A systematic review of animal based indicators of sheep welfare on farm, at market and during transport, and qualitative appraisal of their validity and feasibility for use in UK abattoirs

P. Llonch a, c, E.M. King a, K.A. Clarke b, J.M. Downes b, L.E. Green a

a School of Life Sciences, University of Warwick, Coventry, CV4 7AL, UK
b One Health Veterinary Services, Newcastle upon Tyne, NE20 9JU, UK
c Current address: SRUC, Roslin Institute Building, Easter Bush, Edinburgh, EH25 9RG, UK

* Corresponding author. Tel.: +44 131 6519362.
E-mail address: pol.llonch@sruc.ac.uk (P. Llonch).
Abstract

In the UK, it has been suggested that abattoirs are ideal locations to assess the welfare of sheep because most sheep are slaughtered at abattoirs, either as finished lambs or cull ewes. Data from abattoirs could therefore provide benchmarks for welfare indicators at a national level, as well as demonstrating how these change over time. Additionally, feedback could be provided to farmers and regulatory authorities to help improve welfare and identify high or low standards for quality assurance or risk-based inspections. A systematic review of the scientific literature was conducted, which identified 48 animal-based indicators of sheep welfare that were categorised by the Five Freedoms. Their validity as measures of welfare and feasibility for use in abattoirs were evaluated as potential measures of prior sheep welfare on the farm of origin, at market, or during transportation to the abattoir.

A total of 19 indicators were considered valid, of which nine were considered theoretically feasible to assess sheep welfare at abattoirs. These were body cleanliness, carcass bruising, diarrhoea, skin lesions, skin irritation, castration, ear notching, tail docking and ‘obviously sick’. Further investigation of these indicators is required to test their reliability and repeatability in abattoirs. Novel welfare indicators are needed to assess short-term hunger and thirst, prior normal behaviour and long-term fear and distress.

Keywords: Abattoir; Animal-based welfare indicators; Sheep; Systematic review; Validity
In the UK, most sheep are slaughtered in abattoirs. Observation of sheep in abattoirs, using indicators that are transparent and fair, might provide an assessment of prior health and welfare on farm, at market and during transportation. While such inspections would not replace all inspections elsewhere, data could be used to benchmark the prevalence of welfare indicators (Farm Animal Welfare Council, 1993), to inform risk-based selection for inspections measuring compliance with animal welfare legislation, evaluate whether an assurance scheme’s welfare conditions are met (Kilbride et al., 2012), or provide farmers with information on the health and welfare of their livestock to assist in health planning.

Assessment of sheep in abattoirs requires valid welfare indicators. Animal-based (outcome-based) indicators of welfare use direct assessment of an animal’s mental and physical welfare. They are considered the most valid method of assessing animal welfare because the animals themselves are assessed, not their resources, and comparisons can be made across all systems of husbandry (Main et al., 2007). For industry to use such indicators in abattoirs, they must be valid (measure what they intend, i.e. animal welfare), repeatable (the same result for repeated observations of the same animal by the same and different observers), reliable (consistent results across observation of different animals) and feasible (in terms of speed, cost and not compromising normal operating procedures; Knierim and Winckler, 2009; Napolitano et al., 2009).

With these parameters in mind, we conducted a systematic review of sheep welfare indicators. We categorised these into the Five Freedoms and within each freedom grouped indicators that measured a similar aspect of welfare. Finally, we used the published literature to inform on the validity of each indicator and qualitatively assessed their feasibility for use
in abattoirs, to assess the prior welfare of sheep on farm, at market and during transportation to the abattoir.

**Materials and methods**

**Search criteria and strategy**

We conducted a systematic review of peer-reviewed scientific literature published from 1 January 1995 to 15 December 2012. All experimental and observational studies of sheep welfare (including research papers, conference proceedings and literature reviews) referring to welfare assessment of adult sheep or lambs (*Ovis aries*) were included. Searches were performed using the same search terms in four search engines: (1) Pubmed; (2) ScienceDirect; (3) Scopus; and (4) Web of Knowledge.

The search terms used (including all titles, abstracts and keywords) were:

'(assess* OR indicator* OR monitor* OR audit OR evaluation OR "animal based" OR clinical AND "animal welfare" OR "sheep welfare" OR welfare AND slaughter* OR abattoir OR mortem* OR farm OR on-farm AND ovine OR “ovis aries” OR sheep OR ram OR “dairy sheep” OR “sheep farm” OR “sheep flock” OR ewe OR lamb) AND PUBYEAR > 1994’

---


Only documents written in English were included. Duplicates and documents not
directly related to sheep welfare were removed. A second filter was applied to remove
references containing no information about the methodology of assessment of sheep, which
resulted in 349 articles remaining.

Criteria for selection of animal-based welfare indicators

A total of 349 articles were retrieved and read to identify all animal-based indicators
of sheep welfare. Physiological measurements (e.g. serum cortisol concentration, heart rate)
and resource-based observations (e.g. water availability, bedding quality) were rejected. One
hundred and twenty-one papers on animal-based indicators were then reduced to papers with
indicators made by visual inspection. A total of 218 animal-based indicators in 53 papers
were obtained. Related indicators assessing the same welfare problem e.g. gait score and
lameness score, were combined to give 48 separate indicators (Table 1). Each indicator was
then allocated, using the Five Freedoms (Farm Animal Welfare Council, 1993), to the aspect
within a freedom that it measured; the Five Freedoms have been acknowledged as an
appropriate framework to assess all aspects of animal welfare (McCulloch, 2013).

Since none of the indicators addressed freedom from thirst, additional literature was
reviewed to identify potential indicators reported in other species. This was performed by
including the specific term thirst and removing the terms searching species (e.g. sheep, ovine
etc.) from the search criteria.

