J. (2003). Changes in thermal threshold response in eight cats after administration of buprenorphine, butorphanol and morphine. *The Veterinary Record*, *153*(15), 462–465. https://doi.org/10.1136/vr.153.15.462

Robertson, S. A., Wegner, K., & Lascelles, B. D. X. (2009). Antinociceptive and side-effects of hydromorphone after subcutaneous administration in cats. *Journal of Feline Medicine and Surgery*, *11*(2), 76–81. https://doi.org/10.1016/j.jfms.2008.05.003

Rodan, I. (2010). Understanding Feline Behavior and Application for Appropriate Handling and Management. *Topics in Companion Animal Medicine*, *25*(4), 178–188. https://doi.org/10.1053/j.tcam.2010.09.001

Rossi, H. L., See, L. P., Foster, W., Pitake, S., Gibbs, J., Schmidt, B., Mitchell, C. H., & Abdus-

Saboor, I. (2020). Evoked and spontaneous pain assessment during tooth pulp injury. *Scientific Reports*, *10*(1), 1–12. https://doi.org/10.1038/s41598-020-59742-5

Roughan, J. V., Wright-Williams, S. L., & Flecknell, P. A. (2009). Automated analysis of postoperative behaviour: Assessment of homecagescan as a novel method to rapidly identify pain and analgesic effects in mice. *Laboratory Animals*, *43*(1), 17–26. https://doi.org/10.1258/la.2008.007156

Rousseau-Blass, F., O'Toole, E., Marcoux, J., & Pang, D. S. J. (2020). Prevalence and management of pain in dogs in the emergency service of a veterinary teaching hospital. *The Canadian Veterinary Journal = La Revue Veterinaire Canadienne*, *61*(3), 294–300. http://www.ncbi.nlm.nih.gov/pubmed/32165754

Sadeghiyeh, H., Khorrami, A., & Hatami, J. (2012). Gender differences in empathy for pain: A single-pulse transcranial magnetic stimulation study. *Neurophysiologie Clinique/Clinical Neurophysiology*, *42*(1–2), 62. https://doi.org/10.1016/j.neucli.2011.11.029

Samy, A., Elmetwally, M., & El-khodery, S. A. (2020). Antinociceptive Effect of Intravenous Regional Analgesia in Horses Underwent Selected Short-Time Distal Limb Surgeries. *Journal of Equine Veterinary Science*, *91*, 103113. https://doi.org/10.1016/j.jevs.2020.103113

Saritas, Z. K., Saritas, T. B., Pamuk, K., Korkmaz, M., Yaprakci, M. V, Yilmaz, O., & Demirkan, I. (2015). Evaluation of preemptive dexketoprofen trometamol effect on blood chemistry, vital signs and postoperative pain in dogs undergoing ovariohysterectomy. *Bratislavske Lekarske Listy*, *116*(3), 191–195. https://doi.org/10.4149/bll_2015_038

Schneider, L. E., Henley, K. Y., Turner, O. A., Pat, B., Niedzielko, T. L., & Floyd, C. L. (2017). Application of the Rat Grimace Scale as a Marker of Supraspinal Pain Sensation after Cervical Spinal Cord Injury. *Journal of Neurotrauma*, *34*(21), 2982–2993. https://doi.org/10.1089/neu.2016.4665

Shah, M., Yates, D., Hunt, J., & Murrell, J. (2019). Comparison between methadone and buprenorphine within the QUAD protocol for perioperative analgesia in cats undergoing ovariohysterectomy. *Journal of Feline Medicine and Surgery*, *21*(8), 723–731. https://doi.org/10.1177/1098612X18798840

Shansky, R. M., & Woolley, C. S. (2016). Considering sex as a biological variable will be valuable for neuroscience research. *Journal of Neuroscience*, *36*(47), 11817–11822. https://doi.org/10.1523/JNEUROSCI.1390-16.2016

Shariff, A. F., & Tracy, J. L. (2011). What Are Emotion Expressions For? *Current Directions in Psychological Science*, *20*(6), 395–399. https://doi.org/10.1177/0963721411424739

Sherwin, C. M. (2001). Can Invertebrates Suffer? Or, How Robust is Argument-by-Analogy? *Animal Welfare*, *10*(Suppl 1), 103–118.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Siao, P., & Cros, D. P. (2003). Quantitative sensory testing. *Physical Medicine and Rehabilitation Clinics of North America*, *14*(2), 261–286. https://doi.org/10.1016/s1047-9651(02)00122-5

Silva, P. G. de B., de Lima Martins, J. O., de Lima Praxedes Neto, R. A., Mota Lemos, J. V., Machado, L. C., Matos Carlos, A. C. A., Alves, A. P. N. N., & Lima, R. A. (2020). Tumor necrosis factor alpha mediates orofacial discomfort in an occlusal dental interference model in rats: The role of trigeminal ganglion inflammation. *Journal of Oral Pathology and Medicine*, *49*(2), 169–176. https://doi.org/10.1111/jop.12984

Simon, B. T., Scallan, E. M., Carroll, G., & Steagall, P. V. (2017). The lack of analgesic use (oligoanalgesia) in small animal practice. *Journal of Small Animal Practice*, *58*(10), 543–554. https://doi.org/10.1111/jsap.12717

Simon, B. T., Scallan, E. M., Von Pfeil, D. J. F., Boruta, D. T., Wall, R., Nibblett, B. M., Odette, O., Beauchamp, G., & Steagall, P. V. (2018). Perceptions and opinions of pet owners in the United States about surgery, pain management, and anesthesia in dogs and cats. *Veterinary Surgery*, *47*(2), 277–284. https://doi.org/10.1111/vsu.12753

Slingsby, L. S., Jones, A., & Waterman-Pearson, A. E. (2001). Use of a new finger-mounted device to compare mechanical nociceptive thresholds in cats given pethidine or no medication after castration. *Research in Veterinary Science*, *70*(3), 243–246. https://doi.org/10.1053/RVSC.2001.0467

Slingsby, L. S., Lane, E. C., Mears, E. R., Shanson, M. C., & Waterman-Pearson, A. E. (1998). Postoperative pain after ovariohysterectomy in the cat: a comparison of two anaesthetic regimens. *The Veterinary Record*, *143*(21), 589–590. https://doi.org/10.1136/vr.143.21.589

Smith, J. D., Allen, S. W., & Quandt, J. E. (1999). Changes in cortisol concentration in response to stress and postoperative pain in client-owned cats and correlation with objective clinical variables. *American Journal of Veterinary Research*, *60*(4), 432–436. http://www.ncbi.nlm.nih.gov/pubmed/10211685

Smith, J. D., Allen, S. W., Quandt, J. E., & Tackett, R. L. (1996). Indicators of postoperative pain in cats and correlation with clinical criteria. *American Journal of Veterinary Research*, *57*(11), 1674–1678. http://www.ncbi.nlm.nih.gov/pubmed/8915451

Sneddon, L. U. (2018). Comparative Physiology of Nociception and Pain. *Physiology*, *33*(1), 63–73. https://doi.org/10.1152/physiol.00022.2017

Sneddon, L. U., Elwood, R. W., Adamo, S. A., & Leach, M. C. (2014). Defining and assessing animal pain. *Animal Behaviour*, *97*, 201–212. https://doi.org/10.1016/j.anbehav.2014.09.007

Sorge, R. E., Martin, L. J., Isbester, K. A., Sotocinal, S. G., Rosen, S., Tuttle, A. H., Wieskopf, J. S., Acland, E. L., Dokova, A., Kadoura, B., Leger, P., Mapplebeck, J. C. S., McPhail, M., Delaney, A., Wigerblad, G., Schumann, A. P., Quinn, T., Frasnelli, J., Svensson, C. I., … Mogil, J. S. (2014). Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nature Methods*, *11*(6), 629–632. https://doi.org/10.1038/nmeth.2935

