741. McLennan, K.M., Miller, A.L., Dalla Costa, E., Stucke, D., Corke, M.J., Broom, D.M. and Leach M.C. 2019. Conceptual and methodological issues relating to pain assessment in animals: the development and utilisation of pain facial expression scales. *Applied Animal Behaviour Science*, 217, 1-15. doi:org/10.1016/i.applanim.2019.06.001

Pre-publication copy

Conceptual and methodological issues relating to pain assessment in mammals: the development and utilisation of pain facial expression scales.

Krista M McLennan[a*], Amy L Miller[b], Emanuela Dalla Costa[c], Diana Stucke[d], Murray J Corke[e], Donald M Broom[e], Matthew C Leach[b]

[a] Department of Biological Sciences, University of Chester, Parkgate Rd, Chester CH1 4BJ, UK

[b] Newcastle University, School of Natural and Environmental Sciences, Newcastle upon Tyne, United Kingdom

[c] Università degli Studi di Milano, Dipartimento di Mediciana Veterinaria, Milan, Italy

[d] Havelland Equine Clinic, Beetzsee-Brielow, Germany

[e]Centre for Animal Welfare and Anthrozoology, Department of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge, United Kingdom

*Corresponding Author: Krista McLennan

Email: k.mclennan@chester.ac.uk

Tel: 01224 513686

Abstract

Effective management of pain is critical to the improvement of animal welfare. For this to happen, pain must be recognised and assessed in a variety of contexts. Pain is a complex phenomenon, making reliable, valid, and feasible measurement challenging. The use of facial expressions as a technique to assess pain in non-verbal human patients has been widely utilised for many years. More recently this technique has been developed for use in a number of non-human species: rodents, rabbits, ferrets, cats, sheep, pigs and horses. Facial expression scoring has been demonstrated to provide an effective means of identifying animal pain and in assessing its severity, overcoming some of the limitations of other measures for pain assessment in animals. However, there remain limitations and challenges to the use of facial expression as a welfare assessment tool which must be investigated. This paper reviews current facial expression pain scales ("Grimace Scales"), discussing the general conceptual and methodological issues faced when assessing pain, and highlighting the advantages of using facial expression scales over other pain assessment methods. We provide guidance on how facial expression scales should be developed so as to be valid and reliable, but we also provide guidance on how they should be used in clinical practice.

Key words: Facial expression, Pain, Welfare, Methods, Clinical practice

Contents:

1. Introduction
2. Pain in animals remains a welfare issue
    2.1. The anatomy and physiology of pain
    2.2. The concept of pain as an affective state
    2.3. Additional factors affecting pain experience
    2.4. Managing pain
3. Pain assessment
    3.1. Validity
    3.2. Reliability

3.3. Feasibility
4. Facial expression scales as a tool for pain assessment in mammals
4.1. Image capture and modification
4.2. Scale development
4.3. Experimental design and pain stimulus
4.4. Provision of analgesia
4.5. Other pain assessment tools for comparison
4.6. Scale testing
4.7. Feasibility testing
5. Clinical application of facial expression as a method of pain assessment
6. Conclusion
Acknowledgements
References

1. Introduction

Understanding, recognising and managing pain in animals is of critical importance to their welfare; however, our current understanding of pain is limited by its complexity, and the subjective nature of the response to pain. Pain assessment is complicated by the involvement of an affective component as well as the sensory nervous component (Broom, 2014). The similarity in structure and function of nervous systems between humans and other mammals, coupled with the similarity in behavioural responses to painful stimuli, provides evidence that non-human animals feel pain, resulting in suffering (Broom, 2001). This is not accepted by all in the scientific community, some arguing that conscious awareness of pain is required for suffering to occur and that this is limited to humans and a small range of other species (Bermond, 2001; Key, 2016). Key (2016) argued the behavioural and physiological responses to painful stimuli observed in animals not possessing a prefrontal cortex should be viewed as simple nociceptive responses, not an indication of the feeling of pain. In order to properly understand the aversive nature of pain and the extent of suffering, both the sensory and

affective elements of pain need to be assessed in a validated, reliable manner, that takes a functional rather than anatomical approach to pain (Broom, 2016, 2014, Sneddon et al., 2018, 2014).

Facial expressions have long been used to recognise and quantify pain in human patients who are unable to verbalise, such as neonates or patients with verbal or cognitive impairments (Boerner et al., 2013; Prkachin et al., 1994; Prkachin and Solomon, 2008; Schiavenato et al., 2008). Facial expressions have also been demonstrated to encode both the sensory and affective components of pain in humans (Kunz et al., 2012). Langford et al. (2010) were the first to extend this method of assessing pain in humans to non-human animals, mice. These authors showed facial expressions of mice undergoing a painful experience reduced in a dose dependent manner when treated with effective analgesics. The authors were able to separate the typical sensory response (e.g. writhing) from the emotional response (facial expression) to painful stimuli by lesioning the insula. The insula is an area of the human brain associated with emotional reaction to pain, which is also present in mice. Further investigation is required into this phenomenon before it can be considered conclusive due to the low sample size employed (n=6) in this study. These results, and those from Kunz et al. (2012) provide support to the concept that facial expression could be key to demonstrating the affective component of pain in animals, as well as the nociceptive response.

In the last decade, a number of other species have had facial expression scales developed and validated to varying degrees as a pain assessment tool (Dalla Costa et al., 2014; Di Giminiani et al., 2016; Gleerup et al., 2015; Guesgen et al., 2016; Häger et al., 2017; Holden et al., 2014; Keating et al., 2012; MacRae et al., 2018; McLennan et al., 2016; Reijgwart et al., 2017; Sotocinal et al., 2011; van Loon and VanDierendonck, 2015). For this technique of pain assessment to be effective, we must understand the challenges and limitations to the development and use of facial expression as a pain assessment method.

The aim of this paper is to provide a brief overview of pain as a welfare issue in mammals, and to discuss the reasons why assessment of pain is difficult. A recent review by Descovich and colleagues (2017) argued that facial expression is under-utilised as a welfare assessment tool. These authors

briefly mention the limitations and challenges with the use of facial expression as a welfare assessment tool. It is the purpose of this review to further explore these conceptual and methodological difficulties that are characteristic of a new field further, with specific reference to assessing animal pain. We will discuss some of the scales that have been developed to assess pain in animals exploring the methodological issues that they have faced. We investigate how the researchers have attempted to overcome the conceptual problems of pain assessment when validating the effectiveness of these scales. Additionally, we will highlight the advantages of using facial expression scales over other more common methods of pain assessment. We will demonstrate that facial expression scales provide an opportunity to further our understanding of pain assessment.

We suggest a caveat on the future development and utilisation of facial expression scales. However, we also provide guidance on how these scales should be used in both clinical and research settings in order to be effective in pain assessment. Progress in animal pain assessment critically relies upon the development of robust and compelling experimental designs (Panksepp, 1998). Thus, we also aim to provide a framework on how these scales should be developed for other species and for other emotional states so they are valid and reliable.

2. Pain in animals remains a welfare issue

Pain is aversive, and left unmitigated can lead to severe stress with detrimental physical and mental effects on an animal causing suffering (Dawkins, 2008; Flecknell et al., 2011). The presence of pain reduces play (Mintline et al., 2013; Rushen and de Passillé, 2012; Thornton and Waterman-Pearson, 2002), grooming (Dalla Costa et al., 2014; Ellen et al., 2016; Keating et al., 2012), eating (de Oliveira et al., 2014), and disrupts sleep (Andersen and Tufik, 2003; Ohayon, 2005; Schütz et al., 2003). Despite increased awareness of the existence of pain in animals and its detrimental effects on welfare, animals are still subjected to procedures or events in which pain is likely to occur. Routine husbandry procedures in farm animals such as castration and de-horning can result in pain if carried out with inadequate anaesthesia and analgesia (Lomax and Windsor, 2013; Mintline et al., 2013; Stewart et al., 2014, 2007; Walker et al., 2011). Pain in farm animals reduces production and/or growth (e.g. Green

et al., 2002), directly conflicting with the global need for increased sustainable food production (Hunter et al., 2017). Experimental procedures in laboratory animals, and accidental injury, disease or elective surgery in all species may also result in pain (Abu-Serriah et al., 2007; Matsumiya et al., 2012; Waite et al., 2015). Unmitigated pain can result in pathological changes in physiology and behaviour, increasing variability in data collected, thus decreasing validity of scientific studies (Hawkins, 2014). Pain management in domestic animals is not always provided in an effective manner (Bell et al., 2014; Huxley and Whay, 2006; Norring et al., 2014), resulting in suffering. Such procedures and the resulting negative effects on animal welfare, are a major source of concern for the public (Busch et al., 2017; Doughty et al., 2017; Fredriksen et al., 2011; Robbins et al., 2015; Ventura et al., 2014). Effective assessment and alleviation of pain are closely linked. If we cannot effectively identify pain when it occurs or judge its severity, we shall be unable to alleviate it. To understand the obstacles to pain prevention and alleviation, it is necessary to examine our current understanding of pain.

2.1 The anatomy and physiology of pain

Pain involves both sensory and affective components, and is often associated with actual or potential tissue damage (Broom, 2001; IASP, 1994; Sneddon et al., 2014). The sensory aspect of pain refers to nociception, the transmission of information about tissue damage to the brain via peripheral pain receptors (nociceptors), nerve fibres and neurons. Noxious stimuli (mechanical, thermal or chemical) activate free-nerve endings of thinly myelinated A-delta nerve fibres and unmyelinated C fibres. The action potentials produced pass via the dorsal root ganglia (DRG) into the spinal cord. Neurons within the dorsal horn are activated, mediating local withdrawal reflexes as well as relaying the signal via ascending afferent pathways in the gray matter of the spinal cord to synapses in the medulla, midbrain and thalamus (Brooks and Tracey, 2005). From these centres, in mammals, neurons transmit the signal to the cortex where the conscious affective experience of pain is considered to occur (Hofbauer et al., 2001; Lee et al., 2009). However, interneurons local to the dorsal horn can modulate the nociceptive signal, and descending pathways from the mid and hind brain can inhibit or facilitate the signals transmission to the brain and spinal cord (Heinricher et al., 2009; White et al., 2018). These

154 changes that occur in the neurobiology of the transmitted signal lead to complications in our
155 understanding of the sensory component of pain. Moreover, they can have a significant impact on the
156 affective experience of pain and the associated suffering (Rainville, 2002).

157 Pain may be either acute or chronic in nature. Acute pain is generally short lived and is caused mainly
158 by pathological damage to tissue or nerves resulting from injury, inflammation or infection (Viñuela-
159 Fernández et al., 2007). Acute pain tends to respond to pain relief as the inflammation and infection
160 are controlled and tends not to persist beyond the healing process (Woller et al., 2017). Chronic pain
161 can extend beyond the healing process (Lavand'homme, 2011; Ley et al., 1989) and is associated with
162 greater emotional distress (Baliki et al., 2006; Seminowicz et al., 2009). Chronic pain can be complex,
163 being multifaceted and sometimes not originating from peripheral nociception, making diagnosis of
164 the underlying cause and thus treatment of chronic pain difficult. Moreover, sustained activation of
165 nociceptors, nerve damage, or neural dysfunction, can cause neuropathic pain, presenting itself as
166 allodynia, hyperalgesia or spontaneous pain (Gear and Levine, 2011; Miki et al., 2002). It is now well
167 accepted in human medicine, that chronic pain can be a disease in itself and does not need to be
168 associated with another physical disease or injury (Apkarian and Scholz, 2006; Groh et al., 2017;
169 Tracey and Bushnell, 2009).

170 *2.2. The concept of pain as an affective state*

171 The detection, transduction, transmission, modulation and projection of information to the central
172 nervous system (CNS) appears similar within all mammals (Viñuela-Fernández et al., 2007). In
173 principal, noxious stimuli which are painful to humans will also cause pain in other mammals. This
174 does not necessarily mean that they experience pain in the same way as humans, but it justifies the
175 inference that they do experience the aversive nature of pain (Panksepp, 1998; Weary et al., 2006).
176 When considering where and how pain is experienced much of the evidence supports the view that the
177 medial thalamocortical pathways, including the limbic system and insular cortex, play an important
178 role in mammals (Gu et al., 2013; Jasmin et al., 2004; Lu et al., 2016). Human patients with damage
179 to these areas of the brain report asymbolia, a condition that leaves patients being aware of the

sensory qualities of nociception but without experiencing the aversive nature of pain (e.g. Berthier et al., 1988). This suggests that there is some separation of the sensory and affective dimensions of pain. Conversely, Feinstein et al. (2016) recently reported no effect on the emotional awareness of pain in a human patient with extensive damage to the insula, anterior cingulate and amygdala. This would suggest that these regions are not necessary for the conscious experience of pain. The inconsistency in results raises questions regarding our understanding of where and how the brain experiences the aversive nature of pain, adding to the challenge of assessing the impact of pain on an animal's affective state.

2.3 Additional factors affecting pain experience

An animal's previous experience of pain can have a significant impact on how it responds to a noxious stimulus. Long term changes in pain response have been demonstrated to occur when animals have experienced pain at an early age. Pain experienced as a neonate, either associated with chronic inflammation (Benatti et al., 2009; Lim et al., 2009), or tissue insult (Beggs et al., 2012; Clark et al., 2014), significantly reduces pain thresholds and increases the expression of pain-related behaviours as an adult. These changes are also likely to be long lasting when compared with adults that have not been exposed to pain as neonates (Beggs et al., 2012). The decreased pain threshold in these animals could be due to sensitisation of peripheral neurons or nociceptors, or central mediation occurring at the level of the spinal cord (Beggs et al., 2012; Clark et al., 2014).