The validity and feasibility of measuring each indicator in an abattoir was categorised
as high, medium or low. High validity indicators were those validated in previous research,
medium validity indicators were those where the current method of assessment did not
necessarily indicate poor welfare e.g. body condition score (BCS) is a measure of welfare in adults, but might not be a valid measure of welfare in growing lambs. Low validity indicators were those suggested as indicators in the scientific literature, but lacking evidence that they actually assess welfare. Indicators with high feasibility were those that could be recorded in abattoirs, regardless of the number of animals, the space available for the animals, and the speed of the processing line. Medium feasibility indicators were those that needed special requirements (e.g. extra space or time) for appropriate assessment and low feasibility indicators were those that could not be routinely assessed in commercial abattoirs.

Results

The 48 indicators by category and feasibility are presented in Table 1. They were assigned to freedom from hunger and thirst ($n = 5$), freedom from discomfort ($n = 5$), freedom from pain, injury and disease ($n = 17$), freedom to express normal behaviour ($n = 8$) and freedom from fear and distress ($n = 13$).

Discussion

This systematic review of animal-based measures of sheep welfare is the first step in the identification of valid and feasible indicators that could be used in abattoirs to monitor the prior welfare of sheep on farm, at market, or during transportation. There were 19 valid indicators were identified which provided information on long-term hunger, discomfort, injury and disease and short-term distress, but only nine were considered feasible for measurement in abattoirs (body cleanliness, carcass bruising, diarrhoea, skin lesions, skin irritation, castration, ear notching, tail docking and obviously sick). In addition, conformation and fat carcass classification (two medium validity indicators), were considered feasible to measure and useful to take forward. No valid, feasible indicators were identified that
measured short-term hunger or thirst, long-term normal behaviour or long-term fear and
distress, which could be used in abattoirs to assess welfare on the farm of origin, in markets
or in transit. Indicators are discussed below.

In adult sheep, chronic under nutrition can be measured by low bodyweight or low
BCS (Jefferies, 1961; Phythian et al., 2012a). Bodyweight varies by age, sex and breed, and
since the mature weight of adult sheep varies widely depending on breed, only within-animal
comparisons are likely to be valid. Consequently, BCS is generally used as a measure of
nutritional status. Although BCS does not indicate current hunger, it does provide
information on long-term nutritional status. It is assessed by manual palpation of the lumbar
region (Phythian et al., 2012a) or the ribs (Shands et al., 2009) and provides an estimate of
body fat and muscle. BCS is valid and reliable; variability between observers can occur but
this is reduced by training (Phythian et al., 2012a). Very low BCS (<1.5) indicates emaciation
that arises from inadequate feed, chronic disease, or parasitism (Sargison and Scott, 2010),
implying severe consequences for sheep welfare.

Post-mortem, the EUROP carcass classification\(^6\)\(^7\) indicates the shape and volume of
muscle in relation to bone structure; the 1 - 5 fat classification assesses the amount of visible

---


fat (Stubsjøen et al., 2011). This classification system is designed for growing lambs, which
have a lower fat:muscle ratio than mature animals and consequently adult animals appear
‘fatter’ using this system. Currently, the relationship between BCS and carcass classification
has not been assessed; this is an important area for future research.

An expert panel identified rumen fill as an animal based measure of access to feed
(Phythian et al., 2011). Rumen fill can be used to assess nutritional welfare in the short period
before slaughter but it is not useful to assess the long-term nutritional state of the animal.

There is currently no indicator for thirst that fulfils the criteria of direct animal-based
assessment in sheep. In veal calves, dehydration is measured by testing duration of skin
tenting when skin is pinched between thumb and forefinger (Mellor and Stafford 2004);
dehydration is detected by a delay in the skin returning to its normal position. In horses, skin-
tenting time is not a valid measure of thirst (Pritchard et al., 2008). This indicator has not
been evaluated in sheep.

Dirt irritates the skin and attracts bacteria, ectoparasites and other pathogens and
demonstrates the level of hygiene in which an animal has been kept or transported (Stubsjøen
et al., 2011). Assessment can be based on a numerical scale from absolute cleanliness to
complete coverage of the body with dirt or faeces (Napolitano et al., 2009; Stubsjøen et al.,
2011). Phythian et al. (2012b) focused on certain areas of the body, such as the ventral
abdomen and the breech, to give a global score based on visual assessment. For good
repeatability between observers, training with clear instructions of assessment must be
provided (Stubsjøen et al., 2011).
Lying down is frequently related to resting; however, it also relates to other welfare states in sheep. For instance, increased lying time was related to heat stress in Awassi sheep (Dikmen et al., 2011). In lambs, a reduction in lying time is an indicator of pain after castration (Thornton and Waterman-Pearson, 2002). Conversely, lying, due to an inability to stand, has been suggested to be negatively correlated with fitness in newborn lambs (Phythian et al., 2011). The reasons for lying behaviour are diverse and depend on age, management procedures and other factors. Since its relationship with animal welfare is situation-specific, it is not useful in abattoirs.

Thermal stress can arise from extremely low temperatures, provoking hypothermia, or extreme high temperatures, causing hyperthermia. Sheep have behavioural and physiological coping strategies for these conditions. Shivering is an increase in muscular activity to increase body temperature in hypothermic lambs (Mellor and Stafford, 2004), it could, therefore, be a valid indicator of hypothermia, especially in young lambs. An increase in respiratory rate above 40 breaths per min, together with open-mouthed breathing (panting) indicates manageable heat stress in otherwise healthy sheep (Silanikove, 2000). However, 300 breaths per minute indicates severe heat stress (Hales and Brown, 1974; Silanikove, 2000).

According to Phythian et al. (2011), fleece cover could also provide information about thermal welfare. Fleece cover can increase resistance to cold temperatures but also increase body temperature during hot weather. Hence it is likely that unshorn animals experience heat stress in high environmental temperatures.