Sotocinal, S. G., Sorge, R. E., Zaloum, A., Tuttle, A. H., Martin, L. J., Wieskopf, J. S., Mapplebeck, J. C., Wei, P., Zhan, S., Zhang, S., McDougall, J. J., King, O. D., & Mogil, J. S. (2011). The Rat Grimace Scale: A partially automated method for quantifying pain in the laboratory rat via facial expressions. *Molecular Pain*, *7*(1), 55. https://doi.org/10.1186/1744-8069-7-55

Sperry, M. M., Yu, Y., Kartha, S., Ghimire, P., Welch, R. L., Winkelstein, B. A., & Granquist, E. J. (2020). Intra-articular etanercept attenuates pain and hypoxia from TMJ loading in the rat. *Journal of Orthopaedic Research*, *38*(6), 1316–1326. https://doi.org/10.1002/jor.24581

Srithunyarat, T., Hagman, R., Höglund, O. V., Stridsberg, M., Olsson, U., Hanson, J., Nonthakotr, C., Lagerstedt, A.-S., & Pettersson, A. (2017). Catestatin, vasostatin, cortisol, and pain assessments in dogs suffering from traumatic bone fractures. *BMC Research Notes*, *10*(1), 129. https://doi.org/10.1186/s13104-017-2450-y

Stanton, L. A., Sullivan, M. S., & Fazio, J. M. (2015). A standardized ethogram for the felidae: A tool for behavioral researchers. *Applied Animal Behaviour Science*, *173*, 3–16. https://doi.org/10.1016/J.APPLANIM.2015.04.001

Steagall, P. V., Benito, J., Monteiro, B. P., Doodnaught, G. M., Beauchamp, G., & Evangelista, M. C. (2018). Analgesic effects of gabapentin and buprenorphine in cats undergoing ovariohysterectomy using two pain-scoring systems: a randomized clinical trial. *Journal of Feline Medicine and Surgery*, *20*(8), 741–748. https://doi.org/10.1177/1098612X17730173

Steagall, P. V., Carnicelli, P., Taylor, P. M., Luna, S. P. L., Dixon, M., & Ferreira, T. H. (2006). Effects of subcutaneous methadone, morphine, buprenorphine or saline on thermal and pressure thresholds in cats. *Journal of Veterinary Pharmacology and Therapeutics*, *29*(6), 531–537. https://doi.org/10.1111/j.1365-2885.2006.00800.x

Steagall, P. V., Monteiro-Steagall, B. P., & Taylor, P. M. (2014). A review of the studies using buprenorphine in cats. *Journal of Veterinary Internal Medicine / American College of Veterinary Internal Medicine*, *28*(3), 762–770. https://doi.org/10.1111/jvim.12346

Steagall, P. V., & Monteiro, B. P. (2019). Acute pain in cats: Recent advances in clinical assessment. *Journal of Feline Medicine and Surgery*, *21*, 25–34. https://doi.org/10.1177/1098612X18808103

Steagall, P. V., Monteiro, B. P., Lavoie, A. M., Frank, D., Troncy, E., Luna, S. P. L., & Brondani, J. T. (2017). Validation de la version francophone d'une échelle composite multidimensionnelle pour l'évaluation de la douleur postopératoire chez les chats. *Canadian Veterinary Journal*, *58*(1), 56–64.

Steagall, P. V., Monteiro, B. P., Ruel, H. L. M., Beauchamp, G., Luca, G., Berry, J., Little, S., Stiles, E., Hamilton, S., & Pang, D. (2017). Perceptions and opinions of Canadian pet owners about anaesthesia, pain and surgery in small animals. *Journal of Small Animal Practice*, *58*(7), 380–388. https://doi.org/10.1111/jsap.12674

Steagall, P. V., Taylor, P. M., Rodrigues, L. C. C., Ferreira, T. H., Minto, B. W., & Aguiar, A. J. A. (2009). Analgesia for cats after ovariohysterectomy with either buprenorphine or carprofen alone or in combination. *The Veterinary Record*, *164*(12), 359–363.

Streiner, D. L. (1993). A checklist for evaluation the usefulness of rating scales. *Canadian Journal* *of Psychiatry*, *38*(2), 140–148. https://doi.org/10.1177/070674379303800214

Streiner, D. L., & Norman, G. R. (2008). *Heath Measurement Scales: A practical guide to their development and use* (4th ed.). Oxford University Press.

Susskind, J. M., & Anderson, A. K. (2008). Facial expression form and function. *Communicative & Integrative Biology*, *1*(2), 148–149. https://doi.org/10.4161/cib.1.2.6999

Tatlock, S., Gober, M., Williamson, N., & Arbuckle, R. (2017). Development and preliminary psychometric evaluation of an owner-completed measure of feline quality of life. *Veterinary Journal*, *228*, 22–32. https://doi.org/10.1016/j.tvjl.2017.10.005

Tavernier, C., Ahmed, S., Houpt, K. A., & Yeon, S. C. (2020). Feline vocal communication. *Journal of Veterinary Science*, *21*(1), e18. https://doi.org/10.4142/jvs.2020.21.e18

Taylor, P. M., Kirby, J. J., Robinson, C., Watkins, E. A., Clarke, D. D., Ford, M. A., & Church, K. E. (2010). A prospective multi-centre clinical trial to compare buprenorphine and butorphanol for postoperative analgesia in cats. *Journal of Feline Medicine and Surgery*, *12*(4), 247–255. https://doi.org/10.1016/j.jfms.2009.09.004

Tracey, I. (2008). Imaging pain. In *British Journal of Anaesthesia* (Vol. 101, Issue 1, pp. 32–39). Oxford University Press. https://doi.org/10.1093/bja/aen102

Tracey, I., & Mantyh, P. W. (2007). The Cerebral Signature for Pain Perception and Its Modulation. In *Neuron* (Vol. 55, Issue 3, pp. 377–391). Cell Press. https://doi.org/10.1016/j.neuron.2007.07.012

Troia, R., Mascalzoni, G., Agnoli, C., Lalonde-Paul, D., Giunti, M., & Goggs, R. (2020). Cytokine and Chemokine Profiling in Cats With Sepsis and Septic Shock. *Frontiers in Veterinary Science*, *7*, 305. https://doi.org/10.3389/fvets.2020.00305

Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. In *Annual Review of Neuroscience* (Vol. 31, pp. 411–437). Annu Rev Neurosci. https://doi.org/10.1146/annurev.neuro.30.051606.094238

Turk, D. C., & Okifuji, A. (2001). Pain Terms and Taxonomies of Pain. In J. D. Loeser (Ed.), *Bonica's Management of Pain* (3rd ed., pp. 17–25). Lippincott Williams & Wilkins.

Tuttle, A. H., Molinaro, M. J., Jethwa, J. F., Sotocinal, S. G., Prieto, J. C., Styner, M. A., Mogil, J.