Early life stress can also have significant effects on pain experienced as an adult. Animals born to mothers experiencing high stress levels whilst pregnant, have an amplified pain response (Rutherford et al., 2009; Sandercock et al., 2011). In addition, a mother's neonatal experience of pain has been shown to affect her offspring's response to pain (Clark et al., 2014). The changes in pain response seen in offspring are likely to be an adaptive response to the environment that the mother experiences, with programming of gene expression preparing the offspring for a better chance of survival (Benatti et al., 2009; Clark et al., 2014; Rutherford et al., 2009; Sandercock et al., 2011).

205 Differences in reactivity to pain also occur between sexes within a species (Guesgen et al., 2011;
206 Prusator and Greenwood-Van Meerveld, 2016; Sorge et al., 2014; Winston et al., 2014). Even if pain
207 responses in males and females are the same at birth, males have been shown to have a reduced
208 sensitivity to pain in comparison with females as they age (Guesgen et al., 2011), suggesting a
209 divergence in the ontogeny of pain processing systems. Other factors, such as the animal's personality
210 (Ijichi et al., 2014), whether there is social support (Guesgen et al., 2014), whether the animal has had
211 previous experience with the context of the pain such as with handling (Guesgen et al., 2013), or if
212 there is a presence of a human, particularly a male (Sorge et al., 2014), can affect how an animal deals
213 with and responds to pain. These additional influences add a layer of complexity when trying to assess
214 and manage animal pain.

215 *2.4 Managing pain*

216 Understanding of the major pathways and mediators involved in the transmission of nociceptive
217 information allow a number of pharmacological interventions to be employed in pain management
218 (see Viñuela-Fernández et al., (2007) for review). There are many licensed products available to
219 professionals intended to be used to mitigate pain in particular species, including local and regional
220 anaesthetics, opioids and non-steroidal anti-inflammatory drugs (NSAIDs) (Veterinary Medicines
221 Directorate., 2018). However, in some species such as sheep, licenced pain relief products are
222 currently not available in the UK (Veterinary Medicines Directorate., 2018), and so any pain relief
223 provided is given off-label (Lizarraga and Chambers, 2012), reducing the use of such drugs. For
224 species for which licensed drugs are available, use in practice is still limited (Becker et al., 2013; Bell
225 et al., 2014; Ison and Rutherford, 2014; Richardson and Flecknell, 2005; Weber et al., 2012).
226 Commonly reported barriers to the use of pharmaceuticals include lack of knowledge of pain
227 recognition and assessment, as well as cost, residues in production animals, and uncertainties of their
228 impact on scientific studies in research animals (Bell et al., 2014; Huxley and Whay, 2006; Ison and
229 Rutherford, 2014; Lizarraga and Chambers, 2012; Richardson and Flecknell, 2005). Being able to
230 recognise, assess and evaluate pain in animals is thus critical to preventing and alleviating pain

effectively in order to improve the welfare of animals under human care (Flecknell, 2000; Gentle, 2001).

3. Pain assessment

For any pain assessment method to be of value it must allow for the recognition, assessment and alleviation of pain in a sensitive and specific manner. Current scoring systems for recognising and assessing pain in non-human animals often use a combination of assessing the physiological response, and measuring the general functioning of the body, as well as observing behaviour (Brondani et al., 2013; Bussières et al., 2008; Molony et al., 2002; van Loon and VanDierendonck, 2015). These measures have a number of limitations, sometimes producing contradictory results (Molony et al., 2002) (See Weary et al. (2006) and Sneddon et al. (2014) for full reviews on other pain assessment measures and their interpretation). Physiological responses such as changes in the heart rate, body temperature and level of circulating cortisol provide measures of the sympathetic-adrenomedullary system and the hypothalamus-pituitary-adrenocortical systems. These systems are not specific to pain, but are also influenced by positive and other negative affective states such as stress (Carlson et al., 2006; Jaremka and Collins, 2017; Villani et al., 2006). Moreover, physical restraint is often required to obtain these measures leading to a general stress response, further confusing interpretation of the data (see Mormède et al., (2007) for review). Poor nutrition (Ingvartsen and Moyes, 2013; Lean et al., 2013), lack of physical and mental stimulation (Matur et al., 2016; McCreary and Metz, 2016), and disease (Raaperi et al., 2012; Šavc et al., 2016) are also possible causes of changes in general body function making these measures unreliable for pain assessment. Although longer term changes in behaviour can be objectively measured, they reflect the changes between two time-point observations rather than what the animal is experiencing at any particular time (Weary et al., 2006). Monitoring acute behavioural signs of pain provide a better indication of the current welfare of the animal and the pain they are experiencing. These behaviours, however, are often not pain specific and affected by other factors such as fear or stress (Gougoulis et al., 2010; Rutherford, 2002) creating problems with validity and reliability. Obvious behavioural signs of pain are also not common to all mammals as stoical species do not overtly express their affective state. More subtle signs of behaviour that can

indicate how an animal might feel are required for pain assessment for these species (Flecknell et al., 2011). Moreover, pain has many dimensions, and the measure should be able to consider the intensity, frequency, duration and quality of the pain (Ashley et al., 2005). Essentially, the measure must be valid, reliable and feasible (Bussières et al., 2008; Molony and Kent, 1997).

3.1 Validity

A fundamental attribute of any measure is its validity. There are a number of different types of validity, including; construct, convergent, discriminant, and internal. Construct validity refers to how accurate the measure is at measuring that specific construct (Calvo et al., 2014), in this case, pain. Differences recorded by the measure must be due to the true extent of differences between a painful and non-painful state (discriminant validity); the measure should be both sensitive (be able to correctly identify animals in pain) and specific (be able to correctly identify animals that are not in pain) (Brondani et al., 2013). A good measure of internal validity for pain assessment is to measure the changes that occur in response to analgesic provision in a dose dependent manner (Weary et al., 2006). Thus, after analgesia, the animal should either no longer be in pain, or be in significantly reduced pain, and the measure should be able to identify this correctly. A limitation to this approach is in species for which there is no licenced analgesic, and hence no information about likely effectiveness of that pain relief, making internal validity difficult to test. Additionally, some analgesia may not be effective due to modulation at the nociceptor level (Fleetwood-Walker et al., 2012). Consideration must be given to testing within subjects, assessing at both a baseline level when no pain is present, and again at a separate time point when pain is present. A new measure should also be tested against an already validated measure for that construct (convergent validity) (Battini et al., 2016). It is also critical that during validation of both new and established indicators that observers are blind to the state of the animal (e.g. pain or no pain) to prevent observer bias (Tuyttens et al., 2016).

3.2 Reliability

For a test to be valid it must also be reliable (Dalla Costa et al., 2018). Reliability refers to a measure's ability to generate the same result each time it is used on the same participant in a

consistent and stable manner, independent of the identity of the observer employing the measure (Neuman, 2014; Oliver et al., 2014). The test-retest approach can be implemented to test the consistency of a measure at producing the same result each time it is implemented, provided nothing has changed within the context that the first measurement was made (Napolitano et al., 2011; Prkachin and Solomon, 2008). It would also be expected that a consistent measure be repeatable, yielding the same result each time an observer implemented it (intra-observer reliability) (Oliver et al., 2014); however, the measure should also be repeatable and consistent across different observers (inter-observer reliability) (Oliver et al., 2014; Sotocinal et al., 2011). Consideration of the time interval between observations must be given as it can affect the reliability of measurements; too short a time and observers may remember their original answers (Martin and Bateson, 2007). Observers can also suffer fatigue causing their assessments to be inconsistent between the beginning and end of the test (Kiddie and Collins, 2014).

3.3 Feasibility

For a pain assessment method to be useful it needs to be feasible (Solomon et al., 1997). Feasibility is a measure of external validity, whereby the test must do what it is designed to do in the real-world, outside the context of developing and testing. For pain assessment methods, the measure must be useable on the farm, in the veterinary surgery, in the home, in the field or within a laboratory animal facility and yield the same result, ideally using no specialist apparatus or equipment (Battini et al., 2016). The measure should be quick and easy to use by people with different previous experience, following minimal training, to be of maximum use (Solomon et al., 1997). Being able to use the measure in real-time is essential to get the best assessment of the animal's current pain state. Additionally, being able to link the measure to an intervention score enhances its usefulness (McLennan et al., 2016; Oliver et al., 2014). These are very difficult criteria to achieve for any measure, but they should be considered fully when developing or evaluating any new measure of pain, such as facial expression scales.

4. Facial expression scales as a tool for pain assessment in mammals

The use of facial expressions to assess pain has become frequent in human medicine and research (Boerner et al., 2013; Prkachin et al., 1994; Prkachin and Solomon, 2008; Schiavenato et al., 2008). The Facial Action Coding System (FACS) was originally developed by Ekman and Friesen (1978) to measure changes of the face or groups of muscles, known as "action units" (AUs), to an emotional stimulus. Prkachin (1992) was the first to apply the FACS to assess the facial expressions of pain in humans. Since then, there have been a number of advanced studies addressing the possible uses and limitations of scoring facial expressions to assess pain in humans. Schiavenato et al. (2008), noted that despite commonality of facial pain expressions across different ethnicities and sexes, there were inconsistencies in expression across age groups leading to slightly revised versions of the FACS for neonates (Neonates Facial Coding System (NFCS) (Ahola Kohut and Pillai Riddell, 2009; Schiavenato et al., 2008), infants (Baby FACS) (Ahola Kohut et al., 2012) and children (Child Facial Coding System (CFCS) (Vervoort et al., 2011, 2008; Vlaeyen et al., 2009). Ahola Kohut and Pillai Riddell (2009) investigated the ability to discriminate between pain-related and non-pain related distress in neonates by means of the NFCS; however, it was only possible to distinguish different intensities of distress, rather than between states (pain and no-pain). Conversely, Kunz et al. (2013) identified distinct aversive feelings through differing combinations of AUs being expressed at different strengths. These studies made some key changes to the original FACS (separate FACS for differing age categories, and differing combinations of units for different constructs) refining the technique for use in humans making it valid, reliable and feasible pain assessment tool in a variety of contexts.

A number of FACS for animals have been recently developed (cats (Caeiro et al., 2017), horses (Wathan et al., 2015), chimpanzees (Parr et al., 2007), and macaques (Julle-Danière et al., 2015; Parr et al., 2010)), detailing all possible individual facial movements that can occur across the face. FACS provide a method of objectively identifying facial areas that may be affected by particular contexts (Wathan et al., 2015); however, these systems can be quite complex with as many as 17 different AUs to assess (Wathan et al., 2015). They have not yet been applied to particular contexts such as pain. Having a scale that contains specific actions or group of actions shown to appear in relation to pain is

likely to be more feasible in the clinical setting. Within the last twelve years a number of facial expression scales (also known as “Grimace Scales”) have been designed specifically to assess pain in animals. Many of these scales focus on just four or five facial areas and consider a particular action as present (score 2), partially present (score 1) or not present (score 0) (Dalla Costa et al., 2014; Di Giminiani et al., 2016; Gleerup et al., 2015; Guesgen et al., 2016; Häger et al., 2017; Holden et al., 2014; Keating et al., 2012; Langford et al., 2010; McLennan et al., 2016; Reijgwart et al., 2017; Sotocinal et al., 2011; van Loon and VanDierendonck, 2015). Animals were tested in pain and non-pain states, and a comparison made of their facial expression scores in each context. Careful consideration of a number of factors is required when developing and testing a facial expression scale, in order for the scale to be a valid pain assessment tool. These factors include the experimental design, the pain stimulus used, provision of analgesia, and another known pain assessment tool for comparison, as a minimum. The following scales are now considered: the Mouse Grimace Scale (MGS) (Langford et al., 2010), the Rat Grimace Scale (RGS) (Sotocinal et al., 2011), the Rabbit Grimace Scale (RbtGS) (Keating et al., 2012), the Ferret Grimace Scale (FGS) (Reijgwart et al., 2017), the Horse Grimace Scale (HGS) (Dalla Costa et al., 2016), the Equine Utrecht University Scale for Facial Assessment of Pain (EQUUS-FAP) (van Loon and VanDierendonck, 2015), the Equine Pain Face (Gleerup et al., 2015), the Sheep Pain Facial Expression Scale (SPFES) (McLennan et al., 2016), the Sheep Grimace Scale (SGS) (Häger et al., 2017), the Lamb Grimace Scale (LGS) (Guesgen et al., 2016), the Piglet Grimace Scale (PGS) (Di Giminiani et al., 2016) and the facial pain assessment tool for cats developed by Holden et al., (2014).

4.1 Image capture and modification

Facial expression scales are developed through the analysis of multiple images taken of animals during pain and non-pain states. The majority of scales use high definition video footage with multiple cameras to capture the facial expression of animals pre- and post-pain stimulus exposure (Dalla Costa et al., 2014; Di Giminiani et al., 2016; Gleerup et al., 2015; Guesgen et al., 2016; Häger et al., 2017; Keating et al., 2012; Langford et al., 2010; Leach et al., 2012; Leung et al., 2016; Matsumiya et al., 2012; Miller et al., 2016b, 2015; Miller and Leach, 2016; Sotocinal et al., 2011).