Disease can have a major adverse impact on animal welfare, and some diseases (e.g. clostridial diseases or maedi-visna) are linked to welfare in the published literature (Fitzpatrick et al., 2006). Injury typically leads to inflammation, which is painful. Chronic
pain can lead to hyperalgesia or allodynia (Dolan and Nolan, 2000), which contribute to poor
to contribute to poor welfare. Pain and sub-clinical disease can be difficult to assess in animals. Animal-based
equally contribute to poor welfare. Pain and sub-clinical disease can be difficult to assess in animals. Animal-based
indicators of injury and disease need to include a visible physical abnormality that can be
detected ante- or post-mortem. Bruises can be used as a measure of trauma during handling
(Jarvis and Cockram, 1994; Miranda-de la Lama et al., 2009), and thus constitute a sign of
detected ante- or post-mortem. Bruises can be used as a measure of trauma during handling
poor welfare. Assessment can be performed once the hide has been removed. The location of
poor welfare. Assessment can be performed once the hide has been removed. The location of
each bruise can be used to assess prior welfare; lesions on highly sensitive areas (e.g. face,
each bruise can be used to assess prior welfare; lesions on highly sensitive areas (e.g. face,
abdomen) could have highly deleterious effects on animal welfare. According to Cockram
abdomen) could have highly deleterious effects on animal welfare. According to Cockram
and Lee (1991), the age of a bruise can be estimated by colour and consistency: a reddish and
and Lee (1991), the age of a bruise can be estimated by colour and consistency: a reddish and
haemorrhagic bruise is recent (0 – 10 h), dark-coloured and watery consistency is older (24-48 h), yellowing >48 h. Hence, bruising might be a valid measure for assessing short-term
haemorrhagic bruise is recent (0 – 10 h), dark-coloured and watery consistency is older (24-48 h), yellowing >48 h. Hence, bruising might be a valid measure for assessing short-term
trauma that occurred in the few days or hours before death.

Diarrhoea is a clinical sign of gastrointestinal disease, especially in lambs, and can be
caused by diet or pathogens. Sweeney et al. (2012) define diarrhoea in sheep as the presence of
caused by diet or pathogens. Sweeney et al. (2012) define diarrhoea in sheep as the presence of
loose or liquid faeces, or faecal soiling on the breech fleece. Faecal consistency or staining of
loose or liquid faeces, or faecal soiling on the breech fleece. Faecal consistency or staining of
the perineum can be used to assess diarrhoea. Cabaret et al. (2006) suggested the following
the perineum can be used to assess diarrhoea. Cabaret et al. (2006) suggested the following
scoring system: 1, normal sheep faeces in pellets; 2, ‘soft’ faeces (similar to cow pat); and 3,
scoring system: 1, normal sheep faeces in pellets; 2, ‘soft’ faeces (similar to cow pat); and 3,
diarrhoea (semi-liquid faeces). Parasitic infections lead to reduced growth rate and wool
diarrhoea (semi-liquid faeces). Parasitic infections lead to reduced growth rate and wool
production, increased mortality (especially in young sheep), reduced reproductive success
production, increased mortality (especially in young sheep), reduced reproductive success
and increased susceptibility to other parasites (Coop, 1979; Festa-Bianchet, 1988; Dwyer and
and increased susceptibility to other parasites (Coop, 1979; Festa-Bianchet, 1988; Dwyer and
Bornett, 2004). Some parasitic infections, including Echinococcus granulosus, Fasciola
Bornett, 2004). Some parasitic infections, including Echinococcus granulosus, Fasciola
hepatica, Dicrocoelium dendriticum, Cysticercus ovis, and Sarcocystis spp., can affect the
hepatica, Dicrocoelium dendriticum, Cysticercus ovis, and Sarcocystis spp., can affect the
muscles and viscera of sheep and can be detected by examination of the carcass and viscera
post-mortem (Borji et al., 2012). Round worms, however, are within the gastrointestinal lumen and so they are not readily detectable without opening the stomach or intestines.

The integument comprises the skin and modified skin structures, including head and hoof horn. Most integumentary structures are innervated. Damage is frequently caused by agonistic interactions with conspecifics and abrasions or collisions with physical structures (e.g. barbed wire fences, slatted doors), or by ingrowing head horn. New and old lesions will present differently e.g. dried blood and an open sore in recent injuries, to hyperkeratosis or hairless patches in older lesions. Stubsjøen et al. (2011) suggested two scoring systems for integumentary alterations, one for skin lesions and one for skin irritation. The skin lesion score is based on the following grading scale: 1, no skin lesions; 2, more than one lesion of >1 cm; and 3, ulceration present.

Ectoparasites can cause intense irritation to the skin and heavily infested sheep rub and bite affected areas (Plant, 2006). Dwyer and Bornett (2004) and Plant (2006) suggested that intense and regular rubbing and biting in localized areas could provoke skin lesions that help identify infected animals. For instance, Psoroptes ovis (sheep scab) produces intensely pruritic lesions and wool loss (Wells et al., 2013), while the lesions of the myiasis fly larvae (flystrike) induce inflammation, ulceration and wool loss (Hall and Wall, 1995). The scoring system for skin irritation validated by Stubsjøen et al. (2011) can be used to assess parasite-induced lesions and is as follows: 1, normal skin; 2, loss of wool regions; 3, redness/swelling of regions; and 4, presence of parasites or flies.

Lameness is one of the major welfare concerns of sheep according to farmers (Goddard et al., 2006). Eight papers (Table 1) describe gait assessment systems to categorise
locomotion in sheep. Kaler et al. (2009) proposed a valid and highly repeatable scale from 0 (normal) to 6 (unable to stand or walk) to assess lameness in sheep. Seven papers (Table 1) state that lameness can be assessed post-mortem by examining feet for lesions, once the lower limbs have been removed from the carcass. Hodgkinson (2010) described a systematic foot examination and compared all feet for subtle abnormalities and deformities. Scoring of foot lesions is repeatable between and within observers (Foddai et al., 2012); however, scoring mild abnormalities could overestimate the prevalence of lameness, because not all sheep with foot lesions are lame (Kaler et al., 2011).

It is self-evident that animals experience acute pain and distress at the time of mutilation and medium-term pain arising from tissue damage, with chronic pain also possible (Lomax et al., 2009; Edwards and Bennett, 2014). Mutilations can be assessed by visual inspection and in most cases provide reliable data (EFSA, 2012).

Eye condition was mentioned in two studies (Table 1). Assessment can be based on either inspection of the eyes, or ocular discharge. Blind sheep can only be detected ante-mortem. After death the eye becomes glazed and the eyelid droops, this restricts observation of some abnormalities including trauma, tumours or phthisis, which could be assessed post-mortem. Eye condition has not been validated as a welfare measure in sheep and so needs further investigation.