S., & Zylka, M. J. (2018). A deep neural network to assess spontaneous pain from mouse facial expressions. *Molecular Pain*, *14*, 1744806918763658. https://doi.org/10.1177/1744806918763658

Ueoka, N., & Hikasa, Y. (2015). Effects in cats of atipamezole, flumazenil and 4-aminopyridine on stress-related neurohormonal and metabolic responses induced by medetomidine, midazolam and ketamine. *Journal of Feline Medicine and Surgery*, *17*(8), 711–718. https://doi.org/10.1177/1098612X14556557

Ulrich-Lai, Y. M., & Herman, J. P. (2009). Neural regulation of endocrine and autonomic stress responses. *Nature Reviews. Neuroscience*, *10*(6), 397–409. https://doi.org/10.1038/nrn2647

Väisänen, M. A. M., Tuomikoski-Alin, S. K., Brodbelt, D. C., & Vainio, O. M. (2008). Opinions of Finnish small animal owners about surgery and pain management in small animals: PAPER. *Journal of Small Animal Practice*, *49*(12), 626–632. https://doi.org/10.1111/j.1748-5827.2008.00626.x

Valros, A., & Hänninen, L. (2018). Animal ethical views and perception of animal pain in veterinary students. *Animals*, *8*(12). https://doi.org/10.3390/ani8120220

van Dierendonck, M. C., Burden, F. A., Rickards, K., & van Loon, J. P. A. M. (2020). Monitoring Acute Pain in Donkeys with the Equine Utrecht University Scale for Donkeys Composite Pain Assessment (EQUUS-DONKEY-COMPASS) and the Equine Utrecht University Scale for Donkey Facial Assessment of Pain (EQUUS-DONKEY-FAP). *Animals : An Open Access Journal from MDPI*, *10*(2). https://doi.org/10.3390/ani10020354

van Dierendonck, M. C., & van Loon, J. P. A. M. (2016). Monitoring acute equine visceral pain with the Equine Utrecht University Scale for Composite Pain Assessment (EQUUS-COMPASS) and the Equine Utrecht University Scale for Facial Assessment of Pain (EQUUS-FAP): A validation study. *Veterinary Journal*, *216*, 175–177. https://doi.org/10.1016/j.tvjl.2016.08.004

van Loon, J. P. A. M., & van Dierendonck, M. C. (2015). Monitoring acute equine visceral pain with the Equine Utrecht University Scale for Composite Pain Assessment (EQUUS-COMPASS) and the Equine Utrecht University Scale for Facial Assessment of Pain (EQUUS-FAP): A scale-construction study. *Veterinary Journal*, *206*, 356–364.

van Loon, J. P. A. M., & van Dierendonck, M. C. (2017). Monitoring equine head-related pain with the Equine Utrecht University scale for facial assessment of pain (EQUUS-FAP). *Veterinary Journal*, *220*(January), 88–90. https://doi.org/10.1016/j.tvjl.2017.01.006

van Loon, J. P. A. M., & van Dierendonck, M. C. (2019). Pain assessment in horses after orthopaedic surgery and with orthopaedic trauma. *Veterinary Journal*, *246*, 85–91. https://doi.org/10.1016/j.tvjl.2019.02.001

Vick, S. J., Waller, B. M., Parr, L. A., Pasqualini, M. C. S., & Bard, K. A. (2007). A cross-species comparison of facial morphology and movement in humans and chimpanzees using the Facial Action Coding System (FACS). *Journal of Nonverbal Behavior*, *31*(1), 1–20. https://doi.org/10.1007/s10919-006-0017-z

Viscardi, A. V., Hunniford, M., Lawlis, P., Leach, M. C., & Turner, P. V. (2017). Development of a Piglet Grimace Scale to Evaluate Piglet Pain Using Facial Expressions Following Castration and Tail Docking: A Pilot Study. *Frontiers in Veterinary Science*, *4*, 51. https://doi.org/10.3389/fvets.2017.00051

Viscardi, A. V., & Turner, P. V. (2018). Efficacy of buprenorphine for management of surgical castration pain in piglets. *BMC Veterinary Research*, *14*(1), 1–12. https://doi.org/10.1186/s12917-018-1643-5

von Korff, M., Ormel, J., Keefe, F. J., & Dworkin, S. F. (1992). Grading the severity of chronic pain. *Pain*, *50*, 133–149.

von Muggenthaler, E. (2001). The felid purr: A healing mechanism? *The Journal of the Acoustical Society of America*, *110*(5), 2666–2666. https://doi.org/10.1121/1.4777098

Vullo, C., Barbieri, S., Catone, G., Graïc, J. M., Magaletti, M., Di Rosa, A., Motta, A., Tremolada, C., Canali, E., & Costa, E. D. (2020). Is the Piglet Grimace Scale (PGS) a useful welfare indicator to assess pain after cryptorchidectomy in growing pigs? *Animals*, *10*(3). https://doi.org/10.3390/ani10030412

Vytal, K., & Hamann, S. (2010). Neuroimaging support for discrete neural correlates of basic emotions: A voxel-based meta-analysis. *Journal of Cognitive Neuroscience*, *22*(12), 2864–2885. https://doi.org/10.1162/jocn.2009.21366

Wall, P. D. (1978). The gate control theory of pain mechanisms: A re-examination and re-statement. *Brain*, *101*(1), 1–18. https://doi.org/10.1093/brain/101.1.1

Waller, B. M., Lembeck, M., Kuchenbuch, P., Burrows, A. M., & Liebal, K. (2012). GibbonFACS: A Muscle-Based Facial Movement Coding System for Hylobatids. *International Journal of Primatology*, *33*(4), 809–821. https://doi.org/10.1007/s10764-012-9611-6

Waller, B. M., & Micheletta, J. (2013). Facial expression in nonhuman animals. *Emotion Review*, *5*(1), 54–59. https://doi.org/10.1177/1754073912451503

Waller, B. M., Peirce, K., Caeiro, C. C., Scheider, L., Burrows, A. M., McCune, S., & Kaminski, J. (2013). Paedomorphic facial expressions give dogs a selective advantage. *PLoS ONE*, *8*(12), e82686. https://doi.org/10.1371/journal.pone.0082686

Waran, N., Best, L., Williams, V., Salinsky, J., Dale, A., & Clarke, N. (2007). A preliminary study of behaviour-based indicators of pain in cats. *Animal Welfare*, *16*(Suppl. 1), 105–108. https://www.ingentaconnect.com/content/ufaw/aw/2007/00000016/A00102s1/art00015

Watanabe, R., Doodnaught, G. M., Evangelista, M. C., Monteiro, B. P., Ruel, H. L. M., & Steagall, P. V. (2020). Inter-Rater Reliability of the Feline Grimace Scale in Cats Undergoing Dental Extractions. *Frontiers in Veterinary Science*, *7*, 302. https://doi.org/10.3389/fvets.2020.00302

Watanabe, R., Doodnaught, G., Proulx, C., Auger, J.-P., Monteiro, B., Dumais, Y., Beauchamp, G., Segura, M., & Steagall, P. (2019). A multidisciplinary study of pain in cats undergoing dental extractions: A prospective, blinded, clinical trial. *PLOS ONE*, *14*(3), e0213195. https://doi.org/10.1371/journal.pone.0213195

Watanabe, R., Frank, D., & Steagall, P. V. (2020). Pain behaviors before and after treatment of oral disease in cats using video assessment: a prospective, blinded, randomized clinical trial. *BMC Veterinary Research*, *16*(1), 100. https://doi.org/10.1186/s12917-020-02302-w

Wathan, J., Burrows, A. M., Waller, B. M., & McComb, K. (2015). EquiFACS: The equine facial action coding system. *PLoS ONE*, *10*(8), e0131738. https://doi.org/10.1371/journal.pone.0131738

White, W. B., Berson, A. S., Robbins, C., Jamieson, M. J., Prisant, L. M., Roccella, E., & Sheps, S. G. (1993). National standard for measurement of resting and ambulatory blood pressures with automated sphygmomanometers. *Hypertension*, *21*(4), 504–509. https://doi.org/10.1161/01.HYP.21.4.504

Wiese, A. J., Muir, W. W., & Wittum, T. E. (2005). Characteristics of pain and response to analgesic treatment in dogs and cats examined at a veterinary teaching hospital emergency service. *Journal of the American Veterinary Medical Association*, *226*(12), 2004–2009. https://doi.org/10.2460/javma.2005.226.2004

Williams, V. M., Lascelles, B. D. X., & Robson, M. C. (2005). Current attitudes to, and use of, peri-operative analgesia in dogs and cats by veterinarians in New Zealand. *New Zealand Veterinary Journal*, *53*(3), 193–202. https://doi.org/10.1080/00480169.2005.36504