Some have been developed or tested using still photographs alone (Finlayson et al., 2016; Holden et al., 2014; McLennan et al., 2016; Miller and Leach, 2016, 2015, 2014; Reijgwart et al., 2017). Those that used video footage obtained still images from the video, either manually or through a specific piece of software called "Rodent Face Finder®" which selects frames when there is a clear view of the rodents' face in front of the camera (Sotocinal et al., 2011). This type of technology helps to reduce the bias of collecting and selecting images manually (Tuttle et al., 2018). Where this technology has not been available, assistants blind to treatments and time points have been utilised to select images from footage where there is a clear view of the face of the animal. Images can be selected at certain time points throughout the filming; for example, Langford et al. (2010) collected images at 3-minute intervals, whilst Guesgen et al. (2016) selected images every 15 seconds pre-docking and every 75 secs post-docking. Others have selected randomly throughout the time when any clear view of the animals' face was in front of the camera (Dalla Costa et al., 2014; Di Giminiani et al., 2016; Häger et al., 2017; Leung et al., 2016; Miller et al., 2015). The latter technique ensures that there is a large cohort of images from which to choose randomly those of the highest quality, but it can be time consuming and difficult to replicate in other studies if there are no set parameters of when and how to collect images. Having a more structured approach such as that of Langford et al. (2010) or Guesgen et al. (2016), can improve this.

There is a lack of consistency between studies in collecting images for the facial expression scoring in the length of each recording, or how many photographs were taken. Durations of video recordings used by researchers ranged from just 1-minute pre-pain stimulus (Guesgen et al., 2016) up to 30 minutes of footage (Häger et al., 2017; Matsumiya et al., 2012), with after pain-stimulus footage lasting for 5 minutes (Di Giminiani et al., 2016) or up to 30 minutes (Sotocinal et al., 2011). The length of footage or number of photographs taken should be sufficient to allow for the capture of the most appropriate images for facial expression analysis. This is likely to vary between species as well as between pain stimuli. Some studies may also have a number of constraints, such as time or field location that prevent long durations of video capture. A major advantage of facial expression is that it

390 is a tool for rapid assessment of pain, therefore videos of shorter duration may be more practical,
391 especially when testing the scale.

392 The time intervals to the pain at which the footage was recorded also varied between studies. Baseline
393 images were taken either immediately before the intervention (Gleerup et al., 2015), or up to one
394 week before intervention (Miller and Leach, 2014). For those looking at naturally occurring diseases,
395 baseline values have been captured much later after the initial pain images (one week for horses with
396 acute laminitis (Dalla Costa et al., 2016), and up to 90 days after initial treatment for sheep with
397 footrot (McLennan et al., 2016)). Once the pain stimulus was applied some immediately started
398 recording (Guesgen et al., 2016) whilst others waited for varying lengths of time (up to 8 hours (Dalla
399 Costa et al., 2014)), especially when waiting for any effects of anaesthesia to wear off, or for the
400 severity of pain or benefit of analgesia to become evident. The experimental design is likely to dictate
401 the most appropriate time to capture images.

402 The quality of the image used is highlighted by a number of researchers as being an important part of
403 ensuring reliability of the scales (Dalla Costa et al., 2014; Di Giminiani et al., 2016; Keating et al.,
404 2012; McLennan et al., 2016). Many of the papers have clearly stated the need to use high definition
405 video cameras or still cameras to ensure the best quality image (Dalla Costa et al., 2014; Di Giminiani
406 et al., 2016; Häger et al., 2017; Keating et al., 2012; Langford et al., 2010; McLennan et al., 2016;
407 Sotocinal et al., 2011). Another key point is to ensure that shadows are not present on the face; the use
408 of good lighting in the area where images are taken can help reduce this (Finlayson et al., 2016;
409 Reijgwart et al., 2017). The use of bright light, or camera flashes should be avoided as they may be
410 aversive (Holden et al., 2014). Langford et al. (2010), carried out retrospective adjustment of
411 brightness and contrast on their images to overcome some of the quality issues. In addition, the angle
412 at which photographs are taken is important, and the set-up of each image capture technique needs to
413 be carefully considered. Reijgwart et al. (2017) used a tunnel for each ferret to exit from at the same
414 height in line with the camera, whilst Di Giminiani et al. (2016) had four cameras around the edge of
415 the pen at a set height of 19cm (piglet head height). Dalla Costa et al. (2014), placed cameras at a
416 height above the horse to have the greatest chance of collecting both behavioural and facial images,

417 impacting on the angle that the images were taken. Others have had to handle the animals during the
418 procedure (Di Giminiani et al., 2016; Guesgen et al., 2016; Keating et al., 2012). This had an effect on
419 the facial expression scores given by observers in lambs (Guesgen et al., 2016), whilst in mice, the
420 type of handling was found to have no effect on the facial expression score given (Miller and Leach,
421 2016). Avoiding handling or close contact with the animal is to be preferred during image capture, as
422 many prey species do not overtly express signs of pain and distress when potential predators such as
423 humans are present (Sorge et al., 2014). Leaving an animal to perform the behaviours in conditions
424 that meet their needs is likely to yield the best results during development stages.

425 Most researchers clearly state that the images were cropped so that other postures or behaviours, or
426 indicators of any surgery or disease, are not visible (Dalla Costa et al., 2014; Di Giminiani et al.,
427 2016; Häger et al., 2017; Holden et al., 2014; Keating et al., 2012; Langford et al., 2010; McLennan et
428 al., 2016; Reijgwart et al., 2017; Sotocinal et al., 2011). Cropping images ensures that only the face of
429 the animal is studied and that the rest of the body does not influence the scorer. Reijgwart et al.
430 (2017), and Dalla Costa et al. (2017), also removed the background of the animal's face and displayed
431 all images with a uniform background. It has not been tested whether the background information
432 provided in images has an effect on an observers' scores by providing information about the context
433 in which the animal is photographed. Although it seems unlikely that the background in cropped
434 images could provide such information, until this is tested, removal of the background information
435 from the image is to be encouraged.

436 *4.2. Scale development*

437 Images taken are compared using a collage of multiple images from the pre- and post-pain stimulus to
438 identify specific AUs (e.g. ear position, cheek tightening, or eye closure) that change in the facial
439 expression of an individual animal. The images or stills are often analysed by time-blinded assistants
440 or experts in the field that have been blinded to time and treatment (Dalla Costa et al., 2014; Keating
441 et al., 2012; Langford et al., 2010). Many of the scales do not detail exactly how the AUs are selected;
442 however, the FGS (Reijgwart et al., 2017) and the PGS studies (Di Giminiani et al., 2016) state that

for selection of specific AUs to be included in the scale they had to have consistently changed within animal at 25% and 50% of observations, respectively. Stating clearly the number of times a change must occur before it becomes part of a facial expression scale helps to reduce the number of items within the scale and provides clear justification for its inclusion.

The RGS (Sotocinal et al., 2011) was developed after trying to use the MGS for rats; as observers became more experienced with the MGS, it was noted that rat's facial expression differed from mice. The area of the nose and cheek would flatten in the rats rather than bulge as it does in mice (Sotocinal et al., 2011; Langford et al., 2010). The RGS also uses only four AUs rather than the five from the MGS, combining nose and cheek flattening as they correlated best with the occurrence of pain. Although different scales (i.e. for different species) share interspecies generic AUs (i.e. orbital tightening), it is important that each species has their own scale developed, specific to them. Using a scale from other species is likely to reduce the validity and reliability of the scale. Even within species, the AUs may be slightly different across ages; the LGS (Guesgen et al., 2016) and SPFES (McLennan et al., 2016) both have the same five areas, but the changes that occur in some areas are slightly different. The ears of lambs point backwards when in pain, whereas in adult sheep the ears rotate ventrally and caudally, and the cheek area in lambs being flattened, whereas in adult sheep the masseter muscle becomes more prominent. These differences may be due to different stimuli being used, but there is also the possibility that animals' express pain differently across life stages. More validation work is required to assess the facial expressions of animals across life stages, between sexes and across different phenotypes in order to ensure the consistency of the scales.

Most scales have followed a scoring system of zero to two for each AU, with zero being an AU "not present", one being "partially or moderately present" and two being "obviously present or present". Häger et al. (2017), included a score of three for the action "Flehming", as the response was not mutually exclusive from the action "head position" (score 2) and so a higher score indicated the severity of the pain expressed through this behaviour. Individual AU scores are assessed to provide an overall facial expression score for each animal at each time point. Some scales use a total pain score, adding up all the individual AU scores at any time point (Dalla Costa et al., 2014; Häger et al., 2017;

McLennan et al., 2016), whilst others, such as the RGS (Sotocinal et al., 2011) use an average score of all units reducing the amount of variation, but limiting the total difference between baseline and pain stimulus. Using a total pain score results in a clearer ability to measure the extent of pain experienced by the animal at a time point and to assess how this might change over time in more detail, thus providing higher sensitivity. The advantage of using the average of the AUs is that it is less sensitive to missing values than the total score (Leach et al., 2012). Missing values can often occur when AUs are not visible due to the orientation of the animals or the contrast in the image or video is too low to distinguish particular AUs (Leach et al., 2012).

4.3 Experimental design and pain stimulus

Progress in animal pain assessment relies upon the development of robust experimental designs (Panksepp, 1998). The facial expression scales developed for different species have varied in experimental design. Better designs have allowed for within-animal comparisons, collecting a baseline score before intervention and then comparing to a known pain state that can be assessed and monitored in response to pain relief given at increasing doses (Langford et al., 2010; Sotocinal et al., 2011). Within-animal designs allow for true changes to be monitored without other variables (e.g. personality, genetics, and previous experience) confounding the results (Dawkins, 2007). There are likely to be differences within each individual's baseline image; many scales note that baseline values are not zero suggesting that certain AUs are more prominent in some individuals than in others. It also suggests that at least some AUs may be visible only momentarily in a non-pain state (e.g. orbital tightening and blinking) (Dalla Costa et al., 2014; Di Giminiani et al., 2016; Guesgen et al., 2016; Keating et al., 2012; MacRae et al., 2018; McLennan et al., 2016; Reijgwart et al., 2017; Sotocinal et al., 2011). Differences between strains and sexes of mice have also been noted (Miller et al., 2015; Miller and Leach, 2015), but this needs to be explored in other species.

Between-animal designs have been used alongside within-animal designs by carefully matching animals with a control group to help further validate the scales (Dalla Costa et al., 2014; Guesgen et al., 2016; Keating et al., 2012; MacRae et al., 2018; McLennan et al., 2016). Establishing that control

496 animals do not change their facial expression over time ensures that changes observed in pain state
497 animals were due to the pain, or the pain being relieved, and not due to general changes in facial
498 expression. This is especially useful in surgical designs where anaesthesia may have an effect on the
499 facial expression, as noted by Dalla Costa et al. (2014) and Miller et al. (2016a, 2015). Scales that
500 have been developed solely based on between-animal designs, such as that used by Holden et al.
501 (2014), allow for differences in general facial expression between pain and non-pain states to be
502 determined. They do not allow however, for a full assessment of the pain experienced in an individual
503 and therefore a number of the measures for validity cannot be effectively assessed.

504 The pain stimulus used when developing the scales has also varied; for many of the laboratory
505 animals the use of validated nociceptive assays such as application of Complete Freund's Adjuvant
506 (CFA), Kaolin, and intra-plantar carrageenan have been routinely utilised (e.g. Langford et al., 2010;
507 Sotocinal et al., 2011). Using an already validated painful stimulus allows for a good level of
508 construct and convergent validity to be assessed (Battini et al., 2016; Calvo et al., 2014). Surgical
509 interventions have also been used to develop facial expression scales, such as laparotomy (Langford et
510 al., 2010), and surgical castration in horses (Dalla Costa et al., 2014). Others have used common
511 husbandry practices such as tattooing (Keating et al., 2012), or tail docking (Di Giminiani et al., 2016;
512 Viscardi et al., 2017), or have used a naturally occurring pain state such as that of a disease
513 (McLennan et al., 2016). It can be argued that the use of natural pain states is better than induced
514 laboratory methods in providing face and predictive validity of a pain assessment measure (Mogil,
515 2009). However, there is likely to be better control in laboratory-based settings with more consistency
516 in the pain stimulus provided, as well as better overall experimental designs that are free from
517 practical restrictions or factors that are unavoidable when working in the field. Di Giminiani et al.
518 (2016), for example, had to collect baseline data from piglets that had already been tooth-clipped a
519 few days before the tail-docking experiment. This could have affected the development of the scale,
520 as the values may not have provided a true baseline if any pain was present due to the tooth-clipping
521 procedure. In their second experiment, animals had not been exposed to any painful stimulus before
522 castration.

Where more than one pain stimulus has been used, or where there is a need to assess the effect of handling, a cross over design can be useful. Gleerup et al. (2015) used a semi-randomised, controlled, cross-over trial to test multiple pain stimuli which included a tourniquet and a topical application of capsaicin. Each horse received noxious stimuli in the same sequence, but with an observer present or not. This allowed for any observer effect on facial expression to be monitored and assessed, as well as testing the effect of different stimuli. Keating et al. (2012) also used a cross-over design to account for the effect of tattooing, handling, and analgesic administration; eight New Zealand rabbits each underwent four different treatments of actual or sham tattooing, with and without prior application of a topical local anaesthetic.

The type of stimulus chosen when developing facial expression scales should be carefully considered. Langford et al. (2010) showed that the action units comprising the mouse grimace scale appeared to be sensitive to 'noxious stimuli of moderate duration' (i.e. more than 10 minutes), and therefore we should be cautious when using this method to assess very acute painful stimuli. Miller and Leach (2014) used the MGS to assess the pain associated with routine ear notching in C57BL6 mice. The authors found no difference between groups that underwent ear notching or not (all animals received the same handling and restraint and the noise of the clipper closing) compared to baseline. They suggested that the lack of change in the MGS may have been due to the potentially acute nature of this noxious stimulus. A similar finding was seen by Williams et al. (2008) when using ultrasonic vocalisations to assess pain following ear notching in C57BL6 mice. Sotocinal et al. (2011), also noted that the pain facial expression in rats did not last for more than 48 hours, which they suggested was a natural limiting factor imposed by facial expression itself, especially in chronic pain. This was further supported by Whittaker and colleagues who showed no change in the rat grimace scale when used to assess the more chronic pain associated with chemotherapy-induced mucositis (Whittaker et al., 2016). Animals suffering from chronic pain are unlikely to maintain a certain expression in the long term as pain can fluctuate over time (Baliki et al., 2006; Kunz et al., 2011). Additionally, other factors such as the presence of a male observer or even simply a t-shirt worn by a male observer the previous night, has been shown to inhibit the facial expression of pain (Sorge et al., 2014).