Coughing, dyspnoea and nasal discharge are signs of respiratory disease detected during clinical examination (Table 1). Post-mortem signs of respiratory disease include pulmonary inflammation or necrosis. These parameters have not been validated as welfare indicators, although lung lesions have been associated with increased age at finishing (Green...
et al., 1995). Protocols scoring lung lesion type and severity post-mortem have been
developed for cattle and pigs (Dalmau et al., 2009; Leruste et al., 2012) and such a system is
required for sheep.

Napolitano et al. (2011) considered vulvar discharge as a measure of compromised
health in female sheep. According to Aitken and Longbottom (2007), purulent vulvar
discharge indicates pathology in reproductive organs and so it might be an indicator of
reduced reproductive health. Although vulvar abnormalities have not been identified as
measures of poor welfare (and so were not included as animal-based indicators of welfare;
Table 1), the vulva is a highly sensitive area. Lovatt (2010) suggested that vulvar swelling,
prolapse and injury were abnormalities to be considered during clinical examination of the
reproductive organs in sheep. More evidence is needed to evaluate these as indicators of poor
welfare.

Changes in general demeanour including lethargy and apathy are clinical signs of pain
and systemic disease (Gougoulis et al., 2010). When these signs become severe, authors refer
to animals that are ‘obviously sick’ (Mellor and Stafford, 2003; Stubsjoen et al., 2011), and
this can be identified in a visual inspection. According to Gregory (1998), sickness is
associated with listlessness, fatigue, reduced social interaction, inappetance, discomfort and
mental confusion.

Experts suggest that any change from normal behaviour of an individual sheep can
indicate a health or welfare problem (Phythian et al., 2011). The location and duration of
assessments of behaviour affect the observations recorded and must be considered to avoid
misinterpretation. The frequency of abnormal behaviour in farm animals can provide
information about their emotional state and welfare (Mason, 1991; Miranda-de la Lama et al., 2012). Abnormal behaviours occur more frequently in animals living in confinement that does not allow the expression of natural behaviours. Consequently, abnormal behaviours are less frequent in sheep because they are less frequently confined (Dwyer and Bornett, 2004).

Forms of abnormal behaviour include stereotypies and redirected behaviour (Dwyer and Bornett, 2004; Gougoulis et al., 2010). Wool biting or pulling are redirected behaviours (Dwyer and Bornett, 2004), which could indicate a lack of environmental enrichment. Stereotypic behaviour (i.e. mouthing bars, biting and chewing pen fixtures) is more frequent in animals undergoing stress caused by maladaptation to their environment (Rushen and de Passillé, 1992; Dwyer and Bornett, 2004).

Behaviours such as aggression or threats to pen mates have a negative impact on welfare and misdirected behaviour and attacks on conspecifics are indicators of poor welfare (Broom, 1988). Increased aggression (e.g. butting or chasing episodes), can be observed during sudden environmental or social change in food, feed space or living space restrictions and in large social group size (Dwyer and Bornett, 2004). Actions that harm other animals can be recorded at the abattoir and might be an indicator of prior poor welfare. A sheep showing normal behaviour for its sex, maturity (lamb vs. adult), or season, provides valuable information about its current welfare. For example, lambs would be expected to show play behaviour when not feeding or resting, while adult sheep would routinely ruminate for one third of the day (Moquin et al., 2010). An interruption in expected behavioural pattern could suggest welfare problems (Gougoulis et al., 2010). Qualitative Behaviour Assessment (QBA) is a subjective list of behaviours that have been validated in sheep (Wemelsfelder and Farish, 2004), giving a final holistic estimate that takes into account all behaviours expressed. In
sheep, QBA has high repeatability between assessors (Phythian et al., 2013) and correlates significantly with physiological variables, including heart rate and stress leukogram results (Wickham et al., 2012).

A review of the relationship between social behaviour and welfare in goats suggests that in contrast to negative behaviours, affiliative behaviours, defined as positive, reciprocal behaviours between two or more individuals without reproductive interest, can improve the welfare state of a flock by helping to reduce aggression (Miranda-de la Lama and Mattiello, 2010). Two papers (Table 1) have proposed the identification of positive behaviours, such as nibbling or licking conspecifics, as a valid measure of good welfare in sheep.

Fear is one of the emotions that can severely influence the state of welfare of an animal. Stress is the biological response elicited when an individual perceives a threat to its homeostasis. The consequences of stress can be non-harmful, often referred as ‘good stress’ (e.g. caloric restriction in chronic hunger can promote longevity and better health), or distress, or negative stress, which weakens the immune system (Moberg, 2000). Direct observation of animal behaviour can provide a practical approach to the measurement of fear. In episodes of fear and anxiety, sheep might increase their vigilance behaviour, defined as head in an upright position and ears perpendicular to the head (Wemelsfelder and Farish, 2004; Deiss, et al., 2009). Fear has also been associated with frequency and duration of episodes of immobility, often referred as ‘freezing’ (Bouissou and Vandenneede, 1995; Cockram, 2004), and with ear-posture changes (Reefmann et al., 2009). Hemsworth et al. (2011) reported a correlation between head position and serum cortisol concentration when sheep were approached by a stockperson. The most frequently cited measure to assess fear in sheep is increased vocalisation (Table 1). In livestock species other than sheep (pigs, poultry
and cattle), high-pitched vocalisations are thought to signal appeasement in fear-associated contexts, whereas low-pitched vocalisations are attributed to more aggressive emotions (Manteuffel et al., 2004). In sheep, vocalisations can occur in response to numerous situations including social isolation, social attraction, and the presence of humans (Cockram, 2004; Boissy et al., 2005 and Deiss et al., 2009). Since vocalisation is performed in numerous situations other than fear, its suitability as a measure of fear requires further analysis.