Wiseman, M. L., Nolan, A. M., Reid, J., & Scott, E. M. (2001). Preliminary study on owner-reported behaviour changes associated with chronic pain in dogs. *The Veterinary Record*, *149*(14), 423–424. https://doi.org/10.1136/vr.149.14.423

Wolski, T. R. (1982). Social behavior of the cat. *The Veterinary Clinics of North America. Small Animal Practice*, *12*(4), 693–706. https://doi.org/10.1016/S0195-5616(82)50113-1

Woolf, C. J. (2010). What is this thing called pain? In *Journal of Clinical Investigation* (Vol. 120, Issue 11, pp. 3742–3744). American Society for Clinical Investigation. https://doi.org/10.1172/JCI45178

Yang, S., & Chang, M. C. (2019). Chronic pain: Structural and functional changes in brain structures and associated negative affective states. In *International Journal of Molecular Sciences* (Vol. 20, Issue 13). MDPI AG. https://doi.org/10.3390/ijms20133130

Zamprogno, H., Hansen, B. D., Bondell, H. D., Sumrell, A. T., Simpson, W., Robertson, I. D., Brown, J., Pease, A. P., Roe, S. C., Hardie, E. M., Wheeler, S. J., & Lascelles, D. X. (2010). Item generation and design testing of a questionnaire to assess degenerative joint disease-associated pain in cats. *American Journal of Veterinary Research*, *71*(12), 1417–1424. https://doi.org/10.2460/ajvr.71.12.1417

Zhang, E. Q., Leung, V. S. Y., & Pang, D. S. J. (2019). Influence of Rater Training on Inter- and Intrarater Reliability When Using the Rat Grimace Scale. *Journal of the American Association for Laboratory Animal Science*, *58*(2), 178–183. https://doi.org/10.30802/AALAS-JAALAS-18-000044

Zhou, D., Kusnecov, A. W., Shurin, M. R., DePaoli, M., & Rabin, B. S. (1993). Exposure to physical and psychological stressors elevates plasma interleukin 6: relationship to the activation of hypothalamic-pituitary-adrenal axis. *Endocrinology*, *133*(6), 2523–2530. https://doi.org/10.1210/endo.133.6.8243274

Appendices

This section presents the FGS training manual with pictures, drawing and instructions on how to use the scale and a study performed in our laboratory testing the FGS in cats undergoing dental extractions.

The FGS training manual was published as Supplementary material with the first article presented (Facial expressions of pain in cats: the development and validation of a feline grimace scale, by Evangelista et al., 2019). It illustrates how the FGS is used to score the cat images and also can be used as an education tool.

The reliability of the FGS was in another type of pain stimulus (dental extractions) and the effect of the caregiver presence on FGS scores, were assessed in the study: “Inter-rater Reliability of the Feline Grimace Scale in Cats Undergoing Dental Extractions” by Ryota Watanabe, Graeme M. Doodnaught, Marina C. Evangelista, Beatriz P. Monteiro, Hélène L. M. Ruel and Paulo V. Steagall; published in Frontiers in Veterinary Science (2020; 7: 302). This was an exploratory study that confirmed the good reliability of the FGS for assessment of oral pain in cats. It also demonstrated that the caregiver's presence did not affect the FGS scores.

Appendix A. FGS Training Manual

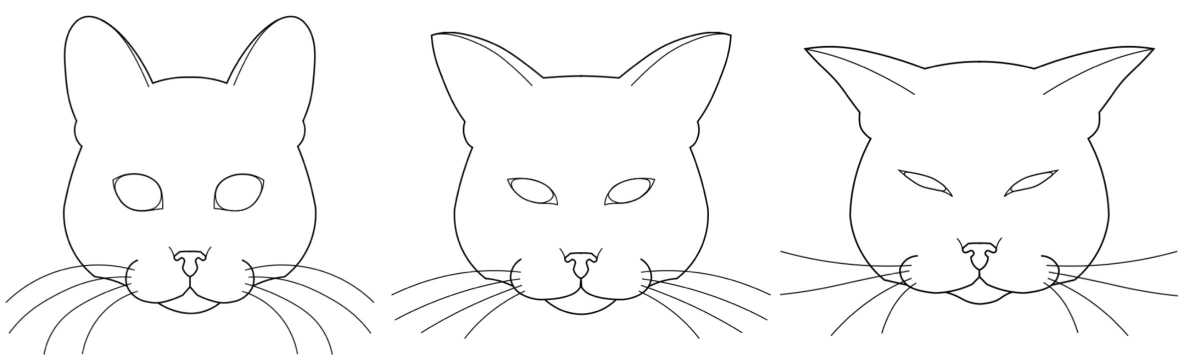
Supplementary material

Facial expressions of pain in cats: the development and validation of a Feline Grimace Scale

Marina C Evangelista, Ryota Watanabe, Vivian SY Leung, Beatriz Monteiro, Elizabeth O'Toole, Daniel SJ Pang, Paulo V Steagall

TRAINING MANUAL

FELINE GRIMACE SCALE



Instructions for using the scale

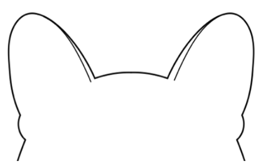
Rate each action unit from 0 to 2:

0 = action unit is absent
1 = moderate appearance of the action unit, or uncertainty over its presence or absence
2 = obvious appearance of the action unit

If the action unit is not visible, please mark the option "not possible to score"

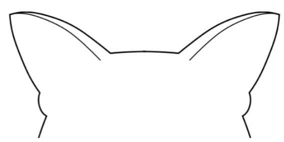
Ear position

0 = absent

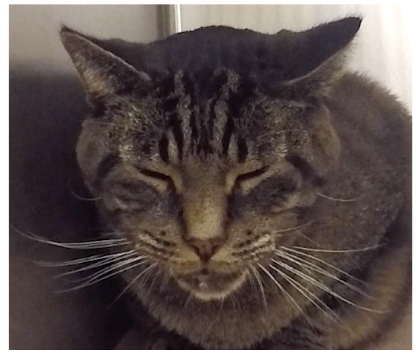


Ears facing forward

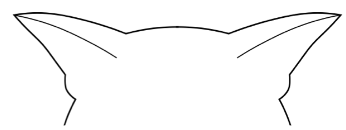
1 = moderately present



Ears slightly pulled apart



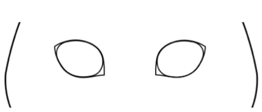
2 = markedly present



Ears rotated outwards

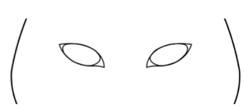
Orbital tightening

0 = absent



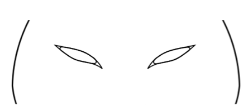
Eyes opened

1 = moderately present



Partially closed eyes

2 = markedly present



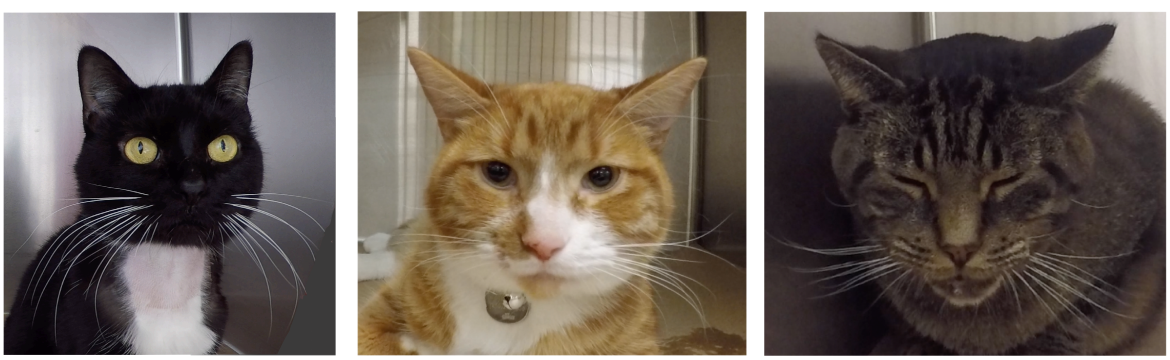
Squinted eyes

Muzzle tension

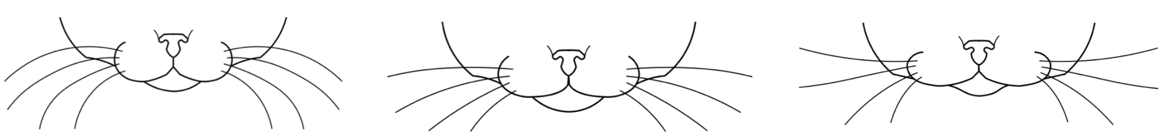
0 = absent	1 = moderately present	2 = markedly present