The correct emotional construct should also be assessed with a particular scale; Finlayson et al. (2016) used the RGS to assess for positive indicators in rats compared with a contrast stimulus, but did not employ painful stimuli. There were no differences in RGS scores between the conditions, which shows that the RGS has good discriminant validity as it did not increase in intensity in non-painful situations. However, this was an incorrect use of the scale as it was used to measure something for which it had not been designed. Dalla Costa et al. (2017) found the HGS score was not influenced by positive or negative emotional states other than pain, inferring that the HGS is a specific tool for assessing pain. Further testing of many of the facial expression scales is still required to ensure discriminant validity.

4.4 Provision of analgesia

A key component in assessing internal validity of a pain assessment tool is to assess the effect of analgesia in a dose dependent manner, and to measure the changes that occur in the measurement tool (Sotocinal et al., 2011). These changes should show that by providing analgesia in this manner, there is a predicted consistent gradual decrease in the facial expression score as the dose of pain relief increases. During the development of facial expression scales there have been a range of ways in which analgesia has been administered, and only two of the scales, the MGS, and the RGS, have provided pain relief in this dose dependent manner during developmental stages (Langford et al., 2010; Sotocinal et al., 2011). Both of these scales observed significant dose dependent changes in the expression of mice (Langford et al., 2010; Matsumiya et al., 2012) and rats (Sotocinal et al., 2011). These results demonstrate that these scales are effective and valid at measuring the pain experienced by these animals.

It is not always possible to test the effect of analgesia in a dose-dependent manner for several reasons. Obtaining ethical approval for dose-dependent facial expression testing and observation of pain may not be possible. This is especially true in non-laboratory contexts where scales may be developed as part of observations of naturally occurring pain states (McLennan et al., 2016; van Loon and VanDierendonck, 2015), or in experiments where protocols and procedures cannot be changed (Häger

et al., 2017; Reijgwart et al., 2017). This has resulted in a variety of protocols used when giving analgesia in the different studies. In these circumstances careful consideration must be given as to when analgesia is provided, and when to best capture facial images so that a true baseline and a true pain state are available. Different groups of animals may be needed to receive differing levels of analgesia or other classes of analgesics to help maintain good animal welfare. This can also provide information about the effects of analgesics on facial expression. The HGS (Dalla Costa et al., 2014) for example, was developed with horses undergoing surgical castration. For ethical and welfare reasons perioperative analgesia was provided to two groups of horses, with one group provided with additional analgesia orally 6 hours after the surgery. Pain-free images obtained before the surgery were compared with images captured 8 hours after the surgery and after analgesia had been administered for both groups. Although there were significant differences between control and castrated horses' facial expression scores, there were no differences in facial expression between the two post-castration groups despite the additional analgesia provided to one group 2 hours before the image capture. These authors state that it is not currently possible to differentiate between post-procedure pain and distress, meaning validation of the scale is not complete. This is also the case for the EQUUS-FAP and SGS, in which images for pain states were taken after analgesia had been provided. The EQUUS-FAP used images from horses suffering from colic that had been provided with NSAIDs upon arrival at the hospital, and horses were only removed from the study if they needed further analgesia (van Loon and VanDierendonck, 2015). The SGS (Häger et al., 2017) was unable to assess the true pain state of the sheep after surgery as animals were provided with analgesia on a daily basis for up to 13 days after surgery. Such methods make it impossible to compare fully painful and pain free states to validate the scales.

For species which have no known effective analgesic drug, or where the evidence of the effectiveness of the drug is contradictory, full validation can be difficult. During the development of the SPFES, half the diseased sheep were treated with antibiotics and a NSAID, whilst the other half received antibiotics only, in line with current industry practices (McLennan et al., 2016). Images were collected before the analgesia was provided when pain was expected to be at its highest, and again at

42 or 90 days after initial treatment by which time the disease had resolved. No differences in facial expression were found between the two different treatment groups, although there were differences between control and diseased sheep. Although a true pain state and baseline were captured, the effect of the NSAID was not captured as this is considered to be most effective for only 72 hours (Shukla et al., 2007). There is a need for more research for species where there is a lack of information on the most effective pain relief, and what dosages and time intervals to use. Indeed, Matsumiya et al. (2012) found that the dose of a drug required to make maximal change in facial expression of mice was higher than that currently advised, and some drugs not effective at reducing the facial expression of pain in these animals.

4.5 Other pain assessment tools for comparison

To show convergent validity of the scales, many researchers have incorporated other pain assessment tools for comparison with the facial expression scores (Dalla Costa et al., 2014; Langford et al., 2010; McLennan et al., 2016; Sotocinal et al., 2011). This has allowed assessment of degree of correlation. Behavioural and physiological measures, including spontaneous behaviours and cortisol concentrations, were the most frequent other pain assessment tools utilised (Dalla Costa et al., 2014; Di Giminiani et al., 2016; Gleerup et al., 2015; Häger et al., 2017; Keating et al., 2012; Langford et al., 2010; MacRae et al., 2018; Sotocinal et al., 2011). Each of these measures has its own validity issues which should be considered when interpreting correlations between measures. For scales developed not using contrived pain states, veterinary and other subjective assessments of pain experienced by the animal were utilised instead. Subjective assessments such as those carried out by veterinarians, although useful, have limited validity in correlation studies (Weary et al., 2006). It is important to use measures that have already been tested and validated for animal pain. Good correlation between measures with the facial expression scale will support validity, therefore it is essential that careful consideration is given to what measures are the most suitable. For example, Leach and colleagues showed a high positive correlation between changes in validated spontaneous pain behaviours and the MGS (Leach et al., 2012), whilst McLennan et al. (2016) correlated lameness and lesion scores of footrot, previously validated as painful by Ley et al. (1995), with the SPFES.

630 The reliability and repeatability of the tool is assessed during testing of the scale (see table 1 for testing values of current scales). This is carried out by using
631 time- and treatment-blind observers who have undergone some level of training. These scorers are asked to assess one or two photographs or video stills of
632 each animal for each AU of the scale, providing a score of 0 to 2 as detailed previously. The scores from each observer are compared for consistency between
633 observers in most studies by using an Intraclass Correlation Coefficient (ICC) (Dalla Costa et al., 2014; Di Giminiani et al., 2016; Häger et al., 2017; Keating
634 et al., 2012; Langford et al., 2010; McLennan et al., 2016; Reijgwart et al., 2017; Sotocinal et al., 2011; van Loon and VanDierendonck, 2015). In addition to
635 scoring each individual unit, observers are often also asked to provide a global pain assessment of the facial expression based on their own experience and
636 expertise, and to make a judgement about how much pain they think the animal has (Dalla Costa et al., 2014; Holden et al., 2014; Keating et al., 2012;
637 Langford et al., 2010; McLennan et al., 2016; Reijgwart et al., 2017; Sotocinal et al., 2011). These subjective decisions are often used to calculate an overall
638 degree of accuracy, testing for how many of the images were assessed correctly as being in pain or not. McLennan et al. (2016) found that accuracy improved
639 greatly when using the total pain score rather than the global assessment of pain. Removing the need to make a decision about an animal's affective state and
640 simply assessing the individual AUs was a more accurate way of identifying sheep in pain. McLennan et al. (2016) were also able to provide guidance on
641 when analgesia should be considered by analysing the sensitivity and specificity of each level of total pain score against a lameness score (a valid pain
642 indicator for sheep with footrot (Ley et al., 1995)), something which is missing from other scales. In 2018, Dalla Costa and colleagues proposed a statistical
643 approach to identifying a classifier that can estimate the pain status of the animal based on AUs included in HGS and MGS. They found that AUs can be
644 weighted to best estimate the pain condition of an animal (Dalla Costa et al., 2018). These results provide support for using the facial expression scales and
645 not relying on an overall judgement of pain simply based on experience.

The way in which testing has been carried out is fairly standard across the scales, except for the number of blinded observers used which has ranged from 2 (MacRae et al., 2018) to 68 (Holden et al., 2014). Developing a scale based on initial observations and subsequent scoring by a small number of observers is unlikely to represent an objective or valid scale (due to the risk of observer bias) as it may not provide effective indices visible to all. The more observers that can be utilised during testing, the more likely that the scale represents actual objective changes and also allows any problems requiring further development of the scale to be identified. The FGS (Reijgwart et al., 2017) employed a new testing technique by providing each of their 11 blinded observers with a seven-part survey which included a week between AUs. This meant that each observer carried out a global assessment before any training, and was then trained for each AU separately before scoring that AU. They then carried out an additional global assessment. They were able to test both inter- and intra-observer reliability, with the effect of training. They had good results for both the inter-observer (ICC=0.89 pre-training, and ICC=0.89 post-training) and for the intra-observer reliability (ICC=0.67). The lower level of ICC for intra-observer reliability test was attributed to the effect of training on the observers. The authors also stated that there were fewer missing overall pain scores suggesting there was an improved confidence by the observers to assign a pain score after viewing examples of each AU providing support for the need for careful training when using these scales.

Reijgwart et al. (2017) also discuss the effect of ferret coat type on observers' ability to assign scores to certain areas of the face. They report that observers had more difficulty assigning scores to long-haired ferrets. The current authors have all experienced difficulties in assessing the facial expression of longer-haired, muscular, darker-haired animals and animals that have had a mixed colouring on their face, as it can be difficult to determine if certain features are changing due to shadow or some other element. Further research is needed into the effect of hair length, muscularity, and coat colour on the ability of observers to assess the facial expression of animals.

4.7 Feasibility testing

Evidence of feasibility of the facial expression scales is, in our opinion, delaying the full utilisation of the scales as a pain assessment tool. The MGS and the RGS have been the only scales so far tested for feasibility with live-scoring compared with retrospective footage analysis. Miller and Leach (2015) compared live-scoring with a photographic data collection of baseline images of different strains and sexes of mice on three separate occasions. They directly compared the 10-minute live-scoring with the photographic data that were collected at the same time. They found that for the female mice photographic scores were significantly higher than the live-scores for all four strains tested. For the male mice there were differences between strains, with C57BL/6 mice scoring higher in photographs than in live observations. C3H/He male mice did not have significantly different scores between the two methods. These were baseline scores in which there was no pain so there should not be a significant difference; however, there may be particular phenotypic features of certain strains that are more difficult to score live than through photographs. When live-scoring the observer would look at the mouse for 5 seconds and then award the appropriate score for each facial AU. This was carried out on three occasions at the beginning, mid-point and end of a ten-minute period. In contrast, the photographs which were used for retrospective analysis were taken across the same ten-minute period whenever the mouse was facing the camera. The different methods and timing of collection may have led to observer bias (systematic timings compared with right position), or possibly the facial expression changed due to different activities (chewing, sniffing, exploration, walking, etc.) being performed between live scoring and the photographic image scoring.

Leung et al. (2016) tested the feasibility of the RGS to accurately assess pain in rats by comparing the standard method of image assessment with real-time observations (interval and point). Real-time observations were carried out at the same time as the video footage to allow for direct comparisons as in the study by Miller and Leach (2015). Leung et al. (2016) used two different methods that were repeated every 30 seconds for 10-minutes of observations: 1) a point observation that was alternated with, 2) 15 second interval observations where the animal was observed for 15 seconds and assigned a single score for the period. Scores were averaged at three-minute intervals to produce three single scores which were then averaged again to produce a single score across the 10-

680 minute period, as in the standard method for the RGS (Sotocinal et al., 2011) allowing for direct comparison. To assess whether the length of observation
681 period also made a difference, real-time observation scores were averaged from the first five and two minutes of observations. They found that there was good
682 agreement between the real-time scores and the standard method, with most of the real-time observations able to discriminate between treatment groups.
683 Interval observations were found to be more sensitive than point observations, and multiple observations were better at correctly predicting treatment groups
684 than single observations. These results suggest a single observation should not be relied upon when making treatment decisions. Longer observation periods
685 (5-minute) were found to also provide a better assessment of the pain, and were considered to provide a good practical balance to assessment of pain. In
686 horses, short video-clips (15-seconds) were scored using the HGS and then compared to HGS scores from still images (Dalla Costa et al., 2016). No
687 significant differences in HGS total scores between the scoring of still images and video sequences were found. However, the 15-second video clips were
688 reported as being more difficult to score, with a high level of variation between the observers.

689 These results demonstrate that facial expression scales could be utilised in real-time pain assessment. More research into this area is required to fully
690 understand whether this difference between live scores and retrospective scoring is due to difficulty in live scoring, or whether there are advantages over
691 being able to pause and choose the right moment to observe the facial images. There is also a clear advantage to making multiple observations over a longer
692 period of time rather than using just one short observation period. When it comes to assessing pain, its fluctuating nature and the influence of a number of
693 factors that cannot be controlled for, need to be considered.

694 Table 2 provides guidelines for best practice in developing and validating any future facial expression scales, with particular consideration of pain. Scales
695 need to be developed for each species, across key life stages and potentially with the inclusion of differing phenotypic features such as those found across

breeds, especially if these differ significantly. The majority of scales require full feasibility testing and this should be incorporated into the further development of current and future scales.