Wemelsfelder and Farish (2004) reviewed a set of qualitative categories, including fearfulness, for the assessment of sheep behaviour by direct observation of the whole flock. This was developed into a QBA protocol, which has been validated for the assessment of sheep welfare (Wickham et al., 2012). Although good reliability and repeatability have been obtained in overall QBA scores between assessors (Wemelsfelder and Farish, 2004), there is no information on whether this tool is reliable when only one category is assessed, because this was not its purpose.

The human-animal relationship (HAR) is a major determinant of sheep welfare because it is an important source of fear in farmed sheep (Waiblinger et al., 2006). This is particularly pertinent to extensive systems with limited interactions with people (Turner and Dwyer, 2007). The degree of aversion to human handling can also be influenced by the quality and sensitivity of the animals’ interaction with the farm stockperson (Dwyer, 2009) and therefore, presumably, abattoir staff. Thus, inferences can also be drawn about social attachment to humans, the nature (positive, neutral or negative) of past experiences with people and the quality of stockmanship (Waiblinger et al., 2006). In sheep, HAR has been measured using alertness to human approach in the field, escape attempts and an avoidance distance test (Table 1). In a comprehensive review of fear tests in farm animals, Forkman et al. (2007) demonstrated that fear of humans correlated with increased heart rate. While these
tests had good repeatability, the review also concluded that more evidence was needed to confirm their validity. The HAR is also affected by other variables, including the type of farming (intensive vs. extensive; Turner and Dwyer, 2007). This impairs its validity unless it is carried out in a controlled environment, which is not compatible with commercial abattoir conditions.

Published reviews have reported the deleterious effects of pre-slaughter stress on meat quality in ruminants (Ferguson and Warner, 2008), including sheep (Sañudo et al., 1998). The organic changes occurring during pre-slaughter stress can lead to rapid decline in pH in muscle due to increased ATPase activity and lactate accumulation (Monin, 1988; Liste et al., 2011). This increases the rate of protein denaturation post-mortem and reduces water holding capacity (WHC) of muscle, leading to dark, dry meat (Bond et al., 2004; Ferguson and Warner, 2008). The most cited meat quality indicators of pre-slaughter stress are pH and meat colour (Table 1), although the assessment of tenderness and WHC have also been used (Liste et al., 2011). Further research is needed to validate meat tenderness and WHC as indicators of prior distress in sheep.

The animal-based indicators described above and listed in Table 1 were developed for the assessment of sheep welfare at farm, during transport, or at the abattoir. From our assessment, 19 indicators can be regarded as high validity indicators from previous research work (Table 1); 13 are of medium validity and need further research, but four of these have high feasibility for measurement in abattoirs (conformation carcass classification, fat carcass classification, meat colour, meat pH); 15 are of low validity. Welfare indicators validated in environments other than the abattoir might be invalid in an abattoir. A summary of the factors that might affect validity or feasibility of the measurements is discussed below.
Live sheep in abattoirs are highly likely to be experiencing some degree of stress. Stress probably occurs during transport and market and might be increased further by mixing or close proximity of unfamiliar sheep. On arrival at the abattoir, during unloading and penning, sheep move rapidly en masse because of their flocking instinct and this increases stress, thereby reducing the validity of indicators that require observation of animals in a consistent environment (e.g. stereotypies, QBA). Additionally, there is often a restricted view of the whole animal (i.e. the torso, belly and legs are difficult to observe in tightly packed sheep), both when sheep are being moved and when penned, reducing the feasibility of measuring indicators that require observation of the whole animal (e.g. gait, vulvar discharge). Abattoir policy is to avoid handling sheep ante-mortem to minimise bruising, reducing the feasibility of taking measurements that require touching sheep (e.g. body condition scoring, close inspection of fleece, feet, eyes etc.).

Post-mortem, there is the potential to inspect the external surface of the carcass, with and without the hide, and the internal organs and carcass. However, the dressing line often moves rapidly and there is separation of carcass from hide and internal organs early in processing that can reduce the traceability between parts of the sheep. Abattoirs minimise handling of animals post-mortem, especially hide, lower limb, head and gastrointestinal organs, to maintain high standards of hygiene. Therefore, although one possible advantage of post-mortem inspection is that carcasses, organs, hide, head and feet could be examined thoroughly after a batch of animals has been processed, this increases the risk of meat contamination and would require a separate space from the line and different personnel to make the inspections. This reduces the feasibility of observing these at abattoirs.
Finally, to assess prior welfare, an indicator needs to be present for a period of time, so indicators that are highly variable temporarily are unlikely to be valid indicators of prior welfare (e.g. panting, shivering and indicators of fear and distress and normal behaviour). In this study, high validity indicators were initially considered by category to minimise repetition and maximise critical comparison; where an indicator was defined as having high validity in its original setting, but did not appear likely to be valid or feasible when measured in an abattoir, we considered the possible use of medium validity indicators with high feasibility, or alternative novel technologies.

Low BCS indicates prior long term poor welfare. Scoring BC is not possible in abattoirs because handling of lambs and ewes ante-mortem is not permitted. Carcass classification is an alternative to BCS and is likely to be more reliable than BCS in lambs. Video image analysis (VIA) technologies have been tested to provide an automated carcass classification (Rius-Vilarrasa et al., 2009; Einarsson et al., 2014), which might improve feasibility and validity of the assessment compared with more subjective evaluation using the EUROP classification system. Carcass classification was rated as medium validity in our study, because while it can be used to assess body condition of adult sheep (typically cull ewes), it was originally developed for slaughter lambs. Therefore, the method needs more validation, as would VIA, before being used to assess adults in abattoirs. Such validation is important because emaciation (defined as BCS<1.5) should be correlated with carcass conformation and fat grades or VIA. In cattle, there is significant correlation between BCS and carcass grade (Emenheiser et al., 2014), but this correlation is lower at low and high BCS

values (Apple, 1999; Apple et al., 1999). Carcass classification has the advantage of not causing bruising or distress to a live animal. However, assessment must take place on the processing line because carcasses are trimmed before being passed fit for human consumption and chilled.