Relaxed (round shape)	Mild tension	Tense (elliptical shape)

Whiskers change



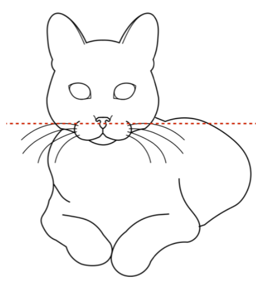
0 = absent	1 = moderately present	2 = markedly present



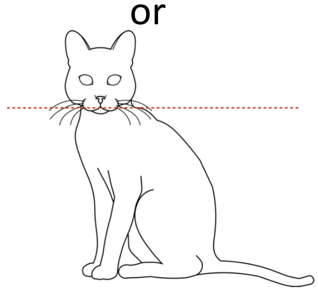
Loose (relaxed) and curved	Slightly curved or straight (closer together)	Straight and moving forward (rostrally, away from the face)

Head position

0 = absent

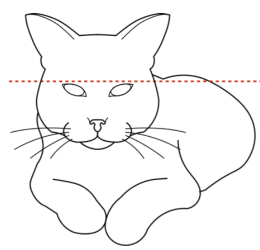


or



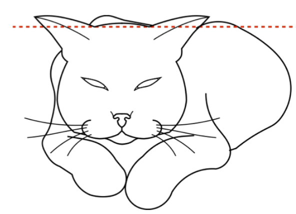
Head above the shoulder line

1 = moderately present

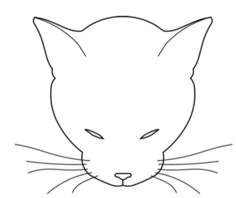


Head aligned with the shoulder line

2 = markedly present



or



Head below the shoulder line or tilted down (chin toward the chest)

Appendix B. Inter-rater Reliability of the Feline Grimace Scale in Cats Undergoing Dental Extractions

Ryota Watanabe[1], Graeme M. Doodnaught[1,2], Marina C. Evangelista[1], Beatriz P. Monteiro[1], Hélène L. M. Ruel[1] and Paulo V. Steagall[1]

Article identifier

This article was published in Frontiers in Veterinary Science 7:302 (2020)

https://doi.org/10.3389/fvets.2020.00302

[1] Department of Clinical Sciences, Faculty of Veterinary Medicine, Université de Montréal, Saint-Hyacinthe, QC, Canada
[2] Department of Veterinary Clinical Medicine, College of Veterinary Medicine, University of Illinois, Urbana, IL, United States

Abstract

This study aimed to evaluate the inter-rater reliability of the Feline Grimace Scale (FGS) in cats undergoing dental extractions and the effects of the caregiver's presence on the FGS scores. Twenty-four cats (6 ± 3.3 years old; 4.9 ± 1.7 kg) undergoing oral treatment were included in a prospective, blinded, randomized, clinical study. They underwent treatment under general anesthesia (acepromazine-hydromorphone-propofol-isoflurane-meloxicam-local anesthetic blocks) at day 1 and were discharged at day 6. Images of cat faces were captured from video recordings with or without the caregiver's presence at 6 h postoperatively (day 1), day 6, and before and after rescue analgesia. Images were randomized and independently evaluated by four raters using the FGS [five action units (AU): ear position, orbital tightening, muzzle tension, whiskers change, and head position; score 0–2 for each]. Inter-rater reliability and the effects of the caregiver's presence were analyzed with intraclass correlation coefficient [single measures (95% confidence interval)] and the Wilcoxon signed-rank test, respectively ($p < 0.05$). A total of 91 images were scored. Total FGS scores showed good inter-rater reliability [0.84 (0.77–0.89)]. Reliability for each AU was: ears [0.68 (0.55–0.78)], orbital tightening [0.76 (0.65–0.84)], muzzle [0.56 (0.43–0.69)], whiskers [0.64 (0.50–0.76)], and head position [0.74 (0.63–0.82)]. The FGS scores were not different with [0.075 (0–0.325)] or without [0.088 (0–0.525)] the caregivers' presence ($p = 0.12$). The FGS is a reliable tool for pain assessment in cats undergoing dental extractions. The caregiver's presence did not affect FGS scores.

Introduction

Oral disease is often observed in veterinary medicine (1). Our laboratory revealed that cats with severe oral disease requiring multiple tooth extractions had specific pain-induced behaviors, higher pain scores, changes in serum inflammatory cytokines, and lower food intake when compared with cats with no/minimal oral disease (2, 3). There are three pain scales with validation for feline pain assessment: Glasgow Composite Measure Pain Scale-Feline (CMPS-F) (4), UNESP-Botucatu multidimensional composite pain scale (5) and the recent Feline Grimace Scale (6, 7). However, these tools have not been used specifically in the context of pain caused by oral disease. The main challenge related to the use of the first two pain scales is that some questions are not applicable to cats with oral pain. For example, cats with oral pain often do not pay attention to the surgical area and it is often difficult to palpate a painful area (i.e., inside the oral cavity), which would be key behaviors in cats with other sources of pain including the abdomen and limbs (2). Thus, oral pain could be underestimated resulting in delays for analgesic intervention.

The Feline Grimace Scale (FGS) has been recently published and it comprises five action units (AU): eyes, ears, muzzle, whiskers, and head position. The instrument was developed and validated for naturally occurring pain of different sources and intensities (6). The clinical applicability of the FGS has been confirmed by comparing image with real-time assessment. In brief, minimal bias and narrow limits of agreement were observed between both methods of assessment (7). However, the FGS has not been specifically tested for assessment of oral pain, yet. In the authors' experience, multiple dental extractions can lead to facial edema which might influence the FGS scores. Therefore, there is an interest to understand the application and reliability of the FGS in cats undergoing oral treatment including dental extractions.

Pain assessment in the clinical setting requires real-time evaluation for early analgesic intervention. In laboratory animals (i.e., mice and rats), it is known that the presence of male evaluators affects pain scores, producing stress-induced pain inhibition (8). It is not known if a similar phenomenon also happens in cats.

The objectives of this study were to evaluate the inter-rater reliability of the FGS after oral treatment and the effect of the caregiver's presence on FGS scores. Our hypotheses were that the scores from different raters would be reliable and the presence of the caregiver would decrease the FGS score.

Materials and Methods

Ethics Statement

The study was approved by the Institutional Animal Care and Use Committee of the Université de Montréal (protocol 17-Rech-1890). Written informed consent was obtained from the owners for the participation of their animals in this study.

Study Design

Data for this study were obtained from a previously reported clinical trial involving dentistry, nutrition, pain management and behavior in cats before and after dental extractions (2, 3). The study was performed at the Centre hospitalier universitaire vétérinaire (CHUV), Faculty of Veterinary Medicine, Université de Montréal, between July 2017 and February 2018. The study is reported according to the CONSORT guidelines (http://www.consort-statement.org). The study design was a prospective, blinded, randomized clinical trial.