5. Clinical applications of facial expression as a method of pain assessment

When making a decision about a patient, whether it be in clinical practice or a research setting, being able to assess the severity of the pain is vital to improving their welfare (Ashley et al., 2005). Many of the current methods of pain assessment are not clinically relevant; many are retrospective, time consuming, and require the caregiver to make a subjective judgement about whether pain relief should be provided or not (Egger et al., 2014; Leach et al., 2009). Variations among clinicians on the level of pain they believe an animal may suffer and differences in empathy levels, which play a role in whether or not pain relief is provided to these animals, make such assessments unreliable (Bell et al., 2014; Huxley and Whay, 2006; Ison and Rutherford, 2014; Norring et al., 2014). Inconsistent care due to a lack of ability to recognise and evaluate pain is a significant factor reported by veterinarians as a reason why pain relief is not provided (Richardson and Flecknell, 2005), resulting in poor welfare.

Much of the research into facial expression has focused on the development of scales within a research setting. Few have undergone full feasibility testing and many of the scales are not yet widely utilised in clinical practice; however, they have been developed with clinical relevance in mind and have been demonstrated to be reliable and valid measures of pain. Many scales were developed using experimental designs based on clinical procedures (Dalla Costa et al., 2014; Di Giminiani et al., 2016; Häger et al., 2017; Keating et al., 2012; Langford et al., 2010; MacRae et al., 2018; Reijgwart et al., 2017; Sotocinal et al., 2011), or naturally occurring diseases (Holden et al., 2014; McLennan et al., 2016) suggesting feasibility, but testing is still required. Testing feasibility can be carried out in the research setting, but the true feasibility and value of the scales will come from clinicians using them in real life settings and feeding

back to authors. Uptake of the facial expression scales to assess pain in real-time within clinical practice is likely to increase as more data demonstrating their feasibility become available.

Despite the current lack of full feasibility testing, there are numerous advantages to using facial expression over other pain assessment methods in clinical and research practice. Facial expression has been shown to be a tool that can help to alleviate a number of the problems associated with other pain assessment techniques. Minimal training (simply providing the scale and descriptions for observers to read themselves) is all that is required to be effective and reliable at using facial expression to recognise and evaluate pain in animals (Dalla Costa et al., 2014; Keating et al., 2012; Langford et al., 2010; McLennan et al., 2016; Sotocinal et al., 2011); however, sensitivity and specificity is likely to improve with more detailed and structured training (Reijgwart et al., 2017). Continued training and reliability testing within a practice will help to ensure that the scale is sensitive, and will increase the confidence of staff concerning the uniformity of its application and the provision of pain relief. The training, and guides placed in pertinent areas would encourage all those involved in animal care to assess the pain on a regular basis. This is irrespective of whether they be in veterinary practice or others working with animals in different settings.

Assessing pain using facial expression does not require any specialist equipment to be bought, and should be possible to carry out in real time. Many of the scales are concise with only a few measurements needed; the majority have five areas of the face to assess for three possible outcomes. This should mean that the assessment is quick and effective to carry out. Leach et al. (2011), highlight that observers are naturally drawn to the face, and so facial expression scoring takes advantage of this. Once clinical staff are familiar with facial expression scales, the scales should become quicker and easier to use in assessing pain. It is likely that once a practitioner has become familiar with one or two different species facial expression scales, they will be able to apply similar principles to other patients that they treat because of the consistency in facial expression across mammals (Chambers and Mogil, 2015).

It is important that the facial expression scale should initially form part of a wider assessment of pain, although it could be used alone. Observing the individual AUs of the face gives an indication of potential pain severity, particularly when a total pain score is used. Observing other areas of the body and the behaviour as well as monitoring physiological signs will give a more rounded picture of the pain experience, including potential causes and the site of pain. Once the cause and site of pain has been identified, subsequent assessments should focus on the facial expression of pain, assessing the emotional impact of the experience on the animal. It is important to assess the ongoing state of the animal over time as pain fluctuates (Baliki et al., 2006). Monitoring how frequent this fluctuation is, or whether there is a high level of constant pain is key to improved welfare. It should be remembered that individuals experience pain differently and have different thresholds (Bateson, 1991; Gentle, 2001; Martuscello et al., 2013), different coping abilities (Koolhaas et al., 1999), and different genetics that can alter the effectiveness of analgesia (Mogil, 1999; Mogil et al., 1996). Carrying out assessment of the facial expression over a period of time and displaying records of scores alongside the animal, will help to build a better understanding of how the animal is coping and whether there needs to be change in pain management strategy. The current authors recommend that facial expression scoring is not used for patients with head injuries or pathological changes to the head or face, as the AUs displayed may be affected by the trauma itself, and so other measures are needed.

Intervention scores which provide guidance on when to give analgesia are still lacking for many of the scales, which again may limit their use in practice and thus limiting when and how pain is managed. McLennan et al. (2016) and Oliver et al. (2014) demonstrated that it is possible to provide guidance of when to consider providing analgesia in sheep and rats respectively, identifying a total pain score that when reached is highly suggestive that the animal is in pain. When assessing the ongoing treatment or monitoring of a patient there also needs to be some guidance as to when a change in score would be relevant. In humans, a change in a total pain score by 2 points or more is considered to be clinically important (Farrar et al., 2001). Further research should effectively determine a threshold and intervention score for each species.

745 Table 3 provides guidelines for best practice for use of facial expression scales as a pain assessment tool in clinical practice. This guide is to ensure the
746 validity and reliability of the current scales that have been developed, and to encourage the uptake of the scales in clinical practice, as well as by those who
747 are involved in the day to day care of animals. If all staff consistently assess the pain of a patient and record the pain score each time they are involved in any
748 sort of care, it is more likely that signs of spontaneous pain will be recognised (Mogil and Crager, 2004). It will also allow for a continued assessment of the
749 animal's recovery and the effectiveness of any analgesia provided. If clinical, research and animal care staff can be encouraged to use and provide feedback
750 on current facial expression scales, scales will be improved and key areas for further research will be identified.

751 **6. Conclusion**

752 The accurate assessment and management of animal pain is essential in ensuring good animal welfare. The inability of animals to articulate experience of pain
753 means that the nature of pain in animals remains controversial for some people. However, there is increasing acceptance that vertebrates and some
754 invertebrate animals are sentient beings, capable of experiencing affective states such as pain. Yet pain remains a significant welfare issue. The recognition
755 and evaluation of pain remains a major limiting factor in pain management for humans and non-humans. There is good evidence that facial expression can be
756 a useful, valid and reliable tool for recognising and evaluating pain in humans and other animals. Both the sensory and emotional components of pain have
757 been demonstrated to affect facial expression, which thus gives a true representation of the affective state of the animal. Many of the mammalian species
758 studied to date have similar facial expression responses to pain. Animal care staff need to be trained to use the appropriate scale for the species under their
759 care.

There is a need for continued development of the currently available facial expression scales. Further testing is required to ensure the validity of the current scales. In particular, many of the scales need feasibility testing and refinement before they can be fully utilised in clinical and field settings. It is imperative that scientists work closely with clinicians in this testing to ensure continued reliability. Most scales developed to date have been developed using only one or two causes of pain. It is important that the scales are validated for other causes of pain before they are applied in different clinical conditions. There may, for example, be different responses to acute pain and to chronic pain in the same species. Young animals may have different responses from adults. The effect of other affective states such as fear or malaise also needs to be assessed, as these may interact with or obscure facial expression of pain. Future work should also consider whether facial expressions of pain have any communicative function. The development of new scales is needed for other species under the care of humans. Facial expression pain scales are already being used in assessment of animal welfare, but further work on facial expression is likely to see many new applications for this approach.

ACKNOWLEDGEMENTS

Research presented in this paper by KMM, DMB, MJC, DS, EDC was funded by the EU VII Framework program (FP7-KBBE-2010-4), Animal Welfare Indicators (AWIN) project. ALM and MCL are funded by the NC3Rs. The funders had no role in the preparation of this manuscript or decision to publish.

We would also like to thank the two anonymous reviewers for the original manuscript that provided positive, constructive comments on how to improve it. Your comments, and time taken to review our work, were much appreciated.

CONFLICT OF INTEREST

There are no conflicts of interest.

References

Abu-Serriah, M., Nolan, A.M., Dolan, S., 2007. Pain assessment following experimental maxillofacial surgical procedure in sheep. Lab. Anim. 41, 345–52. https://doi.org/10.1258/002367707781282794

Ahola Kohut, S., Pillai Riddell, R., 2009. Does the Neonatal Facial Coding System Differentiate Between Infants Experiencing Pain-Related and Non-Pain-Related Distress? J. Pain 10, 214–220. https://doi.org/10.1016/j.jpain.2008.08.010

Ahola Kohut, S., Pillai Riddell, R., Flora, D.B., Oster, H., 2012. A longitudinal analysis of the development of infant facial expressions in response to acute pain: immediate and regulatory expressions. Pain 153, 2458–65. https://doi.org/10.1016/j.pain.2012.09.005

Andersen, M.L., Tufik, S., 2003. Sleep patterns over 21-day period in rats with chronic constriction of sciatic nerve. Brain Res. 984, 84–92. https://doi.org/10.1016/S0006-8993(03)03095-6

Apkarian, A.V., Scholz, J., 2006. Shared mechanisms between chronic pain and neurodegenerative disease. Drug Discov. Today Dis. Mech. 3, 319–326. https://doi.org/10.1016/j.ddmec.2006.09.006

Ashley, F.H., Waterman-Pearson, A.E., Whay, H.R., 2005. Review Article. Behavioural assessment of pain in horses and donkeys: application to clinical

practice and future studies. Equine Vet. J. 37, 565–575.

Baliki, M.N., Chialvo, D.R., Geha, P.Y., Levy, R.M., Harden, R.N., Parrish, T.B., Apkarian, A.V., 2006. Chronic pain and the emotional brain: specific brain activity associated with spontaneous fluctuations of intensity of chronic back pain. J. Neurosci. 26, 12165–73. https://doi.org/10.1523/JNEUROSCI.3576-06.2006

Bateson, P., 1991. Assessment of pain in animals. Anim. Behav. 42, 827–839. https://doi.org/10.1016/S0003-3472(05)80127-7

Battini, M., Barbieri, S., Waiblinger, S., Mattiello, S., 2016. Validity and feasibility of Human-Animal Relationship tests for on-farm welfare assessment in dairy goats. Appl. Anim. Behav. Sci. 178, 32–39. https://doi.org/10.1016/j.applanim.2016.03.012

Becker, J., Reist, M., Friedli, K., Strabel, D., Wüthrich, M., Steiner, A., 2013. Current attitudes of bovine practitioners, claw-trimmers and farmers in Switzerland to pain and painful interventions in the feet in dairy cattle. Vet. J. 196, 467–476. https://doi.org/10.1016/j.tvjl.2012.12.021

Beggs, S., Currie, G., Salter, M.W., Fitzgerald, M., Walker, S.M., 2012. Priming of adult pain responses by neonatal pain experience: Maintenance by central neuroimmune activity. Brain 135, 404–417. https://doi.org/10.1093/brain/awr288

Bell, A., Helm, J., Reid, J., 2014. Veterinarians' attitudes to chronic pain in dogs. Vet. Rec. 175, 428. https://doi.org/10.1136/vr.102352

Benatti, C., Alboni, S., Capone, G., Corsini, D., Caggia, F., Brunello, N., Tascedda, F., Blom, J.M.C., 2009. Early neonatal inflammation affects adult pain reactivity and anxiety related traits in mice: genetic background counts. Int. J. Dev. Neurosci. 27, 661–668. https://doi.org/10.1016/j.ijdevneu.2009.07.009

804 Bermond, B., 2001. A Neuropsychological and Evolutionary Approach to Animal Consciousness and Animal Suffering. Anim. Welf. 10, S47-62.

805 Berthier, M., Starkstein, A., Leiguarda, R., 1988. Asymbolia for Pain: A Sensory-Limbic Disconnection Syndrome. Ann. Neurol. 24, 41–49.

806 Boerner, K.E., Chambers, C.T., Craig, K.D., Pillai Riddell, R.R., Parker, J.A., 2013. Caregiver accuracy in detecting deception in facial expressions of pain in
807 children. Pain 154, 525–33. https://doi.org/10.1016/j.pain.2012.12.015

808 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
809 UNESP-Botucatu multidimensional composite pain scale for assessing postoperative pain in cats. BMC Vet. Res. 9, 143. https://doi.org/10.1186/1746-
810 6148-9-143

811 Brooks, J., Tracey, I., 2005. From nociception to pain perception: Imaging the spinal and supraspinal pathways. J. Anat. 207, 19–33.
812 https://doi.org/10.1111/j.1469-7580.2005.00428.x

813 Broom, D.M., 2001. The evolution of pain. Vlaams Diergeneeskd. Tijdschr. 70, 17–21.

814 Broom, D.M., 2014. Sentience and Animal Welfare. CABI Publishing, Wallingford.

815 Broom, D.M., 2016. Fish brains and behaviour indicate capacity for feeling pain. Anim. Sentience 2016.010.

816 Busch, G., Weary, D.M., Spiller, A., Von Keyserlingk, M.A.G., 2017. American and German attitudes towards cow-calf separation on dairy farms. PLoS One
817 12, 1–20. https://doi.org/10.1371/journal.pone.0174013

818 Bussières, G., Jacques, C., Lainay, O., Beauchamp, G., Leblond, A., Cadoré, J.-L., Desmaizières, L.-M., Cuvelliez, S.G., Troncy, E., 2008. Development of a
819 composite orthopaedic pain scale in horses. Res. Vet. Sci. 85, 294–306. https://doi.org/10.1016/j.rvsc.2007.10.011

820 Caeiro, C.C., Burrows, A.M., Waller, B.M., 2017. Development and application of CatFACS: Are human cat adopters influenced by cat facial expressions?
821 Appl. Anim. Behav. Sci. 189, 66–78. https://doi.org/10.1016/j.applanim.2017.01.005

822 Calvo, G., Holden, E., Reid, J., Scott, E.M., Firth, A., Bell, A., Robertson, S., Nolan, A.M., 2014. Development of a behaviour-based measurement tool with
823 defined intervention level for assessing acute pain in cats. J. Small Anim. Pract. 55, 622–629. https://doi.org/10.1111/jsap.12280

824 Carlson, A.A., Manser, M.B., Young, A.J., Russell, A.F., Jordan, N.R., McNeilly, A.S., Clutton-Brock, T., 2006. Cortisol levels are positively associated with
825 pup-feeding rates in male meerkats. Proc. R. Soc. B Biol. Sci. 273, 571–577. https://doi.org/10.1098/rspb.2005.3087

826 Chambers, C.T., Mogil, J.S., 2015. Ontogeny and phylogeny of facial expression of pain. Pain 156, 798–799.
827 https://doi.org/10.1097/j.pain.0000000000000133

828 Clark, C., Murrell, J., Fernyhough, M., O'Rourke, T., Mendl, M., 2014. Long-term and trans-generational effects of neonatal experience on sheep behaviour.
829 Biol. Lett. 10, 20140273–20140273. https://doi.org/10.1098/rsbl.2014.0273

830 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. J.
831 Equine Vet. Sci. 54, 114–117. https://doi.org/10.1016/j.jevs.2017.03.221

832 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, 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, e0200339. 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. https://doi.org/10.3390/ani6080047

Dawkins, M.S., 2007. Observing Animal Behaviour. Design and Analysis of Quantitative Data. Oxford University Press, Inc., Oxford, UK.