Rumen fill indicates whether sheep were hungry prior to slaughter. It can be assessed by palpation of the left abdominal wall in live sheep or by inspection of the rumen or weighing rumen contents post-mortem. Palpation of live animals in abattoirs is not acceptable, as explained previously. Inspecting the rumen requires handling intestinal material, which is also avoided in abattoirs to avoid contamination. Therefore, to assess short-term hunger in the abattoir, either new indicators would need to be developed or a dedicated individual would have to be present to handle intestinal material in a separate area from the line. As mentioned previously, there is currently no validated measure to assess thirst in sheep and skin-tenting time would not be acceptable in abattoirs because it requires handling of sheep.

Body cleanliness and ectoparasite infestations could be observed both ante- and post-mortem. Body cleanliness could indicate poor welfare at any point prior to arrival and at the abattoir, while ectoparasite infestations are most likely to have occurred on farm.

Indicators such as shivering or panting provide information on the thermal comfort at the moment they are assessed, but they are not valid measures of prior thermal comfort. Therefore, this highly valid indicator is not useful to assess prior welfare. The possible exception is fleece cover, specifically excess or insufficient fleece observed in sheep in very hot or very cold weather, respectively. This low validity indicator might indicate prior
thermal discomfort on farm, at market or during transit to the abattoir. Further research is required to evaluate fleece cover as an indicator of prior thermal comfort.

Disease states can be acute or chronic. Some signs of disease will indicate prior welfare (e.g. dried faecal staining, skin irritation), but many signs of disease might be acute in onset (e.g. lameness, dyspnoea) and so of limited use in the investigation of prior welfare.

Some clinical signs of disease must be assessed ante-mortem (e.g. general demeanour, obviously sick, coughing, nasal discharge, dyspnoea, diarrhoea, lameness [gait score], ocular health, scratching or rubbing, vulvar discharge). These could be assessed without touching the animal and so could be feasible to assess at abattoirs, but sufficient time and space are needed to make appropriate observations of the whole animal. Obviously sick animals would also be detected post-mortem because the carcass would not set (undergo normal post-mortem changes, including rigor mortis).

Disease indicators that require inspection of the skin or fleece (skin lesions or irritation, faecal staining) could be performed ante-mortem without handling, immediately post-mortem, or after the fleece has been removed in a separate area. Large skin lesions could be observed, but small lesions (<2 cm diameter) and lesion depth are more difficult to identify; however, they have important implications for animal welfare. Such indicators are valid to assess prior sheep welfare.

Old injuries and bruising would indicate poor prior welfare. Some injuries would be detectable ante-mortem by examining animals, subject to good visibility. The extent of injuries might only become apparent once the hide is removed. Bruising could be recorded
post-mortem. Since it is included in assessment protocols for welfare in cattle \(^9\) (WQ, 2009), it is likely that it could be assessed in routine carcass assessments in sheep at abattoirs.

Recording of bruising would require dedicated staff on the line because carcasses are trimmed before grading and many bruises would be removed.

Mutilations would typically be incurred on farm and so are indicators of prior poor welfare. They can be observed ante-mortem (tail docking, short tail docking, ear notching), or early post-mortem (castration, tail docking, ear notching).

While some behavioural assessments are considered more valid than others, all assessments of behaviour, including QBA, are unlikely to reflect past welfare. This is because animals arriving from dealers, collection centres and markets might have been mixed and so their behaviour will alter. Moreover, an abattoir is a novel situation and it is likely that in such conditions sheep behaviour would not correspond to their normal behaviour on farm. Thus, all the indicators reviewed that assess behaviour can be considered valid indicators for welfare at the time of inspection at the abattoir, but it is unlikely they would provide valid information about prior welfare. Similarly, indicators that assess the animal’s current mental state (typically those described under freedom from fear and distress) are influenced by the animal’s response to a new environment, novel humans and sheep. Consequently, when assessed at the abattoir they might assess current welfare, but they are unlikely to provide an accurate reflection of past freedom from fear and distress. In contrast, indicators of meat

---

quality as a measure of distress in carcasses might be more indicative of recent past distress, although further research is required to validate these.

This study is the first phase of a project that aims to identify animal-based welfare indicators of prior welfare of sheep that could be measured in commercial abattoirs in the UK. The Five Freedoms was chosen as a framework to explore existing animal-based indicators because it covers all aspects of welfare. We hypothesised that if at least one indicator within a freedom could be identified for use in abattoirs, then prior welfare on that indicator could be assessed at abattoirs (e.g. freedom from hunger and thirst would require indicators for hunger and thirst). While some indicators for freedom from pain, injury and disease could be measured, others were not considered feasible, and some of those might be important. Gait score is one example; lameness is common in sheep but the current layout and movement of sheep in some abattoirs makes observation of sheep walking impossible.

The next stage in the development of a complete list of valid, feasible indicators that could be used to assess prior welfare in abattoirs is to test the indicators listed above to investigate whether they can be recorded with high reliability and repeatability at abattoirs as at their site of development and to investigate whether there are novel indicators that can be developed to assess prior animal welfare for all Five Freedoms. The ultimate aim is to have a complete set of indicators to benchmark the prevalence of welfare indicators to inform risk-based selection for inspections measuring compliance with animal welfare legislation, to evaluate whether an assurance scheme’s welfare conditions are met and to provide farmers with information on the health and welfare of their livestock to assist in health planning.

Conclusions
This systematic review identified some existing high validity indicators that might be useful to assess prior welfare of sheep at abattoirs. These were body cleanliness, carcass bruising, diarrhoea, skin lesions, skin irritation, castration, ear notching, tail docking and obviously sick. In addition, four medium validity, high feasibility, indicators were considered and two of those (carcass and fat classification) were considered useful to take forward. Other high validity indicators could not be used to measure prior welfare because either they were situation-specific (freedom to express normal behaviour or freedom from fear and distress) and measure current or recent welfare, or they were not feasible to measure in abattoirs. No indicators were identified that measured prior freedom to express normal behaviour or freedom from fear and distress. Freedom from discomfort could only be identified through scoring body cleanliness. For a set of indicators for all Five Freedoms to be completely robust, the potentially useful welfare indicators identified need further investigation to test their validity at abattoirs. Some novel welfare indicators are also necessary, assuming that measurement of historic fear and distress or historic ability to express normal behaviour could be measured.