Animals

Twenty-four healthy cats (6 ± 3.3 years old; 4.9 ± 1.7 kg, 11 and 13 neutered males and females, respectively) with or without naturally occurring oral disease were included. Cats were considered healthy based on history, medical records, physical examination, complete blood count and biochemical panel. Recruitment of cats from shelter facilities was performed by two investigators (PS and BM) after informed written consent. All cats were admitted the day before dental procedures (day 0), and they underwent dental treatment under general anesthesia on day 1 and were discharged on day 6. They were housed in stainless steel cages in a cat-only ward and had free access to water, litter box and bedding. The amount of food offered was calculated based on caloric requirement as previously reported (2).

Inclusion and Exclusion Criteria

Cats were divided in one of two groups according to the severity of oral disease: no/minimal oral disease (n = 12) or severe oral disease requiring dental treatment (n = 12) (2). Diagnostic and treatments including dental examination (evaluation of gingival and calculus index, periodontal disease staging, and the number of missing tooth and tooth resorption), radiography, scaling, polishing, and/or extractions were performed as needed. Enrollment into either no/minimal or severe oral disease group in each cat was determined after dental treatment based on the size and number of extracted teeth (2). Cats with a body condition score of <3 or more than seven out of nine were not included. Cats with fearful behaviors, concurrent medical conditions, systemic disease, and the use of analgesics and/or antibiotics within a period up to 10 days before presentation were also not included.

Anesthesia, Analgesia and Dental Treatment

Detailed description of anesthetic and monitoring procedures is available elsewhere (2). Briefly, premedication included the intramuscular (IM) administration of acepromazine (0.02 mg/kg; 1 mg/mL, Acepromazine maleate, Gentès & Bolduc, Saint-Hyacinthe, QC, Canada) and hydromorphone (0.1 mg/kg; 2 mg/mL, Hydromorphone hydrochloride, Sandoz, Boucherville, QC, Canada). Anesthesia was induced with intravenous (IV) propofol (10 mg/mL, Propoflo 28, Zoetis, Kirkland, QC, Canada) and maintained with isoflurane (Isoflurane USP, Fresenius Kabi, Toronto, ON, Canada) in oxygen. Under general anesthesia, complete dental examination, radiography, scaling/polishing and tooth extractions (if needed) were performed by a board-certified individual and a 3rd-year resident of the American Veterinary Dental College. Cats requiring tooth extraction received local anesthetic blocks with bupivacaine (5 mg/mL, Sensorcaine, AstraZeneca, ON, Canada) using a 1 mL syringe and a 25-G needle (up to a total of 2 mg/kg) as needed including infraorbital, maxillary, and/or inferior alveolar mandibular nerve blocks ~20 min before extractions. At the end of dental treatment, all cats received meloxicam (0.2 mg/kg; Metacam 5 mg/mL Solution for Injection; Boehringer Ingelheim, Burlington, ON, Canada) subcutaneously. Oral administration of meloxicam (0.05 mg/kg, Metacam 0.5 mg/mL Oral Suspension for Cats; Boehringer Ingelheim, Burlington, ON, Canada) was continued at 24, 48, and 72 h after the first dose according to label recommendations in Canada.

Real-Time Pain Assessment, Video Recording and Video Editing

Real-time pain assessment was performed by one male observer [RW] using the CMPS-F at 23 different time-points from day 0 to 6. This observer was unaware of the oral condition and/or treatment of the cat. Video recordings were performed at 9 different time-points from day 0 to 6 for the study of orofacial pain-related behaviors using a wide-angle lens camera (GoPro Hero 5, GoPro, Riverside, CA, USA) set between the cage bars and remotely controlled by a smartphone (iPhone7, Apple Inc, Cupertino, CA, USA) (3). Cats were moved to a specific cage for video recording that included better lighting. After a 5-min acclimation to the new cage, 10-min videos were recorded for assessment of general (without the observer in the ward), playing, feeding and post-feeding behaviors (with the observer in room) for the purpose of studying different aspects of oral pain-induced pain behaviors (3). Briefly, the recordings of general and playing behavior were aimed to observe behaviors without interaction with the observer and the behaviors during playing with the observer using a ribbon toy, respectively. Data from selected time-points in which both real-time pain assessment and video recording had been performed were used in this study. These included the following four time-points: at 6 h postoperatively on day 1, at 8 am on day 6 and those recorded before and after rescue analgesia. These time points were chosen to represent a wide range of images of painful and non-painful cats. Video editing (trimming) was performed by the same observer [RW] using a video player software (QuickTime Player 10.5, Apple Inc, Cupertino, CA, USA) to obtain videos without the presence of the caregiver during recordings of general behaviors, and videos with the presence of the caregiver during recordings of playing behaviors. For the latter case, only recordings performed when the caregiver had entered the room but before playing with the cat using a ribbon toy were used.

During real-time pain assessment, if a cat had CMPS-F scores ≥ 5/20, rescue analgesia was administered with hydromorphone [0.05 mg/kg IV, if the IV catheter was in place (i.e., first 24 h after surgery) or 0.1 mg/kg IM, if the IV catheter had been removed]. CMPS-F scores were re-assessed 30 min after rescue analgesia. Additional 5-min videos were recorded immediately before rescue analgesia and 30 min after the administration of hydromorphone without the caregiver in the room.

Image Collection

Following video editing (trimming), a total of 124 videos were randomized using a random permutation generator (http://www.randomization.com) and renamed to consecutive numbers. Image capture (i.e., screenshots) of cat faces was performed for each video by a different investigator [GD] who was not involved with image scoring. Screenshots were performed when the cat was facing the camera and the entire face was visible. Then, the screenshot that was considered the most representative on the entire video for that timepoint was selected. Images were not captured if the cat did not face the camera at any time during the video (no frontal image).

Quality assessment of each screenshot was performed by the same individual who edited the videos [RW]. Image quality was assessed based on the angle of the face, brightness, blur, and whether the entire face including ear tips, whiskers and part of the proximal scapula were visible (Figure 1).