Dawkins, M.S., 2008. The science of animal suffering. Ethology 114, 937–945. https://doi.org/10.1111/j.1439-0310.2008.01557.x

de Oliveira, F.A., Luna, S.P.L., do Amaral, J.B., Rodrigues, K.A., Sant'Anna, A.C., Daolio, M., Brondani, J.T., 2014. Validation of the UNESP-Botucatu unidimensional composite pain scale for assessing postoperative pain in cattle. BMC Vet. Res. 10, 200. https://doi.org/10.1186/s12917-014-0200-0

Descovich, K.A., Wathan, J., Leach, M.C., Buchanan-smith, H.M., Flecknell, P., Farningham, D., Vick, S., 2017. Review article Facial expression : An under-utilised tool for the assessment of welfare in mammals. ALTEX 34, 409–429. 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. Front. Vet. Sci. 3, 1–10. https://doi.org/10.3389/fvets.2016.00100

847 Doughty, A.K., Coleman, G.J., Hinch, G.N., Doyle, R.E., 2017. Stakeholder perceptions of welfare issues and indicators for extensively managed sheep in
848 Australia. Animals 7, 1–24. https://doi.org/10.3390/ani7040028

849 Egger, C.M., Love, L., Doherty, T., 2014. Pain Management in Veterinary Practice. Wiley-Blackwell, Iowa, USA.

850 Ekman, P., Friesen, W., 1978. Facial action coding system: a technique for the measurement of facial action. Consulting Psychologists Press, Palo Alton,
851 Californi.

852 Ellen, Y., Flecknell, P., Leach, M., 2016. Evaluation of using behavioural changes to assess post-operative pain in the Guinea pig (cavia porcellus). PLoS One
853 11, 1–20. https://doi.org/10.1371/journal.pone.0161941

854 Farrar, J.T., Young, J.P., La Moreaux, L., Werth, J.L., Poole, R.M., 2001. Clinical importance of changes in chronic pain intensity measures on an 11-point
855 numerical scale. Pain 94, 149–158.

856 Feinstein, J.S., Khalsa, S.S., Salomons, T. V., Prkachin, K.M., Frey-Law, L.A., Lee, J.E., Tranel, D., Rudrauf, D., 2016. Preserved emotional awareness of
857 pain in a patient with extensive bilateral damage to the insula, anterior cingulate, and amygdala. Brain Struct. Funct. 221, 1499–1511.
858 https://doi.org/10.1007/s00429-014-0986-3

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

861 Flecknell, P., Leach, M., Bateson, M., 2011. Affective state and quality of life in mice. Pain 152, 963–4. https://doi.org/10.1016/j.pain.2011.01.030

Flecknell, P.A., 2000. Animal Pain - an Introduction, in: Pain Management in Animals. W B Saunders, London, United Kingdom, pp. 1–7.

Fleetwood-Walker, S., Sun, L., Jerina, H., Mitchell, R., 2012. Assessment of animal pain and mechanism-based strategies for its reversal. Vet. J. 193, 305–6.

Fredriksen, B., Johnsen, A.M.S., Skuterud, E., 2011. Consumer attitudes towards castration of piglets and alternatives to surgical castration. Res. Vet. Sci. 90, 352–357. https://doi.org/10.1016/j.rvsc.2010.06.018

Gear, R.W., Levine, J.D., 2011. Nucleus accumbens facilitates nociception. Exp. Neurol. 229, 502–506. https://doi.org/10.1016/j.expneurol.2011.03.021

Gentle, M.J., 2001. Attentional Shifts Alter Pain Perception in the Chicken. Anim. Welf. 10, 187–194.

Gleerup, K.B., Forkman, B., Lindegaard, C., Andersen, P.H., 2015. An equine pain face. Vet. Anaesth. Analg. 42, 103–114. https://doi.org/10.1111/vaa.12212

Gougoulis, D., Kyriazakis, I., Fthenakis, G., 2010. Diagnostic significance of behaviour changes of sheep: A selected review. Small Rumin. Res. 92, 52–56. https://doi.org/10.1016/j.smallrumres.2010.04.018

Green, L.E., Hedges, V.J., Schukken, Y.H., Blowey, R.W., Packington, A.J., 2002. The impact of clinical lameness on the milk yield of dairy cows. J. Dairy Sci. 85, 2250–2256.

Groh, A., Krieger, P., Mease, R.A., Henderson, L., 2017. Acute and chronic pain processing in the thalamocortical system of humans and animal models. Neuroscience 387, 58–71. https://doi.org/10.1016/j.neuroscience.2017.09.042

876 Gu, X., Hof, P.R., Friston, K.J., Fan, J., 2013. Anterior insular cortex and emotional awareness. J. Comp. Neurol. 521, 3371–3388.
877 https://doi.org/10.1002/cne.23368

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

880 Guesgen, M.J., Beausoleil, N.J., Minot, E.O., Stewart, M., Jones, G., Stafford, K.J., 2011. The effects of age and sex on pain sensitivity in young lambs. Appl.
881 Anim. Behav. Sci. 135, 51–56. https://doi.org/10.1016/j.applanim.2011.09.008

882 Guesgen, M.J., Beausoleil, N.J., Minot, E.O., Stewart, M., Stafford, K.J., 2014. Social context and other factors influence the behavioural expression of pain
883 by lambs. Appl. Anim. Behav. Sci. 159, 41–49. https://doi.org/10.1016/j.applanim.2014.07.008

884 Guesgen, M.J., Beausoleil, N.J., Stewart, M., 2013. Effects of early human handling on the pain sensitivity of young lambs. Vet. Anaesth. Analg. 40, 55–62.
885 https://doi.org/10.1111/j.1467-2995.2012.00746.x

886 Häger, C., Biernot, S., Buettner, M., Glage, S., Keubler, L.M., Held, N., Bleich, E.M., Otto, K., Müller, C.W., Decker, S., Talbot, S.R., Bleich, A., 2017. The
887 Sheep Grimace Scale as an indicator of post-operative distress and pain in laboratory sheep. PLoS One 12, 1–15.
888 https://doi.org/10.1371/journal.pone.0175839

889 Hawkins, P., 2014. Refining housing, husbandry and care for animals used in studies involving biotelemetry. Animals 4, 361–373.
890 https://doi.org/10.3390/ani4020361

Heinricher, M.M., Tavares, I., Leith, J.L., Lumb, B.M., 2009. Descending control of nociception: Specificity, recruitment and plasticity. Brain Res. Rev. 60, 214–225. https://doi.org/10.1016/j.brainresrev.2008.12.009

Hofbauer, R.K., Rainville, P., Duncan, G.H., Bushnell, M.C., 2001. Cortical Representation of the Sensory Dimension of Pain. J. Neurophysiol. 86, 402-.

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. J. Small Anim. Pract. 55, 615–21. https://doi.org/10.1111/jsap.12283

Hunter, M.C., Smith, R.G., Schipanski, M.E., Atwood, L.W., Mortensen, D.A., 2017. Agriculture in 2050: Recalibrating Targets for Sustainable Intensification. Bioscience 67, 386–391. https://doi.org/https://doi.org/10.1093/biosci/bix010

Huxley, J.N., Whay, H.R., 2006. Current attitudes of cattle practitioners to pain and the use of analgesics in cattle. Vet. Rec. 159, 662–668.

IASP, 1994. Part III: Pain Terms, A Current List with Definitions and Notes on Usage, in: Merskey, H., Bogduk, N. (Eds.), Classification of Chronic Pain. IASP Press, Seattle, pp. 209–214.

Ijichi, C., Collins, L.M., Elwood, R.W., 2014. Pain expression is linked to personality in horses. Appl. Anim. Behav. Sci. 152, 38–43. https://doi.org/10.1016/j.applanim.2013.12.007

Ingvartsen, K.L., Moyes, K., 2013. Nutrition, immune function and health of dairy cattle. Animal 7, 112–122. https://doi.org/10.1017/S175173111200170X

Ison, S.H., Rutherford, K.M.D., 2014. Attitudes of farmers and veterinarians towards pain and the use of pain relief in pigs. Vet. J. 202, 622–7.

https://doi.org/10.1016/j.tvjl.2014.10.003

Jaremka, L.M., Collins, N.L., 2017. Cortisol increases in response to brief social exchanges with opposite sex partners. Biol. Psychol. 124, 39–46. https://doi.org/10.1016/j.biopsycho.2017.01.001

Jasmin, L., Granato, A., Ohara, P.T., 2004. Rostral Agranular Insular Cortex and Pain Areas of the Central Nervous System: A Tract-Tracing Study in the Rat. J. Comp. Neurol. 468, 425–440. https://doi.org/10.1002/cne.10978

Julle-Danière, É., Micheletta, J., Whitehouse, J., Joly, M., Gass, C., Burrows, A.M., Waller, B.M., 2015. *MaqFACS* (Macaque Facial Action Coding System) can be used to document facial movements in Barbary macaques (*Macaca sylvanus)*. PeerJ 3, e1248. https://doi.org/10.7717/peerj.1248

Keating, S.C.J., Thomas, A., Flecknell, P., 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, e44437. https://doi.org/10.1371/journal.pone.0044437

Key, B., 2016. Why fish do not feel pain. Anim. Sentience 3, 1–33.

Kiddie, J.L., Collins, L.M., 2014. Development and validation of a quality of life assessment tool for use in kennelled dogs (Canis familiaris). Appl. Anim. Behav. Sci. 158, 57–68. https://doi.org/10.1016/j.applanim.2014.05.008

Koolhaas, J.M., Korte, S.M., De Boer, S.F., Van Der Vegt, B.J., Van Reenen, C.G., Hopster, H., De Jong, I.C., Ruis, M.A., Blokhuis, H.J., 1999. Coping styles in animals: current status in behavior and stress-physiology. Neurosci. Biobehav. Rev. 23, 925–35.

Kunz, M., Chen, J.-I., Lautenbacher, S., Vachon-Presseau, E., Rainville, P., 2011. Cerebral Regulation of Facial Expressions of Pain. J. Neurosci. 31, 8730–8738. https://doi.org/10.1523/JNEUROSCI.0217-11.2011

Kunz, M., Lautenbacher, S., LeBlanc, N., Rainville, P., 2012. Are both the sensory and the affective dimensions of pain encoded in the face? Pain 153, 350–8. https://doi.org/10.1016/j.pain.2011.10.027

Kunz, M., Peter, J., Huster, S., Lautenbacher, S., 2013. Pain and disgust: The facial signaling of two aversive bodily experiences. PLoS One 8, 1–11. https://doi.org/10.1371/journal.pone.0083277

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.J.M., Ferrari, M.D., Craig, K.D., Mogil, J.S., 2010. Coding of facial expressions of pain in the laboratory mouse. Nat. Methods 7, 447–9.

Lavand'homme, P., 2011. The progression from acute to chronic pain. Curr. Opin. Anaesthesiol. 24, 545–50. https://doi.org/10.1097/ACO.0b013e32834a4f74

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. Res. Vet. Sci. 87, 336–47. https://doi.org/10.1016/j.rvsc.2009.02.001

Leach, M.C., Coulter, C., Richardson, C., Flecknell, P., 2011. Are we looking in the wrong place? Implications for behavioural-based pain assessment in rabbits (Oryctolagus cuniculi) and beyond? PLoS One 6, e13347. https://doi.org/10.1371/journal.pone.0013347

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, e35656. https://doi.org/10.1371/journal.pone.0035656

Lean, I.J., Westwood, C.T., Golder, H.M., Vermunt, J.J., 2013. Impact of nutrition on lameness and claw health in cattle. Livest. Sci. 156, 71–87. https://doi.org/10.1016/j.livsci.2013.06.006

Lee, M.C., Mouraux, A., Iannetti, G.D., 2009. Characterizing the Cortical Activity through Which Pain Emerges from Nociception. J. Neurosci. 29, 7909–7916. https://doi.org/10.1523/JNEUROSCI.0014-09.2009

Leung, V., Zhang, E., Pang, D.S., 2016. Real-time application of the Rat Grimace Scale as a welfare refinement in laboratory rats. Sci. Rep. 6, 31667. https://doi.org/10.1038/srep31667

Ley, S.J., Livingston, A., Waterman, A.E., 1989. The effect of chronic clinical pain on thermal and mechanical thresholds in sheep. Pain 39, 353–357. https://doi.org/10.1016/0304-3959(89)90049-3

Ley, S.J., Waterman, A.E., Livingston, A., 1995. A field study of the effect of lameness on mechanical nociceptive thresholds in sheep. Vet. Rec. 137, 85–7.