Acknowledgements

This research was funded by the Department for Environment, Food and Rural Affairs (Grant code AW1028). We thank the two anonymous reviewers for the useful comments and suggestions and Dr Corinna Clark for proof reading the paper.

Conflict of interest statement

None of the authors of this paper has a financial or personal relationship with other people or organisations that could inappropriately influence or bias the content of the paper.
References


Jarvis, A.M. Cockram, M.S., 1994. Effects of handling and transport on bruising of sheep sent directly from farms to slaughter. Veterinary Record 135, 523-527.


Sargison, N.D., Scott, P.R., 2010. The implementation and value of diagnostic procedures in sheep health management. Small Ruminant Research 92, 2-9.


Table 1: Animal-based indicators of sheep welfare derived from a systematic review, classified by the Five Freedoms and their validity and likely feasibility for use in abattoirs to assess prior sheep welfare.

<table>
<thead>
<tr>
<th>Freedom</th>
<th>Category</th>
<th>Indicator</th>
<th>Observation (ante-mortem, A; post-mortem, P)</th>
<th>Validity a</th>
<th>Feasibility in abattoir b</th>
<th>References (first author and year of publication) c</th>
</tr>
</thead>
<tbody>
<tr>
<td>From hunger and thirst d</td>
<td>Body condition</td>
<td>Bodyweight</td>
<td>A or P</td>
<td>Low</td>
<td>High</td>
<td>Phythian, 2011</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Body condition score (BCS)</td>
<td>A or P</td>
<td>High</td>
<td>Low</td>
<td>Morgan-Davies, 2008; Caroprese, 2009; Dwyer, 2009; Napolitano, 2009; Phythian, 2011; Stubsjoen, 2011; van Burgel, 2011; Phythian, 2012a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Conformation carcass classification</td>
<td>P</td>
<td>Medium</td>
<td>High</td>
<td>Stubsjoen, 2011</td>
</tr>
<tr>
<td>From discomfort</td>
<td>Access to feed</td>
<td>Rumen fill</td>
<td>P</td>
<td>Medium</td>
<td>Low</td>
<td>Caroprese, 2009; Napolitano, 2009; Phythian, 2011; Stubsjoen, 2011; Phythian, 2012a; Phythian, 2012b</td>
</tr>
<tr>
<td></td>
<td>Cleanliness</td>
<td>Body cleanliness d</td>
<td>A or P</td>
<td>High</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lying behaviour</td>
<td>A</td>
<td>Low</td>
<td>Low</td>
<td>Dikmen, 2011; Phythian, 2011; Stubsjoen, 2011; Cockram, 2012</td>
</tr>
<tr>
<td></td>
<td>Thermal comfort</td>
<td>Fleece cover</td>
<td>A or P</td>
<td>Low</td>
<td>High</td>
<td>Phythian, 2011</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Shivering</td>
<td>A</td>
<td>High in lambs</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Panting</td>
<td>A</td>
<td>High</td>
<td>Low</td>
<td>Thornton, 2002; Cockram, 2004; Gougoulis, 2010; Phythian, 2011; Phythian, 2012b</td>
</tr>
<tr>
<td>From pain, injury and disease</td>
<td>Bruises</td>
<td>Carcass bruising e</td>
<td>P</td>
<td>High</td>
<td>High</td>
<td>Jarvis, 1996; Miranda-de la Lama, 2010; Liste, 2011; Teixeira, 2012</td>
</tr>
<tr>
<td></td>
<td>Gastrointestinal health</td>
<td>Diarrhoea e</td>
<td>A</td>
<td>High in lambs</td>
<td>High</td>
<td>Napolitano, 2011; Stubsjoen, 2011</td>
</tr>
<tr>
<td></td>
<td>Integument alterations</td>
<td>Endoparasitism</td>
<td>P</td>
<td>High</td>
<td>Medium</td>
<td>Dwyer, 2004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Skin lesions e</td>
<td>A or P</td>
<td>High</td>
<td>High</td>
<td>Scott, 2003; Fitzpatrick, 2006; Dwyer, 2008; Caroprese, 2009; Napolitano, 2009, 2011; Phythian, 2011; Stubsjoen, 2011 Fitzpatrick, 2006; Plant, 2006; Stubsjoen, 2011; Phythian, 2012b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Skin irritation e</td>
<td>A or P</td>
<td>High</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Lameness</td>
<td>Foot lesions</td>
<td>A or P</td>
<td>Medium</td>
<td>Medium</td>
<td>Fitzpatrick, 2006; Caroprese, 2009; Hodgkinson, 2010; Scott, 2003; Napolitano, 2009; Stubsjoen, 2011; Foddai, 2012</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>--------------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Gait assessment</td>
<td>A or P</td>
<td>High</td>
<td>Medium</td>
<td>Fitzpatrick, 2006; Caroprese, 2009; Kaler, 2009; King, 2011; Napolitano, 2009; Gougoulis, 2010; Phythian, 2011; Phythian, 2012b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mutilations</td>
<td>Castration</td>
<td>A or P</td>
<td>High</td>
<td>High</td>
<td>Scott, 2003; Fitzpatrick, 2006; Stubsjoen, 2011</td>
<td></td>
</tr>
<tr>
<td>Ear notch e</td>
<td>A or P</td>
<td>High</td>
<td>High</td>
<td>Phythian, 2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tail docking</td>
<td>A or P</td>
<td>High</td>
<td>High</td>
<td>Scott, 2003; Fitzpatrick, 2006; Stubsjoen, 2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ocular health</td>
<td>Eye condition</td>
<td>A</td>
<td>Low</td>
<td>Medium</td>
<td>Napolitano, 2011; Scott, 2011; Stubsjoen, 2011; Phythian, 2012b</td>
<td></td>
</tr>
<tr>
<td>Respiratory health</td>
<td>Cough</td>
<td>A</td>
<td>High</td>
<td>Medium</td>
<td>Napolitano, 2011; Scott, 2011; Stubsjoen, 2011; Phythian, 2012b</td>
<td></td>
</tr>
<tr>
<td>Dyspnoea</td>
<td>A</td>
<td>High</td>
<td>Medium</td>
<td>Mellor, 2003; Scott, 2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal discharge</td>
<td>A</td>
<td>Medium</td>
<td>Medium</td>
<td>Napolitano, 2009; Phythian, 2011; Scott, 2011; Stubsjoen, 2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reproductive health</td>
<td>Vulvar discharge</td>
<td>A or P</td>