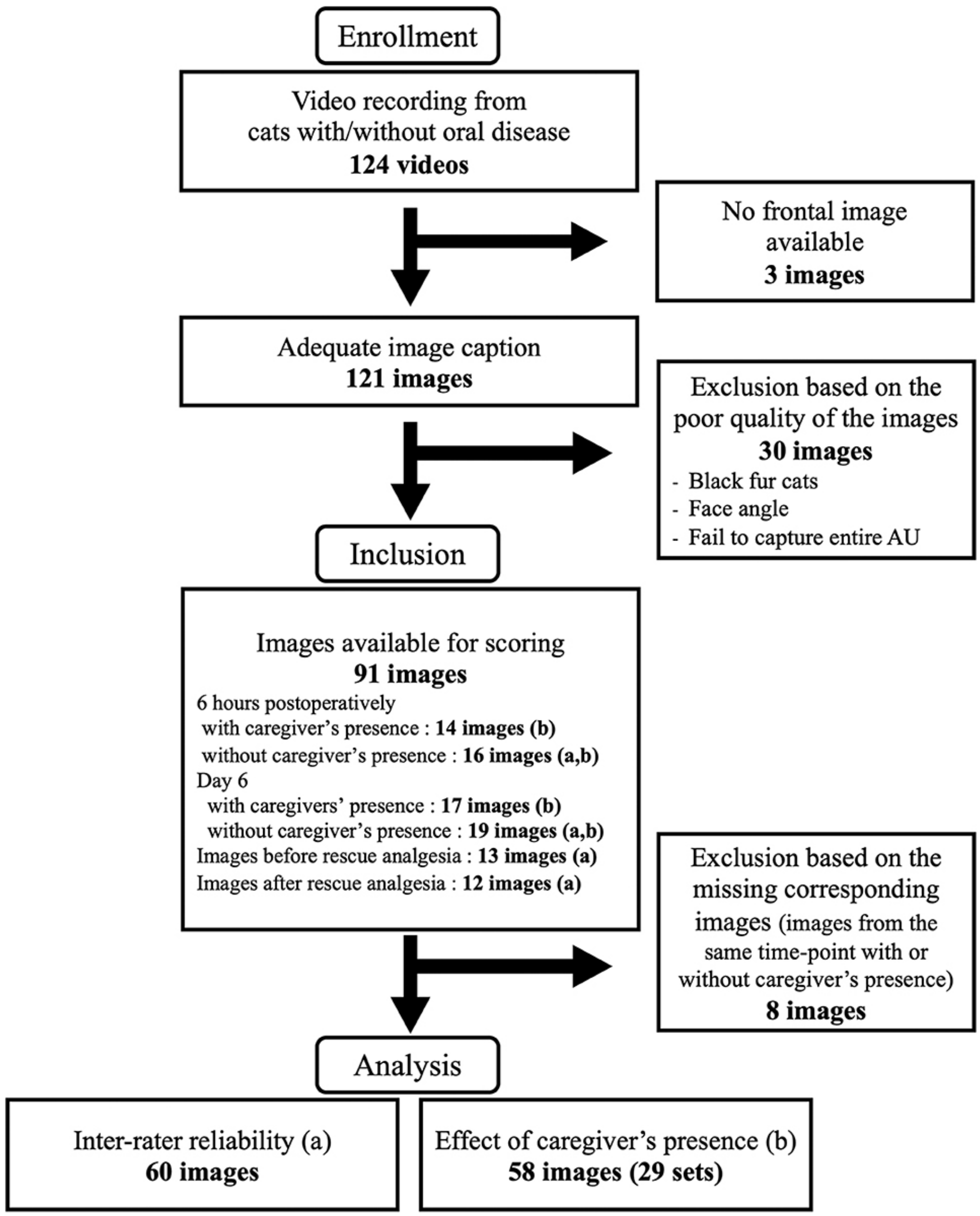


Figure 1. Flowchart of images captured from 24 cats with oral disease included in the study

Images with (a) and (b) were included for the analyses of inter-rater reliability and the effect of caregiver's presence, respectively.

Image Scoring

A total of 91 images were independently scored by 4 raters [ME, BM, HR, PS, three Ph.D. candidates (female) and one board-certified veterinary anesthesiologist (male)] who were blinded to the oral conditions of cats and timing of the recording (Figure 1). The raters were supplied with the training manual published with the original article (6) (https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-019-55693-8/MediaObjects/41598_2019_55693_MOESM1_ESM.pdf). Each image was evaluated using the FGS for scoring of five action units (AU): ears, eyes, muzzle, whiskers, and head position. The AUs were scored as following: 0 = AU is absent; 1 = moderate appearance of the AU, or uncertainty over its presence or absence; 2 = obvious appearance of the AU; or "not possible to score" = e.g., if the AU was not clearly visible (6). A total score was calculated by the sum of the scores of the AUs divided by the total possible score, excluding those marked as not possible to score (e.g., 3/8 = 0.375). The images were scored using an online survey (SurveyMonkey, https://www.surveymonkey.com) and divided into two sets. There was a minimum of 24 h and maximum of 48 h between scoring of the first and second set of images to avoid rater's fatigue. Scoring was performed between May 21st and 24th, 2019. Images receiving "not possible to score" for two or more AUs were excluded from statistical analyses.

Statistical Analyses

Statistical analyses were performed using SPSS software (version 25.0 IBM SPSS Statistics, Armonk, NY, USA). Images from days 1 and 6 without the caregiver's presence and images before and after rescue analgesia were used for the analysis of inter-rater reliability. Images from days 1 and 6 with and without caregiver's presence were used for the analysis of effect of caregiver. Inter-rater reliability was calculated for each AU and for the total FGS score using intraclass correlation coefficients (ICC) with 2-way random effects ICC model for absolute agreement. ICC was interpreted according to a previously described scale (9): <0.5 = poor, 0.5–0.75 = moderate, 0.75–0.9 = good, and > 0.90 = excellent reliability. The ICC was calculated based on single measures (ICC single) which is an index for the reliability of the rating for one rater and the average of the measures (ICC average) which is an index for the reliability of mean of k raters as recommendation of the guideline (9). The effect of the caregiver's presence was assessed by comparing FGS scores of images with and without caregiver's presence using a Wilcoxon signed rank test. The FGS scores with and without the caregiver's presence were compared between no/minimal and severe oral disease cats using a Mann-Whitney U-test, and within each group using a Wilcoxon signed rank test. Normality of the distribution of the scores was assessed using a Shapiro-Wilk test. Values of $p < 0.05$ were considered statistically significant.

Results

Inter-rater Reliability

Sixty images without the caregiver's presence were included in the analysis. Images were available from days 1 and 6 (n = 16 and n = 19, respectively) and from before and after rescue analgesia from days 1, 2 and 3 (n = 13 and n =12, respectively) (Figure 1). Inter-rater reliability is presented in Table 1. ICC single was moderate for ears, muzzle, whiskers, and head position and good for eyes. The ICC average was good for muzzle and excellent for ears, eyes, whiskers and head position. Reliability of total FGS scores was good and excellent, based on ICC single and ICC average, respectively.

Table 1. Inter-rater reliability of the Feline Grimace Scale in cats with oral disease

Action unit		ICC (95% CI)
Ears	ICC single	0.68 (0.55 - 0.78)
	ICC average	0.89 (0.83 - 0.94)
Eyes	ICC single	0.76 (0.65 - 0.84)
	ICC average	0.93 (0.88 - 0.95)
Muzzle	ICC single	0.56 (0.43 - 0.69)
	ICC average	0.84 (0.75 - 0.90)
Whiskers	ICC single	0.64 (0.50 - 0.76)
	ICC average	0.88 (0.80 - 0.93)
Head position	ICC single	0.74 (0.63 - 0.82)
	ICC average	0.92 (0.87 - 0.95)
FGS total score	ICC single	0.84 (0.77 - 0.89)
	ICC average	0.95 (0.93 - 0.97)

A total of 91 images were independently scored by 4 raters who were blinded to the oral conditions of cats and timing of the recording. Intraclass correlation coefficient (ICC) estimates with 95% confidence intervals (95% CI) were calculated based on single measure (ICC single) and average (ICC average) of measures, using a 2-way random effects model for absolute agreement. Interpretation of ICC was performed as following: ICC < 0.5 = poor, 05-0.75 = moderate, 0.75-0.9 = good, and > 0.90 = excellent reliability.

Effect of Caregiver's Presence

A total of 66 images were collected. From these, 29 images (13 and 16 sets from male and female cats, respectively) had a corresponding match (i.e., image from the same time-point with or without caregiver's presence), resulting in 58 images to be scored (day 1, n = 28 and day 6, n = 30). A total of 8 images did not have the corresponding match and were excluded (Figure 1). Median (range) of total FGS score without and with caregiver's presence were 0.088 (0–0.525) and 0.075 (0–0.325), respectively. Overall, there were not significant differences between scores with and without the caregiver's presence ($p = 0.12$). Median (range) of FGS scores without the caregiver's presence was 0.088 (0–0.325) in the minimal and 0.088 (0–0.525) in the severe group ($p = 1.000$). Median (range) FGS scores with the caregiver's presence in each group was 0.075 (0–0.325) in the minimal and 0.063 (0–0.250) in the severe group ($p = 0.711$). The FGS scores were not significantly different with or without the caregiver's presence within the no/minimal group ($p = 0.195$) or severe group ($p = 0.398$).