Lim, E.J., Back, S.K., Kim, M.A., Li, C., Lee, J., Jeong, K.Y., Na, H.S., 2009. Long-lasting neonatal inflammation enhances pain responses to subsequent inflammation, but not peripheral nerve injury in adult rats. Int. J. Dev. Neurosci. 27, 215–222. https://doi.org/10.1016/j.ijdevneu.2009.01.005

Lizarraga, I., Chambers, J.P., 2012. Use of analgesic drugs for pain management in sheep. N. Z. Vet. J. 60, 87–94. https://doi.org/10.1080/00480169.2011.642772

948 Lomax, S., Windsor, P.A., 2013. Topical anesthesia mitigates the pain of castration in beef calves. J. Anim. Sci. https://doi.org/10.2527/jas.2012-5984

949 Lu, C., Yang, T., Zhao, H., Zhang, M., Meng, F., Fu, H., Xie, Y., Xu, H., 2016. Insular Cortex is Critical for the Perception, Modulation, and Chronification
950 of Pain. Neurosci. Bull. 32, 191–201. https://doi.org/10.1007/s12264-016-0016-y

951 MacRae, A.M., Joanna Makowska, I., Fraser, D., 2018. Initial evaluation of facial expressions and behaviours of harbour seal pups (Phoca vitulina) in
952 response to tagging and microchipping. Appl. Anim. Behav. Sci. 205, 167–174. https://doi.org/10.1016/j.applanim.2018.05.001

953 Martin, P., Bateson, P., 2007. Measuring Behaviour: an introductory guide, 3rd ed. Cambridge University Press, Cambridge.

954 Martuscello, R.T., Spengler, R.N., Bonoiu, A.C., Davidson, B.A., Helinski, J., Ding, H., Mahajan, S., Kumar, R., Bergey, E.J., Knight, R., Prasad, P.N.,
955 Ignatowski, T.A., 2013. Increasing TNF levels solely in the rat hippocampus produces persistant pain like symptoms 153, 1871–1882.
956 https://doi.org/10.1016/j.pain.2012.05.028.

957 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
958 reevaluate the efficacy of postoperative analgesics in laboratory mice. J. Am. Assoc. Lab. Anim. Sci. 51, 42–9.

959 Matur, E., Akyazi, I., Eraslan, E., Ergul Ekiz, E., Eseceli, H., Keten, M., Metiner, K., Aktaran Bala, D., 2016. The effects of environmental enrichment and
960 transport stress on the weights of lymphoid organs, cell-mediated immune response, heterophil functions and antibody production in laying hens. Anim.
961 Sci. J. 87, 284–292. https://doi.org/10.1111/asj.12411

962 McCreary, J.K., Metz, G.A.S., 2016. Environmental enrichment as an intervention for adverse health outcomes of prenatal stress. Environ. Epigenetics 2,

963 dvw013. https://doi.org/10.1093/eep/dvw013

964 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
965 footrot and mastitis as models of pain in sheep. Appl. Anim. Behav. Sci. 176, 19–26. https://doi.org/10.1016/j.applanim.2016.01.007

966 Miki, K., Zhou, Q.Q., Guo, W., Guan, Y., Terayama, R., Dubner, R., Ren, K., 2002. Changes in gene expression and neuronal phenotype in brain stem pain
967 modulatory circuitry after inflammation. J Neurophysiol 87, 750–760. https://doi.org/10.1152/jn.00534.2001

968 Miller, A.L., Golledge, H.D.R., Leach, M.C., 2016a. The influence of isoflourane anaesthesia on the rat grimace scale. PLoS One 11, e0166652.

969 Miller, A.L., Kitson, G., Skalkoyannis, B., Leach, M., 2015. The effect of isoflurane anaesthesia and buprenorphine on the mouse grimace scale and
970 behaviour in CBA and DBA/2 mice. Appl. Anim. Behav. Sci. 172, 58–62. https://doi.org/10.1016/j.applanim.2015.08.038

971 Miller, A.L., Kitson, G.L., Skalkoyannis, B., Flecknell, P.A., Leach, M.C., 2016b. Using the mouse grimace scale and behaviour to assess pain in CBA mice
972 following vasectomy. Appl. Anim. Behav. Sci. 181, 160–165. https://doi.org/10.1016/j.applanim.2016.05.020

973 Miller, A.L., Leach, M.C., 2016. The effect of handling method on the mouse grimace scale in two strains of laboratory mice. Lab. Anim. 50, 305–307.
974 https://doi.org/10.1177/0023677215622144

975 Miller, A.L., Leach, M.C., 2015. The mouse grimace scale: A clinically useful tool? PLoS One 10, 1–11. https://doi.org/10.1371/journal.pone.0136000

976 Miller, A.L., Leach, M.C., 2014. Using the mouse grimace scale to assess pain associated with routine ear notching and the effect of analgesia in laboratory

977 mice. Lab. Anim. 49, 117–20. https://doi.org/10.1177/0023677214559084

978 Mintline, E.M., Stewart, M., Rogers, A.R., Cox, N.R., Verkerk, G.A., Stookey, J.M., Webster, J.R., Tucker, C.B., 2013. Play behavior as an indicator of
979 animal welfare: Disbudding in dairy calves. Appl. Anim. Behav. Sci. 144, 22–30. https://doi.org/10.1016/j.applanim.2012.12.008

980 Mogil, J.S., 2009. Animal models of pain: progress and challenges. Nat. Rev. Neurosci. 10, 283–94. https://doi.org/10.1038/nrn2606

981 Mogil, J.S., 1999. The genetic mediation of individual differences in sensitivity to pain and its inhibition. Proc. Natl. Acad. Sci. U.S.A. 96, 7744–51.
982 https://doi.org/10.1073/pnas.96.14.7744

983 Mogil, J.S., Crager, S.E., 2004. What should we be measuring in behavioral studies of chronic pain in animals? Pain 112, 12–5.
984 https://doi.org/10.1016/j.pain.2004.09.028

985 Mogil, J.S., Sternberg, W.F., Marek, P., Sadowski, B., Belknap, J.K., Liebeskind, J.C., 1996. The genetics of pain and pain inhibition. Proc. Natl. Acad. Sci.
986 U.S.A. 93, 3048–55. https://doi.org/10.1073/pnas.93.7.3048

987 Molony, V., Kent, J.E., 1997. Assessment of acute pain in farm animals using behavioral and physiological measurements T. J. Anim. Sci. 75, 266–272.

988 Molony, V., Kent, J.E., McKendrick, I.J., 2002. Validation of a method for assessment of an acute pain in lambs. Appl. Anim. Behav. Sci. 76, 215–238.
989 https://doi.org/10.1016/S0168-1591(02)00014-X

990 Mormède, P., Andanson, S., Aupérin, B., Beerda, B., Guémené, D., Malmkvist, J., Manteca, X., Manteuffel, G., Prunet, P., van Reenen, C.G., Richard, S.,

Veissier, I., 2007. Exploration of the hypothalamic-pituitary-adrenal function as a tool to evaluate animal welfare. Physiol. Behav. 92, 317–39. https://doi.org/10.1016/j.physbeh.2006.12.003

Napolitano, F., De Rosa, G., Girolami, A., Scavone, M., Braghieri, A., 2011. Avoidance distance in sheep: Test-retest reliability and relationship with stockmen attitude. Small Rumin. Res. 99, 81–86. https://doi.org/10.1016/j.smallrumres.2011.03.044

Neuman, W.L., 2014. Qualitative and Quantitative Measurement, in: Social Research Methods: Qualitative and Quantitative Approaches. Pearson Education Limited, Essex.

Norring, M., Wikman, I., Hokkanen, A.-H., Kujala, M. V, Hänninen, L., 2014. Empathic veterinarians score cattle pain higher. Vet. J. 200, 186–90. https://doi.org/10.1016/j.tvjl.2014.02.005

Ohayon, M.M., 2005. Relationship between chronic painful physical condition and insomnia. J. Psychiatr. Res. 39, 151–159. https://doi.org/10.1016/j.jpsychires.2004.07.001

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. https://doi.org/10.1371/journal.pone.0097882

Panksepp, J., 1998. Affective Neuroscience. The Foundations of Human and Animal Emotions. Oxford University Press, Inc., Oxford, UK.

Parr, L., Waller, B., Burrows, A., Gothard, K., Vick, S., 2010. MaqFACS: A Muscle-Based Facial Movement Coding System for the Rhesus Macaque. Am. J. Phys. Anthropol. 143, 625–630. https://doi.org/10.1002/ajpa.21401.MaqFACS

1006 Parr, L.A., Waller, B.M., Vick, S.-J., Bard, K.A., 2007. Classifying Chimpanzee Facial Expressions Using Muscle Action. Emotion 7, 172–181.
1007 https://doi.org/10.1037/1528-3542.7.1.172.Classifying

1008 Prkachin, K.M., 1992. The consistency of facial expressions of pain: a comparison across modalities. Pain 51, 297–306.

1009 Prkachin, K.M., Berzins, S., Mercer, S.R., 1994. Encoding and decoding of pain expressions: a judgement study. Pain 58, 253–259.

1010 Prkachin, K.M., Solomon, P.E., 2008. The structure, reliability and validity of pain expression: evidence from patients with shoulder pain. Pain 139, 267–74.
1011 https://doi.org/10.1016/j.pain.2008.04.010

1012 Prusator, D.K., Greenwood-Van Meerveld, B., 2016. Sex-related differences in pain behaviors following three early life stress paradigms. Biol. Sex Differ. 7,
1013 29. https://doi.org/10.1186/s13293-016-0082-x

1014 Raaperi, K., Bougeard, S., Aleksejev, A., Orro, T., Viltrop, A., 2012. Association of herd BHV-1 seroprevalence with respiratory disease in youngstock in
1015 Estonian dairy cattle. Res. Vet. Sci. 93, 641–648. https://doi.org/10.1016/j.rvsc.2011.10.015

1016 Rainville, P., 2002. Brain mechanisms of pain affect and pain modulation. Curr. Opin. Neurobiol. 12, 195–204. https://doi.org/10.1016/S0959-
1017 4388(02)00313-6

1018 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,
1019 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,
1020 e0187986.

1021 Richardson, C., Flecknell, P.A., 2005. Anaesthesia and post-operative analgesia following experimental surgery in laboratory rodents: are we making
1022 progress? Altern. Lab. Anim. 33, 119–27.

1023 Robbins, J.A., Weary, D.M., Schuppli, C.A., Von Keyserlingk, M.A.G., 2015. Stakeholder views on treating pain due to dehorning dairy calves. Anim. Welf.
1024 24, 399–406. https://doi.org/10.7120/09627286.24.4.399

1025 Rushen, J., de Passillé, A.M., 2012. Automated measurement of acceleration can detect effects of age, dehorning and weaning on locomotor play of calves.
1026 Appl. Anim. Behav. Sci. 139, 169–174. https://doi.org/10.1016/j.applanim.2012.04.011

1027 Rutherford, K.M., Robson, S.K., Donald, R.D., Jarvis, S., Sandercock, D.A., Scott, E.M., Nolan, A.M., Lawrence, A.B., 2009. Pre-natal stress amplifies the
1028 immediate behavioural responses to acute pain in piglets. Biol Lett 5, 452–454. https://doi.org/10.1098/rsbl.2009.0175

1029 Rutherford, K.M.D., 2002. Assessing pain in animals. Anim. Welf. 11, 31–53.

1030 Sandercock, D.A., Gibson, I.F., Rutherford, K.M.D., Donald, R.D., Lawrence, A.B., Brash, H.M., Scott, E.M., Nolan, A.M., 2011. The impact of prenatal
1031 stress on basal nociception and evoked responses to tail-docking and inflammatory challenge in juvenile pigs. Physiol. Behav. 104, 728–37.
1032 https://doi.org/10.1016/j.physbeh.2011.07.018

1033 Šavc, M., Duane, M., O'Grady, L.E., Somers, J.R., Beltman, M.E., 2016. Uterine disease and its effect on subsequent reproductive performance of dairy
1034 cattle: a comparison of two cow-side diagnostic methods. Theriogenology 86, 1983–1988. https://doi.org/10.1016/j.theriogenology.2016.06.018

1035 Schiavenato, M., Byers, J.F., Scovanner, P., McMahon, J.M., Xia, Y., Lu, N., He, H., 2008. Neonatal pain facial expression: evaluating the primal face of

pain. Pain 138, 460–71. https://doi.org/10.1016/j.pain.2008.07.009

Schütz, T.C.B., Andersen, M.L., Tufik, S., 2003. Sleep alterations in an experimental orofacial pain model in rats. Brain Res. 993, 164–171. https://doi.org/10.1016/j.brainres.2003.09.006

Seminowicz, D.A., Laferriere, A.L., Millecamps, M., Yu, J.S.C., Coderre, T.J., Bushnell, M.C., 2009. MRI structural brain changes associated with sensory and emotional function in a rat model of long-term neuropathic pain. Neuroimage 47, 1007–14. https://doi.org/10.1016/j.neuroimage.2009.05.068

Shukla, M., Singh, G., Sindhura, B.G., Telang, A.G., Rao, G.S., Malik, J.K., 2007. Comparative plasma pharmacokinetics of meloxicam in sheep and goats following intravenous administration. Comp. Biochesmistry Physiol. 145, 528–532. https://doi.org/10.1016/j.cbpc.2007.01.020

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

Sneddon, L.U., Lopez-Luna, I.J., Wolfenden, D.C.C., Leach, M.C., Valentim, A.M., Steenbergen, P.J., Bardine, N., Currie, A.D., Broom, D.M., Brown, C., 2018. Fish sentience denial: muddying the waters. Anim. Sentience 21.