<td>Medium</td>
<td>Medium</td>
<td>Napolitano, 2011</td>
<td></td>
</tr>
<tr>
<td>Systemic disease</td>
<td>General demeanour</td>
<td>A</td>
<td>Medium</td>
<td>Low</td>
<td>Gougoulis, 2010; Phythian, 2011; Phythian, 2012b</td>
<td></td>
</tr>
<tr>
<td>Obviously sick e</td>
<td>A</td>
<td>High</td>
<td>High</td>
<td>Mellor, 2003; Stubsjoen, 2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal behaviour</td>
<td>Redirected behaviour</td>
<td>A</td>
<td>High</td>
<td>Low</td>
<td>Dwyer, 2004; Dwyer, 2008; Nowak, 2008; Teixeira, 2012</td>
<td></td>
</tr>
<tr>
<td>Stereotypies</td>
<td>A</td>
<td>High</td>
<td>Low</td>
<td>Dwyer, 2004; Dwyer, 2008; Nowak, 2008; Teixeira, 2012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agonistic behaviour</td>
<td>Aggression</td>
<td>A or P</td>
<td>Low</td>
<td>Low</td>
<td>Dwyer, 2004; Weir, 2004; Gougoulis, 2010; Weir, 2012</td>
<td></td>
</tr>
<tr>
<td>Threats</td>
<td>A</td>
<td>Low</td>
<td>Low</td>
<td>Teixeira, 2012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal behaviour</td>
<td>Play behaviour</td>
<td>A</td>
<td>Low</td>
<td>Low</td>
<td>Boissy, 2000; Dwyer, 2008; Phythian, 2011</td>
<td></td>
</tr>
<tr>
<td>Ruminatory behaviour</td>
<td>A</td>
<td>Low</td>
<td>Low</td>
<td>Phythian, 2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qualitative Behaviour Assessment (QBA)</td>
<td>Qualitative behaviour</td>
<td>A</td>
<td>High</td>
<td>Low</td>
<td>Wemelsfelder, 2004; Dwyer, 2008; Phythian, 2011</td>
<td></td>
</tr>
<tr>
<td>From fear and distress</td>
<td>Affiliative behaviour</td>
<td>A</td>
<td>Low</td>
<td>Low</td>
<td>Boissy, 2007; Teixeira, 2012</td>
<td></td>
</tr>
<tr>
<td>Positive behaviour</td>
<td>Ear posture</td>
<td>A</td>
<td>Medium</td>
<td>Low</td>
<td>Reefmann, 2009</td>
<td></td>
</tr>
<tr>
<td>Animal behaviour</td>
<td>Freezing</td>
<td>A</td>
<td>Medium</td>
<td>Low</td>
<td>Bouissou, 1995; Cockram, 2004</td>
<td></td>
</tr>
<tr>
<td>Head position</td>
<td>A</td>
<td>Low</td>
<td>Low</td>
<td>Cockram, 2004; Hesworth, 2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QBA</td>
<td>A</td>
<td>Low</td>
<td>Low</td>
<td>Wemelsfelder, 2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>Indicator</td>
<td>Grade 1</td>
<td>Grade 2</td>
<td>Grade 3</td>
<td>Reference(s)</td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-----------------------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>----------------------------------</td>
<td></td>
</tr>
<tr>
<td>Vigilance behaviour</td>
<td>A</td>
<td>Low</td>
<td>Low</td>
<td>Dwyer, 2008; Deiss, 2009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocalisations</td>
<td>A</td>
<td>Low</td>
<td>Low</td>
<td>Bouissou, 1995; Cockram, 2004;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Boissy, 2007; Greiveldinger, 2007;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Nowak, 2008; Deiss, 2009;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gougoulis, 2010; Hemsworth, 2011;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Phythian, 2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human-animal relationship</td>
<td>Alertness to approach in the field</td>
<td>A</td>
<td>Low</td>
<td>Low</td>
<td>Napolitano, 2006; Waiblinger, 2006;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dwyer, 2008; Caroprese, 2009;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mattiello, 2010; Napolitano, 2011;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Stubsjoen, 2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bouissou, 1995; Cockram, 2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Avoidance distance test (ADT)</td>
<td>A</td>
<td>Low</td>
<td>Low</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Escape attempts</td>
<td>A</td>
<td>Low</td>
<td>Low</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meat quality</td>
<td>Meat colour</td>
<td>P</td>
<td>Medium</td>
<td>High</td>
<td>Dwyer, 2008; Liste, 2011;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Napolitano, 2011; Teixeira, 2012</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pH meat</td>
<td>P</td>
<td>Medium</td>
<td>High</td>
<td>Dwyer, 2004; Napolitano, 2006;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Deiss, 2009; Liste, 2011;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Teixeira, 2012</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tenderness</td>
<td>P</td>
<td>Medium</td>
<td>Low</td>
<td>Liste, 2011</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Water holding capacity</td>
<td>P</td>
<td>Medium</td>
<td>Low</td>
<td>Liste, 2011</td>
<td></td>
</tr>
</tbody>
</table>

a Validity: High validity indicators were those validated in previous research, medium validity indicators were those without a reliable method of assessment and low validity indicators were those that have been suggested as indicators in scientific literature but without evidence that they actually assess welfare.

b Feasibility is based on the likely ability to make the observation of the indicator in an abattoir.

c Reference column shows publications that identify the proposed indicator as a measure to assess sheep welfare. Only first author and year are presented (full references are in the reference list).

d No indicators of thirst were identified in the systematic literature review.

e Indicators graded as both highly valid and feasible and which might be used in abattoirs to assess prior sheep welfare (on farm, at market and during transportation).