Discussion

This study evaluated the inter-rater reliability of the FGS for pain assessment in cats with naturally occurring oral disease and the effect of the caregiver's presence on FGS scores. Overall, the results indicate that the reliability of each AU and total FGS scores based on ICC single were moderate to good and that the presence of a male caregiver had no significant effect on the FGS scores.

Inter-rater reliability of total FGS scores was good to excellent considering ICC single and ICC average. The estimate ICC single is commonly used when a decision is made based on the scores of a single rater, however values of ICC average are usually higher (9). In the current study, the inter-rater reliability for each AU was moderate (ears, muzzle, whiskers, and head position) to good (eyes). Reliability of scores of the muzzle and whiskers were lower than other AUs (ICC single for muzzle and whiskers were 0.56 and 0.64, respectively). It is possible that dental extractions caused inflammation and facial edema likely impacting the scoring of muzzle and whiskers (i.e., difficulty of distinction between postoperative inflammation and the painful facial expression). Nevertheless, similar results were observed in the previous study in cats (0.63 and 0.55 for the muzzle and whiskers, respectively) (6). Reliability of the AUs ears and head position (0.68 and 0.74, respectively) were lower than the previous study (0.87 and 0.90, respectively) (6). In the present study, the camera was positioned to film the cats' behaviors for another study (3), and the height and angle of the video camera (set higher in the cage) may have not been ideal to capture the frontal image of the cat and further FGS scoring. If the camera angle is not optimal, the visualization, and interpretation of AUs could change between raters. However, ICC single of total scores were good, and the result indicates that the raters could still identify the changes associated to pain in these cats.

In this study, 51.7% (15/29) of images with the caregiver's presence had lower, yet not significant, scores than those without caregiver's presence. Indeed, the caregiver's presence did not significantly affect the FGS scores either when data for each group were analyzed together or independently. On the other hand, a previous study reported that the presence of a male experimenter produced a stress-induced pain inhibition response in mice and rats (8). This previous study reported that this response disappears within 30–60 min and it is not known if longer acclimation periods would change the FGS scores with or without the presence of a caregiver in cats. Furthermore, in the present study, a male observer scored male and female cats during the study whereas male and female raters scored the images. The sex of the observer is known to affect real-time pain assessment in rodents (i.e., male pheromone induces analgesic effect) (8); similar findings have been reported with video-assessment in small animals (10). Although the present study was not specifically designed to evaluate the effect of sex on pain assessment, the presence of a male caregiver did not affect FGS scores via image assessment. However, it is not known if the sex of raters could have influenced FGS scores.

There are some limitations in this study. First, this was an exploratory study and the materials were obtained from previous reports (2, 3). As a result, sub-optimal image quality played an important role as discussed above. Indeed, 26.6% of the images were excluded. Additionally, power analysis and sample size calculation were not performed before the experiment because there is no consensus to determine the sample size a priori in the validation studies (11). Second, the order of video recording could not be randomized, and the videos without the caregiver's presence were always obtained before those with the caregiver's presence. However, this order bias was not present during image assessment because the videos were trimmed, and images were randomized before image selection and scoring by an observer not involved with image scoring. Third, images of cats presenting moderate (nine images) to severe (13 images) pain based on CMPS-F (3–4, and ≥ 5, respectively) were underrepresented. This could represent an important limitation to study the effects of caregiver's presence. If the images of painful cats were underrepresented, it is possible that some of these patients had low FGS scores which could not be significantly reduced during a stress-induced pain inhibition response with the caregiver's presence, as observed previously (8). One of the reasons for the lack of good quality images was that three black cats required rescue analgesia, and five of these images were excluded from analysis because identification of muzzle and whiskers were not possible in these individuals. This issue was also reported in previous studies in horses and cats (6, 12), and a possible solution would be the use of artificial lighting sources during recordings. The other possible way to balance the distribution of pain intensity across the images might be to obtain several screenshots from same painful time points (i.e., videos filmed before rescue analgesia). However, the increase of number of images from same cats could bias the raters' scores. Finally, images of days 1 and 6 were included for the analysis of the effect of caregiver's presence. The images obtained on day 6 might have biased the results since perhaps cats were no longer painful. However, the pain scores (CMPS-F) in the severe group were significantly higher than the minimal group on day 6 (2), which made the authors believe cats in severe group could still be in mild pain.

In conclusion, the FGS is a reliable tool for assessment of oral pain in cats, though some action units were difficult to identify due to poor image quality and facial edema and inflammation. The caregiver's presence did not affect the FGS scores. The influence of sex in the FGS scores should be a subject of future investigations.

Acknowledgments

RW is a recipient of a scholarship from the Doctoral research scholarships program for foreign students of the Ministère de l'Éducation et de l'Enseignement Supérieur du Québec. PS's laboratory is funded by a Discovery Grant by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2018-03831).

Additional information

Author Contributions

RW and PS designed, conducted the study, and drafted the manuscript. RW performed postoperative care and pain assessment, the video filming, the image selection and the statistical analyses. GD performed the general anesthesia and the image captures. ME, HR, BM, and PS scored the images. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Bellows J, Berg ML, Dennis S, Harvey R, Lobprise HB, Snyder CJ, et al. AAHA dental care guidelines for dogs and cats. J Am AnimHosp Assoc. (2019) 55:49–69. doi: 10.5326/JAAHA-MS-6933

2. Watanabe R, Doodnaught G, Proulx C, Auger JP, Monteiro B, Dumais Y, et al. A multidisciplinary study of pain in cats undergoing dental extractions: a prospective, blinded, clinical trial. PLoS ONE. (2019) 14:e0213195. doi: 10.1371/journal.pone.0213195

3. Watanabe R, Frank D, Steagall P. Pain behaviors before and after treatment of oral disease in cats using video assessment: a prospective, blinded, randomized clinical trial. BMC Vet Res. (2020) 16:100. doi: 10.1186/s12917-020-02302-w

4. Reid J, Scott EM, Calvo G, Nolan AM. Definitive Glasgow acute pain scale for cats: validation and intervention level. Vet Rec. (2017) 180:449. doi: 10.1136/vr.104208

5. Brondani JT, Mama KR, Luna SP, 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) 9:143. doi: 10.1186/1746-6148-9-143

6. Evangelista MC, Watanabe R, Leung VSY, Monteiro BP, O'Toole E, Pang DSJ, et al. Facial expressions of pain in cats: the development and validation of a Feline Grimace Scale. Sci Rep. (2019) 9:19128. doi: 10.1038/s41598-019-55693-8

7. Evangelista M, Benito J, Monteiro B, Watanabe R, Doodnaught GM, Pang DSJ, et al. Clinical applicability of the Feline Grimace Scale: real-time versus image scoring and the influence of sedation and surgery. Peer J. (2020) 8:e8967. doi: 10.7717/peerj.8967

8. 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

9. Koo TK, Li MYA. Guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. (2016) 15:155–63. doi: 10.1016/j.jcm.2016.02.012

10. Doodnaught GM, Benito J, Monteiro BP, Beauchamp G, Grasso SC, Steagall PV. Agreement among undergraduate and graduate veterinary students and veterinary anesthesiologists on pain assessment in cats and dogs: a preliminary study. Can Vet J. (2017) 58:805–8.

11. Anthoine E, Moret L, Regnault A, Sébille V, Hardouin JB. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. Health Qual Life Outcomes. (2014) 12:176. doi: 10.1186/s12955-014-0176-2

12. Dalla Costa E, Minero M, Lebelt D, Stucke D, Canali E, Leach MC. Development of the Horse Grimace Scale (HGS) as a pain assessment tool in horses undergoing routine castration. PLoS ONE. (2014) 9:e92281. doi: 10.1371/journal.pone.0092281