Solomon, P.E., Prkachin, K.M., Farewell, V., 1997. Enhancing sensitivity to facial expression of pain. Pain 71, 279–84.

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., Sternberg, W.F., Mogil, J.S., 2014. Olfactory exposure to males, inluding men, causes stress and related analgesia in rodents. Nat. Methods 11, 629–634.

https://doi.org/doi:10.1038/nmeth.2935

Sotocinal, S.G., Sorge, R.E., Zaloum, A., Tuttle, A.H., Martin, L.J., Wieskopf, J.S., Mapplebeck, J.C.S., 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. Mol. Pain 7, 55. https://doi.org/10.1186/1744-8069-7-55

Stewart, M., Beausoleil, N.J., Johnson, C.B., Webster, J.R., Schütz, K.E., Cox, N., Stafford, K.J., 2014. Do rubber rings coated with lignocaine reduce the pain associated with ring castration of lambs? Appl. Anim. Behav. Sci. 160, 56–63. https://doi.org/10.1016/j.applanim.2014.08.007

Stewart, M., Webster, J.R., Verkerk, G.A., Schaefer, A.L., Colyn, J.J., Stafford, K.J., 2007. Non-invasive measurement of stress in dairy cows using infrared thermography. Physiol. Behav. 92, 520–5. https://doi.org/10.1016/j.physbeh.2007.04.034

Thornton, P.D., Waterman-Pearson, A.E., 2002. Behavioural responses to castration in lambs. Anim. Welf. 11, 203–212.

Tracey, I., Bushnell, M.C., 2009. How neuroimaging studies have challenged us to rethink: is chronic pain a disease? J. Pain 10, 1113–20. https://doi.org/10.1016/j.jpain.2009.09.001

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. Mol. Pain 14, 174480691876365. https://doi.org/10.1177/1744806918763658

Tuyttens, F.A.M., Stadig, L., Heerkens, J.L.T., Van laer, E., Buijs, S., Ampe, B., 2016. Opinion of applied ethologists on expectation bias, blinding observers and other debiasing techniques. Appl. Anim. Behav. Sci. 181, 27–33. https://doi.org/10.1016/j.applanim.2016.04.019

van Loon, J.P.A.M., VanDierendonck, 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. Vet. J. 206, 356–364. https://doi.org/10.1016/j.tvjl.2015.08.023

Ventura, B.A., von Keyserlingk, M.A.G., Weary, D.M., 2014. Animal Welfare Concerns and Values of Stakeholders Within the Dairy Industry. J. Agric. Environ. Ethics 28, 109–126. https://doi.org/10.1007/s10806-014-9523-x

Vervoort, T., Caes, L., Trost, Z., Sullivan, M., Vangronsveld, K., Goubert, L., 2011. Social modulation of facial pain display in high-catastrophizing children: an observational study in schoolchildren and their parents. Pain 152, 1591–9. https://doi.org/10.1016/j.pain.2011.02.048

Vervoort, T., Goubert, L., Eccleston, C., Verhoeven, K., De Clercq, A., Buysse, A., Crombez, G., 2008. The effects of parental presence upon the facial expression of pain: the moderating role of child pain catastrophizing. Pain 138, 277–85. https://doi.org/10.1016/j.pain.2007.12.013

Veterinary Medicines Directorate., 2018. Product Information Database [WWW Document]. URL https://www.vmd.defra.gov.uk/ProductInformationDatabase/ (accessed 11.29.18).

Villani, M., Cairoli, F., Kindahl, H., Galeati, G., Faustini, M., Carluccio, A., Veronesi, M.C., 2006. Effects of mating on plasma concentrations of testosterone, cortisol, oestrone sulphate and 15-ketodihydro-PGF2α in stallions. Reprod. Domest. Anim. 41, 544–548. https://doi.org/10.1111/j.1439-0531.2006.00711.x

Viñuela-Fernández, I., Jones, E., Welsh, E.M., Fleetwood-Walker, S.M., 2007. Pain mechanisms and their implication for the management of pain in farm

and companion animals. Vet. J. 174, 227–39. https://doi.org/10.1016/j.tvjl.2007.02.002

Viscardi, A. V., Hunniford, M., Lawlis, P., Leach, M., 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. Front. Vet. Sci. 4, 1–9. https://doi.org/10.3389/fvets.2017.00051

Vlaeyen, J.W.S., Hanssen, M., Goubert, L., Vervoort, T., Peters, M., van Breukelen, G., Sullivan, M.J.L., Morley, S., 2009. Threat of pain influences social context effects on verbal pain report and facial expression. Behav. Res. Ther. 47, 774–82. https://doi.org/10.1016/j.brat.2009.05.008

Waite, M.E., Tomkovich, A., Quinn, T.L., Schumann, A.P., Dewberry, L.S., Totsch, S.K., Sorge, R.E., 2015. Efficacy of Common Analgesics for Postsurgical Pain in Rats. J. Am. Assoc. Lab. Anim. Sci. 54, 420–5.

Walker, K.A., Duffield, T.F., Weary, D.M., 2011. Identifying and preventing pain during and after surgery in farm animals. Appl. Anim. Behav. Sci. 135, 259–265. https://doi.org/10.1016/j.applanim.2011.10.021

Wathan, J., Burrows, A.M., Waller, B.M., McComb, K., 2015. EquiFACS: The equine facial action coding system. PLoS One 10, 1–35. https://doi.org/10.1371/journal.pone.0131738

Weary, D.M., Niel, L., Flower, F.C., Fraser, D., 2006. Identifying and preventing pain in animals. Appl. Anim. Behav. Sci. 100, 64–76. https://doi.org/10.1016/j.applanim.2006.04.013

Weber, G.H., Morton, J.M., Keates, H., 2012. Postoperative pain and perioperative analgesic administration in dogs: Practices, attitudes and beliefs of Queensland veterinarians. Aust. Vet. J. 90, 186–193. https://doi.org/10.1111/j.1751-0813.2012.00901.x

White, K., Targett, M., Harris, J., 2018. Gainfully employing descending controls in acute and chronic pain management. Vet. J. 237, 16–25. https://doi.org/https://doi.org/10.1016/j.tvjl.2018.05.005

Whittaker, A.L., Leach, M.C., Preston, F.L., Lymn, K.A., Howarth, G.S., 2016. Effects of acute chemotherapy-induced mucositis on spontaneous behaviour and the grimace scale in laboratory rats. Lab. Anim. 50, 108–118.

Williams, W.O., Riskin, D.K., Mott, K.., 2008. Ultrasonic sound as an indicator of acute pain in laboratory mice. J. Am. Assoc. Lab. Anim. Sci. 47, 8–10.

Winston, J.H., Li, Q., Sarna, S.K., 2014. Chronic prenatal stress epigenetically modifies spinal cord BDNF expression to induce sex-specific visceral hypersensitivity in offspring. Neurogastroenterol. Motil. 26, 715–30. https://doi.org/10.1111/nmo.12326

Woller, S.A., Eddinger, K.A., Corr, M., Yaksh, T.L., Corr, M., Yaksh, T.L., 2017. An overview of pathways encoding nociception. Clin. Exp. Rheumatol. 35, S40–S46.

Table 1

Table 1. **The validity, sensitivity, specificity and reliability of current facial expression scales.**

Species	Scale name	Reference	Validity (accuracy)	Sensitivity and specificity	Reliability
Mice	Mouse Grimace Scale (MGS)	Langford et al., 2010	Global assessment 72% (trained) and 81% (with experience)	Details not provided	ICC = 0.90
Rats	Rat Grimace Scale (RGS)	Sotocinal et al., 2011	81.6% (76-87.5)	FN=10.3%, FP=8.2%	ICC = 0.90 (ranging from 0.86-0.96 for AU's)
Rabbits	Rabbit Grimace Scale (RbtGS)	Keating et al., 2012	Global assessment 83.6%	FN=10.6%, FP=5.8%	ICC = 0.91 (ranging from 0.84-0.94 for AU's)
Ferrets	Ferret Grimace Scale (FGS)	Reijgwart et al., 2017	Highest accuracy 80%	SN=85%, SP=74%	ICC=0.97 (ranging from 0.85-0.89 for AU's)
Horses	Horse Grimace Scale (HGS)	Dalla Costa et al., 2014	Global assessment 73.3%	FN = 9.8%, FP=17.0%	ICC = 0.92 (ranging from 0.72-0.97 for AU's)
	Equine Utrecht University Scale for facial assessment of pain (EQUUS-FAP)	Van Loon and Van Dierendonck, 2015	-	SN=87.5, SF=88.0, PPV=87.5, NPV=88.0 (For colic versus controls). Conservative treatment versus surgical treatment SN=30.0, SF=64.3, PPV=37.5, NPV = 56.3.	ICC=0.93
Sheep	Sheep Pain Facial Expression Scale (SPFES)	McLennan et al., 2016	Global assessment 67% (ranging from 60-75%) AUC 0.81	FN=6.3%, FP=26.3%	ICC=0.86 (ranging from 0.63-0.90 for AU's)
	Sheep Grimace Scale (SGS)	Häger et al., 2017	Accuracy 68.2%	FN=9.1, FP = 22.7%	ICC=0.92
Lambs	Lamb Grimace Scale (LGS)	Guesgen et al., 2016	-	-	W=0.60-0.66 (ranging from 0.34-0.79 for AU's)
Piglets	Piglet Grimace Scale (PGS)	Di Giminiani et al., 2016	-	-	ICC=0.97 (ranging from 0.82-0.97 for AU's)
Cats	Facial Pain Assessment Tool	Holden et al., 2014	Ranged from 18-94%, weak correlation (Pearson=0.214) with Numerical rating scale	NT	NT

Table X: Best practice guidelines for facial expression scale development

Factor	Best practice guidelines
Experimental design	• Ensure you have a baseline for comparison. • Within rather than between animal designs are preferable. Consider cross-over designs if confounding factors cannot be eliminated. • Pain stimulus carefully considered and previously validated. • Pain relief provided and tested for dose dependent changes and observations carried out when pain relief is to be at its most effective. • If anaesthesia used, take note of wearing off time. Consider a control group of anaesthesia treatment only. • Consider carefully the use of an already valid pain indicator for comparison that is used to measure the same construct. • Consider sex, age and breed differences - how are they going to be controlled for? What sample size is required to account for this?
Capturing data	• Higher definition video footage preferable over photographs. • Undisturbed footage. • Ideally a different person for collecting, selecting and analysing images. • Camera set-up should consider lighting, angle and space use. Multiple cameras enable more footage and will cover all facial areas. • Reduce stress inducing factors such removal from home cages or separation from others. • Avoid handling the animal at time of image capture, and before as far as is reasonably possible. • A longer data collection period over multiple time points is preferred over short, or single time points of assessment. • Collection of baseline and pain images should be as close in time as possible.
Development	• Selection of multiple images in a systematic manner using automated or blinded to treatment and time assistants. • Within animal changes should be assessed. • Cropped images with background removed. • Choose good quality images with a clear view of face. • Systematic decisions of action unit choice, consider those that occur between 25% and 50% of the time. • Ensure action units are species specific and for that age group. Consider additional information and examples for breed differences.
Testing	• Ensure adequate training has been given to blinded observers. • Consider using total pain scores over average scores. • Global assessment of pain no longer required. • Use as many time and treatment blinded observers as possible, and test for both the intra- and inter-observer reliability. • Analyse the sensitivity and specificity of each pain score level to provide guidance on when intervention is required.

- Test for feasibility over a longer period of time, with multiple observations. Avoid the effect of observer presence by scoring live from video footage, as well as taking images from that same footage. Systematic collection of data is required to ensure repeatability.

1119

1120

1121 Table 3

Table Z. Advantages of using facial expression scales to assess pain in humans and other animals

- Provides information on the emotional component of pain (Kunz et al., 2012; Langford et al., 2010)
- Humans readily look at the area of the body where there is facial expression is (Leach et al., 2011)
- The use of the scale provides a more objective assessment of pain than a global assessment of pain (McLennan et al., 2016b)
- Minimal training is required and can be quickly taught to caretakers
- The scales are quick and simple to use once familiar
- Has the potential to be developed for other affective states
- Aids recognition of pain, the key to its management and for rescue analgesia
- Allows assessment of pain severity
- Increases sensitivity of assessment
- Can be used as part of a composite pain scale
- Allows for continued assessment – assess effectivity of drug treatment and pain relief
- Spontaneous and immediate observations of pain, which can naturally fluctuate.
- Facial expression of pain is consistent across modalities in humans (Prkachin, 1992).
- Honest signals of the intensity of pain (Poole and Craig, 1992; Porter et al., 2012; Porter and ten Brinke, 2008).
- Faked pain expressions easily identified in humans (Boerner et al., 2013; Larochette et al., 2006).
- Expression of and sensitivity to pain expression, conserved across development (Deyo et al., 2004; Prkachin, 2009).

- Expression of pain in animals appears conserved across species (mammals) (Chambers and Mogil, 2015).
- Shows high validity and reliability
- Some have shown good feasibility (Leung et al., 2016)
- Potential for an automatic recording of facial expression, removing subjectivity (Lu et al., 2017)